

**RERconverge Expansion: Using Relative Evolutionary Rates to Study
Complex Categorical Trait Evolution**

Ruby Redlich

Advisor: Dr. Andreas Pfenning

Mentors: Dr. Amanda Kowalczyk, Heather Sestili

Abstract

Comparative genomics approaches seek to associate molecular evolution with the evolution of phenotypes across a phylogeny. Many of these methods, including our evolutionary rates-based method, RERconverge, lack the ability to analyze non-ordinal, multicategorical traits. To address this limitation, we introduce an expansion to RERconverge that associates shifts in evolutionary rates with the convergent evolution of multi-categorical traits. The categorical RERconverge expansion includes methods for performing categorical ancestral state reconstruction, statistical tests for associating relative evolutionary rates with categorical variables, and a new method for performing phylogeny-aware permutations, “permutations”, on multi-categorical traits. In addition to demonstrating our new method on a three-category diet phenotype, we compare its performance to binary RERconverge analyses and two existing methods for comparative genomic analyses of categorical traits: phylogenetic simulations and a phylogenetic signal based method. Our results show that our new categorical method outperforms phylogenetic simulations at identifying genes and enriched pathways significantly associated with the diet phenotypes and that the categorical ancestral state reconstruction drives an improvement in our ability to capture diet-related enriched pathways compared to binary RERconverge when implemented without user input on phenotype evolution. Through investigation of the *PIEZO1* gene, we also illustrate how diet-relevant genes detected by our method can possess convergent patterns of amino acid sequence change. An additional case study using the binary pair bonding phenotype illustrates how our categorical expansion can still be applied successfully to binary traits as indicated by our identification of relevant biological pathways related to male gametes, ovarian follicles, and behavioral response to drugs. The categorical expansion to RERconverge will provide a strong foundation for applying the comparative method to categorical traits on larger data sets with more species and more complex trait evolution than have previously been analyzed.

Introduction

A fundamental question in biology is identifying the genomic elements underlying complex phenotypes. One way to address this question is through the lens of convergent evolution, or the process in which more than one evolutionary lineage independently acquires similar phenotypes. These independent evolutionary events are natural replicates of phenotype acquisition in which similar genetic changes in correspondence with similar phenotypic changes may indicate a genotype-phenotype relationship (Kowalczyk et al. 2019; Li et al. 2010; Partha et al. 2017). RERconverge identifies genes whose evolutionary rates, the number of amino acid substitutions per unit time, are changing in accordance with the phenotype (Kowalczyk et al., 2019).

More specifically, the rate of substitution along a branch is quantified as an evolutionary rate and, by correcting for factors affecting rates at all genes or on all branches of a phylogeny, we can compute relative evolutionary rates (RERs) that reflect whether a given gene is evolving slower or faster than expected along a branch of the phylogeny (Kowalczyk et al. 2019). Comparing the distributions of these RERs across phenotypes enables us to identify regions of the genome showing convergent rate shifts in association with the phenotype, thus making use of convergent evolution as a tool in comparative genomic analyses. RERconverge has been used with great success to discover both coding and non-coding elements related to the evolution of continuous and binary traits including mammalian lifespan, hairlessness, and marine habitation (Kowalczyk, Chikina, and Clark 2022; Kowalczyk et al. 2020; Chikina, Robinson, and Clark 2016).

Nonetheless, like many phylogenetic software packages, RERconverge was compatible with only binary or continuous traits, but did not support non-ordinal, multicategorical phenotypes despite their prevalence in nature e.g circadian rhythm, diet type. Here I present an extension to RERconverge that provides three major updates enabling the use of the software for analyses of multicategorical traits: (1) it infers the ancestral history of categorical traits; (2) it provides a “permutation” strategy for categorical traits to return reliable, phylogeny aware, corrected p-values; and (3) it has options for both parametric and non-parametric statistical tests to associate RERs with categorical phenotypes, including post hoc pairwise testing. As a test case, we applied these new methods to a set of categories representing variation in mammalian diet phenotypes. Applying our new methodology, we found many diet-related pathways enriched for genes with significant differences in RER distributions across diet categories, highlighting the potential of this method for genome-wide analyses. This work is described in full detail in our manuscript, *RERconverge Expansion: Using Relative Evolutionary Rates to Study Complex Categorical Trait Evolution* (Redlich et al., 2023) and summarized in this thesis report.

One of the key contributions of the RERconverge extension is the incorporation of maximum likelihood inference methods for ancestral state reconstruction. Pair bonding, the prolonged preference for a single mating partner, was previously analyzed with the binary RERconverge method with limited success (unpublished data). We revisit the pair bonding phenotype using our new methods for ancestral inference and permutations and demonstrate our ability to capture true signal in the data through the identification of many relevant biological pathways enriched with genes evolving at different rates in pair bonding mammals.

Background/Prior work

Changes in selective pressure can frequently be observed as changes in the rate of substitutions along the branches of a phylogeny due to increased constraint, relaxation of constraint, or positive selection (Yang 2007; Pond, Frost, and Muse 2005). There are a variety of methods that have been developed to detect convergent evolution at the same individual nucleotides (Hasselmann et al. 2008), individual amino acids ((Rey et al. 2018; Thomas and Hahn 2015; Fukushima and Pollock 2023), regulatory elements (Partha et al. 2017; Kowalczyk, Chikina, and Clark 2022; Kaplow et al. 2023; Yan et al. 2023; Hu et al. 2019), and genes (Kosakovsky Pond et al. 2020; Kowalczyk et al. 2019). These methods have been applied to identify convergent molecular evolution associated with a wide variety of traits in a large number of clades (Yusuf et al. 2023; Jin et al. 2023; Wang et al. 2020; Espindola-Hernandez, Mueller, and Kempnaers 2022; Bodawatta et al. 2023).

Categorical Methods

A limitation to most phylogenetic comparative methods is the inability to analyze non-ordinal, multicategorical traits. Consider, in the simplest case, a three category phenotype. Excluding any one category to perform a binary analysis may significantly reduce the number of species in the analysis, and therefore the power to detect genetic elements. Combining categories instead of excluding them presents potentially arbitrary choices that could have significant impacts on the results. These problems are likely to arise for any non-ordinal, multicategorical trait. Moreover, it becomes increasingly challenging to encode a categorical trait as a binary trait as the number of categories increases.

Some methods do exist to specifically study categorical data. Garland et al. (Garland et al. 1993) introduced the phylogenetic simulations method, which uses an ANOVA test to compare means of a continuous phenotype across categories and computes p-values empirically from simulations to account for phylogenetic relationships between species. Multiple phylogenetic packages including phytools (Revell 2012) and geiger (Pennell et al. 2014) have implemented this phylogenetic ANOVA test. ANOVA is a parametric test, with much stricter assumptions than the non-parametric alternatives, and phylogenetic data often do not meet these assumptions. Moreover, the method is technically designed to compare *continuous* traits among species grouped into certain categories.

Ribeiro and Borges (Borges et al. 2019; Ribeiro et al. 2023) presented the delta statistic for calculating phylogenetic signal for categorical traits. One of many applications of the delta statistic is identifying genetic elements associated with a categorical phenotype. This statistic has been successfully applied to find genes associated with the evolution of mammalian activity patterns (Borges et al. 2019). However, we still note a few limitations to this method. Currently, the only option is to calculate p-values by random permutation of the trait vector. Random permutation represents the null hypothesis that there is no phylogenetic signal, but this is not a good representation of the null hypothesis that a genetic element is not associated with a phenotype since we expect even null phenotypes to maintain phylogenetic relationships (Saputra et al. 2021). Additionally, the delta statistic does not take into account or provide any information on evolutionary rates, so it cannot be used in the specific context where our goal is to find genetic elements that are convergently accelerated or conserved.

Phenotypes

One common way of categorizing mammalian diets is by trophic level: carnivory, herbivory, and omnivory (Kim et al. 2016). We considered mammals that almost exclusively consume plant matter as herbivores, mammals that almost exclusively consume vertebrates or invertebrates as carnivores, and mammals that consume more than one type of food source as omnivores (Redlich et al., 2023).

Lukas, Clutton-Brock 2013 classify pair bonding, or social monogamy, as when a single male and female share a territory and remain together for greater than one breeding season whether or not they have offspring (Lukas and Clutton-Brock 2013).

Methods

Continuous time Markov model (CTMM) Ancestral Reconstruction

Categorical RERconverge uses a continuous time markov model of evolution in order to perform ancestral state reconstruction, a framework that has been commonly applied to categorical traits (Pagel 1994; Pupko et al. 2000; Paradis, Claude, and Strimmer 2004). This is an important step because it allows RERconverge to associate relative evolutionary rates with the convergent evolution of the phenotype across the entire phylogeny, not just extant branches. Maximum likelihood estimation is used to infer a transition rate matrix, Q , from the user supplied phylogeny, rate model, and extant phenotype data. The phylogeny is the master tree which includes all species in the analysis and has branch lengths representing average genome-wide evolutionary rates. The transition matrix is inferred using the *fit_mk* function from the *castor* package (Louca and Doebeli 2018), and is used to compute the marginal ancestral likelihoods at each node. Code for the computation of ancestral likelihoods is heavily based on *ace* from the package *ape*, using the same double pass algorithm but is modified to work with unrooted, non-dichotomous trees (Paradis, Claude, and Strimmer 2004). Each node is then assigned the state with the maximum marginal likelihood. Though this is not the same as the assignment of states that maximizes the joint likelihood, marginal ancestral likelihoods can be computed rapidly even for large phylogenies.

The rate model describes the number and position of free rate parameters in the transition rate matrix. In order to compare rate models, RERconverge implements a likelihood ratio test which computes the log likelihood of the fitted transition matrix under each user supplied rate model. Pairwise comparisons are made between each rate model with its more complex rate models and the likelihood ratio and p-value is computed for each comparison. If the two rate models are nested (the simpler one is a special case of the more complex one), then the likelihood ratio is distributed as a chi-squared with degrees of freedom equal to the difference in the number of free parameters between the simpler and more complex model and the p-value is determined accordingly (Pagel 1994). Other packages implement a similar test for nested models including *anova* in the *ape* package (Paradis et al., n.d.). However, the likelihood ratio test in RERconverge works with both nested and non-nested models and will automatically detect whether the models are nested or not. For non-nested models, Monte Carlo simulations are used to determine an empirical p-value for the likelihood ratio (Pagel 1994).

For binary phenotypes, the user can choose whether branches are considered foreground based on the state of possessing the convergent trait or based on the transition from not possessing to gaining the trait.

Both methods were tested for hairless species with no large impact found (Kowalczyk, Chikina, and Clark 2022). However, with categorical traits especially when the ancestral trait is unclear, the number of possible transition types is quadratic with respect to the number of categories. Thus we decided to assign edges based on state rather than transition. Edges were assigned the state of their descendant nodes such that all edges leading to extant species are assigned based on the observed phenotype.

In order to handle missing species, RERconverge computes “paths” (Kowalczyk et al. 2019). When species are missing, certain nodes are no longer necessary so edges will be combined into “composite” edges and the paths describe what state values to use for these composite edges. Categorical RERconverge assigns the composite edges the state of the most recent edge in the master tree. This ensures that composite edges leading to extant species always use the observed state.

Permutations

Permutations is an important part of RERconverge because it allows users to calculate reliable p-values for the association of genomic elements with convergent traits despite unknown sources of dependence in the data leading to non-uniform null p-value distributions (Saputra et al. 2021). Permutations generate permuted phenotype trees that maintain the phylogenetic relationships in the data. For instance, binary permutations maintain the same number of foreground species and the same structural relationships between foreground species (Saputra et al. 2021).

Categorical permutations are accomplished slightly differently than binary and continuous permutations because the simulation step does not use a Brownian motion model. Instead, phenotypes are simulated from the continuous time markov model that was used to reconstruct the ancestral history of the trait. As with binary and continuous permutations, the simulation is based on a phylogeny with branch lengths representing the average genome-wide evolutionary rate along that branch.

The categorical permutations algorithm employs three steps to ensure that the permuted phenotype contains the same number of species with each trait value as the original phenotype:

- 1) Rejection sampling: any simulated phenotype in which there are not the same number of extant species with each trait value as the original phenotype is rejected.
- 2) Permutation of internal traits: the simulated values for internal species are ignored. Instead, the originally inferred internal trait values are permuted and assigned to internal species in the permuted phenotype. Assignment is weighted by the ancestral likelihoods calculated from the simulated tip values. This ensures that the initial permutation of internal traits is more optimal than a completely random shuffle.
- 3) Re-organize internal traits: a search technique similar to simulated annealing is used to reorganize the internal states relative to the simulated extant states to improve the likelihood of the permuted phenotype. This is accomplished through a series of swaps. Pairs of candidate nodes are suggested based on which internal nodes from step (2) disagree most with the ancestral likelihoods at that node. A swap is made, and if the swap improves the likelihood of the tree, then the swap is kept. To avoid getting stuck, the swap is also made with a small probability even if it does not improve the likelihood of the tree. This

generates a plausible trait history that exactly matches trait category counts and has a comparable likelihood to the original simulation.

Algorithmically step 3 is computed as follows: Let Q be the instantaneous transition rate matrix that was fit on the phenotype data. Let A be the matrix of ancestral likelihoods where each row is a node and each column is a phenotype state. The initial likelihood of the tree after step 2 is computed.

$$likelihood = \prod_{i=1}^{num\ edges} (e^{Qt_i})_{x_i y_i} \quad (1)$$

where t_i is the length of edge i , e^{Qt_i} is the transition probability matrix, x_i is the ancestor on edge i , and y_i is the descendant on edge i . For each internal node i currently assigned state y , ratios are computed for all other states x .

$$ratio_{i,x} = \frac{A_{i,x}}{A_{i,y}}$$

Thus a large ratio indicates that node i prefers to be in state x over its current state y .

A ratio, $r1$, is selected at random with more weight placed on larger ratios so that poorly assigned nodes are more likely to be swapped. Let $r1 = ratio_{i,x}$ where node i is currently in state y . Then, a second ratio, $r2$, is selected from a list of ratios, $ratio_{j,y}$, where each node j is currently in state x . Once again, more weight is placed on larger ratios. A swap in which node i is switched to state x and node j is switched to state y is then proposed.

The likelihood of the tree is recomputed under the swap. If the likelihood improves, then the swap is kept. If the likelihood does not improve, then the swap is made with probability $u = \exp(-dh/T_k)$ where $dh = \log(\frac{likelihood\ before\ swap}{likelihood\ after\ swap})$. Thus if dh is large (the swap is very unfavorable), u will be small and it is less likely to make the swap. T_k represents the ‘‘temperature’’ which is a common feature of simulated annealing algorithms. The temperature begins high during early iterations and decreases as the iterations go on. Thus, in the early iterations u is larger, and unfavorable swaps are made with a higher probability. This allows the algorithm to try more swaps, especially early on, to increase the overall number of potential state configurations it explores, even if some of those swaps initially decrease the likelihood. Only allowing swaps that improve the likelihood may cause the trees to get stuck before they reach a more favorable state assignment.

At the end of each cycle k , the new temperature is calculated as $T_k = \frac{T_0}{1 + \alpha k}$ where T_0 and α were chosen semi-arbitrarily and are currently 10 and 0.9 respectively. 100 cycles were run, each with 10 iterations for a total of 1000 iterations.

In order to determine the effectiveness and importance of each step in the above algorithm, we plotted the distributions of tree log likelihoods after each step (**Fig. 1**). After step 2, the log likelihoods decrease because we replace the simulated internal states with a permutation of the original internal states. However, after step 3, the completed permulated trees (tan) return to having the same or improved likelihoods compared to the simulations (blue). Thus, plotting the log likelihood distributions after each

step demonstrates why step 3 is extremely important for ensuring the overall plausibility of the permulated phenotypes and thus maintenance of the types of phylogenetic dependencies present in the original phenotype tree.

Once the permulated phenotypes are generated, the correlation between RERs and each permulated phenotype is computed for each gene. Empirical p-values are then computed as the number of null effect sizes more extreme than the observed effect size for that gene. One sided empirical p-values are computed for the omnibus test and two-sided empirical p-values are computed for the pairwise tests.

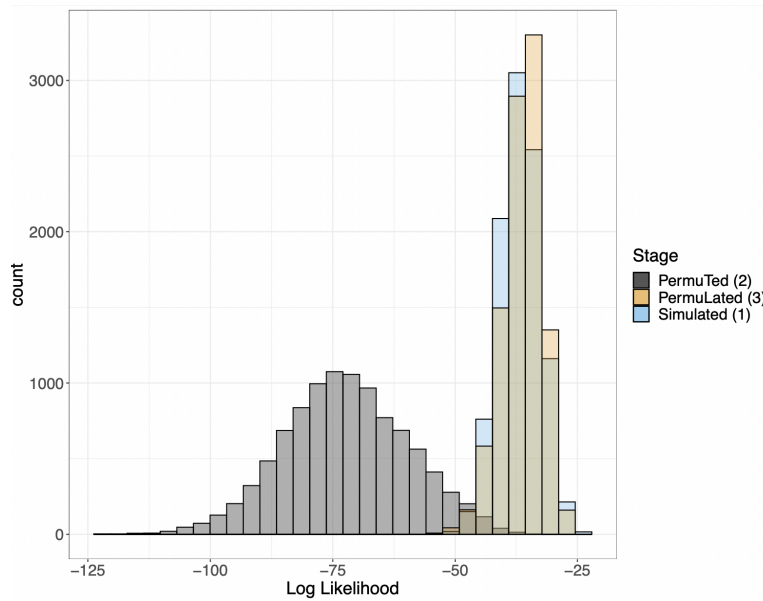


Figure 1. In blue are the log likelihoods of the 10,000 simulated trees (after step 1), in gray are the log likelihoods of the 10,000 permuted trees (after step 2), and in tan are the log likelihoods of the 10,000 finished permulated trees (after step 3).

Categorical Statistical Tests

Relative evolutionary rates are associated with the categorical phenotype in one of two ways. The default approach is to use a nonparametric Kruskal Wallis test followed by pairwise Dunn tests (Ogle et al. 2023). There is also an option to use an ANOVA test followed by pairwise Tukey tests (Foundation for Statistical Computing, n.d.). The Kruskal Wallis test reports the epsilon squared effect size computed as $\frac{H}{n-1}$ where H is the Kruskal Wallis test statistic and n is the total number of observations (species), and the corresponding p-value (King, Rosopa, and Minium 2018). The Dunn test reports the Z statistic for each pairwise test and the p-value after adjusting for multiple comparisons. The ANOVA test reports the eta squared effect size computed as $\frac{\text{sum squares effect}}{\text{sum squares effect} + \text{sum square residual}}$, and the corresponding p-value. The Tukey test reports the difference between each pair of means and the p-value corrected for multiple comparisons.

Results

Diet Phenotype Ancestral Reconstruction

To test our extension of the RERconverge method to categorical traits, we used the phenotype of diet (carnivore, omnivore, and herbivore) for 115 mammals, along with gene trees inferred from coding sequence alignments from these species' high quality genomes (Hecker and Hiller 2020). The diet phenotype was annotated using prior literature for those species (Nowak 1999) (**Fig. 2**).

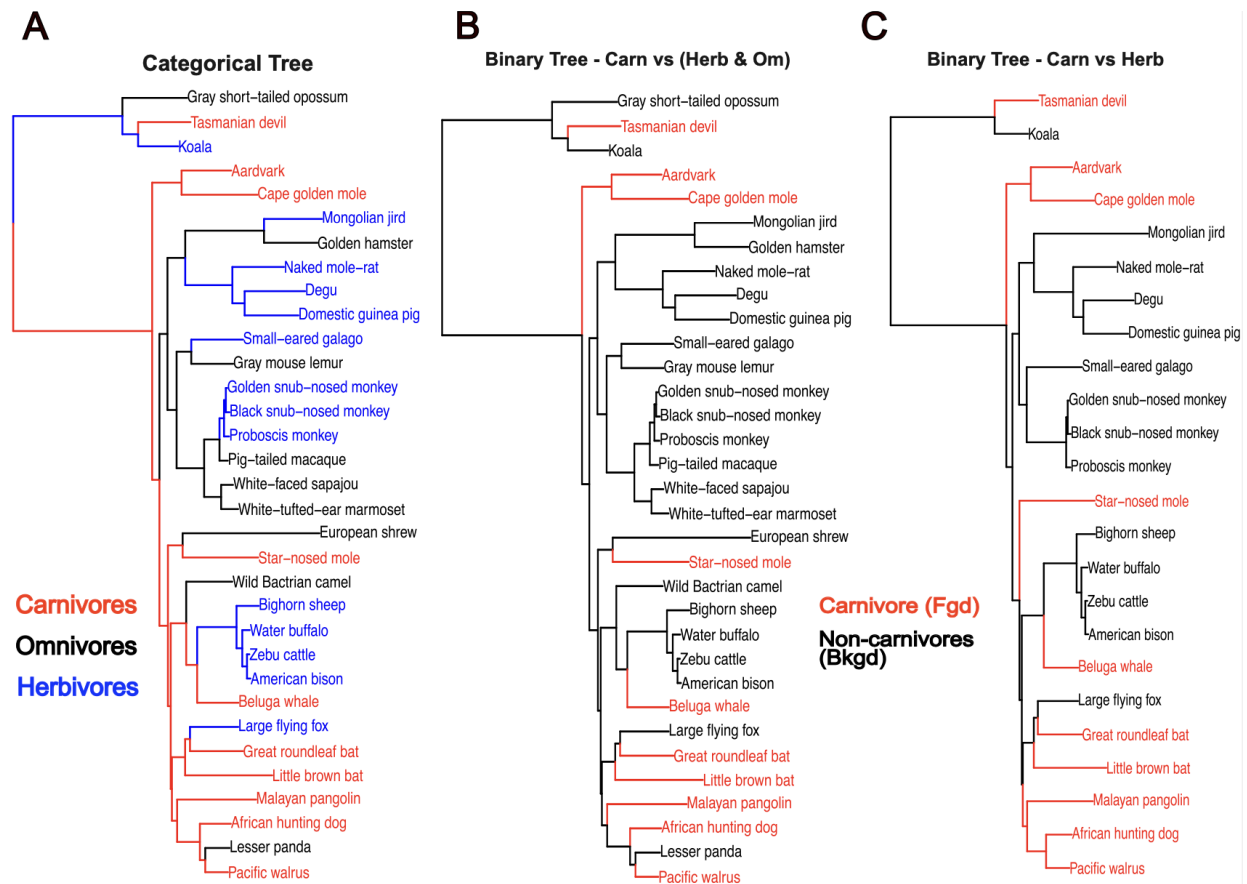


Figure 2. Example of a categorical trait reconstruction on a subset of mammals included in the full analyses. A) Categorical trait reconstruction using maximum likelihood applied to a continuous time Markov model. B) Default binary trait reconstruction with carnivore foreground and herbivore/omnivore background. Uses an approximate maximum parsimony based approach and assumes trait evolution can only occur from background to foreground. C) Default binary reconstruction on phylogeny of carnivores (foreground) and herbivores (background) with omnivores removed. Uses an approximate maximum parsimony based approach and assumes trait evolution can only occur from background to foreground.

In order to perform the ancestral reconstruction, we first considered three commonly applied rate models. These are equal rates (ER), symmetric (SYM), and all rates different (ARD) (Paradis, Claude, and Strimmer 2004) (**Fig. 3**).

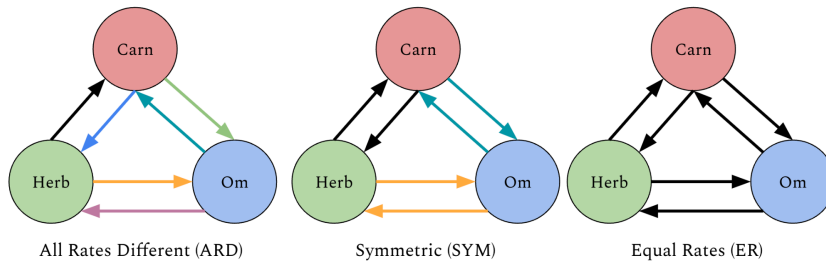


Figure 3. The circles represent the three categories (states) - carnivore (Carn), herbivore (Herb), omnivore (Om). The colored arrows represent the transition rates where arrows of the same color represent transitions with the same rate.

The rate model can have a significant impact on the results of the ancestral trait/state reconstruction; thus, we compared the fit of each of these three rate models to our observed phenotype data using a likelihood ratio test, as implemented in the expanded categorical RERconverge (Kowalczyk et al. 2019; Pagel 1994). The ARD model provided a significantly better fit to the data than the other models ($p = 0.00952$ compared to ER, $p = 0.02354$ compared to SYM); thus, this model was used to infer the ancestral states. Under the ARD model, the ancestral mammal was inferred to be a carnivore, while omnivores and herbivores were inferred to have evolved independently multiple times. The resulting pattern of evolution of the three phenotypes on the phylogeny, inferred using the ARD model, included multiple transitions between categories, suggesting sufficient power to identify diet-associated convergent molecular evolution. Specifically, this inference included 4 direct transitions between carnivore and herbivore, 12 direct transitions between carnivore and omnivore, and 19 direct transitions between herbivore and omnivore, where each transition is interpreted as a potential independent convergent event.

Categorical RERconverge analysis results

The ability to detect diet relevant genes with our categorical methods are highlighted by the RERs of two genes within the digestive tract development pathway, ITGB4 and ITGA6 (**Fig. 4**). They encode integrin subunits which tend to associate to form a heterodimer (O’Leary et al. 2016), and interestingly both show acceleration among carnivores compared to herbivores (**Fig. 4A,B**; ITGB4 $p = 5.662 \times 10^{-14}$, adjusted $p = 1.0719 \times 10^{-9}$, permutation $p = 0.0$; ITGA6 $p = 5.74 \times 10^{-5}$, adjusted $p = 0.00407$, permutation $p = 0.006$). These genes show different patterns relative to omnivores; ITGB4 illustrates that the distributions of relative evolutionary rates may differ significantly between all three categories while ITGA6 is not significantly different between carnivores and omnivores.

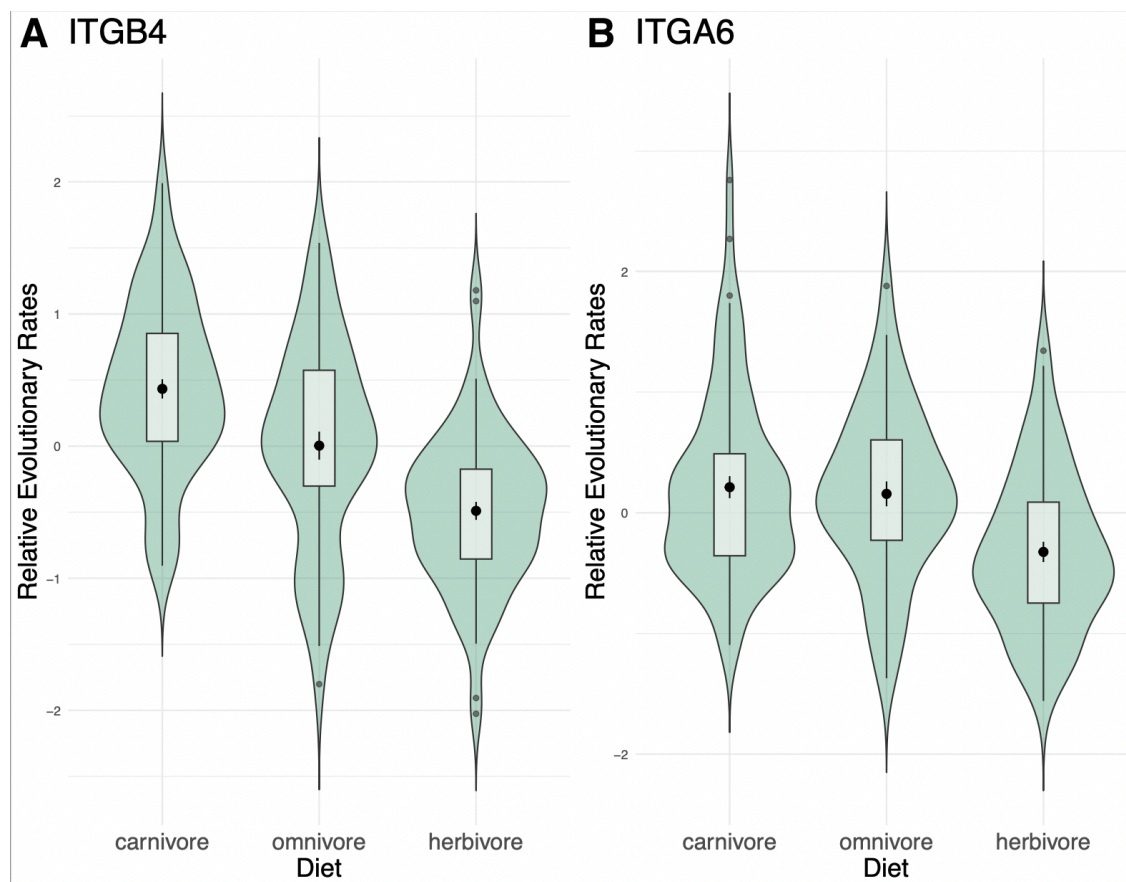


Figure 4. Violin plots of the distribution of relative evolutionary rates across each diet category for A) ITGB4 and B) ITGA6.

In addition to the ability to detect diet relevant genes with different patterns of evolutionary rate shifts across categories, we demonstrate that detected genes can also show convergent patterns of amino acid sequence change. We decided to focus on the PIEZO1 gene due to its important role in the digestive system and significant difference in evolutionary rates across groups as detected by our categorical RERconverge method (omnibus test adjusted $p = 2.7 \times 10^{-4}$; permutations $p = 0.029$) (He et al. 2023) (**Fig. 5A**). PIEZO1 is a mechanosensitive ion channel which opens in response to mechanical forces. The ensuing influx of cations, especially Ca^{2+} , is involved in many downstream processes including activation of integrin, ERK1/2-MAPK, and other signaling pathways involved in cell differentiation during digestive system development (He et al. 2023). Interestingly, integrin genes like those above were among the most significant genes identified by the categorical RERconverge analysis and both positive regulation of ERK1/2 cascade ($p = 3.87 \times 10^{-5}$; adjusted $p = 0.0027$; permutation $p = 0.0175$) and MAPK cascade ($p = 5.64 \times 10^{-4}$; adjusted $p = 0.0188$; permutation $p = 0.0143$) pathways were enriched among the carnivore-herbivore pairwise results (**Fig. 5C**). Together, this may imply a candidate role for PIEZO1 mediated cell differentiation during digestive system development in the evolution of diet type, though more follow up research would be necessary. PIEZO1 has also been implicated in the stimulation of the inflammatory response in the digestive system, sensory conduction in teeth, the stimulation of bile secretion through calcium-mediated pathways, iron release in the liver, intestinal motility, and maintenance of the intestinal epithelium through promotion of apical extrusion (He et al. 2023).

An analysis of the PIEZO1 alignment identified at least one position with convergent patterns in amino acid change. A convergent pattern was defined by instances in which the set of amino acids possessed by the species in one diet category have minimal overlap with the set of amino acids possessed by species in one or more of the other diet categories. The residue corresponding to residue 1835 in the human PIEZO1 amino acid sequence is glycine in all but two of the 29 carnivore lineages (ignoring lineages with a gap character). While 8 of the 44 herbivore lineages and 5 of the 23 omnivore lineages also possess a glycine, this residue is much more variable among herbivores and omnivores with the majority of lineages possessing alanine, glutamic acid, arginine, threonine, valine, or leucine (**Fig. 5A**). Furthermore, we verified that this pattern was independent of clade membership in the phylogeny and truly represented differences between diet types (**Fig. 5D**). This suggests that carnivores may be less tolerant to changes in this position compared to herbivores and omnivores. In human PIEZO1, this residue falls within a disordered region between two transmembrane domains (Surhone, Tennoe, and Hensonow 2010). It is also the site of a known, likely benign, missense variant within human populations that changes alanine to valine (“rs762530150” 2022).

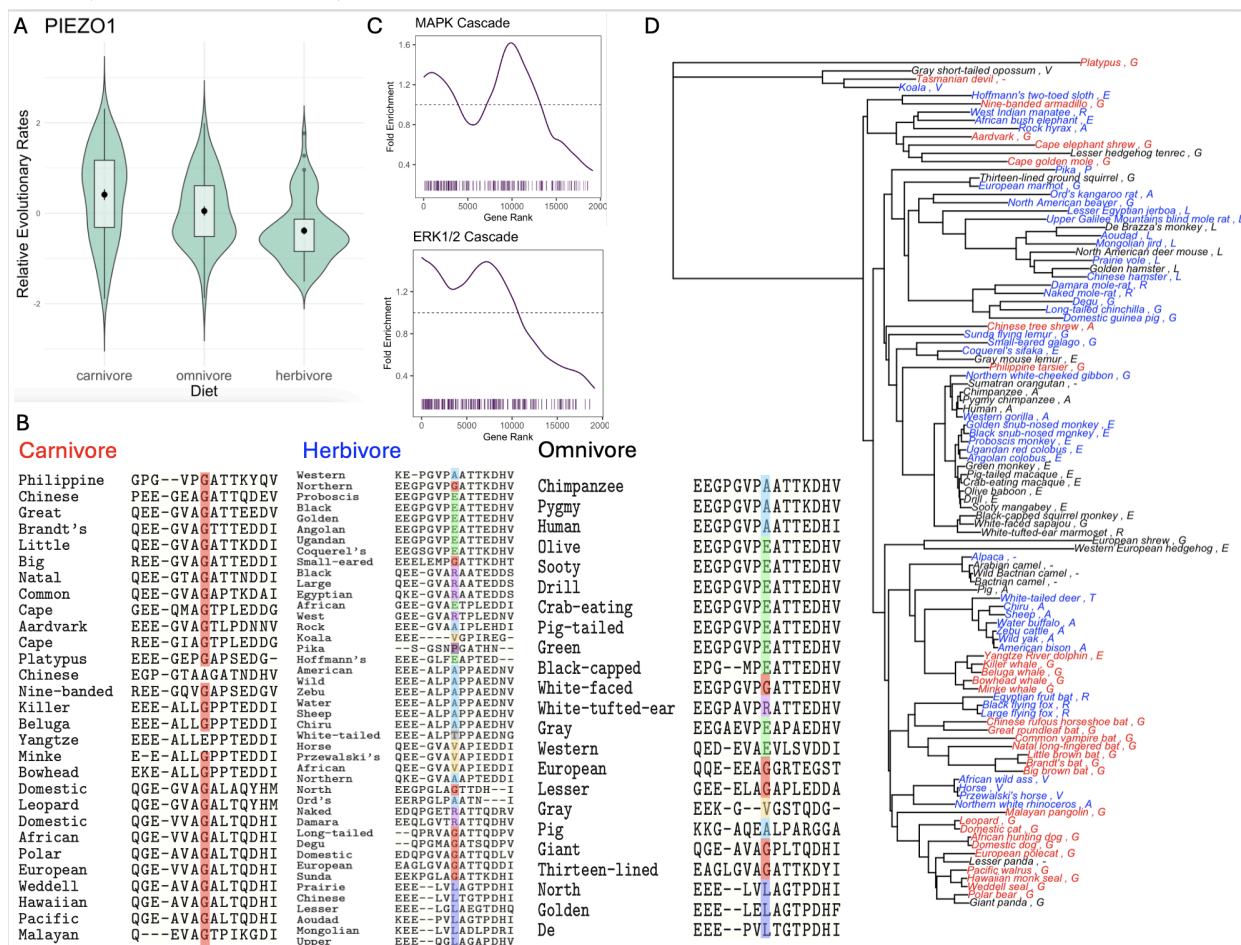


Figure 5. A) Violin plot of relative evolutionary rates of PIEZO1 showing a significant difference in rates between diet categories. B) Segment of the multiple sequence alignment with the convergent residue highlighted. Most carnivores possess a glycine whereas there is much more variety at this position in herbivores and omnivores. C) Fold enrichment and bar code plots of the MAPK cascade and positive

regulation of ERK1/2 cascade GO pathways which had significant enrichment under the categorical carnivore-herbivore pairwise test. D) Phylogeny with carnivores (red), herbivores (blue), and omnivores (black). The identity of the highlighted residue is written next to the species name.

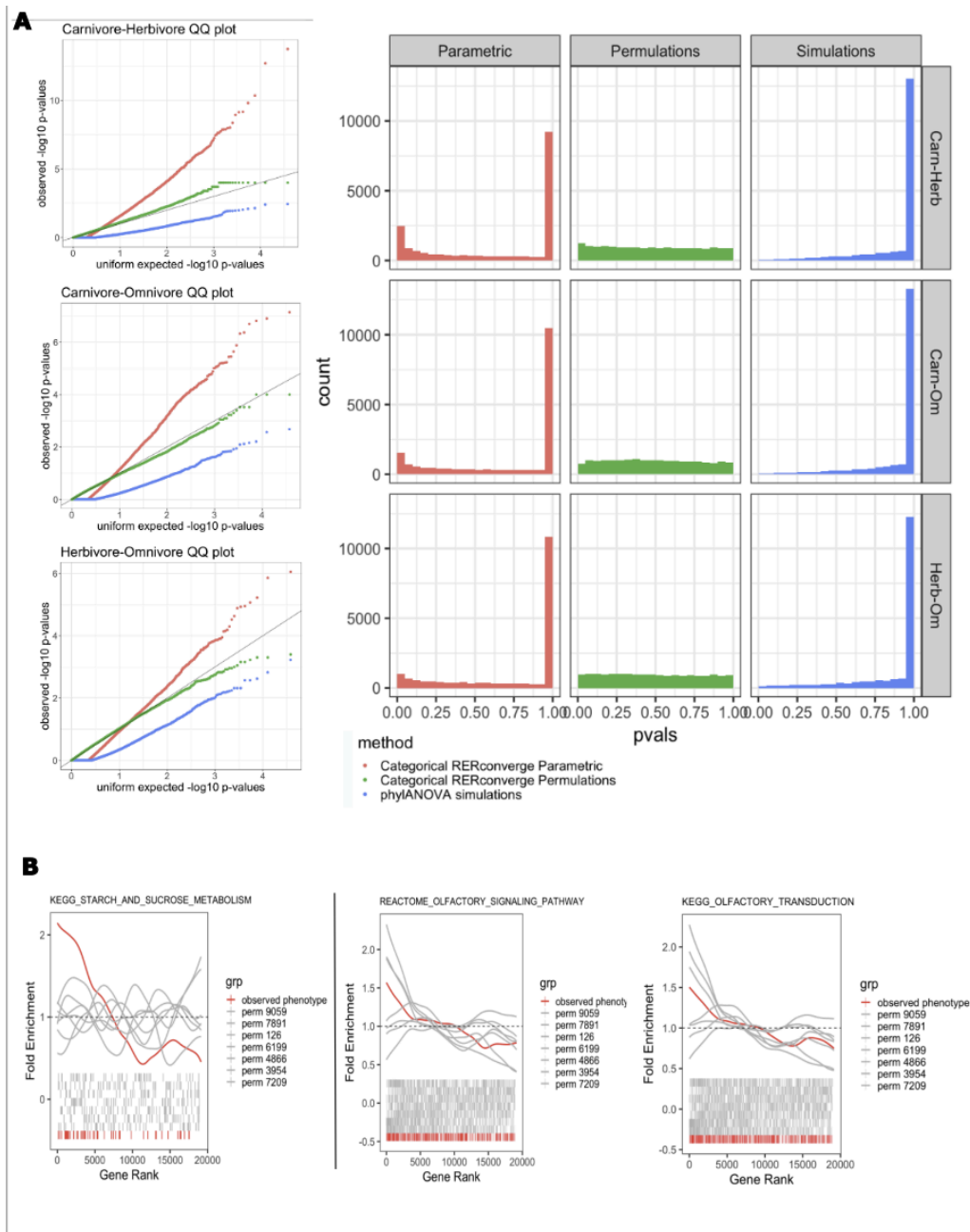
Permutations Results

Previously, binary and continuous permutations was demonstrated to refine pathway enrichment results by taking nonindependence of gene ranks into account (Saputra et al. 2021). We demonstrate that categorical permutations has the same ability to refine pathway enrichment results. For example, we observed that olfactory genes commonly cluster together in rank. As a result, olfactory signaling and olfactory transduction were two of the most significant pathway enrichment results according to non-permuted p-values. These had more significant p-values than the starch and sucrose metabolism pathway from the canonical gene set (Subramanian and Others 2005; Liberzon et al. 2011) which has a clear relationship to the diet phenotype. After permutations, starch and sucrose metabolism was retained as the top result (based on permutation p-values), while olfactory signaling and olfactory transduction were no longer among the top enriched pathways (**Fig. 6B**).

We also expect categorical permutations to account for non-uniform null distributions. To determine if this was the case, we plotted histograms and quantile-quantile plots of the parametric and permutation p-values (**Fig. 6A**). The categorical pairwise tests all showed a large enrichment for high parametric p-values of around 1. This enrichment of high p-values is also detected in the permutations and is thus no longer present among the permutation p-values. This demonstrates that the enrichment of high p-values is most likely specific to this statistical test and sources of nonindependence in the data rather than representing a true pattern specific to the diet phenotype.

Notably, the simulation p-values calculated by phylANOVA still have an enrichment of high p-values near one (**Fig. 6A**). In fact, this enrichment is even more extreme than that of the categorical pairwise parametric p-values. PhylANOVA simulates the RERs using a Brownian Motion model, so while this takes phylogenetic dependence into account, it doesn't account for other systematic variation such as nucleotide content or genome quality that can affect the RER values themselves and lead to non-uniform null distributions. The enrichment of parametric and phylANOVA simulation p-values near 1 and removal of such enrichment by permutations was observed across all the pairwise tests (**Fig. 6A**). Thus, unlike permutations, phylogenetic simulations were not able to correct for this pattern.

Additionally, though the same number of simulations and permutations (10,000) were performed, the categorical RERconverge pairwise test permutation p-values show more power, which may be because phylANOVA only uses extant species while RERconverge uses ancestral species as well.



Pairbonding Results

We obtained pair bonding phenotype annotations from Lukas, Clutton-Brock 2013. We used a phylogeny of 173 mammals constructed from genomes from the Zoonomia Consortium (Zoonomia Consortium 2020) to run an RERconverge analysis using our new categorical methods on a binary encoding of the pair bonding phenotype. There were 24 contemporary socially monogamous mammals across multiple clades and 149 non-pair bonding mammals. We performed ancestral reconstruction using the maximum likelihood inference methods included in the RERconverge categorical expansion. We inferred 11 independent transitions to pair bonding from a non-pair bonding ancestral mammal, providing statistical power for the comparative analysis. In order to calculate phylogeny corrected p-values, we performed 10,000 permutations for both the gene correlation results and the biological pathway enrichment results.

To interpret the results at a more functional level, we performed clustering on the enriched pathways from the mouse genome informatics (MGI) annotations (Liberzon et al. 2011; Subramanian and Others 2005), resulting in clusters of significant pathways (adjusted p-value < 0.05, permutation p-value < 0.01) that share common genes and thus similar biological functions. Notably, we identified clusters relating to male germ cells, ovarian follicles, the pituitary and luteinizing hormone, and drug response (**Fig. 7**). These clusters highlight important changes occurring during the convergent evolution of pair bonding, many of which are supported by experimental results in pair bonding species including changes due to lower sperm competition, hormones involved in female reproduction, and potential parallels between neural regions involved in drug responses and partner bonding.

Sperm competition refers to when sperm of two or more males compete to fertilize the oocytes of one female (van der Horst and Maree 2014). Socially monogamous (pair bonding) species experience much lower sperm competition because, with the exception of extra-pair mating which generally occurs at low rates in mammals (Lukas and Clutton-Brock 2013), only one male is mating with a single female. Studies in diverse species that experience little to no sperm competition, including the Eurasian bullfinch, bandicoot rat, and naked mole rat, identified slower sperm velocity, retention of immature sperm morphology, abnormalities in sperm morphology, smaller testis, and lower quality ejaculate (van der Horst and Maree 2014). Though naked mole rats are eusocial and live in large colonies, the female queen breeds with only one selected mate, so like socially monogamous species, they experience low sperm competition (van der Horst and Maree 2014). Immature sperm morphology is a result of not undergoing the final stages of spermiogenesis (last stage of spermatogenesis) and is consistent with our finding of significant enrichment of arrest of spermatogenesis (adjusted p-value = 2.55×10^{-5} , permutation p-value = 0.0090). We also found significant enrichment of abnormal male germ cell morphology (adjusted p-value = 2.53×10^{-7} , permutation p-value = 0.0035) and abnormal testis size (adjusted p-value = 3.04×10^{-5} , permutation p-value = 0.0079). The studies confirmed that the abnormalities were the result of decreased sperm competition and were most likely due to energy saving benefits since high quality sperm are energetically costly (van der Horst and Maree 2014). It has been hypothesized that in response to poorer sperm quality, females have evolved larger numbers of high quality oocytes that are capable of selecting for the best quality sperm (van der Horst and Maree 2014). Interestingly, two of the three pathways in the ovarian follicle cluster were related to ovarian follicle number. Finally, though not a comparative study, prairie voles, a model organism for studying pair bonding, were found to have increased levels of luteinizing hormone-releasing hormone (LHRH) and serum levels of luteinizing hormone (LH) in

response to contact with male urine, suggesting that LH may play an important role in activation of female reproduction in pair bonding mammals (Carter, Getz, and Cohen-Parsons 1986).

We also identified pathways related to drug response, which may be related to the importance of reward and motivation pathways in the evolution of pair bonding since pair bonding mammals must learn a preference for their mating partner. While I could not find a direct study comparing the neural mechanisms involved in pair bonding to those involved in response to drugs, there is a study comparing the neural response to cocaine vs. maternal nurturing in postpartum rats (Mattson and Morrell 2005). This is relevant because there are strong parallels in neural circuitry between mother-infant bonding and pair bonding which suggest the evolution of prolonged social bonds between mating partners built upon these same maternal-infant bonding mechanisms (Numan and Young 2016). The study found that a learned preference for cocaine or preference for pups resulted in increased activity compared to controls in the same, motivational processing-related brain regions including the prefrontal cortex, basal-lateral amygdala, and medial preoptic area, though at different levels between the cocaine and pup groups and most likely in only partially overlapping or non-overlapping sub populations of neurons within these brain regions (Mattson and Morrell 2005). Among the significantly enriched drug response pathways, we identified abnormal behavioral response to cocaine (adjusted p-value = 0.0151, permutation p-value = 0.0002). This suggests that similar mechanisms may be involved in processing maternal rewards and drug response rewards, and further that these may be similar to the reward processing mechanisms essential for formation of pair bonds.

While the studies mentioned identified crucial neural, endocrine, and morphological traits in pair bonding species, they did not identify the genes or genetic mechanisms involved. Our compelling pathway enrichment results are evidence that the association of protein coding regions with the evolution of pair bonding captured true signal in the data, and by looking at top ranked genes in these pathways, we can begin to improve our genetic understanding of the evolution of pair bonding. (See Future Work).

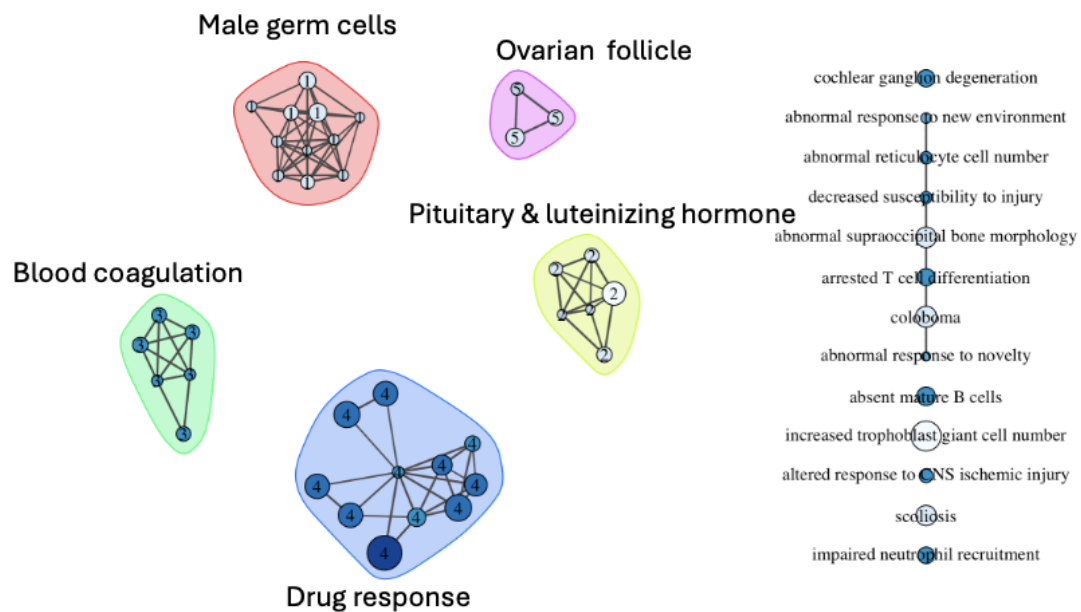


Figure 7. Clusters of pathways significantly enriched for genes associated with the evolution of pair bonding. Individual nodes on the right represent significantly enriched pathways that did not cluster.

Comparison with prior work

In addition to our new categorical RERconverge method we performed two types of binary RERconverge pairwise analyses (Kowalczyk et al. 2019), a categorical phylANOVA analysis using the RERs computed by RERconverge, and a delta statistic analysis (Borges et al. 2019) (**Table 1**). All of these methods were tested on the same set of 19,137 gene trees based on the reference phylogeny of 115 mammals whose phenotypes we could clearly assign (Hecker and Hiller 2020).

Overview of Methods Compared

Method	Works with discordant trees	Includes pairwise comparisons	P-values included (in package/GitHub available code)?	Approximate Runtime ^{1,2} (mins)
Categorical RERconverge	No	Yes	Yes, parametric and permutations	RER computation: 70 Analysis: 0.15 Permutations: 86 Total: 155
Pairwise Binary RERconverge (method I)	No	N/A	Yes, parametric and permutations	RER computation: 70 Analysis: 0.15 Permutations ⁴ : ~210 Total: ~ 280
Pairwise Binary RERconverge (method II)	No	N/A	Yes, parametric and permutations	RER computation: 120 Analysis: 0.75 Permutations ⁵ : ~360 Total: ~ 480
phylANOVA + base RERconverge (phylogenetic)	No ³	Yes	Yes, simulations only	RER computation: 70 Analysis: 114

simulations with Relative Evolutionary Rates) ³				Total: 184
Delta statistic of phylogenetic signal	Yes	No	No	Total: 103

Table 1. Comparison of the methods.

1. Includes time to run 500 permutations or 500 simulations with phylANOVA. Note: for this paper we ran 10,000 permutations/simulations. Does not include the time to perform a pathway enrichment analysis. Does not include the time to estimate trees from the MSA since this is universal to all the methods. The runtime for pairwise binary RERconverge method I was determined using all species with herbivores as foreground and the runtime for pairwise binary RERconverge method II was determined using carnivores/herbivores with carnivore foreground. Note that runtimes for permutations can vary depending on the specific structure of the phenotype.
2. Runtime is for 19,137 genes and a phylogeny with 115 species
3. The base RERconverge package was used to estimate relative evolutionary rates, these were then associated with the phenotype data using phylANOVA. While phylANOVA itself works with discordant trees, computing relative evolutionary rates with RERconverge does not work with discordant trees.
4. The runtime was 70 minutes for the binary phenotype with herbivore foreground. To determine the total runtime for all three tests (one for each category as foreground), this was multiplied by 3 to obtain ~210 minutes. This is approximate since each run of permutations varies depending on the structure of the phenotype.
5. The runtime was 60 minutes for the binary phenotype with carnivore foreground and herbivore background (omnivores removed). To determine the total runtime for all six tests, this was multiplied by 6 to obtain ~360 minutes. This is approximate since each run of permutations varies depending on the structure of the phenotype.

Comparison of ancestral reconstruction results

One important factor distinguishing these methods was the use or assignment of internal branches. PhylANOVA only includes extant species; by default binary RERconverge uses an approximate maximum parsimony approach to infer ancestral states; and based on our analysis above, we chose to apply categorical RERconverge using maximum likelihood with a continuous time Markov model (CTMM) (Pagel 1994; Pupko et al. 2000; Paradis, Claude, and Strimmer 2004) with an ARD rate model to infer ancestral states.

The default inference of ancestral states with the approximate maximum parsimony approach used for binary traits in RERconverge is heavily dependent on the choice of foreground. In the default mode, the trait is assumed to evolve only from background to foreground; hence, the method steps back through the

tree and assigns ancestral branches of foreground species to the foreground until all the descendants are no longer part of the foreground. As a result, this approach is also sensitive to excluding species. The impact of removing omnivores on the assignment of branches to the foreground using this method is illustrated on a subset of the full tree (**Fig. 2B** vs. **Fig. 2C**). When omnivores are removed, the common ancestor of the African hunting dog and Pacific walrus is assigned to the carnivore foreground. However, when omnivores are included, this ancestor is not assigned to the carnivore foreground because the lesser panda is an omnivore and thus not part of the foreground. The categorical reconstruction on the full tree predicts this ancestor is a carnivore, but regardless of the right classification, we take away from this observation the sensitivity of the default binary reconstruction to species presence or absence when choosing how to represent a multicategorical trait as multiple pairwise binary comparisons. The CTMM ancestral reconstruction is not dependent on a choice of foreground, as the ancestral state(s) versus the convergent state(s) are effectively inferred by the maximum likelihood assignment of states. Categorical trait reconstruction infers a carnivorous mammalian ancestor while the default binary reconstructions do not infer as many carnivorous internal branches (**Fig. 2A**). The exact extent of the effect of different inferred ancestral states on the gene and pathway enrichment results is most likely phenotype- and phylogeny-specific.

Comparison of gene results across methods

Association statistics between the evolution of the diet phenotype and RERs were computed for the three methods that use relative evolutionary rates: categorical RERconverge, pairwise binary RERconverge, and phylANOVA with RERs computed from RERconverge. The delta statistic method did not provide an approach for computing p-values so it was excluded from these analysis-wide gene result comparisons. 10,000 permutations were used to obtain permutation p-values for the binary and categorical RERconverge analyses. Similarly, 10,000 simulations were used to obtain simulation p-values for the phylANOVA analysis.

Quantile-quantile plots of the permutation and simulation p-values were made to determine which comparisons are driving the identification of significant genes and assess the signal across the different methods (**Fig. 8A-D**). PhylANOVA was significantly underpowered, with far fewer significant p-values in the QQ plots compared to all other methods (**Fig. 8A-D, gray**), although the same number of permutations and simulations were performed (**Fig. 8A-D, gray**). This lack of power in PhylANOVA could be due to lack of information on internal branches, an important feature included in RERconverge. In addition, it may also be due to differences in the phylogenetic simulations. In the ANOVA approach, RERs are treated as a continuously valued phenotype and simulated with a Brownian Motion model (Garland et al. 1993; Revell 2012). In contrast, the RERconverge approach allows for the use of phylogenetic permutations of the diet phenotypes (Saputra et al. 2021), and these permutations which have been shown to better correct for the tree structure. The categorical RERconverge update, which features methods to include internal branches and perform permutations, offers a significant improvement upon existing methods for comparing relative evolutionary rates across categories of species.

The binary and pairwise comparisons that measure differences between carnivores and herbivores tend to have the greatest signal, suggesting genetic differences between carnivores and herbivores are driving the

results (**Fig. 8A-C, indicated by +**). For instance, among the herbivore and omnivore comparisons, the binary comparison between herbivores in the foreground and carnivores/omnivores in the background has the strongest signal (**Fig. 8C, indicated +**). In contrast, both binary comparisons between herbivores and omnivores with carnivores removed do not show strong signal (**Fig. 8C, pink and purple**). Furthermore, the comparison between omnivores in the foreground and carnivores/herbivores in the background does not exhibit strong signal (**Fig. 8C, open green circle**). Thus, the key feature related to strong signal appears to be the separation of carnivores and herbivores between the foreground and background.

However, the exception to the trend of carnivore vs. herbivore as the most significant comparison is within the carnivore-omnivore signal comparison, where carnivore vs. omnivore have the most signal (**Fig. 8B**). The curves for these comparisons depart from the unity line early on, suggesting that the greater signal may be the result of binary permutations under-correcting for false positives due to failing to fully correct for the phylogenetic structure in the data rather than these methods being more powerful. The effect of removing all species of one category is often to have more full clades inferred together as foreground with fewer convergent gains or losses of the trait (**Fig.2**). This makes it more difficult to correct for phylogenetic relatedness driving the signal, especially since we used a relaxed version of binary permutations that does not enforce the permulated trees to exactly match the original phenotype tree in the structure of the relationships between foreground branches. Such differences in the inferred ancestral states due to removing herbivores may be contributing to a deceptively high degree of signal among these methods in other ways as well. These findings are consistent with our observations that internal branch assignments can have a large impact on genes' significance.

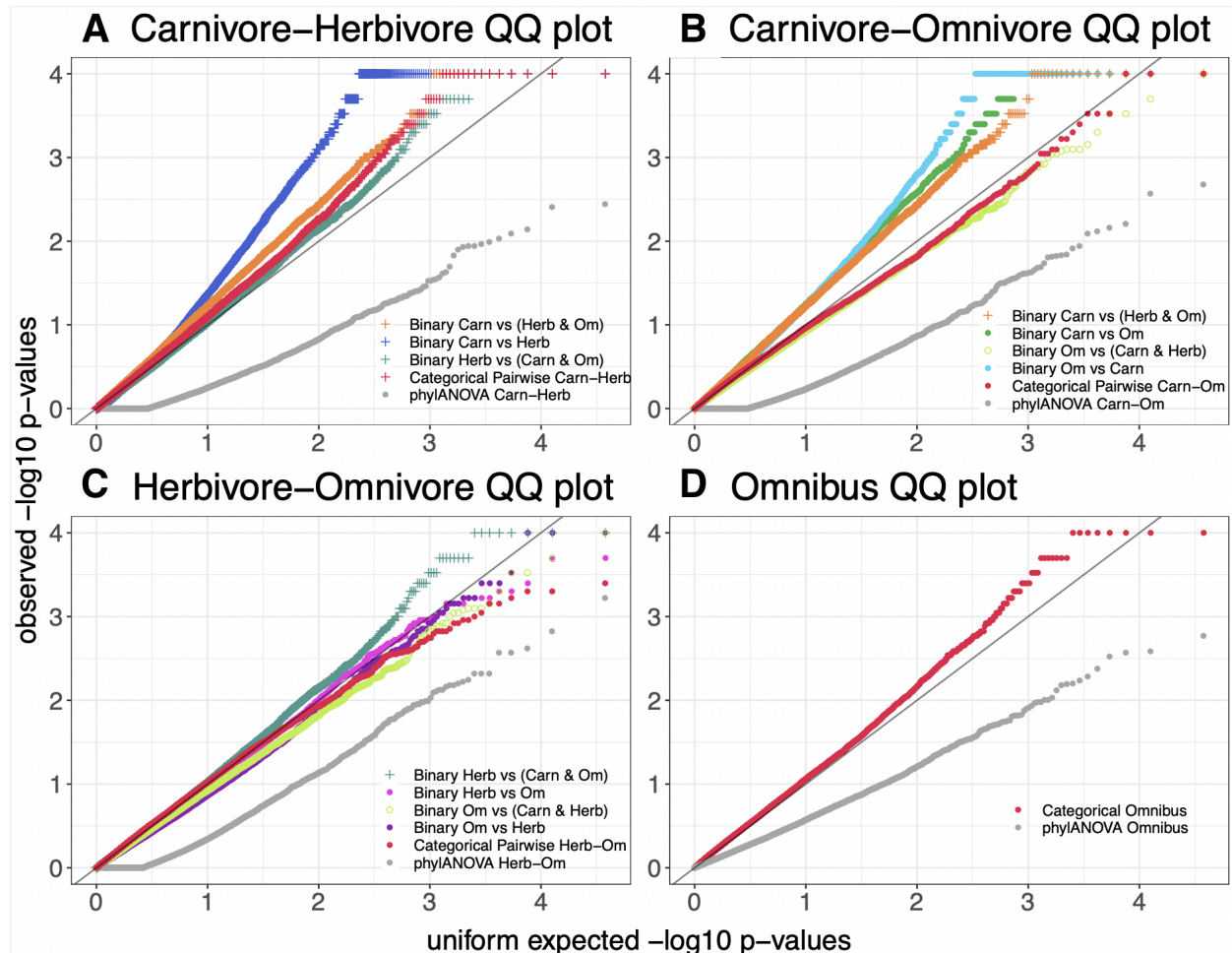


Figure 8. Quantile quantile (QQ) plots of the permutation or simulation p-values for the results of categorical RERconverge, phylANOVA, and pairwise binary RERconverge. A-C) A + symbol is used to denote any comparison in which carnivores and herbivores are separated between the foreground and background. C) An open circle symbol denotes the analysis in which carnivores and herbivores together form the background and omnivores form the foreground.

To further explore the impact of internal branches, we focus on RERs of two genes evolving significantly slower or faster among carnivores than omnivores according to the binary methods, but showing no significant rate shift according to the categorical method (**Fig. 9**). In both cases, whether a rate shift is observed or not is driven by differences in the RER distributions of the internal branches. MADCAM1 is significantly accelerated in omnivores according to the binary analysis with carnivore foreground ($p = 4.987 \times 10^{-7}$, adjusted $p = 0.0005558$, permutation $p = 0.0067$), however this is not the case in the categorical pairwise test ($p = 0.16$, adjusted $p = 1$, permutation $p = 0.2161$). Separating the relative evolutionary rates by extant vs. internal species, we see that in the categorical reconstruction, the RERs of carnivores are driven upwards due to the presence of internal branches with larger RERs assigned as carnivores (**Fig. 9**). PHB2 is significantly accelerated in carnivores according to the binary analysis with omnivore foreground ($p = 9.226 \times 10^{-5}$, adjusted $p = 0.01139$, permutation p -value = 0.0) but not in the categorical pairwise test ($p = 0.3461$, adjusted $p = 1$, permutation $p = 0.0776$). The signal for this gene in the binary analysis is being driven by a group of internal branches with negative RERs. These internal

branches are not assigned as omnivores in the categorical reconstruction. When this group is no longer present, there is no significant difference in RERs between carnivores and omnivores (**Fig. 9**).

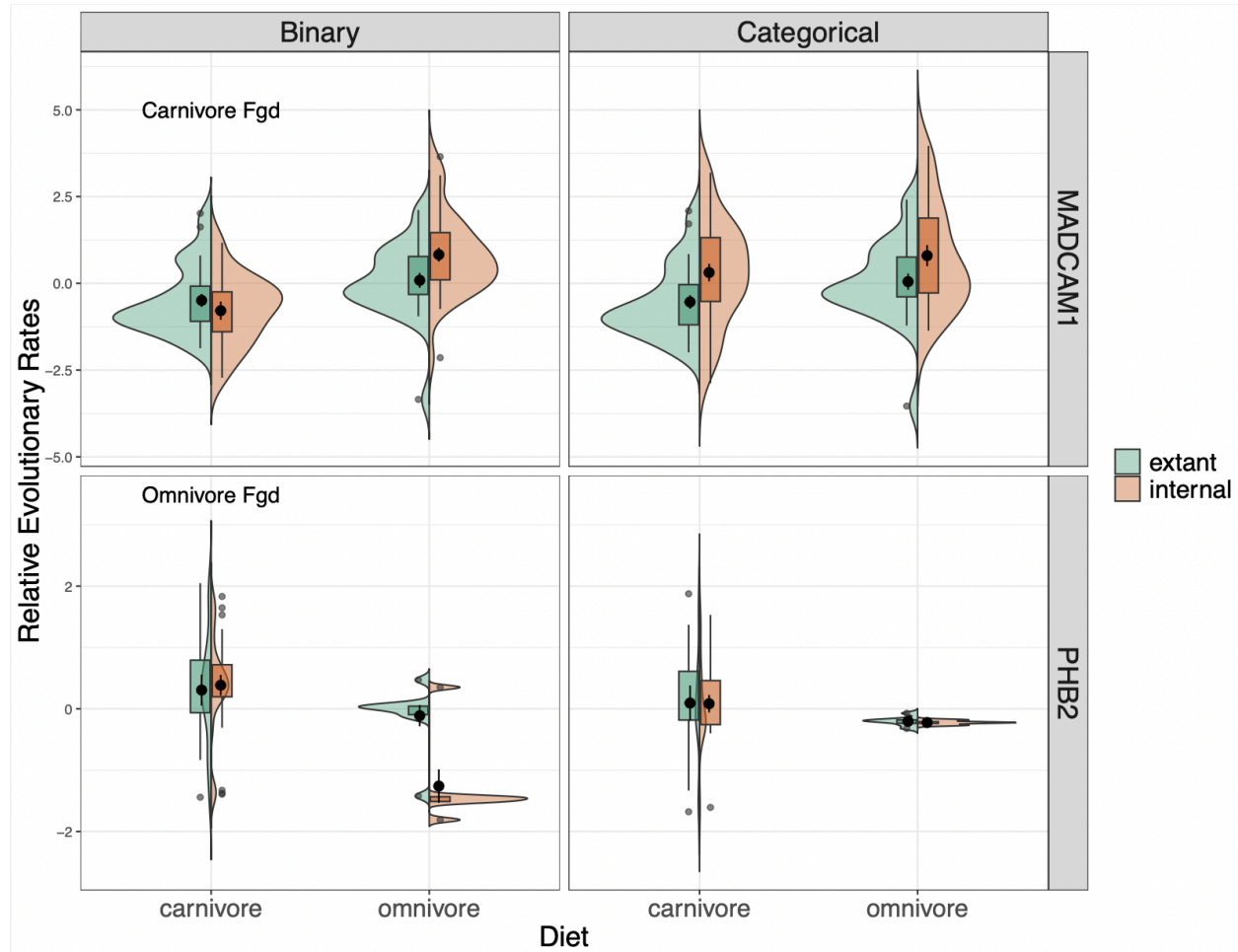


Figure 9. Each violin plot compares the distribution of RERs between carnivores and omnivores for MADCAM1 (top row) and PHB2 (bottom row). In green, on the left, is the distribution of RERs for extant branches and in orange, on the right, is the distribution of RERs for internal branches. The left column shows the RERs from the binary analyses of carnivores vs. omnivores with either carnivore foreground (top) or omnivore foreground (bottom). The right column contains the RERs from the categorical analysis.

Comparison of pathway enrichment results across methods

Effects of the ancestral reconstruction extend beyond the gene level results to the pathway enrichment results as well. One GO pathway for which we may expect to see significant enrichment is digestive tract development. Carnivores have been shown to have shortened digestive tracts (Kim et al. 2016) and we would expect the digestive tract to specialize to a species' diet. Digestive tract development was significantly enriched in the categorical pairwise test between carnivores and herbivores (adjusted $p = 0.044$, permutation $p = 0.0011$) (**Fig. 10A, blue line**). The four different binary pairwise analyses between carnivores and herbivores are also shown. Enrichment is lower among these four methods compared to the categorical method. Notably, the two methods with herbivore foreground (**Fig. 10A, purple and**

green) have greater enrichment than the method with carnivore foreground. This is consistent with the categorical ancestral reconstruction inferring more carnivorous ancestral species than herbivores so these binary ancestral reconstructions are more similar to the categorical ancestral reconstruction. This suggested that the ancestral reconstruction was playing an important part in the enrichment of this pathway.

To specifically test the effect of the ancestral reconstruction method relative to different statistical tests employed, we used the CTMC ancestral reconstruction method to infer the ancestral states for each of the four binary phenotypes. We then performed the rest of the binary analysis in the same way as before (**Fig. 10B**). Note that the two binary methods in which omnivores were removed (red and purple lines) now had the same assignment of ancestral states because the reconstruction is no longer dependent on foreground choice. Using the new ancestral reconstruction caused the enrichment of digestive tract development genes to improve, reaching nearly the same level as the categorical pairwise test. Thus, we infer the differences in enrichment were being driven almost entirely by the ancestral reconstruction. Using the more sophisticated ancestral reconstruction resulted in improved enrichment for a pathway which we would expect to be enriched, suggesting that capturing more complex patterns of evolution and using more reliable reconstructions is beneficial at the pathway enrichment level as well. The new ancestral reconstruction also appears more robust to removing species from the analysis because in the results from the new reconstruction, the red and purple lines, in which omnivores were removed, are much more similar to the orange and green lines respectively, in which omnivores were not removed, compared to the old reconstruction (**Fig. 10B** vs. **Fig. 10A**).

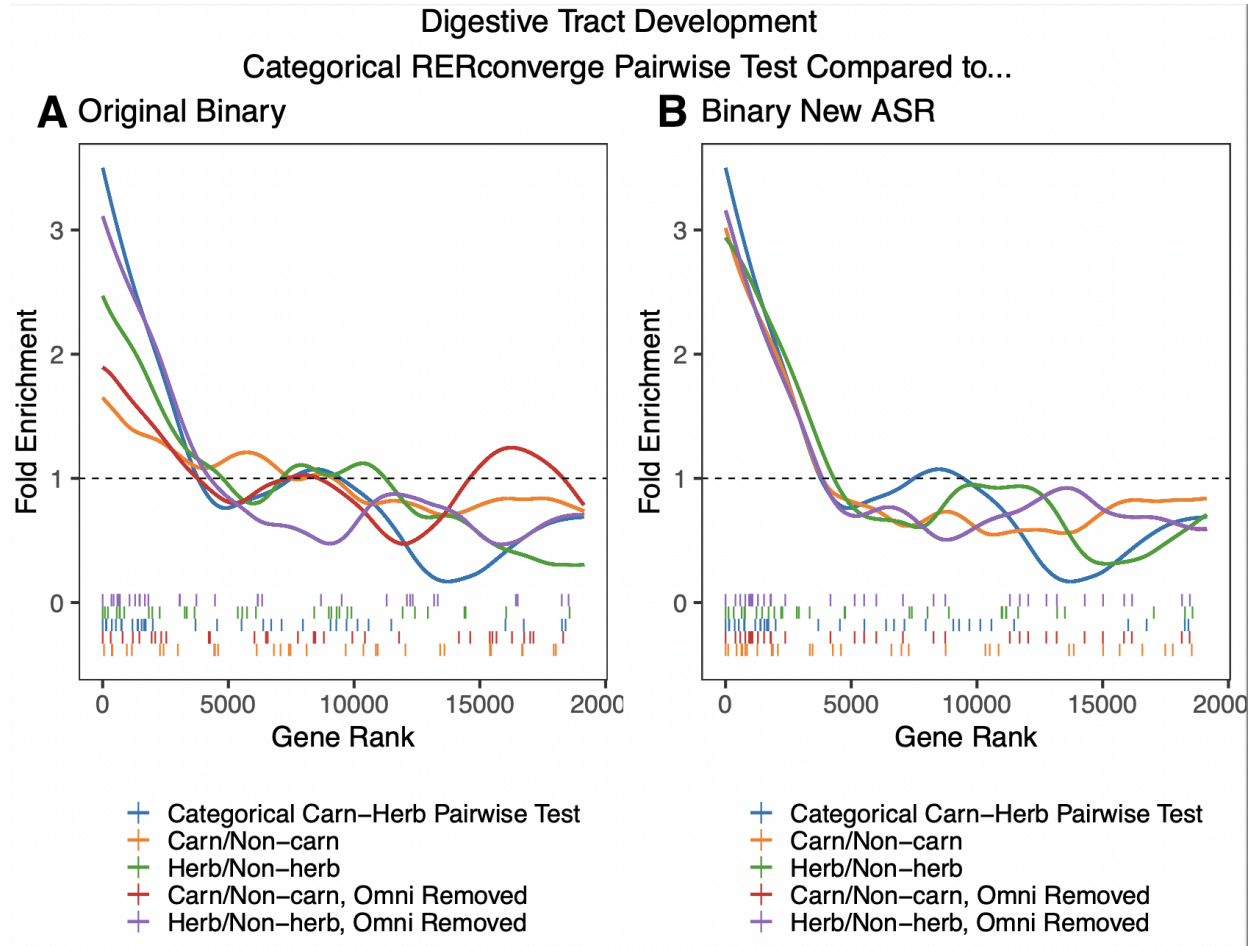


Figure 10. Fold enrichment and barcode plots for the digestive tract development pathway. A) The categorical carnivore-herbivore pairwise test (blue) is compared to the four different original binary analyses which compared carnivores to herbivores. B) The categorical carnivore-herbivore pairwise test (blue) is compared to the four binary RERconverge analyses in which the new ancestral state reconstruction (ASR) method, maximum likelihood applied to a continuous time Markov model, has been used in place of the original maximum parsimony approach.

Comparison to Delta Statistic

Having interpreted the updated RERconverge results, we next conducted a comparison of our method to the delta statistic. Based on its clear diet relevance and strong enrichment, we conducted the comparison on the genes within the abnormal digestion pathway (**Fig. 11**). We found significant enrichment of the categorical omnibus test and the delta statistic results but none of the pairwise or binary analyses (**Fig. 11, blue, orange**; categorical omnibus $p = 6.668 \times 10^{-5}$, adjusted $p = 0.1015$, permutation $p = 9 \times 10^{-4}$; delta statistic $p = 0.0002565$, adjusted $p = 0.05586$). This lack of enrichment remained even when performed with the CTMC ancestral reconstruction as was done with the digestive tract development pathway (**Fig. 10**). This suggests that the categorical test is capturing effects across multiple pairwise comparisons which alone aren't sufficient to detect enrichment for this pathway.

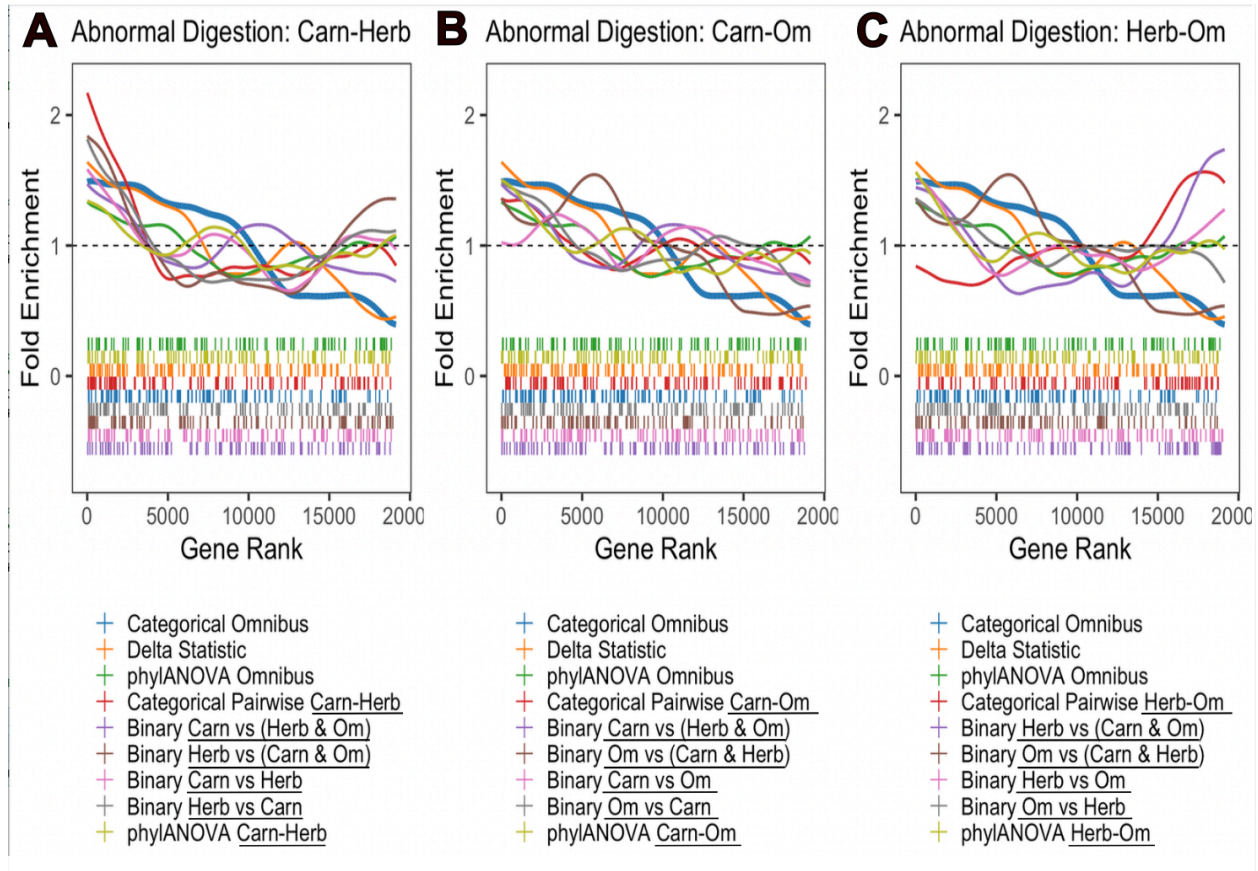


Figure 11. Fold enrichment and barcode plots for the abnormal digestion pathway. All three plots contain the categorical omnibus curve (blue, bold), delta statistic curve (orange), and phylANOVA omnibus curve (green). A) In addition to the omnibus methods, the methods comparing carnivores to herbivores are included. B) In addition to the omnibus methods, the methods comparing carnivores to omnivores are included. C) In addition to the omnibus methods, the methods comparing herbivores to omnivores are included. In A-C, the same color is used for *corresponding* methods. Where the corresponding methods may differ in which specific categories are being compared, the category names are underlined in the legend.

The delta statistic method is markedly different from the other methods in this analysis in that it does not use evolutionary rates. More details on how the delta statistic is calculated are given in the delta statistic paper (Borges et al. 2019; Ribeiro et al. 2023)). In order to determine if the delta statistic and categorical RERconverge identified similar or different signatures of convergent evolution, we first compared both methods on the abnormal digestion pathway, which was significantly enriched according to both. The most significant genes in the abnormal digestion pathway identified by the delta statistic were for the most part different from those identified by omnibus categorical RERconverge (**Fig. 12A**). Though there's no clear relationship between the delta statistic and the omnibus results, patterns emerge when the results are broken down into their pairwise comparisons. The most significant delta statistic genes tended to be those which were evolving slower in carnivores compared to omnivores or herbivores as indicated by negative statistics (**Fig. 12A**). Genes in the pathway which were ranked least significant according to the

delta statistic, but which exhibited a rate shift, tended to be those evolving faster in carnivores compared to herbivores or omnivores (**Fig. 12A**).

To determine if this trend persists across all genes in the analysis, not just those within the abnormal digestion pathway, we plotted the fold enrichment of the top 200 delta statistic genes among the categorical RERconverge tests (**Fig. 12B**). The general results indeed recapitulate what we observed within the abnormal digestion pathway. Genes with greater phylogenetic signal for the diet phenotype tend to evolve slower in carnivores compared to herbivores (**Fig. 12B, green**). To a lesser extent, the same trend is observed for carnivores compared to omnivores (**Fig. 12B, blue**) and herbivores compared to omnivores (**Fig. 12B, pink**). There is no enrichment of the top delta statistic genes among the omnibus results (**Fig. 12B, yellow**), indicating that overall the two methods are identifying different, though potentially complementary, top genes associated with this diet phenotype.

This trend may be the consequence of the delta statistic using reconstruction certainty to identify phylogenetic signal. Reconstruction certainty, among many factors, depends on inferred phylogeny branch lengths which are related to evolutionary rates. We may expect a phylogeny's ancestral history to be inferred with greater or less certainty when the branch lengths also exhibit a certain pattern as a result of allowing more or less change along the branches in order to give rise to the observed phenotype structure. In this case, slower evolutionary rates among carnivores may improve the certainty of inferring carnivorous ancestors corresponding to what we'd expect from our maximum likelihood reconstruction on the master tree, leading to higher delta statistics. However, it is unclear to what extent these patterns reflect the real biology or methodological uncertainty in the inferred branch lengths, gene tree structure, and ancestral likelihoods.

We interpret this to mean that the delta statistic is sometimes able to recognize features distinct from those captured by the evolutionary rates in RERconverge, however that the delta statistic may be more likely to fail to detect phylogenetic signal for certain genes depending on their patterns of convergent evolutionary rate shifts.

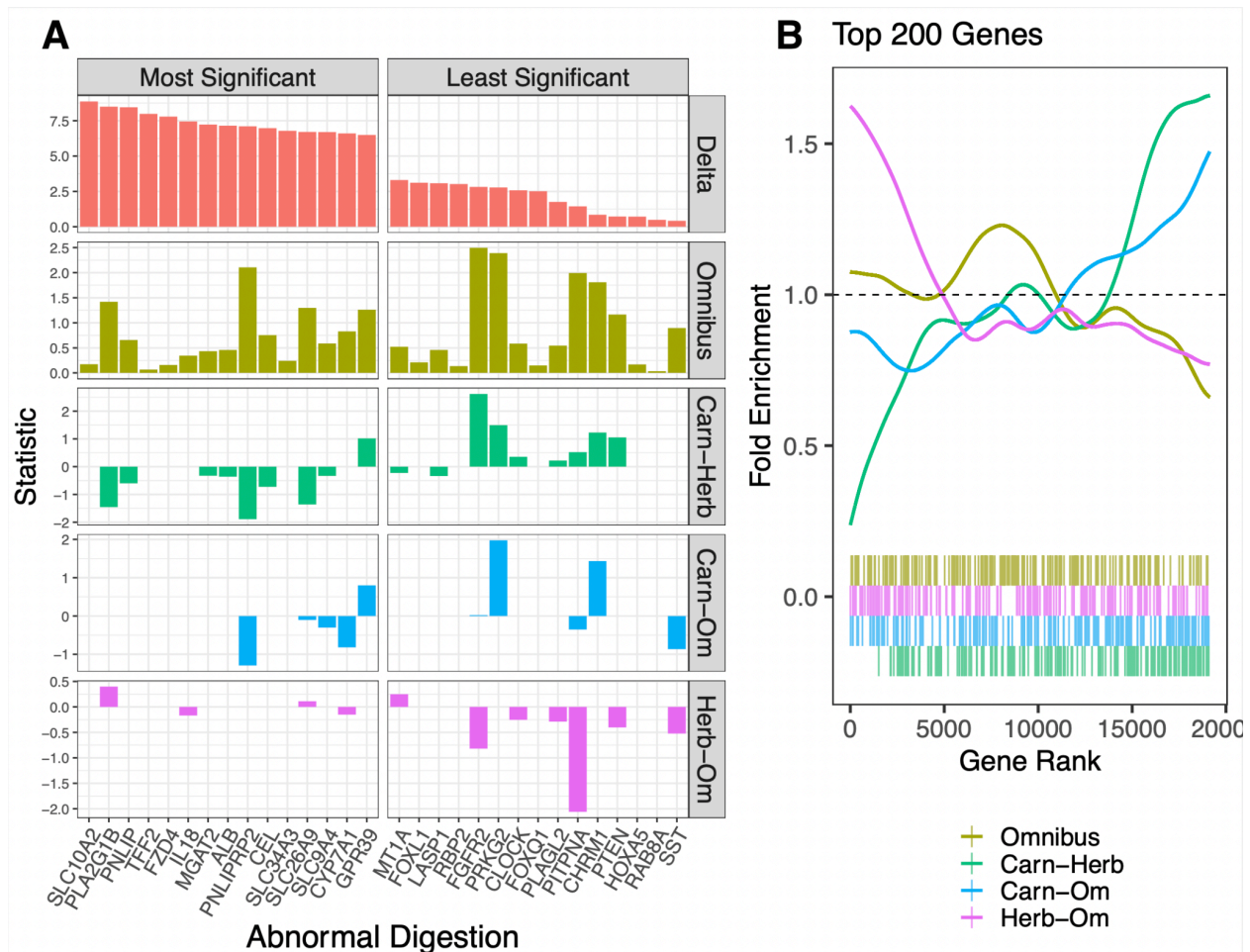


Figure 12. A) Bar plots of the statistic used to calculate enrichments for the top 15 most significant genes in the abnormal digestion pathway (right) and 15 least significant genes in the abnormal digestion pathway (left) as ranked by the delta statistic. For the delta statistic method, the statistic is delta, the measure of phylogenetic signal. For the RERconverge methods, the statistic is the signed log p-value. B) The fold enrichment and barcode plots among the RERconverge omnibus and pairwise tests of the top 200 genes ranked by the delta statistic of phylogenetic signal.

Future work

We submitted our work on the categorical expansion to RERconverge and its application to the diet phenotype to Molecular Biology and Evolution as a manuscript titled, *RERconverge Expansion: Using Relative Evolutionary Rates to Study Complex Categorical Trait Evolution* where it is currently under review. One important direction for future work is to continue investigating the pair bonding phenotype. We identified many relevant enriched biological pathways, so next steps would include analyzing the top ranking genes identified by RERconverge to improve our understanding of the genetic mechanisms underlying the evolution of pair bonding. For instance, we could apply tools from the HyPhy package to determine the types of selective pressure faced by the genes identified by the RERconverge analysis (Wirthlin et al. 2024).

We also believe that, in addition to protein coding genes, gene regulatory elements may play an important role in the evolution of pair bonding. Enhancers are gene regulatory elements that, when active, can influence gene expression levels. We can use the Tissue-Aware Conservation Toolkit (TACIT), a machine learning method developed in the Pfenning lab, to associate machine learning predictions of enhancer activity with a convergent phenotype (Kaplou et al. 2023). Currently, I am running the prediction-phenotype association step of the TACIT pipeline to identify enhancers that are related to pair bonding. If we are able to find such enhancers, we will also compare them to our protein coding RERconverge results to determine whether there are genes undergoing both coding sequence and regulatory changes.

Acknowledgements

Much of the work, figures, and writing for this thesis comes from our manuscript, *RERconverge Expansion: Using Relative Evolutionary Rates to Study Complex Categorical Trait Evolution* (Redlich et al., 2023). I would like to extend a huge thank you to all of my co-authors, especially my advisor, Dr. Andreas Pfenning, and mentors, Dr. Amanda Kowalczyk and Heather Sestili, and acknowledge all of their work and help that has contributed to this thesis: Amanda Kowalczyk, Michael Tene, Heather H. Sestili, Kathleen Foley, Elysia Saputra, Nathan Clark, Maria Chikina, Wynn K. Meyer, Andreas R. Pfenning.

References

- Bodawatta, Kasun H., Haofu Hu, Felix Schalk, Jan-Martin Daniel, Gibson Maiah, Bonny Koane, Bulisa Iova, Christine Beemelmans, Michael Poulsen, and Knud A. Jønsson. 2023. “Multiple Mutations in the Nav1.4 Sodium Channel of New Guinean Toxic Birds Provide Autoresistance to Deadly Batrachotoxin.” *Molecular Ecology*, February. <https://doi.org/10.1111/mec.16878>.
- Borges, Rui, João Paulo Machado, Cidália Gomes, Ana Paula Rocha, and Agostinho Antunes. 2019. “Measuring Phylogenetic Signal between Categorical Traits and Phylogenies.” *Bioinformatics* 35 (11): 1862–69.
- Carter, C. Sue, Lowell L. Getz, and Martha Cohen-Parsons. 1986. “Relationships between Social Organization and Behavioral Endocrinology in a Monogamous Mammal.” In *Advances in the Study of Behavior*, edited by Jay S. Rosenblatt, Colin Beer, Marie-Claire Busnel, and Peter J. B. Slater, 16:109–45. Academic Press.
- Chikina, Maria, Joseph D. Robinson, and Nathan L. Clark. 2016. “Hundreds of Genes Experienced Convergent Shifts in Selective Pressure in Marine Mammals.” *Molecular Biology and Evolution* 33 (9): 2182–92.
- Espindola-Hernandez, Pamela, Jakob C. Mueller, and Bart Kempnaers. 2022. “Genomic Signatures of the Evolution of a Diurnal Lifestyle in Strigiformes.” *G3 Genes|Genomes|Genetics* 12 (8). <https://doi.org/10.1093/G3JOURNAL/JKAC135>.
- Foundation for Statistical Computing, R. R. n.d. “R: A Language and Environment for Statistical Computing.” *RA Lang Environ Stat Comput*.
- Fukushima, Kenji, and David D. Pollock. 2023. “Detecting Macroevolutionary Genotype-Phenotype Associations Using Error-Corrected Rates of Protein Convergence.” *Nature Ecology & Evolution* 7 (1): 155–70.
- Garland, Theodore, Allan W. Dickerman, Christine M. Janis, and Jason A. Jones. 1993. “Phylogenetic Analysis of Covariance by Computer Simulation.” *Systematic Biology* 42 (3): 265–92.
- Hasselmann, Martin, Xavier Vekemans, Jochen Pflugfelder, Nikolaus Koeniger, Gudrun Koeniger, Salim Tingek, and Martin Beye. 2008. “Evidence for Convergent Nucleotide Evolution and High Allelic

- Turnover Rates at the Complementary Sex Determiner Gene of Western and Asian Honeybees.” *Molecular Biology and Evolution* 25 (4): 696–708.
- Hecker, Nikolai, and Michael Hiller. 2020. “A Genome Alignment of 120 Mammals Highlights Ultraconserved Element Variability and Placenta-Associated Enhancers.” *GigaScience* 9 (1). <https://doi.org/10.1093/gigascience/giz159>.
- He, Jing, Xiaotian Xie, Zhuanglong Xiao, Wei Qian, Lei Zhang, and Xiaohua Hou. 2023. “Piezo1 in Digestive System Function and Dysfunction.” *International Journal of Molecular Sciences* 24 (16). <https://doi.org/10.3390/ijms241612953>.
- Horst, Gerhard van der, and Liana Maree. 2014. “Sperm Form and Function in the Absence of Sperm Competition.” *Molecular Reproduction and Development* 81 (3): 204–16.
- Hu, Zhirui, Timothy B. Sackton, Scott V. Edwards, and Jun S. Liu. 2019. “Bayesian Detection of Convergent Rate Changes of Conserved Noncoding Elements on Phylogenetic Trees.” *Molecular Biology and Evolution* 36 (5): 1086–1100.
- Jin, Pengyu, Bingyue Zhu, Yinjun Jia, Yiming Zhang, Wei Wang, Yunxiao Shen, Yu Zhong, et al. 2023. “Single-Cell Transcriptomics Reveals the Brain Evolution of Web-Building Spiders.” *Nature Ecology & Evolution*, November. <https://doi.org/10.1038/s41559-023-02238-y>.
- Kaplow, Irene M., Alyssa J. Lawler, Daniel E. Schäffer, Chaitanya Srinivasan, Heather H. Sestili, Morgan E. Wirthlin, Badoi N. Phan, et al. 2023. “Relating Enhancer Genetic Variation across Mammals to Complex Phenotypes Using Machine Learning.” *Science* 380 (6643): eabm7993.
- Kim, Soonok, Yun Sung Cho, Hak-Min Kim, Oksung Chung, Hyunho Kim, Sungwoong Jho, Hong Seomun, et al. 2016. “Comparison of Carnivore, Omnivore, and Herbivore Mammalian Genomes with a New Leopard Assembly.” *Genome Biology* 17 (1): 211.
- King, Bruce M., Patrick J. Rosopa, and Edward W. Minium. 2018. *Statistical Reasoning in the Behavioral Sciences*. John Wiley & Sons.
- Kosakovsky Pond, Sergei L., Art F. Y. Poon, Ryan Velazquez, Steven Weaver, N. Lance Hepler, Ben Murrell, Stephen D. Shank, et al. 2020. “HyPhy 2.5-A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies.” *Molecular Biology and Evolution* 37 (1): 295–99.
- Kowalczyk, Amanda, Maria Chikina, and Nathan Clark. 2022. “Complementary Evolution of Coding and Noncoding Sequence Underlies Mammalian Hairlessness.” *eLife* 11 (November). <https://doi.org/10.7554/eLife.76911>.
- Kowalczyk, Amanda, Wynn K. Meyer, Raghavendran Partha, Weiguang Mao, Nathan L. Clark, and Maria Chikina. 2019. “RERconverge: An R Package for Associating Evolutionary Rates with Convergent Traits.” *Bioinformatics* 35 (22): 4815–17.
- Kowalczyk, Amanda, Raghavendran Partha, Nathan L. Clark, and Maria Chikina. 2020. “Pan-Mammalian Analysis of Molecular Constraints Underlying Extended Lifespan.” *eLife* 9 (February). <https://doi.org/10.7554/eLife.51089>.
- Liberzon, Arthur, Aravind Subramanian, Reid Pinchback, Helga Thorvaldsdóttir, Pablo Tamayo, and Jill P. Mesirov. 2011. “Molecular Signatures Database (MSigDB) 3.0.” *Bioinformatics* 27 (12): 1739–40.
- Li, Ying, Zhen Liu, Peng Shi, and Jianzhi Zhang. 2010. “The Hearing Gene Prestin Unites Echolocating Bats and Whales.” *Current Biology: CB* 20 (2): R55–56.
- Louca, Stilianos, and Michael Doebeli. 2018. “Efficient Comparative Phylogenetics on Large Trees.” *Bioinformatics* 34 (6): 1053–55.
- Lukas, D., and T. H. Clutton-Brock. 2013. “The Evolution of Social Monogamy in Mammals.” *Science* 341 (6145): 526–30.
- Mattson, B. J., and J. I. Morrell. 2005. “Preference for Cocaine- versus Pup-Associated Cues Differentially Activates Neurons Expressing Either Fos or Cocaine- and Amphetamine-Regulated Transcript in Lactating, Maternal Rodents.” *Neuroscience* 135 (2): 315–28.
- Nowak, Ronald M. 1999. *Walker’s Mammals of the World*. Johns Hopkins University Press.
- Numan, Michael, and Larry J. Young. 2016. “Neural Mechanisms of Mother-Infant Bonding and Pair Bonding: Similarities, Differences, and Broader Implications.” *Hormones and Behavior* 77

- (January): 98–112.
- Ogle, D. H., J. C. Doll, A. P. Wheeler, and A. Dinno. 2023. “FSA: Simple Fisheries Stock Assessment Methods. R Package Version 0.9. 4.” FSA: Tokyo, Japan.
- O’Leary, Nuala A., Mathew W. Wright, J. Rodney Brister, Stacy Ciufu, Diana Haddad, Rich McVeigh, Bhanu Rajput, et al. 2016. “Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation.” *Nucleic Acids Research* 44 (D1): D733–45.
- Pagel, Mark. 1994. “Detecting Correlated Evolution on Phylogenies: A General Method for the Comparative Analysis of Discrete Characters.” *Proceedings of the Royal Society of London. Series B: Biological Sciences* 255 (1342): 37–45.
- Paradis, Emmanuel, Julien Claude, and Korbinian Strimmer. 2004. “APE: Analyses of Phylogenetics and Evolution in R Language.” *Bioinformatics* 20 (2): 289–90.
- Paradis, Strimmer, Claude, and Jobb. n.d. “The Ape Package.” *Of Phylogenetics ...*. https://math.mcmaster.ca/~bolker/eid/2009/Evolution/2009_Day1_phylogenetics/R_phylo_resource/s/ape.pdf.
- Partha, Raghavendran, Bhareesh K. Chauhan, Zelia Ferreira, Joseph D. Robinson, Kira Lathrop, Ken K. Nischal, Maria Chikina, and Nathan L. Clark. 2017. “Subterranean Mammals Show Convergent Regression in Ocular Genes and Enhancers, along with Adaptation to Tunneling.” *eLife* 6 (October). <https://doi.org/10.7554/eLife.25884>.
- Pennell, Matthew W., Jonathan M. Eastman, Graham J. Slater, Joseph W. Brown, Josef C. Uyeda, Richard G. FitzJohn, Michael E. Alfaro, and Luke J. Harmon. 2014. “Geiger v2.0: An Expanded Suite of Methods for Fitting Macroevolutionary Models to Phylogenetic Trees.” *Bioinformatics* 30 (15): 2216–18.
- Pond, Sergei L. Kosakovsky, Simon D. W. Frost, and Spencer V. Muse. 2005. “HyPhy: Hypothesis Testing Using Phylogenies.” *Bioinformatics* 21 (5): 676–79.
- Pupko, T., I. Pe’er, R. Shamir, and D. Graur. 2000. “A Fast Algorithm for Joint Reconstruction of Ancestral Amino Acid Sequences.” *Molecular Biology and Evolution* 17 (6): 890–96.
- Revell, Liam J. 2012. “Phytools: An R Package for Phylogenetic Comparative Biology (and Other Things).” *Methods in Ecology and Evolution / British Ecological Society* 3 (2): 217–23.
- Rey, Carine, Laurent Guéguen, Marie Sémon, and Bastien Boussau. 2018. “Accurate Detection of Convergent Amino-Acid Evolution with PCOC.” *Molecular Biology and Evolution* 35 (9): 2296–2306.
- Ribeiro, Diogo, Rui Borges, Ana Paula Rocha, and Agostinho Antunes. 2023. “Testing Phylogenetic Signal with Categorical Traits and Tree Uncertainty.” *Bioinformatics* 39 (7). <https://doi.org/10.1093/bioinformatics/btad433>.
- “rs762530150.” 2022. https://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?type=rs&rs=rs762530150.
- Saputra, Elysia, Amanda Kowalczyk, Luisa Cusick, Nathan Clark, and Maria Chikina. 2021. “Phylogenetic Permutations: A Statistically Rigorous Approach to Measure Confidence in Associations in a Phylogenetic Context.” *Molecular Biology and Evolution* 38 (7): 3004–21.
- Subramanian, Tamayo, and Others. 2005. “Molecular Signatures Database (MSigDB).” *Proceedings of the National Academy of Sciences of the United States of America* 102: 15545–50.
- Surhone, Lambert M., Mariam T. Tennoe, and Susan F. Henssonow, eds. 2010. *Uniprot*. Betascript Publishing.
- Thomas, Gregg W. C., and Matthew W. Hahn. 2015. “Determining the Null Model for Detecting Adaptive Convergence from Genomic Data: A Case Study Using Echolocating Mammals.” *Molecular Biology and Evolution* 32 (5): 1232–36.
- Wang, Kai, Shilin Tian, Jorge Galindo-González, Liliana M. Dávalos, Yuzhi Zhang, and Huabin Zhao. 2020. “Molecular Adaptation and Convergent Evolution of Frugivory in Old World and Neotropical Fruit Bats.” *Molecular Ecology*, July, mec.15542.
- Wirthlin, Morgan E., Tobias A. Schmid, Julie E. Elie, Xiaomeng Zhang, Amanda Kowalczyk, Ruby Redlich, Varvara A. Shvareva, et al. 2024. “Vocal Learning-Associated Convergent Evolution in Mammalian Proteins and Regulatory Elements.” *Science* 383 (6690): eabn3263.

- Yang, Ziheng. 2007. "PAML 4: Phylogenetic Analysis by Maximum Likelihood." *Molecular Biology and Evolution* 24 (8): 1586–91.
- Yan, Han, Zhirui Hu, Gregg W. C. Thomas, Scott V. Edwards, Timothy B. Sackton, and Jun S. Liu. 2023. "PhyloAcc-GT: A Bayesian Method for Inferring Patterns of Substitution Rate Shifts on Targeted Lineages Accounting for Gene Tree Discordance." *Molecular Biology and Evolution* 40 (9). <https://doi.org/10.1093/molbev/msad195>.
- Yusuf, Leeban H., Yolitzí Saldívar Lemus, Peter Thorpe, Constantino Macías García, and Michael G. Ritchie. 2023. "Genomic Signatures Associated with Transitions to Viviparity in Cyprinodontiformes." *Molecular Biology and Evolution* 40 (10). <https://doi.org/10.1093/molbev/msad208>.
- Zoonomia Consortium. 2020. "A Comparative Genomics Multitool for Scientific Discovery and Conservation." *Nature* 587 (7833): 240–45.