Evolution-based strategies to elucidate genotype-phenotype relationships in traits relevant to human health

by

Amanda Kowalczyk

B.S., Westminster College, 2016

Submitted to the Graduate Faculty of the

School of Medicine in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Amanda Kowalczyk

It was defended on

May 17, 2021

and approved by

Maria Chikina, Assistant Professor, Department of Computational and Systems Biology, University of Pittsburgh

Dennis Kostka, Associate Professor, Department of Developmental Biology, University of Pittsburgh

Andreas Pfenning, Assistant Professor, Computational Biology Department, Carnegie Mellon University

Leah Byrne, Assistant Professor, Department of Ophthalmology, University of Pittsburgh

Dissertation Director: Nathan Clark, Associate Professor, Department of Human Genetics, University of Utah

Evolution-based strategies to elucidate genotype-phenotype relationships in traits relevant to human health

Amanda Kowalczyk, PhD

University of Pittsburgh, 2021

Understanding the link between genotype and phenotype is a key question in biological research. In other words, how do changes in DNA sequence give rise to the enormous diversity of life on Earth? In this work, we address that question by focusing on the genomics of convergent phenotypes, or phenotypes that have evolved independently in unrelated species. These natural biological replicates of phenotype evolution allow us to identify genomic regions, either regulatory or protein coding, that have experienced convergent evolution in concordance with convergent evolution of phenotypes, thus linking genomic regions to phenotypes.

In this work, we calculate evolutionary rates throughout the mammalian phylogeny for numerous genomic sequences to find concordance between species phenotype and evolutionary rate of sequences to link genomic regions to phenotypes. Chapter one describes three methods associated with quantifying the connection between genomic region evolution and phenotype evolution. First, RERconverge connects genomic regions to phenotypes in a linear regression-based framework. Second, permulations are a statistical extension to RERconverge that allow for rigorous calculation of confidence in associations from RERconverge. Third, proper implementation of branch-site models for convergent positive selection allows for identification of genes potentially driving convergent evolution.

Chapter two describes implementation of methods from chapter one to longevity phenotypes in 61 mammal species. We found increased evolutionary constraint in cancer control

genes in large, long-lived species, thus likely conferring additional protection from cancer. Species exceptionally long-lived given their size showed increased evolutionary constraint on DNA repair pathways, indicating that efficient DNA repair is important to evolution of extreme lifespan independent of body size. This work provided insight into pan-mammalian genomic mechanisms underlying lifespan.

Chapter three describes further implementation of chapter one methods to the hairlessness phenotype in 61 mammal species. Although all mammals have some hair at some developmental time point, several mammals, such as cetaceans, naked mole-rats, armadillos, and humans, have relatively little hair. Many genomic elements we identified were known to be hair-related, and many more are valuable candidates for further testing into hair-related functions. This work for the first time provided insights to the natural evolution of mammalian hairlessness.

Table of Contents

List of Tables

List of Figures

Preface

As I construct this document and reflect on my time as a graduate student, I cannot help but feel overwhelmed with gratitude that I have had the privilege of going through this experience. Pursuing higher education is an opportunity that most people in the world will never be able to experience, and I am so grateful that I have not only been able to pursue my dream of earning my PhD in the sciences, but also that I have been able to help others along the way.

I would like to thank my advisors, Nathan Clark and Maria Chikina, for supporting and guiding me from the moment I dropped on their doorstep as an undergraduate researcher until now as I work through the completion of my doctoral research. I would truly have been lost without them. I have presented some wild ideas to them over the years, and I have often strayed far from the traditional path of a PhD student, but they have always stuck with me. I hope that I have made them proud and been able to do my part to contribute to the continued success of their labs.

I would further like to thank Clark and Chikina lab members for supporting me and helping me to become a better scientist. A huge thanks to Wynn Meyer for being a constant sounding board for science, life, career, and beyond. Thank you to my other lab mates in Pittsburgh, Raghavendran Partha, Elysia Saputra, Melissa Plakke, and Wayne Mao. Thanks to my new lab mates in Utah, Allie Graham, Jason Presnell, Jarret Lieberth, Sarah Lucas, and Jordan Little who open my eyes to new areas of inquiry every day and who I hope to eventually meet in person!

Thank you to my committee, Drs. Andreas Pfenning, Dennis Kostka, and Leah Byrne for making my life easier by always being available to answer questions, attend meetings, and provide

invaluable feedback to improve my research. I would also like to thank Dr. Irene Kaplow from the Pfenning lab for excellent discussion and collaborative work on methods development pursuits.

A huge thank you to all the supports of my outreach work, including the American Society for Cell Biology, the Kiwanis Club of Greensburg, the Delmont Lions Club, the CPCB PhD program and the ISB PhD program who sponsored events. Thank you to my team, Adam Watkins, William Dion, and Mische Holland, who got on board to help coordinate events for the 2020-2021 school year.

Finally, I would remiss if I did not thank my friends in the CPCB program and especially Trevor Frisby, Cathy Su, Laura Tung, and Emilee Holtzapple for keeping me company and keeping my spirits up. You all make my life a brighter place, and I don't know if I would have made it through without you all.

Much of the work in my three thesis chapters has been published, and the remainder is in preparation to be published. RERconverge was published at (Amanda Kowalczyk, Partha, Clark, & Chikina, 2020), the permulation method was published at (Saputra, Kowalczyk, Cusick, Clark, & Chikina, 2021), and the branch-site model work is in preparation for submission at *MBE*. The longevity work was published at (Amanda Kowalczyk et al., 2020) and the hairlessness work is in preparation for publication with a preprint available here (Amanda Kowalczyk, Clark, & Chikina, 2021).

1.0 Introduction

Planet Earth is an incredibly diverse ecosystem, with habitats ranging from rainforests to deserts to tundra to grasslands to oceans and everything in between. In part because of diversity of habitats, life on Earth also exhibits a wide array of characteristics, and species have adapted to colonize niches worldwide. While diversity of life may not be particularly surprising given that species are uniquely adapted to survive and thrive in their respective environments and under their respective set of selective pressures, what is perhaps more surprising is when we observe a lack of diversity, or when unrelated species independently acquire similar characteristics. The uncommon instance of unrelated species evolving similar phenotypes is termed convergent evolution. Classic examples of convergent evolution include birds and bats independently developing the ability to fly, species that inhabit caves and subterranean environments losing the ability to see, and three groups of marine mammals (pinnipeds i.e. seals and walruses, cetaceans i.e. whales and dolphins, and sirenians i.e. manatees and dugongs) acquiring characteristics such as smooth skin, blubber, and increased lung capacity to allow them to thrive in the ocean.

While relatively rare, convergent evolution is a highly useful tool to study evolution in general because it provides natural biological replicates of the evolution of often complex phenotypes. If species show convergent evolution of a phenotype, they may also show convergent evolution in genes and regulatory regions associated with the phenotype. By seeking concordant patterns of phenotypic and genetic convergence, we can thereby link genetic elements to phenotypes to both assign functions to those elements and to better understand how phenotypes evolve.

Convergent evolution in genomic sequence can be measured at a wide range of scales, from convergence at individual nucleotides to convergence of whole networks of proteins with similar functions. For very distantly related species that evolve convergent phenotypes, identical changes in specific nucleotide or amino acid positions to give rise to the same phenotype are rare because those sequences experienced a long period of divergence prior to potential convergent changes. In distantly related species, the relative position of sequence on chromosomes, the three-dimensional configuration of the DNA, and the entire network of genes and regulatory machinery evolved distinctly, and thus it is unlikely that changing a single nucleotide or amino acid in different species would give rise to the same phenotypic change. Instead, an alternative approach to studying convergent evolution in genomic sequence is to consider how much change has occurred in sequence during the development of a convergent phenotype. Measured as the number of substitutions that occurred in sequence between a species and its ancestor, the amount of genomic change is also called the evolutionary rate, and it serves as a proxy for the level of selective pressure on the sequence. When a particular sequence demonstrates convergent selective pressure shifts in tandem with convergent phenotypic evolution, that suggests that the sequence may be related to evolution of the phenotype.

Even for phenotypes that do not demonstrate convergent evolution of individual sequence elements, such as genes or regulatory elements, species that exhibit the phenotype may experience genomic convergence at the level of whole networks of elements that work together to give rise to phenotypes. These networks may take the form of gene pathways, networks of regulatory elements, or even complex sequence interactions that remain unknown to the scientific community. By investigating high-level convergence of functionally related sequence, we can better understand how convergent phenotypes arise at a broad scale in addition to a fine-grained

view of individual sequences. Analysis of multiple possible scales of convergence, as well as incorporating information about both regulatory and gene sequence convergence, allows for a holistic view of how convergent traits evolve.

In this dissertation, I first describe methods built, improved, and tested to link genetic elements to phenotypes in the context of convergent evolution. I then demonstrate the use of such methods to study evolution of two health-related phenotypes across the mammalian phylogeny.

In chapter 1, I describe three complementary methods to study convergent evolution in a phylogenetic context. First, RERconverge is a method to link convergently evolving genomic elements to phenotypes, and it serves as the flagship method for analyses in subsequent chapters. Second, an accessory method termed permulations corrects statistical anomalies often observed when performing studies in a phylogenetic context, including when using RERconverge. Finally, I include a note emphasizing and demonstrating the importance of correctly using branch-site models for positive selection to test for convergent positive selection.

In chapter 2, I use my methods to study the evolution of extreme longevity in mammals. My work is the first to use RERconverge to analyze a continuous phenotype, and it also represents first use of permulations in phylogenetic analyses. I used a unique metric to study longevity that incorporated body size and maximum lifespan into two principal components, thereby resolving potential confounding between the two variables. Through my analyses, I found that cancer control mechanisms, such as cell cycle control, are essential to evolution of large body size and long life. On the other hand, DNA repair mechanisms appeared more important for long life independent of body size. These results largely agree with results from previous studies in individual species, and my work for the first time indicates that longevity mechanisms are shared throughout all mammals.

In chapter 3, I further apply my methods to the hairlessness phenotypes in mammals. Although all mammals have at least some hair, several mammals, such as whales, naked mole-rats, elephants, and rhinos, have notably less hair. This work allowed me to extend my permulation method to use in binary phenotypes, and it served as the first instance of using RERconverge on phenotype change not necessarily driven by the same selective pressure. I was able to identify a suite of genes, noncoding regions, and microRNAs associated with convergent hair loss. Interestingly, I observed a striking difference in patterns of convergence between genes and their associated noncoding regions, supporting the hypothesis that complementary evolution of genes and regulatory elements drives phenotypic evolution. Uncharacterized noncoding regions and microRNAs are valuable candidates for further testing in hair-related functions.

My broader impacts section describes outreach work I have conducted while completing my thesis research. I target local schools in a relatively impoverished area for various programs to show them that a career in science is possible for them. Programs include field trips to visit research laboratories, guest lectures to classrooms, mentorship through job shadows and guided research projects, and a yearly science fair. The programs have grown over the years and now encompass numerous schools under the header of the Greensburg Area Science Program (GASP) that I founded.

2.0 Computational methods to study phenotype evolution in a phylogenetic context

Analyzing numerous genome alignments in a statistically rigorous way requires innovative and thoughtful computational tools. The focus of the work in this chapter is to create tools to connect genetic elements to phenotypes in a phylogenetic context using genome alignment data. In its three sections, this chapter describes RERconverge, a tool to link genetic elements to phenotypes, permulations, a statistical method to accurately quantify confidence in associations from RERconverge, and proper use of branch-site models for positive selection to detect convergent positive selection. Used together, these tools provide a full view of connections between convergent evolution of genetic elements and phenotypes.

2.1 RERconverge: a computational method to connect convergently evolving genomic elements to convergent phenotypes

RERconverge is a rigorous statistical method to link convergently evolving genetic elements and phenotypes in a phylogenetic context. Designed to be user-friendly even for researchers not familiar with computational work, RERconverge allows for either binary or continuous trait input and includes all functions and instructions to proceed from alignments of sequences of interest to final statistics and visualizations.

RERconverge was constructed by a team of Clark and Chikina lab members (myself, Wynn Meyer, Raghavendran Partha, and Weiguang Mao). I was the key contributor responsible for

implementing, benchmarking, and testing methods for continuous phenotypes, implementing and testing pathway enrichment functions, and creating vignette walkthroughs for continuous trait and enrichment analyses. I additionally served as the primary point of contact for users after RERconverge was published. The paper describing RERconverge was originally published in *Bioinformatics* at (Amanda Kowalczyk et al., 2019).

2.1.1 Introduction

A major motivation in evolutionary biology is to determine which genetic changes underlie phenotypic adaptations. Convergent evolution, in which the same phenotype arises independently in distinct evolutionary lineages, provides natural replicates of phenotypic adaptation that aid researchers in linking phenotypes to their underlying genetic changes. Selection repeatedly targets the same genes in several known cases of phenotypic convergence, including *Prestin* in echolocation in bats and marine mammals (Y. Li, Liu, Shi, & Zhang, 2010), *Mc1r* in reduced pigmentation in multiple vertebrate species (Kronforst et al., 2012) and Na_v *1.4* in toxin resistance in snakes (Feldman, Brodie, Brodie, & Pfrender, 2012). This pattern leads to the prediction that genes responding selectively to a convergent environmental or phenotypic transition may show convergent shifts in evolutionary rates (i.e. number of nucleotide or amino acid substitutions per unit time), due to reduced or increased selective constraints on those genes. For example, both genes and non-coding elements involved in eye function lose selective constraint in subterranean mammals, and the corresponding rate shift signal can be used genome-wide to identify candidate regions with eye-specific function (Partha et al., 2017a). Thus, evolutionary rates provide a rich

source of genome-wide molecular data that, in combination with convergent phenotypes, can be used to infer targets of convergent selection.

Several research groups have developed methods to identify patterns of convergent evolutionary rate shifts associated with convergent changes in traits or environment, including Coevol (Lartillot & Poujol, 2011), Forward Genomics (Hiller et al., 2012b; Prudent, Parra, Schwede, Roscito, & Hiller, 2016) and PhyloAcc (Hu, Sackton, Edwards, & Liu, 2019). However, existing methods to identify these patterns still have several areas for improvement, including support for both binary and continuous traits, compute times that enable genome-wide analyses, robust statistical treatment (Partha, Kowalczyk, Clark, & Chikina, 2019) and ease of visualization and downstream analyses. To address these issues, here we present RERconverge, a software package that implements tests of association between evolutionary rates of genetic loci and convergent phenotypic traits, both binary and continuous (pipeline illustrated in Figure 1). RERconverge incorporates and corrects for known phylogenetic relationships among included species, performs rapid analyses of genome-scale data and implements appropriate statistical techniques to reduce the effect of outliers and correct for multiple testing. This easy-to-implement software represents a major advance in comparative genomic methods because it enables biologists with extensive organismal knowledge to apply computational techniques to quickly link molecular data with phenotypes.

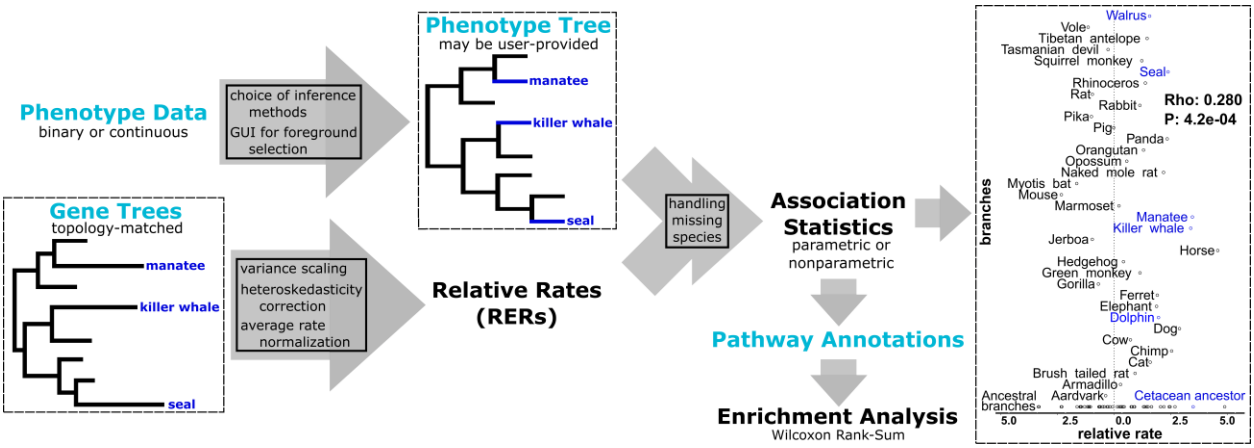


Figure 1 RERconverge Pipeline. Schematic of the RERconverge pipeline focusing on the example discussed in the case study. Light blue text indicates user-provided input. Far right figure shows RER for olfactory receptor *OR9Q2*, a top hit for marine-specfic acceleration. For additional details of the methods (boxes in arrows) see (Partha et al., 2019).

2.1.2 Description and Implementation

2.1.2.1 Basic usage of RERconverge

The RERconverge package runs within an installation of the R software (Team, 2018) on any platform (Linux, Windows and Mac OS). The user provides the following data as input:

- A set of phylogenetic trees, all with the same topology, with branch lengths for each tree calculated from alignments of that 'gene' sequence (these can also represent non-coding sequences such as enhancers) using software such as PAML (Yang, 2007).
- Values for a trait of interest for extant species/tips of the tree. RERconverge can infer values for internal branches of the tree, or the user can provide these values

via a single tree or by using RERconverge interactive branch selection to select foreground branches.

Detailed instructions for RERconverge are available within the Supplementary Material to (A. Kowalczyk et al., 2019) or at https://github.com/nclark-lab/RERconverge/wiki/Vignettes.

2.1.2.2 Rapid estimation and visualization of relative evolutionary rates (RER)

RERconverge offers efficient computation of gene-specific rates of evolution on branches of phylogenetic trees in genome-scale datasets. These gene-specific rates of evolution, termed relative evolutionary rates (RER), reflect the amount of sequence divergence on a particular branch after correcting for non-specific factors affecting divergence on the branch such as time since speciation and mutation rate. Additionally, using a combination of statistical approaches including data transformation and weighted linear regression, RERconverge provides estimates of RER that are robust to several factors introducing outliers in the dataset, such as the presence of distantly related species in the phylogeny (Partha et al., 2019). To accelerate the computations underlying RER calculations, key functions are written in C++ and integrated with the R code through the *Rcpp* package (Debian & François, 2011).

2.1.2.3 Flexible specification of binary and continuous trait evolution

For binary traits, RERconverge provides multiple methods for users to specify which branches are in the 'foreground', meaning those that display the convergent trait of interest. Users can directly supply foreground branch assignments by providing a phylogenetic tree with appropriate branch lengths or by interactively selecting branches on a phylogeny. Alternatively, users can specify which extant species are foreground and allow the software to infer which

internal branches to consider foreground under a variety of scenarios. These include whether to consider only branches at a transition to a convergent phenotype as foreground or to also consider subsequent branches that preserve the phenotype as foreground, whether the foreground trait can be lost or only gained, and whether each foreground branch should be weighted equally.

In addition to supporting analysis of binary traits, RERconverge can perform correlation analysis between evolutionary rates and continuous traits. Trait evolution is modeled on a trait tree constructed from user-provided phenotype values for extant species provided by the user and ancestral nodes inferred through a phylogenetic model. By default, trait tree branch lengths represent change in the trait between a species and its ancestor, which removes phylogenetic dependence between branches. For completeness, other options allow branch lengths to be calculated as the average or terminal trait value along a branch to represent the state of a trait rather than its change.

2.1.2.4 Genome-wide association between RER and traits, with correct for multiple testing

RERconverge rapidly computes the association between gene-specific RER and the user-specified traits for large sets of genes. The program estimates the correlation between gene tree branch lengths and the values for traits along those branches using (by default) Kendall's Tau for binary traits and a Pearson linear correlation for continuous traits. Full gene lists are returned with correlation statistic values, the number of data points used for the correlations and adjusted P-values (FDR) derived using the Benjamini–Hochberg correction (Benjamini & Hochberg, 1995).

2.1.2.5 Assessing gene set enrichment in results

Further analysis of gene lists from correlation analyses can be performed using built-in RERconverge pathway enrichment functions. These functions perform a rank-based enrichment analysis by using a Wilcoxon Rank-Sum Test to detect distribution shifts between a subset of genes and all genes in annotated pathways. Enrichment results are returned with pathway names, enrichment statistics, Benjamini–Hochberg corrected P-values and ranked pathway genes.

2.1.3 Case Study

We demonstrate the features of RERconverge using gene trees derived from the coding sequences of 19,149 genes for 62 mammal species, along with foreground branches that represent lineages of the mammalian phylogeny whose members live predominantly in marine aquatic environments (Chikina, Robinson, & Clark, 2016; W.K. Meyer et al., 2018; Partha et al., 2019). For this dataset, RERconverge ran in 1h and 18min on RStudio with R version 3.3.0 installed on a computer running Windows 10 with 16 GB of RAM and an Intel Core i7-6500U 2.50 GHz CPU. Top categories of functional enrichment included olfactory transduction (FDR<10e−10), GPCR signaling (FDR<10e−10) and immune functions (FDR = 5.4e−7), as in Chikina *et al.* (Chikina et al., 2016). We show the RERconverge workflow and visualization of RERs for a top olfactory receptor, *OR9Q2*, in Figure 1.

2.1.4 Summary

RERconverge is an easy-to-implement method that can quickly test for associations between genes' RER and traits of interest on a phylogeny. This enables researchers studying a wide variety of questions to generate lists of candidate genes associated with evolutionarily important traits and to explore through enrichment analyses the biological functions showing the most evidence of molecular change in association with these traits.

2.2 Permulations: a rigorous statistical strategy to measure confidence in associations in a phylogenetic context

Permulations are a method to accurately quantify confidence in statistical associations from RERconverge and related phylogenetic methods. When working in a phylogenetic context, there are many potential confounders, both known and unknown, that may bias statistics calculated in that context. Permulations empirically quantify confidence and thus correct for potential confounders without the need to explicitly define them.

I was the primary creator of permulation methods, and Elysia Saputra subsequently assisted with benchmarking and improving methods for binary phenotypes. I performed all analyses and benchmarking described for continuous phenotypes, and I created the original implementation of binary trait permulations. Further, I performed tests using permulations with PGLS and analyzed results from Forward Genomics that were generated by Luisa Cusick. The paper describing

permulations was originally published in *Molecular Biology and Evolution* at (Saputra et al., 2021).

2.2.1 Introduction

Despite the availability of complete genomes for many species, identifying the genetic elements responsible for a phenotype of interest is difficult because there are millions of genetic differences between almost every pair of species. One strategy to link genotypes and phenotypes is to take advantage of convergent evolutionary events in which multiple unrelated species have evolved similar characteristics. Such events represent natural biological replicates of evolution during which species may have experienced similar genetic changes driving similar phenotypic changes. When lineages independently evolve or lose a shared phenotype, convergent molecular signals can be used to identify specific genetic elements associated with the phenotypic shift.

Diverse analytic approaches have been developed to use convergent phenotypes to identify specific genetic elements underlying a trait. The methods include analyzing convergent amino acid substitutions (Foote et al., 2015) and convergent shifts in evolutionary rates (Hiller et al., 2012b; Hu et al., 2019; A. Kowalczyk et al., 2019; Prudent et al., 2016; Wertheim, Murrell, Smith, Kosakovsky Pond, & Scheffler, n.d.), as well as investigating convergent gene loss (Hiller et al., 2012b; W.K. Meyer et al., 2018). Methods that analyze convergent shifts in evolutionary rates (rather than convergence to any specific sequence) have been particularly successful. We have previously developed one such method called RERconverge (A. Kowalczyk et al., 2019; Partha et al., 2019) to link genetic elements to convergently evolving phenotypes based on evolution across a sequence of interest. Our method has been successfully used to identify the genetic basis of

adaptation to a marine habitat (Chikina et al., 2016), regression of ocular structures in a subterranean habitat (Partha et al., 2017a), and evolution of extreme lifespan and body size phenotypes (Amanda Kowalczyk et al., 2020) in mammals. Other groups have developed similar methods for identifying convergent shifts in evolutionary pressure. The Forward Genomics algorithm, which correlates percent sequence change along a phylogeny with phenotypic changes (Hiller et al., 2012b; Prudent et al., 2016), has been used to identify genetic elements underlying low levels of biliary phospholipid levels in horses and guinea pigs, the loss of ability to synthesize vitamin C in some primates, bats, and guinea pigs, as well as the loss of ocular structures in two independent subterranean mammals. Both RERconverge and Forward Genomics involve a phylogenetic inference step and a subsequent test for phenotype association. More sophisticated but computationally intensive methods that consider the phenotype at the phylogenetic inference step have also been developed, notably PhyloAcc (Hu et al., 2019), although these methods are difficult to scale to genome-wide analyses. A related but distinct approach is to assess the association between gene loss (the limiting case of relaxed evolutionary pressure) and convergent phenotypes. A recent study used phylogenetic generalized least squares (PGLS) (Grafen, 1989) to compute associations between gene losses and diverse traits and found a large number of significant associations (Prudent et al., 2016).

Importantly, these methods are often applied in a genome-wide discovery context. As such, the general approach can be summarized as using a statistical test to calculate the association between convergent phenotypes and some measure of molecular evolution (evolutionary rate or gene loss) across a large number of genomic regions, followed by multiple hypothesis testing corrections. If an enrichment of small p-values is observed, then it is presumed that some genes (or other genetic elements) are truly associated with the phenotype. This conclusion rests on the

assumption that under the null hypothesis of no association, each data point is sampled independently from a common null distribution, in which case uniform p-values would be observed. However, when applied to genome-scale datasets, phylogenetic methods often show atypical statistical behavior in which the expected uniform distribution of p-values is not observed when using null phenotypes (Figure 2A**)**. For example, the standard RERconverge analysis is anti-conservative when applied to the marine phenotype but conservative when applied to the long-lived large-bodied phenotype. Forward Genomics likewise produces large deviations from the expected null. This issue exists for even the widely used PGLS method, which produces a near-uniform null when applied to gene loss in long-lived large-bodied mammals, but an extremely skewed distribution when applied to loss of transcription factor binding sites in the same phenotype.

PGLS is a method specifically designed to correct for phylogenetic dependence. Therefore, the fact that a non-uniform null is observed for even the PGLS method demonstrates that deviations from the expected null cannot be explained by the phylogenetic structure of the data alone, but can also result from other sources of dependence that arise in the context of large multiple alignment datasets. Differences in genome quality (Hosner, Faircloth, Glenn, Braun, & Kimball, 2016), nucleotide frequencies (Romiguier & Roux, 2017), a mis-specified phylogeny, or other unknown systematic effects all create systematic biases which accumulate when the method is applied to thousands of genomic regions. As such, even if the tests can be proven to be theoretically valid under some assumptions (such as the well-understood PGLS model), they are not guaranteed to produce the expected uniform distribution when applied repeatedly to data from the same multiple sequence alignment. This deviation from the null expectation can result in overestimated statistical confidence and produce spurious genotype-phenotype associations.

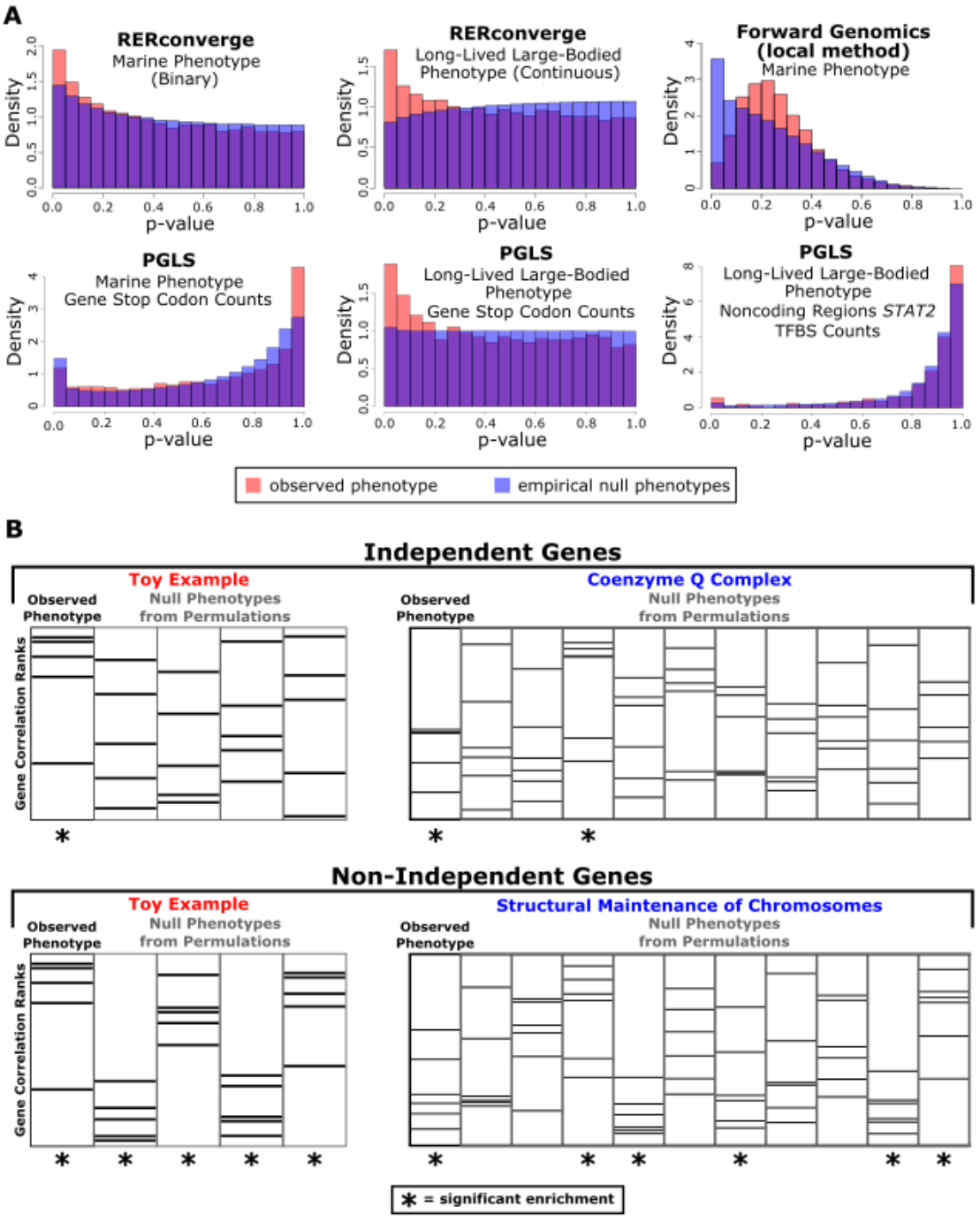


Figure 2 Motivation for Permulations. Permulations reveal statistical anomalies in genetic element- and pathway-level analyses because parametric p-values deviate from the expected uniform distribution when assessed on null phenotypes. (A) p-value histograms comparing p-values obtained using an observed phenotype (red) compared to p-values obtained from 500 (or more, see Results) null phenotypes from

permulations. We evaluate a binary phenotype (marine) and a continuous phenotype (long-lived large-bodied) through RERconverge, a binary phenotype (marine) through Forward Genomics, and a binary phenotype (marine) and a continuous phenotype (long-lived and large-bodied) through PGLS with gene stop codon counts and noncoding element *STAT2* TFBS counts. In all cases, the empirical null from permulations (shown in blue) is non-uniform. Since null p-value distributions are often non-uniform (shown in blue), observed parametric p-values from standard statistical tests (shown in red) cannot be interpreted using traditional strategies. (B) Pathway enrichment statistics from RERconverge long-lived large-bodied analyses demonstrate artificially inflated significance because genes in many pathways are non-independent. Accordingly, null phenotypes from permulations often show false signals of enrichment. Permulations correct for non-independence by quantifying the frequency at which significant pathway enrichment occurs due to chance.

The problem is further compounded when results from genetic elements are aggregated at the pathway level. Beyond the existing biases that arise from the nature of multiple sequence alignments, geneset analyses suffer additional non-independence induced by the evolutionary process itself. It is well established that genes that are functionally related experience correlated evolutionary pressure and thus evolve in a dependent fashion (Clark, Alani, & Aquadro, 2012, 2013; Juan, Pazos, & Valencia, 2008). One extreme example of such coevolution is "reductive evolution", where losing a member of interacting proteins decreases the selection pressure for preserving its interacting partners (Ochoa & Pazos, 2014). As a result of coevolution, many functionally related genes "travel in packs" in association with a phenotype, meaning that if one gene in a group appears to be associated with a phenotype, the other genes in the group will as well because they do not evolve independently. The result is that a function could appear as associated with the phenotype due to random chance instead of actual involvement, causing an erroneous inference of enrichment.

The implication of coevolution is apparent when we apply standard pathway enrichment analysis to gain insight into which groups of functionally related genes are overrepresented among convergently evolving genes, as implemented in standard tools such as Gorilla, GO::TermFinder, and RERconverge enrichment functions (Boyle et al., 2004; Eden, Lipson, Yogev, & Yakhini, 2007; Eden, Navon, Steinfeld, Lipson, & Yakhini, 2009; Amanda Kowalczyk et al., 2019). Figure 2B demonstrates how correlated evolutionary rates can cause problems in pathway enrichment analyses. When genes are ranked based on gene-phenotype associations, coevolving genes tend to have clustered ranks. Such clusters make it easier to observe enrichment of extreme ranks, or coevolving genes that all have either high or low ranks, due to chance alone, and therefore the typical null expectation does not hold. Even when using a null phenotype, genes appear to cluster at the extremes of the ranked list. The clustering, and resulting enrichment, is caused by the genes "traveling in packs", in which case simple enrichment tests assign undue confidence to an essentially spurious enrichment.

Rigorous statistical handling needs to be employed to address these sources of bias. Systematic solutions have been devised to correct issues with non-independence, both in the contexts of quantitative genetics (Allison et al., 2002) and phylogenetics (Stone, Nee, & Felsenstein, 2011). However, these systematic approaches often make assumptions on the evolutionary process or other distributional assumptions, which may not accurately represent the data. We argue that an empirical approach that is grounded in the observed data can provide better calibration against sources of bias. In the context of gene expression, this problem is typically handled by performing label permutations (Majewski et al., 2010; Ritchie et al., 2015; A. Subramanian et al., 2005) and in certain cases parametric adjustments (Wu & Smyth, 2012). However, simple label permutations are not applicable to associations involving a phylogeny as

they would not preserve the underlying phylogenetic relationships, thereby producing false positives.

Here, we develop a novel strategy that combines *permu*tations and phylogenetic simu*lation*s to generate null phenotypes, termed "permulations". The strategy addresses statistical non-independence empirically by generating phenotype permutations from phylogenetic simulations. In this way, the strategy preserves the underlying phylogenetic dependence by sampling permutations from the correct covariance structure. It also more accurately mimics the null expectation for a given phenotype by exactly matching the distribution of observed phenotype values for continuous phenotypes and exactly matching the number and structure of foreground branches (branches on which the phenotype changes) for binary phenotypes. We use these "permulated" phenotypes to calculate empirical p-values for gene-phenotype associations and pathway enrichment related to a phenotype. In doing so, we have created a statistical pipeline that accurately reports confidence in relationships between genetic elements and phenotypes at the level of both individual elements and pathways.

2.2.2 New Approaches

2.2.2.1 Permulations: A Hybrid Approach of Using Permutations and Phylogenetic Simulations to Generate Null Statistics

The goal of permulations is to empirically calibrate p-values from phylogenetic methods by producing permutations of the phenotype tree that account for the structure in the data. The permulation method requires a master species tree and a species phenotype (either continuous or binary). The method then returns a set of phenotypes that are random but preserve the phylogenetic

dependence of the input phenotype. We typically generate 1,000 such permulated phenotypes, which are then used in the framework of a certain phylogenetic method (e.g., RERconverge) to compute gene-trait associations, resulting in 1,000 empirical null statistics for each gene. Similarly, we can also run enrichment analyses using the permulated phenotypes to produce 1,000 empirical null statistics for each pathway. Finally, for each gene or pathway, we calculate the empirical p-value as the proportion of empirical null statistics that are as extreme or more extreme than the observed parametric statistic for that gene or pathway. Since empirical null statistics capture the true null distributions for genes and pathways, the empirical p-values represent the confidence we have to reject the null hypotheses of no association, correlation, or enrichment given the underlying structure of our data. Note that permulations do not eliminate the need for multiple hypothesis correction; even with a corrected null model, the likelihood that false discoveries are made from performing multiple statistical inferences simultaneously still exists. Our permulation methods for binary and continuous phenotypes have been included in the publicly available RERconverge package for R (Kowalczyk et al. 2019) (published on github at https://github.com/nclark-lab/RERconverge), with a supplementary walkthrough (see Supplementary Walkthrough) also available as a vignette included in the RERconverge package.

2.2.2.2 Phylogenetic Permulation for Continuous Phenotypes

For continuous traits, generating permulated phenotypes is a two-step process. First, null phenotype values are simulated. Second, real phenotype values are assigned based on the simulated values. In step one, given the master tree with branch lengths representing average evolutionary rates and phenotype values for each species, we simulate a random phenotype using the Brownian motion model of evolution. The Brownian motion model takes a "random walk"

down the master tree phylogeny to assign phenotype values. Since more closely related species are a shorter "walk" from each other, they are more likely to have more similar phenotype values than more distantly related species. In step two, real phenotype values are assigned to species based on ranks of the simulated values. The species with the highest simulated value is assigned the highest observed value, the species with the second-highest simulated value is assigned the second highest observed value, and so on. By doing so, observed phenotypes are shuffled among species with respect to the underlying phylogenetic relationships among the species. Since simulated values are more similar among more closely related species compare to distantly related species, the newly reassigned real values follow the same pattern (Figure 3).

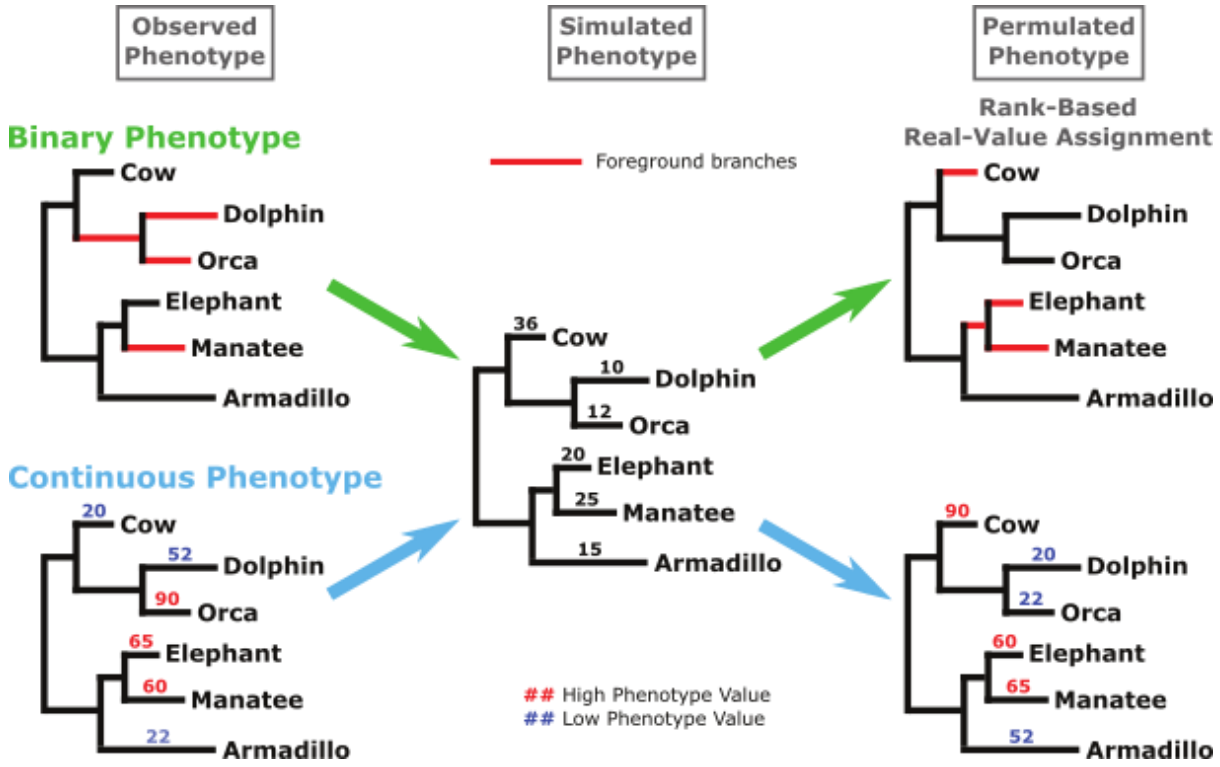


Figure 3 Permulated Phenotype Generation Procedure. Permulated phenotypes were generated by simulating phenotypes and then assigning observed phenotype values based on the rank of simulated values. Simulations were performed using Brownian motion phylogenetic simulations and a phylogeny containing all mammals with branch lengths representing the average evolutionary rate along that branch genome-wide. For binary phenotypes, foreground branches for permulated phenotypes are assigned based on the highest-ranked simulated values while preserving the phylogenetic relationships between foregrounds. For continuous phenotypes, observed numeric values were assigned directly to species based on ranks of simulated values.

2.2.2.3 Phylogenetic Permulation for Binary Phenotypes

For binary traits, the critical feature is the number of foreground species and their exact phylogenetic relationship, and hence the inferred number of phenotype-positive internal nodes or equivalently phenotypic transitions. The two-step process proposed above does not guarantee to perfectly preserve this structure. Instead, we employ a rejection sampling strategy where the

simulation is used to propose phenotypes which are accepted only if they match the stricter requirements. Specifically, species are ranked based on simulated values, and a set of top-ranked species chosen to match the number of foreground species in the observed phenotype are proposed as a null phenotype. The proposed phenotype is only accepted if it preserves the phylogenetic relationships among chosen foregrounds, as observed in the actual foregrounds (Figure 3, Binary Phenotype). Using the simulation as the proposed distribution ensures that phylogenetically dependent phenotypes are generated and thus speeds up the construction of null phenotypes over what can be achieved from random selection.

We present two binary permulation strategies: the complete case (CC) method and the species subset match (SSM) method. The SSM method accounts for the fact that not all genes have orthologs in all species while the CC method ignores species presence/absence for simplicity. The strategies encompass the trade-off between computational feasibility and statistical exactitude—in some cases, it may not be possible to perform the SSM method, in which case the CC method is a viable alternative. The CC method is the first and simpler strategy. The CC method performs permulations using the master tree in which all species are present and therefore generates permulated trees that contain the complete set of species. Since not all species will have sequences available for all genes and the CC method produces one set of permulated phenotypes for all the genes, the exact number of foreground and background species per genetic element may not be preserved because of species presence/absence in those alignments (Figure 4). Thus, the CC method is an imperfect but fast method to generate null phenotypes, but we recommend use of the SSM method whenever feasible.

In contrast, the SSM method accounts for the presence/absence of species in different gene trees. For each permulation, the SSM method generates separate null phenotypes for each tree in

the set of genetic elements. Since genetic element-specific trees contain exactly the species that have that genetic element, the null phenotypes exactly match the observed phenotypes for that genetic element in terms of number of foreground and background species (Figure 4). Additionally, unlike the CC method, null phenotypes for a single permulation iteration are distinct, and potentially unique, from each other because they are generated on a genetic element-by-genetic element basis. Although the SSM method is statistically more ideal than the CC method, it is much more computationally intensive and may not be feasible for very large datasets. For example, the CC method took 7 seconds to produce 50 permulated traits for 200 genes, whereas the SSM method took ~15.5 minutes.

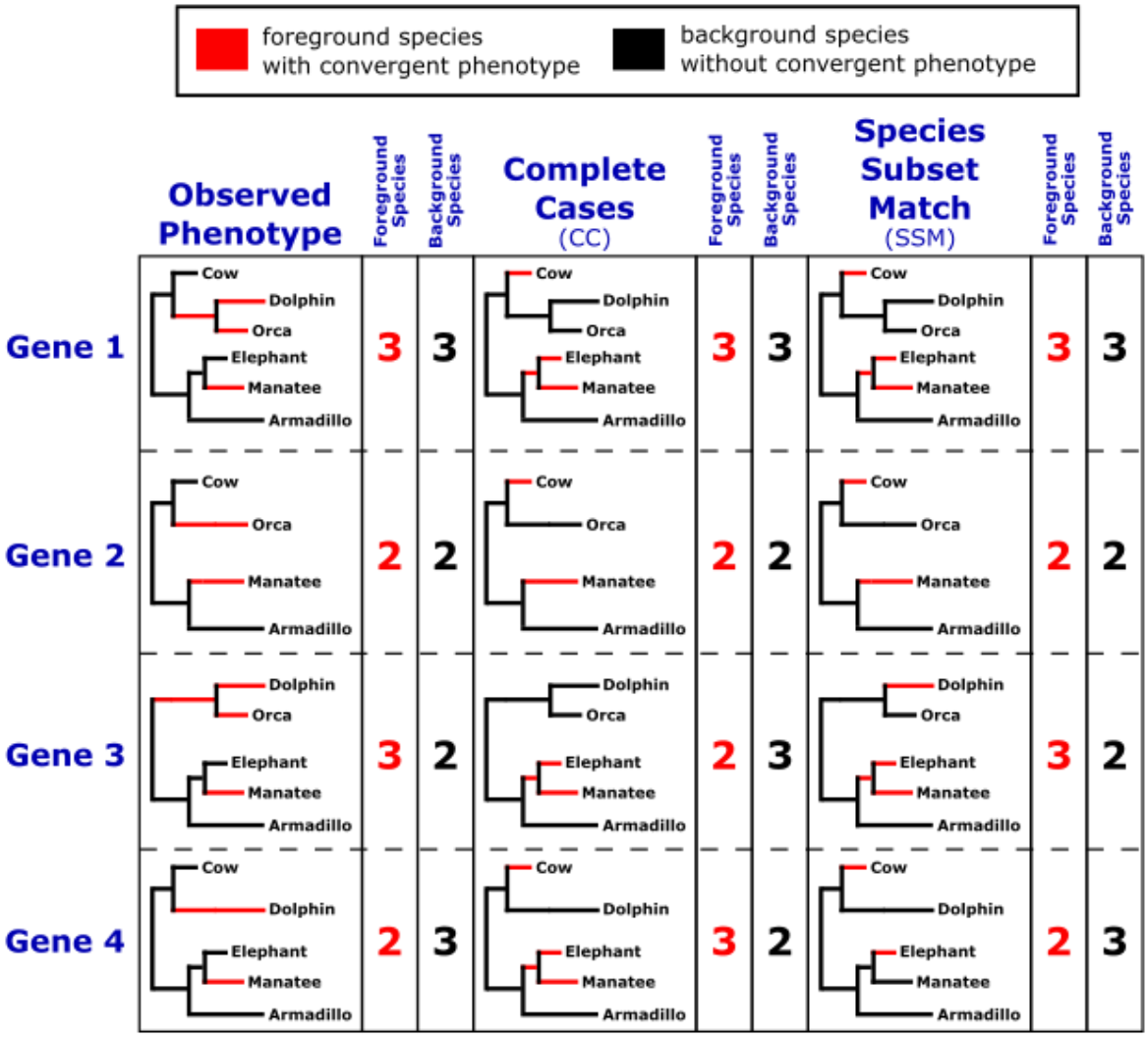


Figure 4 CC vs. SSM Permulations. Examples of toy binary phenotypes permulated using the complete case (CC) method or the species subset match (SSM) method. For the CC method, top-ranked simulated values are assigned as foreground regardless of gene-specific species absence. For the SSM method, top-ranked simulated values are assigned as foreground after considering gene-specific species absence so the number of foreground and background species for each gene is consistent across every permulated phenotype. Note that in the case of genes with all species present (e.g., Gene 1), CC and SSM methods are identical.

2.2.2.4 Datasets for Method Evaluation

We evaluate the performance of our permulation methods by using RERconverge to find genetic elements that demonstrate convergent acceleration of evolutionary rates in association with convergent phenotypic adaptations that are well-characterized, namely the evolution of the marine mammal phenotype (Chikina et al., 2016; Wynn K Meyer et al., 2018), the subterranean mammal phenotype (Partha et al., 2017b), and the long-lived large-bodied mammal phenotype (Amanda Kowalczyk et al., 2020). For the remaining part of this article, we will refer to these phenotypes as the marine phenotype, the subterranean phenotype, and the long-lived large-bodied phenotype, respectively. We use the set of protein-coding gene trees across 63 mammalian species previously computed by Partha et al. (Partha et al., 2019). These trees have the "Meredith+" tree topology (Amanda Kowalczyk et al., 2020) (Figure 5), a modification of the tree topologies published by Meredith et al. (Meredith et al., 2011) and Bininda-Emonds et al. (Bininda-Emonds et al., 2007), resolved for their differences across various studies as originally reported by Meyer et al. (Wynn K Meyer et al., 2018).

For the binary marine phenotype, we set three independent lineages as foreground species that possess the marine trait (blue branches in Figure 5, Binary Phenotype): pinnipeds (Weddell seal, walrus), cetaceans (bottlenose dolphin, killer whale, the cetacean ancestor), and sirenians (West Indian manatee) (Chikina et al., 2016). For the subterranean phenotype, we set as foregrounds three independent subterranean species for which high quality genomes are available in our dataset: naked mole-rat, star-nosed mole, and cape golden mole (red branches in Figure 5, Binary Phenotype).

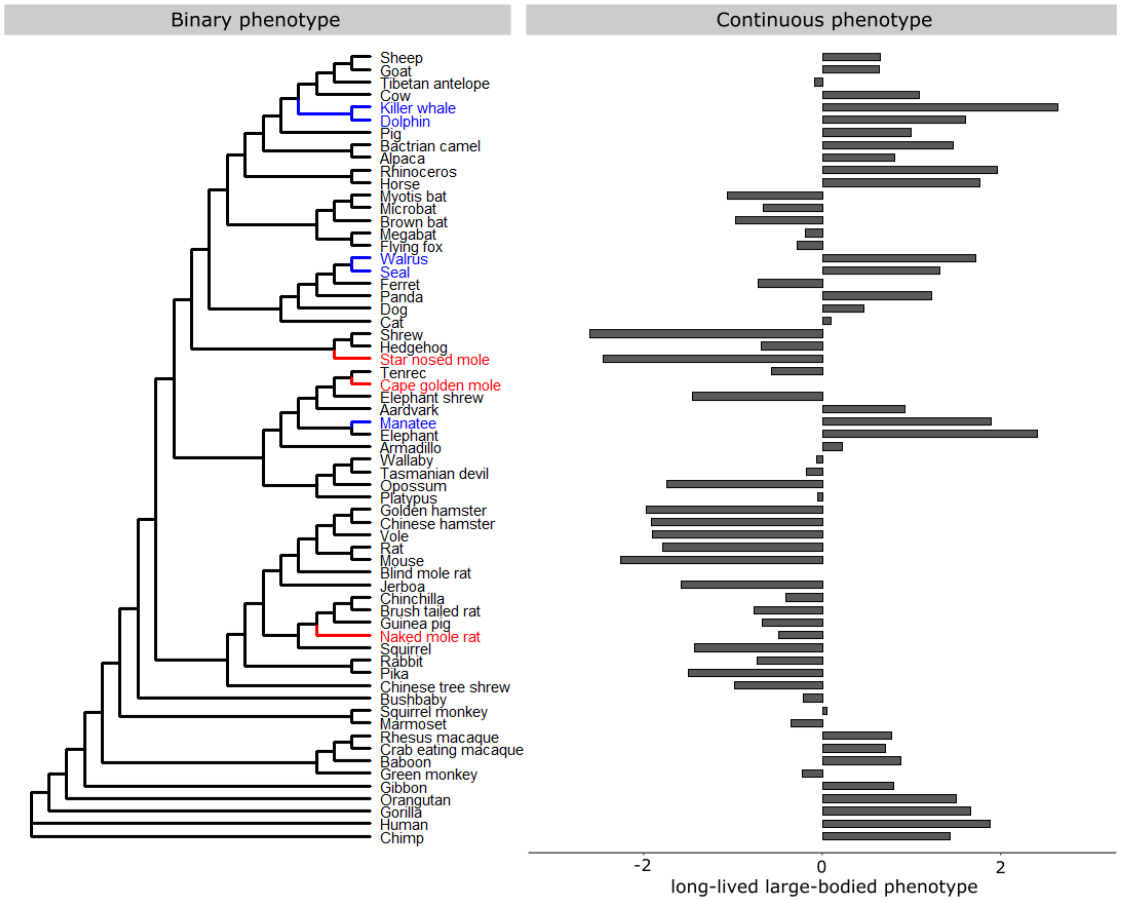


Figure 5 Phenotypes to Test Permulations. Meredith+ tree topology and the binary and continuous phenotypes evaluated. Binary phenotypes include the marine mammal phenotype and the subterranean mammal phenotype (foreground branches are indicated in blue and red, respectively). The continuous phenotype evaluated is the long-lived large-bodied phenotype as constructed based on the first principal component between species body size and maximum longevity (Kowalczyk et al. 2020).

Finally, for the continuous long-lived large-bodied phenotype, we use the "3L" trait as defined in previous work (Amanda Kowalczyk et al., 2020). The numerical phenotype is constructed by calculating the first principal component (PC1) between body size and maximum lifespan across 61 mammal species (Figure 5, Continuous Phenotype). PC1 therefore represents the agreement between body size and lifespan – species like whales with long lifespans and large

sizes have large phenotype values and species like rodents with short lifespans and small sizes have small phenotype values. For example, killer whale, elephant, and rhino have the highest values (2.63, 2.40, and 1.95) because they are both large and long-lived, and shrew, star-nosed mole, and mouse have the smallest values (-2.62, -2.46, and -2.27) because they are small and short-lived. Human, while longest-lived among the mammals included, has the fifth largest value (1.87) because humans are relatively small compared to the other mammals. Likewise, large grazing animals like cow also have smaller PC1 values (1.08, the 15th largest value) because although cows are large, they are not very long-lived given their body size.

2.2.3 Results

2.2.3.1 Permulation of Binary Phenotypes Improved Power and Type I Error Control

To evaluate the performance of the permulation methods compared to the parametric method for binary phenotypes, we first use RERconverge to find genetic elements with convergently accelerated evolutionary rates in species with the marine phenotype. We consider three p-value calculation methods: parametric, complete case (CC) permulations, and species subset match (SSM) permulations. Resulting p-values are corrected for multiple hypothesis testing using Storey's correction (J. D. Storey, Bass, Dabney, & Robinson, 2020; J. D. Storey & Tibshirani, 2003). We see in Figure 2A that the parametric p-values for the association of genes with the observed marine phenotype (red histogram) are enriched for small p-values. According to the standard parametric approach, which assumes a simple null hypothesis with uniformly distributed p-values, the enrichment of low p-values indicates the possible presence of genes with evolutionary rate shifts that are significantly correlated with marine adaptation. However, when

we construct the empirical null p-value distribution using 1,000 permulations of the marine phenotype, we see that the null distribution of parametric p-values is not uniform. In fact, the enrichment of low p-values is also present in the null distribution (blue histogram), albeit a lesser enrichment than the observed, meaning that observing low p-values by chance is more likely than expected. Thus, if we use standard multiple testing procedures directly on the parametric p-values, we will identify more positive genes than the true number of positives, in other words causing an undercorrection of p-values.

To demonstrate that our permulation strategy effectively corrects for the background p-value distribution, we plot similar histograms of the empirical p-values for the marine phenotype versus 1,000 permulated phenotypes, generated from both CC and SSM permulations. With permulations, we can see that while some enrichment of small empirical p-values is observed for the marine phenotype, the empirical p-values for the null phenotypes are almost perfectly uniform, meaning that our permulation methods are able to construct the correct null distribution (Supplementary Figure 1 in (Saputra et al., 2021)). When we overlay the p-value histograms of the parametric and empirical p-values for the marine phenotype, we can see that compared to the parametric method, the histograms for the CC and SSM permulations have steeper slopes at low p-values, indicating that the permulation methods have better Type I error control (Figure 6**A**). Furthermore, the histograms for the permulation methods plateau at higher π_0 than the parametric method, consistent with the postulation that the parametric method would identify more (possibly false) positives. These findings are also observed when we define genes with significant evolutionary acceleration in marine mammals (i.e., "marine-accelerated" genes) by setting a rejection threshold of Storey's false discovery rate (FDR) ≤ 0.4 (the high threshold is set considering the high minimum FDR from the parametric method), as shown in Figure 6B. For the

permulation methods, as the number of permulations increases, the number of identified marine-accelerated genes increases and eventually stabilizes after ~400 permulations. The asymptotic numbers of marine-accelerated genes identified by permulations (~350 genes for CC permulation and ~450 genes for SSM permulation) are much smaller than the ~700 genes identified through parametric statistics, demonstrating improved Type I error control.

Surprisingly, while the permulation methods identify fewer significantly accelerated regions, we have greater confidence in their significance. Figure 6C shows the minimum FDRs achieved by the permulation methods with increasing number of permulations. The figure shows that the permulation methods provide better control of FDRs compared to the parametric method with only a few permulations (above ~125 permulations). With increasing permulations, the minimum FDR continues to drop to reach levels below 0.1 at 1000 permulations, while the minimum FDR from parametric statistics is higher at above 0.3. Use of the permulation null substantially improves the statistical power of the method and provides much higher confidence in detecting true correlations between evolutionary rate shifts and the convergent phenotype of interest.

Lastly, we find that permulation methods can identify marine-accelerated genes that are missing in many species, i.e., genes with phylogenetic trees containing few species. In contrast, the parametric method fails to identify any such gene (Figure 6D).

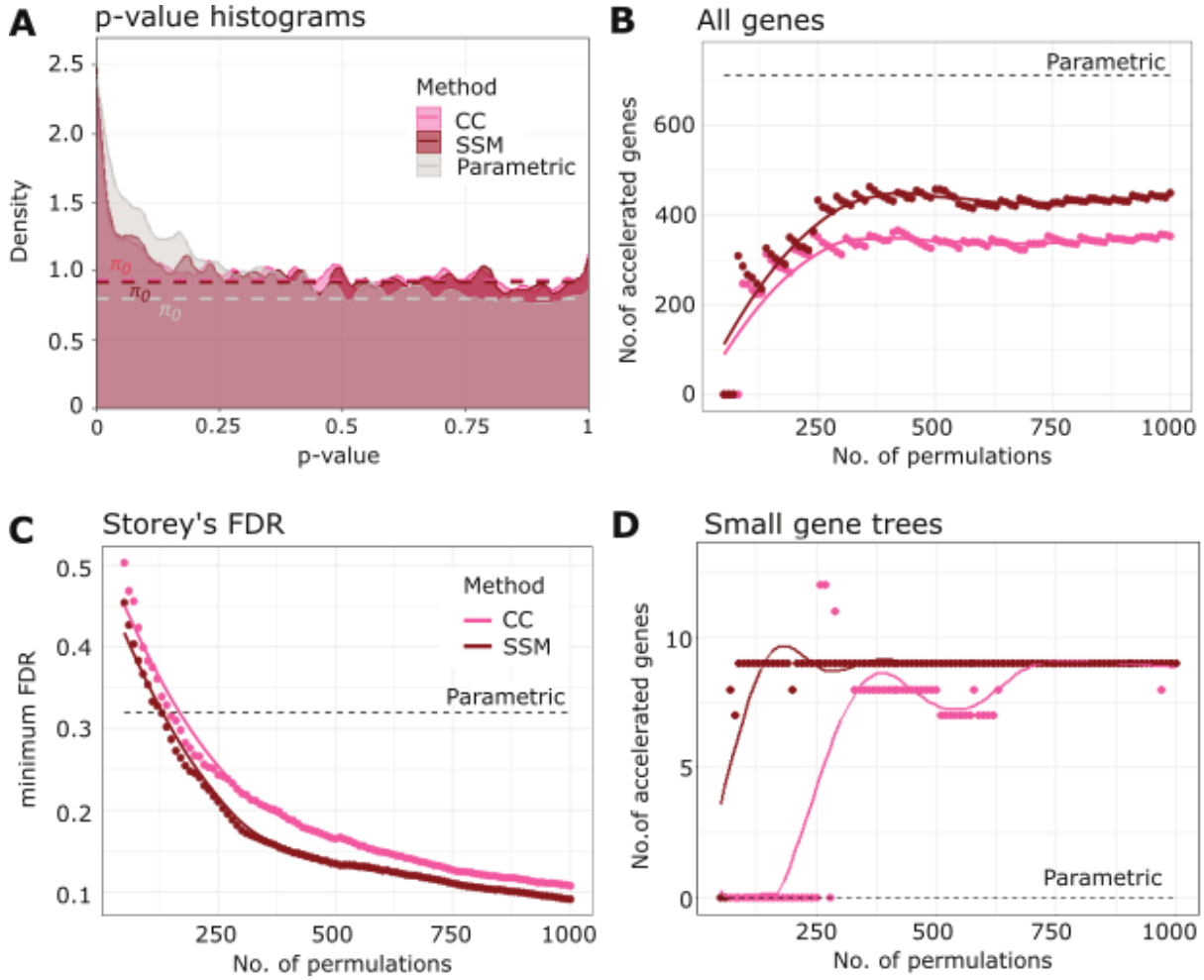


Figure 6 Permulations Correct Inflation of Statistical Significance for a Binary Phenotype. Permulation of binary phenotypes corrects for inflation of statistical significance in finding evolutionarily accelerated genes in marine mammals. (A) Histogram of parametric and permulation p-values for the marine phenotype from the parametric, the complete case (CC) permulation, and the species subset match (SSM) permulation methods. (B) Permulation methods identify fewer accelerated genes in marine mammals compared to the parametric method, correcting for the inflation of significance. The rejection region of the multiple hypothesis testing is set to be Storey's FDR ≤ 0.4, considering the weak power of the parametric method. (C) Binary permulation methods have greater statistical power compared to the parametric method, as shown by the minimum false discovery rate (FDR) calculated using Storey's method. (D) Permulation methods can identify accelerated genes that are missing in many species (gene tree size ≤ 30), whereas the parametric method fails to do so.

2.2.3.2 Binary Permulation Methods Improved Gene-level Detection of Functional Enrichment

We have demonstrated that the permulation methods show favorable statistical properties based on the distribution of p-values. We expect that this approach also improves the biological signal of rate convergence analysis. To address this question, we ask if the marine-accelerated genes identified by binary permulations are enriched for functions that are consistent with the marine phenotype. Our group previously identified marine-specific pseudogenes that should be undergoing accelerated evolution in marine mammals due to relaxation of evolutionary constraint (Wynn K Meyer et al., 2018). Putative pseudogenes associated with marine mammals were identified using Bayes Traits software (Pagel & Meade, 2006) to find signals of coevolution between marine status and pseudogenization. In addition, our group also previously found that marine-accelerated genes that evolve under relaxed constraint are enriched for genes responsible for the loss of olfactory and gustatory functions (Chikina et al., 2016). Thus, to represent the "ground truth", we select a collection of gene sets relevant to olfactory and gustatory functions from the Mouse Genome Informatics (MGI) database and top-ranking marine-specific pseudogenes with Bayes Traits FDR values less than 0.25.

We then perform the one-tailed Fisher's exact test to measure the enrichment of the functions in the marine-accelerated genes from the parametric and permulation methods. The Fisher's exact test odds ratios indeed show that the CC and SSM permulation methods generally magnify or maintain the effect sizes of enrichment across the gene sets compared to the parametric method (Figure 7A). At worst, the permulation methods match the performance of the parametric method (e.g., "taste/olfaction phenotype" gene set). The improved performance of the permulation

methods is also demonstrated in the example precision-recall curves for the marine-associated pseudogenes in Figure 7B.

To see if this observation generalized to other phenotypes, we repeat the whole analysis to find genes that are accelerated in species with the subterranean phenotype. As subterranean-accelerated genes have been found to be enriched in ocular functions (Partha et al., 2017b, 2019; Prudent et al., 2016), we pick gene sets relevant to vision-related functions as the "ground truth". In general, the signals we obtained from RERconverge for the subterranean phenotype are much weaker than in the marine phenotype case, but the enrichment is still captured in the rankings of the genes. Similar to the marine phenotype, permulation methods generally improve or match the performance of the parametric method (Figure 7C-D).

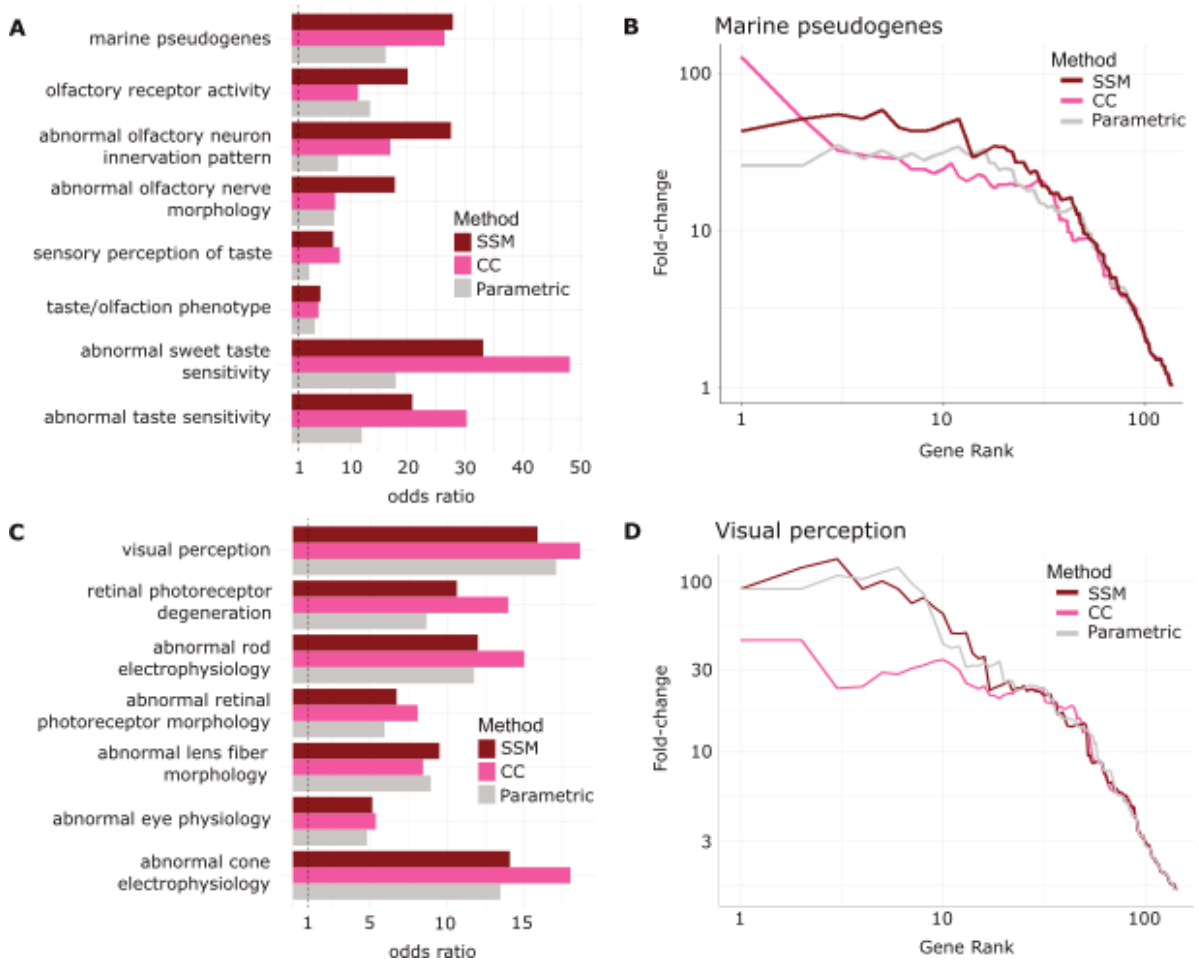


Figure 7 Binary Permulations Match or Improve Power to Detect True Positives. Binary permulation methods have matching or improved power compared to the parametric method in detecting enrichments of functions consistent with known phenotypes. (A) Fisher's exact test odds ratios showing that marine-accelerated genes identified by the permulation methods have greater enrichment of gustatory and olfactory genes, compared to the parametric method. (B) Precision-recall curves for the enrichment of the marine pseudogenes in the identified marine-accelerated genes. Greater area under the curve (curves that have higher values on the left side of the plot) have greater enrichment. (C) Fisher's exact test odds ratios showing that subterranean-accelerated genes identified by the permulation methods have greater or comparable enrichment of ocular genes, compared to the parametric method. (D) Precision-recall curves for the enrichment of the visual perception genes in the identified subterranean-accelerated genes.

2.2.3.3 Binary Permulation Method Corrects for False Positives in Related Approaches

In addition to performing permulations using RERconverge, we test our methods using Forward Genomics and PGLS. Other methods, such as PhyloAcc, would require tens of millions of computational hours to generate 500 permulations (from the analysis with RERconverge, the number of identified accelerated genes plateaus after 400-500 permulations are used (Figure 6B)), and thus permulations are not scalable to those analyses.

Forward Genomics (Hiller et al., 2012a; Prudent et al., 2016), like RERconverge, tests for an accelerated evolutionary rate in a set of foreground species by correlating a normalized substitution rate with phenotypes using Pearson correlation. It works only for binary phenotypes and has demonstrated success in coding and non-coding elements. Forward Genomics' "global method" uses substitution rate with respect to each tree's root to correlate with trait loss and identify convergent relaxed selection; therefore, it does not correct for evolutionary relatedness. The "local branch method", an improvement on the original approach, uses substitution rate with respect to the most recent ancestor to identify relaxed selection, which substantially improves its power (Prudent et al., 2016). We use the most recent version of both the global and the local methods to test for associations between gene evolutionary rates and the binary marine phenotype.

Both global and local Forward Genomics methods had unusual p-value distributions. The local method identified high proportion of positives with significant p-values (Figure 2A), while p-values from the global method were highly concentrated around 0.5 (global p-values not shown). Adjusting for multiple testing further exaggerated this issue. For the global method, due to the number of genes with very low p-values, the lowest possible Benjamini-Hochberg (BH) corrected parametric p-value is 0.531, and for the local method, the lowest possible corrected p-value is 0.465. For the local method, out of 18,797 genes, more than half of the genes (12,438) have the

lowest possible corrected parametric p-value. As such, it is impossible to designate a significance cut-off, because it will either include no genes or include most of the genes. Applying the permulation strategy to Forward Genomics output, we find that of the same set, 889 have a corrected empirical p-value that is less than or equal to 0.465 (the minimum observed corrected parametric p-value), allowing for a more reasonable selection of a rejection threshold. Thus, permulation can improve statistical performance even for a statistic with known flaws.

We further investigate our results from Forward Genomics at the pathway level in addition to analyzing results at the individual gene level. We use the marine pseudogenes as a "ground truth" set of genes that should be undergoing accelerated evolution in marine species, to test our ability to detect pathway enrichment of these genes. As shown in Figure 8A, the global and local parametric test statistics show slight enrichment for elements that are pseudogenized in marine mammals, and the difference is improved when empirical p-values are computed. Figure 8B shows the same data as a precision-recall plots, clearly demonstrating that the permulation correction improves the predictive power of both methods.

Next, we tested the effect of permulations on PGLS results. PGLS tests for association between two traits across species while adjusting for the phylogenetic relationships among those species. In doing so, it numerically corrects for non-independence due to phylogenetic relatedness. Note that unlike RERconverge and Forward Genomics, PGLS does not require evolutionary rate information and is therefore a more generalized phylogenetic analysis. We test PGLS using both the binary marine and the continuous long-lived large-bodied phenotype for coevolution with stop codon counts across genes. We additionally test the continuous phenotype for coevolution with *STAT2* transcription factor binding site counts across noncoding regions.

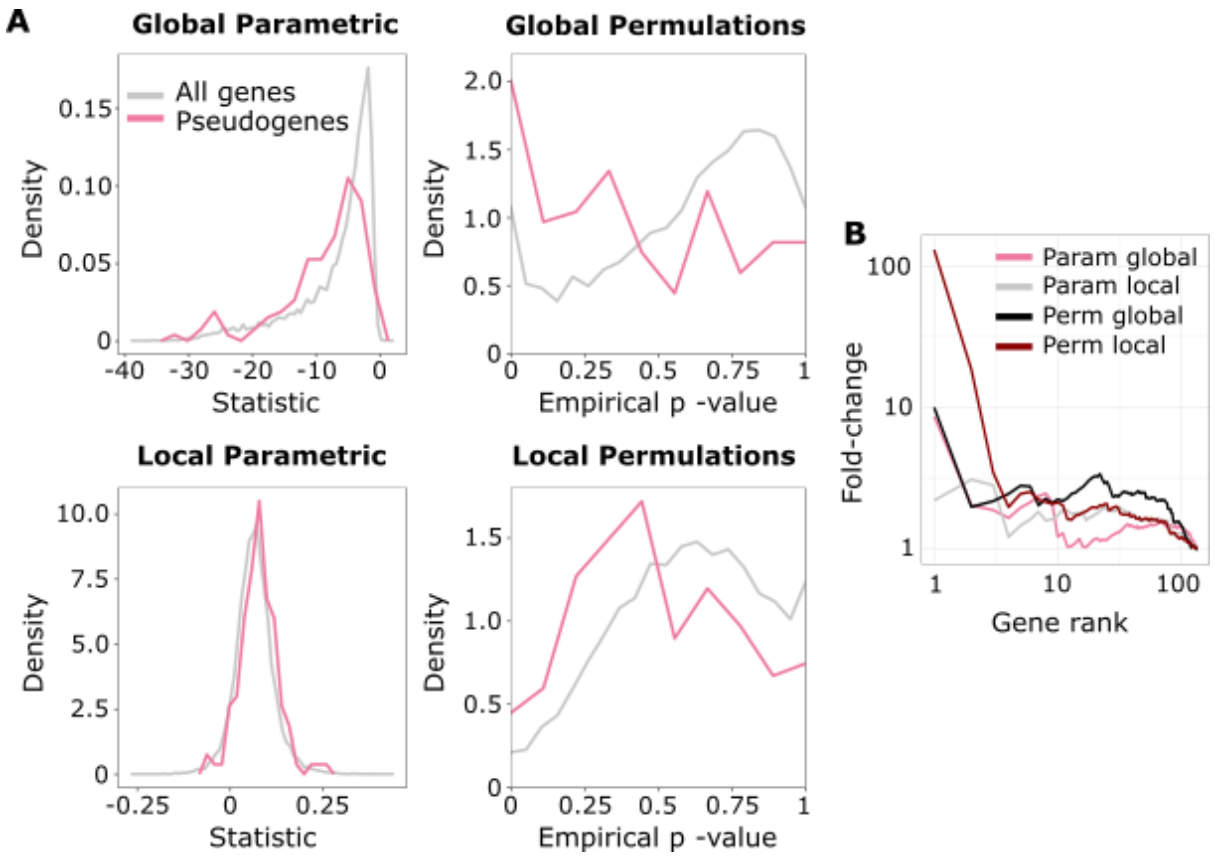


Figure 8 Permulations for Forward Genomics. Binary permulation methods improve Forward Genomics' positive-predictive value and power. (A) Distributions of Forward Genomics statistics and corresponding permulation p-values for local and global methods. Both global and local statistics show slight shifts (to the left for global statistics and to the right for local statistics) indicating enrichment of marine mammal pseudogenes under accelerated evolution (global AUC=0.6235; local AUC=0.6196). Permulation p-values show a more dramatic shift toward significant values for marine pseudogenes under accelerated evolution for the global method (AUC=0.6653) and about the same shift for the local method (AUC=0.6086) compared to parametric statistics. (B) Precision-recall curves for the enrichment of pseudogenes in marine-accelerated genes using parametric statistics and permulation p-values for both local and global methods. Permulated values represent a unique ranking in which ties in permulation p-values for genes are broken based on parametric statistics. Permulation methods perform at least as well as both global and local methods, indicated by curves that are higher at the left side of the plot.

Like other methods, PGLS demonstrates unexpected null behavior that varies across genomic datasets and phenotypes (Figure 2A). Although the null distribution of p-values for associations between the long-lived large-bodied phenotype and the stop codon counts show only a slight inflation of low p-values (5.2% of null p-values below 0.05) and otherwise nearly uniform distribution, tests using the marine phenotype and the transcription factor binding site counts show much different behavior. Permulations for associations between the marine phenotype and stop codon counts reveal that, although there may appear to be a meaningful enrichment of low observed p-values, such enrichment is observed even when analyzing permulated phenotypes. Conversely, although the enrichment of low observed p-values appears relatively less for associations between the long-lived large-bodied phenotype and transcription factor binding site counts in noncoding regions, such enrichment is indeed meaningful because it is greater than observed when analyzing permulated phenotypes. Together, these observations indicate that PGLS may exhibit aberrant statistical behaviors, that the exact nature of the behaviors may vary greatly across datasets, and that permulations are a valid strategy to identify and correct those behaviors.

2.2.3.4 Permulations Improve Power to Detect Genes Correlated with a Continuous Phenotype

When we use RERconverge to evaluate the long-lived large-bodied mammal phenotype, a continuous phenotype, we observe that the Type I error rate is in fact too low. We demonstrate this by performing one thousand permulations to generate 1,000 null statistics and p-values for each gene, calculating empirical p-values as the proportion of null statistics that were as extreme or more extreme than the observed statistic per gene. As shown in Figure 2A, the parametric null

p-value distribution for genes associated with the long-lived large-bodied phenotype is non-uniform, and in fact slopes down at low p-values. This indicates that observing small p-values due to chance alone happens less often in our dataset than we would typically expect compared to the standard uniform expectation. In practice, the result of the non-uniform null is overcorrection of parametric p-values using a standard multiple hypothesis testing correction. In other words, for this dataset, corrected parametric p-values are larger than they should be when using multiple hypothesis testing correction (such as a Benjamini-Hochberg correction) that assumes a uniform null. The null distribution of *empirical* p-values, however, does follow a standard uniform null by construction, so Benjamini-Hochberg corrected empirical p-values represent our true, higher confidence in a correlation between gene evolutionary rate and phenotypic evolution. We observe this increased confidence in our data – after multiple hypothesis testing correction, only 24 parametric p-values remain significant at an alpha threshold of 0.15 while 305 empirical p-values remain significant. Regardless of the increase in power, empirical p-values provide a more accurate representation of confidence in rejecting the null hypothesis, and thus are a more valid metric than parametric p-values.

2.2.3.5 Permulations Correct Pathway Enrichments for Genes with Correlated Evolutionary Rates

After generating null p-values and statistics from permulations for either binary or continuous traits, those values can be used to calculate null pathway enrichment statistics. Empirical p-values for pathways are then calculated as the proportion of null pathway enrichment statistics as extreme or more extreme than the observed statistic. This procedure corrects for gene sets with correlated evolutionary rates, that is genes whose rates will "travel in packs" regardless

of any relation to the phenotype (e.g., Figure 2B). Such groups of genes will tend to show enrichment more often than would be observed if the genes' rates were independent after conditioning on phenotype, resulting in false signals of pathway enrichment.

Permulations account for the non-independence problem by explicitly incorporating it into the null distribution used to calculate empirical p-values. In the demonstrated case of the Coenzyme Q Complex, only one permulation out of the ten depicted shows enrichment due to random chance (indicated by an asterisk * below the vertical bar in Figure 2B), which would correspond to a empirical p-value of 0.1 in this toy example. This interpretation is identical to the standard p-value interpretation—the proportion of times we expect to see a statistic as extreme or more extreme than observed *assuming that the null expectation is true*. In the case of permulations, we simply explicitly calculate the null expectation rather than using a predefined distribution (t-distribution, F-distribution, etc.). In the case of enrichment for a pathway with independent genes, the significance of the empirical p-value will agree with the significance of the parametric p-value because the null expectation from permulations agrees with the typical null expectation.

In the case of a pathway with genes with non-independent evolutionary rates, the empirical p-value will be larger than the parametric p-value because the empirical p-value will penalize for non-independence. An example with "Structural Maintenance of Chromosomes" genes shows that, although there is an apparent enrichment based on the observed phenotype, half (5 out of 10) of permulated phenotypes show at least as strong enrichment for an empirical p-value of 0.5. Therefore, although the pathway does appear to be enriched from parametric statistics, its enrichment is actually not exceptional given the null expectation for that set of genes.

Empirical p-values are calculated for every pathway individually. Table 1 shows top enriched pathways under accelerated evolution and decelerated evolution in association with the

long-lived large-bodied phenotype. While most significantly enriched pathways under decelerated evolution based on parametric p-values also demonstrate significant empirical p-values, many pathways under significant acceleration show non-significant empirical p-values. Thus, this phenotype shows little evidence for accelerated pathway evolution associated with phenotypic evolution.

Table 1 Continuous Phenotype Pathway Enrichment with Permulations. Top-enriched pathways with quickly evolving genes and slowly evolving genes in association with the long-lived large-bodied phenotype according to parametric p-values. Note that due to the number of pathways, the lowest possible Benjamini-Hochberg corrected permulation p-value is 0.0913. Boxes in green show significance at $\alpha = 0.25$. Note that many accelerated pathways that appear to be enriched based on parametric p-values are not enriched based on permulation p-values.

Pathway Enrichment							
Positive				Negative			
Pathway	Statistic	p-adjusted	perm p-adjusted	Pathway	Statistic	p-adjusted	perm p-adjusted
Olfactory Signaling	0.217	9.25e-43	0.199	Cytokine-Cytokine Receptor Interaction	-0.181	3.40e-20	0.0913
GPCR Signaling	0.0606	8.34e-7	0.596	Mitotic Cell Cycle	-0.132	6.03e-12	0.213
Biological Oxidations	0.150	1.10e-6	0.276	Immune System	-0.0600	1.54e-6	0.0913
Valine and Isoleucine Degradation	0.219	3.32e-5	0.354	DNA Replication	-0.122	2.81e-6	0.352
Fatty Acid Metabolism	0.215	8.26e-5	0.352	Fanconi Anemia	-0.212	4.45e-5	0.221

2.2.3.6 Comparison of Phylogenetic Simulations, Permutations, and Permulations

Alternatives to permulations include either permutations or simulations alone. Permutations involve randomly assigning phenotype values to species regardless of the underlying phylogenetic relationships among those species. Meanwhile, simulations refer to the first step of permulations – phenotype values are generated based on predicted phenotype evolution along the phylogenetic tree. However, unlike permulations, simulations do not include reassigning real values based on simulated values and thus do not preserve the distribution of the original phenotype values.

At the pathway level, permulations result in p-values that are about equally as conservative as phylogenetic simulations alone and more conservative than permutations alone (Figure 9). Both permulations and simulations are preferred to permutations because null phenotypes generated from permulations or simulations reflect the underlying phylogenetic relationships among species, while null phenotypes from permutations do not. Therefore, the empirical null generated from permulations or simulations more closely represents the true null expectation for phenotype evolution. Although permulations and simulations show similar performance, we prefer permulations because permulated phenotypes exactly match the distribution of observed phenotypes, and thus create null phenotypes uniquely tailored to a particular continuous phenotype of interest. Such matching eliminates statistical anomalies that can arise due to discrepancies in range and distribution of permulated phenotypes compared to observed phenotypes.

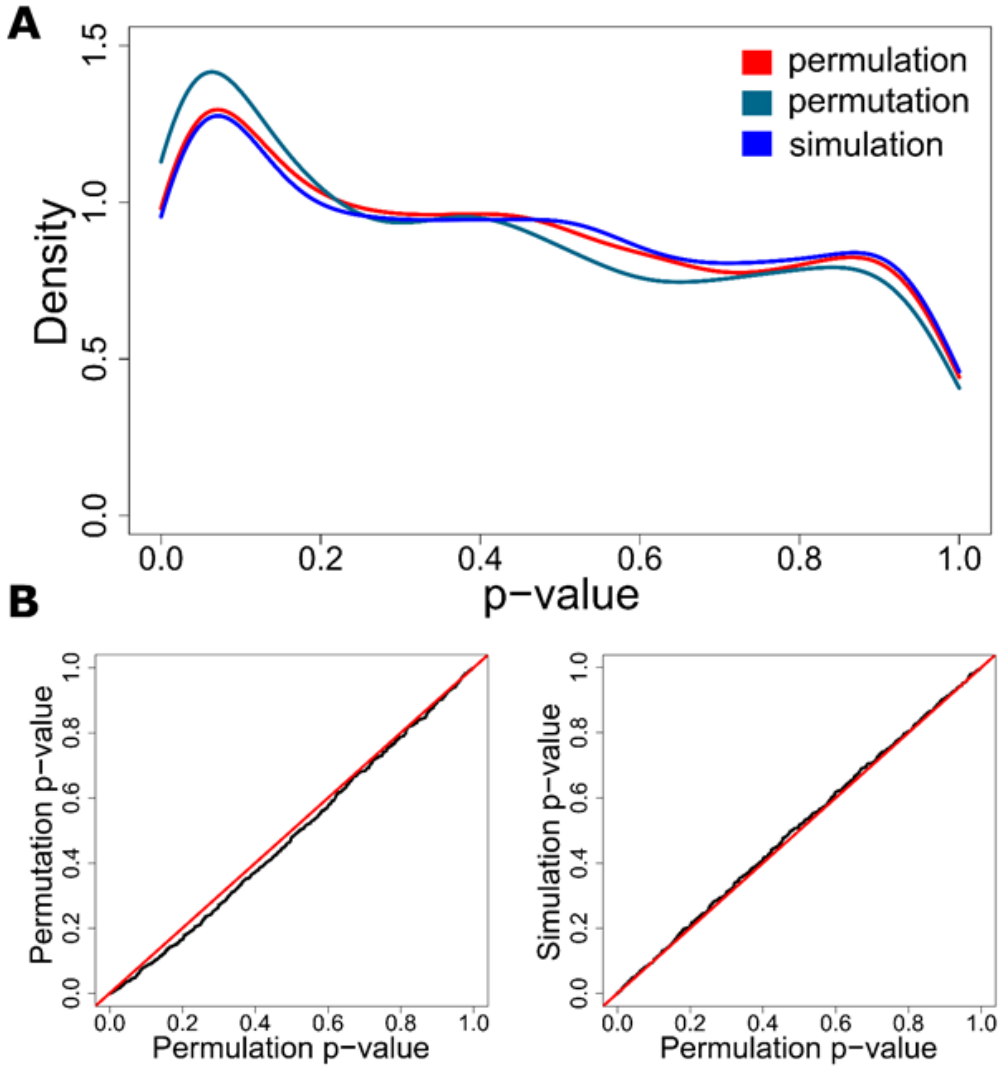


Figure 9 Comparison of Permulations, Permutations, and Simulations. Permulations p-values are more conservative than permutation p-values and about equally as conservative as simulation p-values. All plots demonstrate enrichment for canonical pathways associated with the long-lived large-bodied phenotype. (A) Density plots representing the empirical p-value distributions for the three methods to generate null p-values. Permulation and simulation curves are very similar, while the permutation curve demonstrates a stronger enrichment of low p-values and therefore less conservative p-values. (B) Q-Q plots comparing empirical p-values from permulations to empirical p-values from simulations and permutations also demonstrate that permulation p-values are more conservative than permutation p-values and about equally as conservative as simulation p-values.

2.2.4 Discussion

In the present work, we present permulations, a set of novel empirical methods to address problems of non-independence and bias in phylogenetic analysis. The methods use phylogenetic relationships among species alongside known values of an observed phenotype to inform Brownian motion simulations from which permuted phenotypes are then generated. By doing so, the methods empirically construct the possibly composite null distribution and account for this complexity in multiple hypothesis testing. For permulation of binary phenotypes, the phylogenetic characteristics preserved are the number of foreground branches and the underlying relationships among foreground branches. For continuous phenotypes, the exact distribution of phenotype values is preserved in addition to the underlying phylogenetic relationships among species.

From testing the strategy on binary and continuous phenotypes, we find that our permulation strategy is an effective approach for overcoming challenges in multiple testing with composite nulls in comparative phylogenetic studies. We discuss with examples how our binary and continuous permulation methods fix issues of both undercorrection and overcorrection of p-values for specified phenotypes, and subsequently improve the quality and confidence of prediction. Note that although our examples demonstrate the usefulness of permulations, they are not necessarily representative of how empirical null distributions will deviate from the typical null for all phenotypes over all phylogenies for all sets of genetic elements. In fact, we expect permulations to behave differently as those variables change, and thus the best way to determine how permulations will affect a particular data set is to run the permulation analyses.

Devising a systematic solution for such problems is difficult because the causes of complex null distributions in phylogenetic studies can be confounding. The necessity for incorporating

phylogenetic information to correct for phylogenetic effects is well understood (Felsenstein, 1985; Sakamoto & Venditti, 2018; Stone et al., 2011), and some systematic solutions have been designed to tackle the problem, including Phylogenetic Independent Contrast (PIC) (Felsenstein, 1985), Phylogenetic Generalized Least Squares (PGLS) (Grafen, 1989), phylogenetic autoregression (Cheverud & Dow, 1985; Gittleman & Kot, 1990), and phylogenetic mixed models (Hadfield & Nakagawa, 2010; Housworth, Martins, & Lynch, 2004; Lynch, 1991). However, systematic solutions usually make phylogenetic or distributional assumptions that can lead to inaccuracies if the assumptions do not accurately represent the data. For example, PIC makes an assumption that the observed phenotype evolved by Brownian motion, and it can lead to overcorrection when the selection giving rise to the observed data did not actually cause strong phylogenetic effects (Martins, 2000). In addition, phylogenetic mixed models usually assume that evolution along the phylogeny follows a Brownian motion process and that the resulting phenotype values are normally distributed. With the lack of full understanding of the underlying evolutionary mechanism, incorrect assumptions can lead to overcorrection or undercorrection of statistical confidence. Empirically correcting p-values using permulation methods allows us to circumvent the need to artificially deconstruct this unknown correlation structure in the data. Importantly, while our permulation methods are based on Brownian motion simulations, the simulated trait values themselves are not incorporated in the null phenotypes, and instead are only used as a way to incorporate phylogenetic dependencies in informing how trait values should be permuted across the phylogeny. In this sense, the choice of simulation model is not important.

For binary phenotypes, our permulation methods choose permuted foreground sets by matching the number of foregrounds and their underlying relationships to those observed in the actual phenotype. This approach of defining null phenotypes can be justified by phylogenetic non-

independence, a notion that arises from the implications of shared ancestry (Felsenstein, 1985). At the time of divergence, closely related species diverging from a common ancestor are likely to experience similar selective pressures as the ancestor as well as similar genetic predispositions to respond to the selection pressures. With progressing evolutionary time, the daughter species will evolve independently in response to their respective environments. Such similarities in environmental pressures and genetic predispositions diminish with increasing evolutionary distance between species, meaning that the variance in phenotype values will increase with increasing divergence in evolutionary time. Considering this phylogenetic non-independence and that adaptations to selection pressures are often assumed to be reflected in evolutionary rates, it is reasonable to preserve the pattern of divergence between foreground species to construct hypothetical null phenotypes, in finding correlations between evolutionary rates and phenotypes. It is impossible to pick a new set of foreground branches with perfectly matching divergence times, but matching divergence patterns can serve as a justifiable workaround because the general implications of shared ancestry on phylogenetic non-independence among the new set of foregrounds would apply in a similar way.

We develop two versions of permulation methods for binary phenotypes. The complete case (CC) algorithm produces one permuted phenotype from the master tree to apply for all genes simultaneously, while the species subset match (SSM) algorithm produces distinct permuted trees for each gene, accounting for the differences in species membership in different gene trees. This makes the CC method statistically imperfect. For example, a gene that is missing in some species will have a phylogenetic tree that is missing some branches. As a consequence of producing permuted trees from the master tree that contains all species, the CC method may not conserve the number and relationships of foregrounds across the permulations of the example gene (e.g., genes

3 and 4 in Figure 4 CC vs. SSM Permulations). In contrast, the SSM method accounts for differences in numbers and patterns of foregrounds among different genes and addresses each gene independently. This means that the SSM method is the ideal implementation of our concept of binary permulations. However, the CC method is both computationally much faster and accounts for the fact that existing comparative genomics methods take in phenotype inputs in different forms. For example, Forward Genomics requires one phenotype tree to apply for all genes, while HyPhy RELAX requires multiple phenotype trees with matching topology to each gene. Regardless of the statistical flaw, our results demonstrate that applying the CC method on Forward Genomics is beneficial for improving prediction (Figure 8). The CC method is significantly faster than the SSM method because it only produces one permuted tree for each permulation, instead of a heterogeneous set of permuted trees applying to different genes. Therefore, in the case of limited computational resources or very large datasets in which using the SSM method is infeasible, the CC method can serve as a good alternative.

Our results also demonstrate that binary permulations improve the sensitivity of RERconverge to identify significantly accelerated genes that are missing in many species (Figure 6D), i.e., genes with small trees. Because of lower species numbers, genes with small trees suffer from lower statistical power compared to genes with large trees (for example, the number of ways to permute a small tree is much fewer compared to a large tree). As such, pooling all the p-values together to perform multiple testing correction unfairly penalizes genes with small trees. Calculating empirical p-values from multiple empirical permutations is a way to correct for this imbalance in power by indirectly incorporating important covariates, which accounts for the number of foregrounds, backgrounds, and the ratio and phylogenetic relationship between them. Indeed, the pooled null empirical p-values have a uniform distribution (Supplementary Figure 1 in

(Saputra et al., 2021)), establishing the validity of applying standard multiple testing methods to identify significant divergence in evolutionary rates. Future work can evaluate if such benefits are similarly observed when applied to other comparative genomics methods.

Permulations grant increased power to detect genes associated with a continuous phenotype as suggested by the shape of the empirical null distribution (Figure 2). When p-values from permulations are compared with permutations or simulations of trait values, we find that permulation p-values are more conservative than p-values from permutations alone, and equally as conservative as p-values from simulations alone. This suggests that permulations offer a valid alternative to phylogenetic simulations. Importantly, permulations preserve the exact distribution and range of phenotype values, a critical characteristic related to the power of the correlation calculated between gene evolution and phenotype evolution. Thus, permulations more accurately match the power between observed and permulated statistics compared to observed and simulated statistics.

Although many of our tests of the permulation strategy were performed using RERconverge, permulations are applicable to any similar methods. When using permulations to calculate empirical p-values using Forward Genomics, an alternative evolutionary rates-based method, we show that we can quantify a realistic confidence level at which we believe a gene is under accelerated evolution in a subset of species. Even when using the Forward Genomics global method, a deprecated method that does not account for phylogenetic relationships among species, permulations improved the ability to detect accelerated evolution in marine pseudogenes (Figure 8). The improvement is likely due to permulations indirectly capturing phylogenetic information through their construction. For the Forward Genomics local method, permulations captured realistic confidence levels without losing the ability to detect accelerated evolution in marine

pseudogenes. Theoretical p-values directly from the Forward Genomics method (Figure 2A) show over half of the genome under significantly accelerated evolution related to the marine phenotype (12,438 out of 18,797 genes with the lowest possible Benjamini Hocberg corrected p-value), which is biologically highly unlikely (Eyre-Walker & Keightley, 1999; Eyre-Walker, Keightley, Smith, & Gaffney, 2002; Eyre-Walker, Woolfit, & Phelps, 2006; Kryukov, Pennacchio, & Sunyaev, 2007). Permulations reduce the number of genes under significantly accelerated evolutionary rates to a more modest number (889 genes if using the same confidence level cut-off) to more accurately reflect both the biology of the system and our confidence in identifying genes with significant evolutionary rate shifts.

Our permulations also reveal aberrant statistical behavior in PGLS. Designed to correct for phylogenetic relatedness when testing for coevolution of traits, PGLS indeed demonstrates a near-uniform empirical p-value distribution for one set of tests for coevolution of the long-lived large-bodied phenotype and gene stop codon counts. However, the method's behavior is dramatically different when testing for coevolution of gene stop codon counts with the binary marine phenotype. It likewise shows undesirable behavior when testing for coevolution of *STAT2* transcription factor binding site counts across noncoding regions. In addition to revealing a non-uniform null, the exact identity of noncoding regions with significant observed and permulation p-values is different, completely altering analysis results. These findings suggest that phylogenetic methods may behave in unexpected ways, and permulations are a valid strategy to investigate those behaviors and perform appropriate statistical corrections.

Finally, permulations demonstrate a crucial correction to pathway enrichment statistics that corrects for coevolution among genes in a pathway of interest. Since pathways often contain functionally related genes that evolve at similar rates, performing pathway enrichment treating

each gene as an independent observation is statistically incorrect and will result in erroneous conclusions. Performing permulations at the pathway level identifies pathways that are falsely shown to be enriched and correctly quantifies the confidence at which we may state that a pathway is enriched. We argue that a strategy like permulations is essential in virtually all cases of pathway enrichment calculations to account for gene non-independence driven by correlated evolutionary trends.

Overall, permulations are an important statistical consideration that should be undertaken to accurately report results from evolutionary rates-based analyses as presented here. Regardless of whether permulation allows for greater or fewer null hypothesis rejections at a given threshold, they are an accurate depiction of statistical power given a data structure. In the absence of a known parametric null that accurately represents a data set, a permulation-style approach is an important tool to calculate statistical confidence.

2.2.5 Materials and Methods

2.2.5.1 RERconverge

RERconverge finds associations between genetic elements and phenotypes by detecting convergent evolutionary rate shifts in species with convergent phenotypes. The method operates on any type of genetic element and has been used successfully for both protein-coding and noncoding regions. Prior to running RERconverge, phylogenetic trees for each genetic element are generated using the Phylogenetic Analysis by Maximum Likelihood (PAML) program (Yang, 2007) or related method, with branch lengths that represent the number of substitutions that occurred between a species and its ancestor. Raw evolutionary rates are converted to relative

evolutionary rates (RERs) using RERconverge functions *readTrees* and *getAllResiduals*, which normalize branches for average evolutionary rate along that branch genome-wide and correct for the mean-variance relationship among branch lengths (Partha et al., 2019). RERs and phenotype information are then supplied to *correlateWithBinaryPhenotype* or *correlateWithContinuousPhenotype* functions to calculate element-phenotype associations. Kendall's Tau associations are calculated for binary phenotypes, and Pearson correlation values are calculated for continuous phenotypes, both by default.

After calculating association statistics, signed log p-values for associations are used to calculate pathway enrichment using the rank-based Wilcoxon Rank-Sum test. The *fastWilcoxGMTAll* function in RERconverge calculates pathway enrichment statistics over a list of pathway annotations using all genes in a particular annotation set as the background.

2.2.5.2 Phylogenetic Simulations

As shown in Figure 3, each permulated phenotype is generated by first performing a phylogenetic simulation using an established phylogenetic topology. To generate the master tree, whose branch lengths represent the average evolutionary rates of all genetic elements in the dataset for each species, the function *readTrees* in RERconverge can be used. Next, the master tree and the trait values (binary or continuous) are used to compute the expected variance of the phenotype per unit time, and subsequently perform a Brownian motion simulation to simulate branch lengths; the R package *GEIGER* (Harmon, Weir, Brock, Glor, & Challenger, 2008) is used to perform both operations. Simulated values are then used in different ways for binary and continuous phenotypes to generate permulated phenotypes.

2.2.5.3 Implementation of Permulation Methods

In RERconverge, CC and SSM permulations are performed using the *getPermsBinary* function, by setting the argument "permmode" to "cc" or "ssm", respectively. The function requires the user to supply information on the original foreground species and their relationships by specifying 1) the names of the extant (tip) foreground species and 2) an R list object containing pair(s) of sister species whose common ancestor(s) is to be included in the foreground set as well (see examples in Supplementary Walkthrough in (Saputra et al., 2021)). Using these inputs, the function infers the original phenotype tree and assigns the phenotype values to the correct branches (1 for foreground, 0 for background), which is subsequently used as constraints for the permulation. Phylogenetic simulations are then run using the master tree to assign simulated branch lengths to the tree branches.

For the CC permulation, the *n* tip branches with the highest trait values from the simulation, where *n* is the number of observed tip foregrounds, are selected as the new foregrounds. The function then calls the *foreground2Tree* function in RERconverge with "clade" set to "all" to construct a binary tree with a foreground set that includes all branches (tip and internal) in the foreground clades. A valid permulation has the same number of internal and tip foreground branches as the original phenotype. Thus, permulated phenotypes with an incorrect foreground configuration are rejected and phenotype generation is repeated until the correct number of permulations is achieved. Note that the CC method uses the same permulated phenotype for every genomic element, so statistics for some genes will not be calculated for some permulations because of species presence/absence across genes. In other words, some genes will have fewer total permulations because of the way permulated phenotypes are constructed. The exact number of

foreground and background species may also differ across each permulated phenotype for the same gene.

The SSM permulation matches the tree topology of the permulated phenotypes to the tree of individual genes. To do this, the SSM permulation follows the same steps as described above, with an additional step of trimming off branches that are missing in the gene tree. In this case, the *m* longest tip branches (where *m* is the number of observed tip foregrounds in the *gene* tree) are chosen as new tip foregrounds to run *foreground2Tree*. Thus, in the SSM method, genes with different tree topologies will have different sets of permulations. However, for each unique topology, the number and phylogenetic relationships of the foregrounds are preserved. Figure 4 shows examples of CC- and SSM-permulated trees for 4 genes with distinct topologies.

For the continuous phenotype, the function *simpermvec* generates a permulated phenotype given the original phenotype vector and the underlying phylogeny with appropriate branch lengths. The master tree from the RERconverge *readTrees* function is appropriate to use for simulations. In most cases, the user will not have to use the *simpermvec* function directly – instead, the *getPermsContinuous* function that calculates null empirical p-values for genes correlations and pathway enrichments will call *simpermvec* internally.

After calculating empirical null statistics and p-values, empirical p-values per gene are calculated by finding the proportion of null statistics from permulated phenotypes that are as extreme or more extreme than the statistic calculated using the real phenotype. This proportion represents the proportion of times that random chance produces a concordance between gene and phenotype evolution that is as strong as the observed statistic, given the underlying structure of the data. In RERconverge, the *permpvalcor* function calculates the empirical p-values for a given set of permulation association statistics. Note that since empirical p-values are a proportion of total

permulations, the precision of empirical p-values is based on the total number of permulations performed. For example, with 1,000 permulations, the lowest reportable p-value is 0.001 and empirical p-values calculated as 0 must be reported as <0.001 because we only have precision to report p-values to the thousandths place.

Finally, to determine the number of permulations that can provide sufficient correction for systematic bias, the function *plotPositivesFromPermulations* can be used to plot how the number of significantly accelerated or conserved genetic elements changes with increasing number of permulations (e.g., Figure 6B). From the generated plot, users can determine the minimum number of permulations by evaluating when the number of positives start to stabilize.

2.2.5.4 Empirical p-values for Pathway Enrichment

Empirical null statistics and p-values for pathways are calculated using the empirical null statistics and p-values for individual genes. For each set of empirical null statistics generated from a particular permulated phenotype, genes are assigned the log of the empirical null p-value times the sign of the empirical null statistic for that permulation. Empirical null pathway statistics are calculated for each permulation using those values with the RERconverge function *fastWilcoxGMTall* that performs a Wilcoxon Rank-Sum test comparing values from genes in a pathway to values in background genes. The function *getEnrichPerms* calculates null enrichment statistics given a set of null correlation statistics, or, alternatively, *getPermsBinary* and *getPermsContinuous* calculate both null correlation and null pathway enrichment statistics simultaneously by default for the binary and continuous phenotypes, respectively. Empirical p-values for pathway enrichment are then calculated as the proportion of empirical null statistics that are as extreme or more extreme than the observed enrichment statistic using the *permpvalenrich*

function. Pathways that show significant parametric p-values and non-significant empirical p-values are likely cases of genes “moving in packs” and are not truly significantly enriched.

2.2.5.5 Phylogenetic Generalized Least Squares (PGLS)

PGLS analyses were conducted through R as implemented in the "nlme" package using the *gls* function. Within-group correlation structure was defined using the *corBrownian* function from the "ape" package and a master tree with branch lengths representing genome-wide evolutionary rates per species.

Noncoding regions were identified based on evolutionary convergence from phastCons scores across the 63 mammal species as described here: https://github.com/nclark-lab/RERconverge/blob/master/NoncodingRegionWorkflow. Stop codon calls per gene were obtained from Meyer et al. (Wynn K Meyer et al., 2018) and were based on genome-wide calls across species.

TFBS calls were obtained using the HOCOMOCO *STAT2* binding site motif based on position weight matrix scores. Calls for 29,880 noncoding regions corresponding to human chromosome 1 were used for analyses. Of those regions, 560 had a sufficient number of calls and variation in calls across species to calculate PGLS statistics.

2.3 Proper use of branch-site models to detect convergent positive selection

Branch-site models are a powerful tool to detect positive selection along branches in a phylogeny based on the ratio of nonsynonymous and synonymous mutations. When considering

convergent positive selection using the popular PAML implementation of branch-site models, it is essential to correctly perform tests for background positive selection to correctly distinguish genes under foreground branch-specific positive selection (convergent positive selection) from genes ubiquitously under positive selection on all or many branches of the phylogeny. Since many published manuscripts incorrectly use or incorrectly interpret branch-site models for positive selection, this work is a cautionary example and guide of how to correctly use those tests.

I completely all work in this section independently.

2.3.1 Introduction

Researchers have long used comparative biology to understand adaptations to biological challenges. One powerful comparative approach is to study independent evolutionary lineages that convergently experienced an evolutionary pressure. Traits that evolved in species sharing that pressure, but not in other species, could be inferred to be involved in an adaptive response. This strategy is now frequently applied using genome sequences from multiple species to identify genes or regulatory sequences potentially responsible for their convergent response.

Multiple methods have been developed to find such genes by scanning for shifts in a gene's overall evolutionary rate occurring preferentially in species sharing the evolutionary pressure. Those methods include RERconverge (Amanda Kowalczyk et al., 2019), Forward Genomics (Hiller et al., 2012b), PhyloAcc (Hu et al., 2019), and HyPhy RELAX (Wertheim, Murrell, Smith, Kosakovsky Pond, & Scheffler, 2015). It is important to clarify that rather than looking for shared changes at a particular nucleotide or amino acid site, these methods identify genes that are evolving at different rates in the convergently evolving species compared to all other species. They have

been applied to discover genes involved in phenotypes such as vision loss in subterranean mammals (Partha et al., 2017a; Prudent et al., 2016), transition from a terrestrial to a marine environment (Chikina et al., 2016; Hu et al., 2019), reduction of hair density (Amanda Kowalczyk et al., 2021), increased lifespan in mammals (Amanda Kowalczyk et al., 2020), complex sociality in bees (Kapheim et al., 2015), endosymbiosis in proteobacteria (Wertheim et al., 2015), and loss of flight in birds (Hu et al., 2019). Importantly, these methods have statistical designs that assure statistically significant genes experienced a rate change preferentially in the convergently evolving species. However, we have seen the rise of another approach to study convergent evolution that does not assure the same, and which could lead to false positive claims.

That concerning approach intends to identify genes showing evidence of positive selection on branches leading to the convergently evolving species, i.e., the foreground branches. Positive selection is inferred on foreground branches when a class of codons is found to have a d_N/d_S ratio significantly greater than one, which is a hallmark of positive selection (Hughes & Nei, 1988). Repeating this test for many genes serves as a screen for genes that responded adaptively to the convergent selective pressure. The tests are frequently performed using the branch-site models in the PAML package (Yang, 2007) that test for $d_N/d_S > 1$ on branches of interest.

The problem with this approach is that the results of branch-site models are misinterpreted; while they do test for positive selection on foreground branches, they do **not** test for the **absence** of positive selection on **background** branches. A significant result from a branch-site test could indicate either that the foreground species have $d_N/d_S >1$ and the background species have $d_N/d_S \leq 1$ (Figure 10: Foreground-Specific Positive Selection) or that both foreground and background branches have $d_N/d_S >1$ (Figure 10: Tree-Wide Positive Selection). This is because, in the case of a gene truly experiencing positive selection on all branches, the branch-site model fits better when

allowing positive selection on foreground branches, even though it would fit even better if positive selection were allowed on all branches. Therefore, **significant results from branch-site models alone do not reliably identify genes under positive selection associated with a convergent trait**. This approach runs the risk of inferring that genes evolving under positive selection on most or all branches, such as genes interacting with pathogens, are specifically responding to the convergent selective pressure, when in fact they are not. Unfortunately, branch-site models are frequently misinterpreted in this manner (Fang et al., 2014; Kim et al., 2011; McGowen, Tsagkogeorga, Williamson, Morin, & Rossiter, 2020; Nam et al., 2017; Sahm et al., 2018) and can lead to the erroneous conclusion that vast swaths of genomes are under positive selection in response to the focal convergent trait, when in fact many of those genes likely experienced tree-wide positive selection not specific to foreground branches.

To illustrate, we demonstrate how interpreting the branch-site models in this way leads to false inferences of positive selection in response to a convergent pressure. We then propose a simple "drop-out" test that filters out genes with evidence for positive selection more broadly than in the foreground species. The drop-out test is a sites model test using only non-convergent species to test for tree-wide positive selection (Figure 10: Sites Model Drop Foreground). Genes that demonstrate evidence for positive selection in species of interest from branch-site models **and** show no evidence for tree-wide positive selection are candidates for convergent positive selection unique to foreground species. We provide code and documentation to run our recommended tests (https://github.com/nclark-lab/bsmodels-dropout).

2.3.2 New Approaches

We strongly recommend the drop-out test to detect convergent positive selection using the branch-site analysis and provide a framework with detailed instructions. In brief, after performing branch-site tests to detect positive selection in foreground species, those foreground species should be removed and *codeml* sites models, or HyPhy equivalent (Wertheim et al., n.d.), should be analyzed to determine if positive selection occurs tree-wide or only in foreground species. Without the drop-out test, it is impossible to distinguish tree-wide positive selection from convergent positive selection along foreground branches.

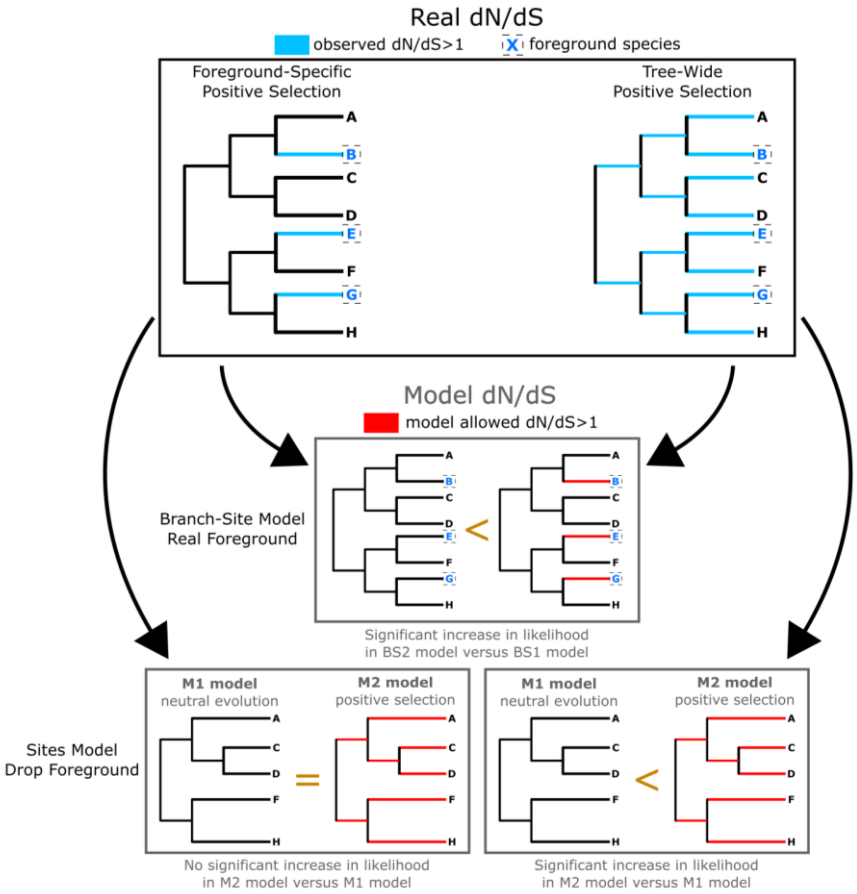


Figure 10 Distinguishing Tree-Wide Positive Selection from Foreground-Specific Positive Selection. Branch-site models alone cannot detect positive selection unique to foreground branches. Consider two cases shown in the top box: the case of positive selection unique to foreground branches (Foreground-Specific Positive Selection) and the case of positive selection on all branches (Tree-Wide Positive Selection). In both cases, branch-site models for positive selection on the foreground branches will show significant signals of positive selection because there is positive selection along those branches (Branch-Site Model Real Foreground). However, if foreground branches are removed and sites models to test for tree-wide positive selection are performed, the two cases will behave differently (Sites Model Drop Foreground). In the case of foreground-specific positive selection, sites models do not show signals of tree-wide positive selection. However, in the case of tree-wide positive selection, sites models will show signals of tree-wide positive selection through the drop foreground analysis.

2.3.3 Results and Discussion

We present branch-site models for positive selection and drop-out site models to test for tree-wide positive selection performed on simulated data and real data. These results demonstrate that the drop-out tests are essential to distinguish convergent positive selection from tree-wide positive selection.

Simulated data depicted in Figure 11 were generated with either a class of positively selected sites specifically along foreground branches (branch-site model simulations) or a class of positively selected sites tree-wide (sites model simulations). Regardless of simulation type, all replicates showed strong signals of positive selection along foreground branches (Fig. 2: BS2-BS1 Real Foreground). Thus, using only branch-site models, it is impossible to distinguish foreground-specific from tree-wide positive selection. Only upon removing foreground branches and running sites models did it become clear that genes simulated under foreground-specific positive selection showed little signal for tree-wide positive selection while genes simulated under tree-wide positive selection showed strong signal for such selection (Figure 11: M2-M1 Drop Foreground).

Tests on real data confirm that such concerns prevail in genuine gene sequences (Figure 12). To demonstrate, we arbitrarily assigned foreground species based on a “large” phenotype calculated from adult mass. It included 14 out of 35 mammals in the dataset, including large marine mammals, large African mammals, gorilla, and panda. Note that our test is to confirm that tree-wide positive selection can be incorrectly interpreted as foreground-specific positive selection, so the choice of foreground species is irrelevant. Any set of foreground species, regardless of phenotypic convergence, could behave similarly. We demonstrate that four randomly selected genes with evidence of positive selection from a branch-site test also show evidence of tree-wide

positive selection when foreground species are removed (Figure 12: Real Foreground *vs.* Drop Foreground). Therefore, none of the genes could be concluded to be under convergent positive selection in conjunction with this “large” phenotype and should not be connected its convergent evolution.

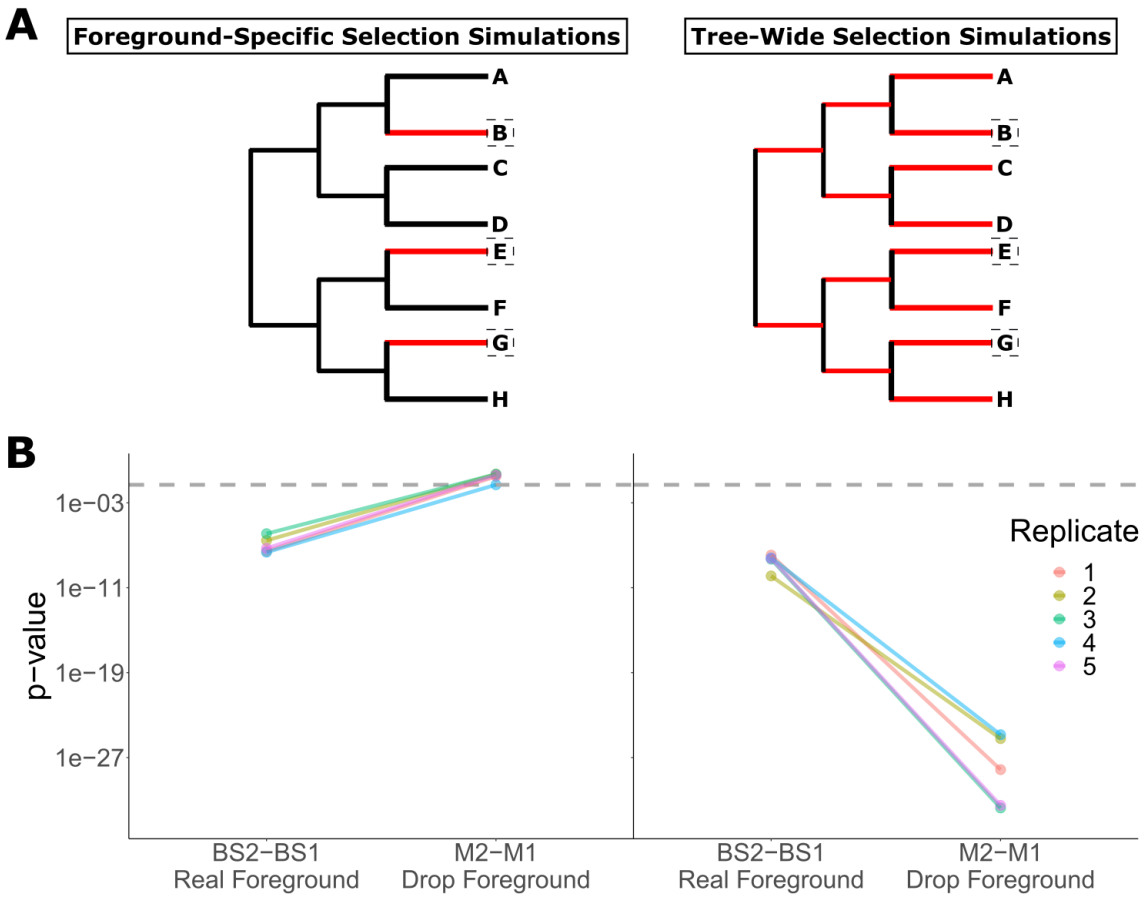


Figure 11 Testing for Positive Selection Using Simulated Data. Simulated data demonstrate that branch-site models for positive selection can not distinguish positive selection unique to foreground branches from tree-wide positive selection. A) Topology of trees used to simulate sequences. In red are branches with sites under positive selection (*dN/dS* > 1). Boxes indicate foreground branches, which were used for branch-site models for positive selection and removed to test for tree-wide positive selection using sites models. B) Sites models distinguish foreground-specific positive selection from tree-wide positive selection. P-values from likelihood ratio tests for BS2-BS1 models (branch-site models) are indistinguishable between data simulated to have foreground-specific versus tree-wide positive selection. After foreground species are removed (Drop Foreground), p-values from M2-M1 model comparisons (sites models) are no longer significant for genes simulated with foreground-specific positive selection, while p-values remain significant for genes simulated under tree-wide positive selection, thus distinguishing the two cases. Dashed line indicates statistical significance threshold at an alpha of 0.05.

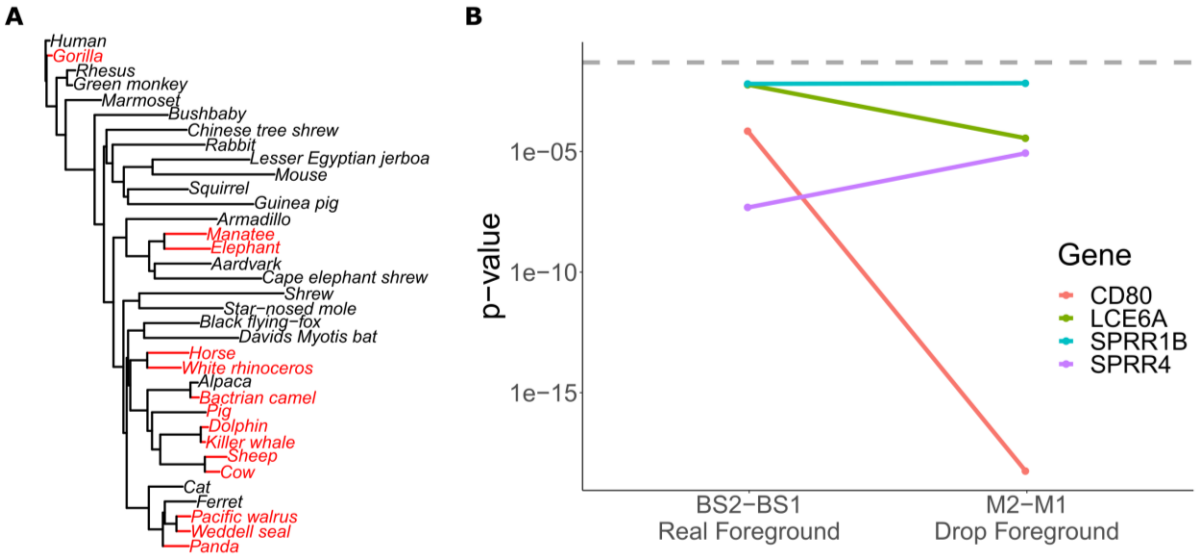


Figure 12 Testing for Positive Selection Using Real Data. Real data demonstrate genes under tree-wide positive selection in mammals. A) Mammal phylogeny. Included are all species used in sites and branch-site models. Foreground species are labeled in red. The arbitrary phenotype "large body size" was used solely for demonstration purposes. B) All genes with signals of positive selection from branch-site models (BS2-BS1) also show evidence of tree-wide positive selection after dropping foreground species (M2-M1). Depicted are four genes that demonstrate patterns of positive selection in foreground branches (according to branch-site models) and tree-wide (according to sites models). Note that without running sites models on non-foreground species, we may have falsely concluded that the genes depicted are under convergent positive selection unique to foreground species. Dashed line indicates statistical significance threshold at an alpha of 0.05.

We again emphasize that to identify genes under convergent positive selection unique to foreground species it is essential to perform branch-site model tests for positive selection and then filter out genes with evidence for tree-wide positive selection using drop-out tests. It is possible that studies that performed branch-site tests without drop-out tests identified genes with foreground-specific positive selection, but without a drop-out test it is unclear which genes those are. Alternatively, the authors could use a method mentioned in the introduction to test for

convergent rate shifts in association with their convergent trait or an additional solution presented by Davies *et al.* (Davies, Bennett, Tsagkogeorga, Rossiter, & Faulkes, 2015), who used clade models to control for tree-wide positive selection (Bielawski & Yang, 2004). Another possible strategy to control for tree-wide selection is to swap the foreground and background branches in the branch-site test. While this swap test is a good idea, its power depends on the number of foreground and background species and the relative heterogeneity of positive selection among those species, and thus it is less reliable than the drop-out test. In our simulations, the swap test sometimes failed to detect tree-wide positive selection and sometimes erroneously detected tree-wide positive selection in simulations of foreground-specific positive selection. On the other hand, the drop-out test always accurately distinguished foreground-specific positive selection from tree-wide positive selection, and thus we recommend the drop-out test.

We postulate that it is unlikely to find large numbers of genes under positive selection in association with a trait because convergent positive selection is difficult to detect. In previous work on various phenotypes, we detected relatively little positive selection (Chikina et al., 2016; Amanda Kowalczyk et al., 2021, 2020; Partha et al., 2017a), suggesting convergent positive selection strong enough to be detected by branch-site models is rare. Previous work on simulated data also suggests that branch-site tests for positive selection are conservative and are likely underpowered in cases where species are distantly-related and near synonymous mutation saturation (Gharib & Robinson-Rechavi, 2013). We interpret such rarity not as a downside, but an exciting indicator of novelty when robust signals of convergent positive selection are identified. Branch-site models are a useful tool to make such discoveries, especially when implemented with a control test to rule out phylogeny-wide positive selection.

2.3.4 Materials and Methods

2.3.4.1 Simulations

Data were simulated using *evolver* with the MCcodonNSbranchsites and MCcodonNSsites control files (Yang, 2007). For both, kappa = 1.7, # sites = 1000 codons, and tree length = 3. All non-stop-codon frequencies were set to 0.01639344 for simplicity. In sites models, three site classes were defined with omega = 0.1 at frequency = 0.15 (purifying selection), omega = 1 at frequency 0.15 (neutral evolution), and omega = 3 at frequency = 0.7 (positive selection). For branch-site models, four site classes were defined as follows: foreground and background omega = 0.1 at frequency 0.075 (uniform purifying selection), foreground and background omega = 1 at frequency 0.075 (uniform neutral evolution), foreground omega = 3 and background omega = 0.1 at frequency 0.035 (foreground positive selection, background purifying selection), and foreground omega = 3 and background omega = 1 at frequency 0.35 (foreground positive selection, background neutral). Note that sites and branch-site simulations were designed such that the same proportion of sites were accelerated either in the foreground (for branch-site models) or tree-wide (for sites models). Five replicates were generated for each type of simulation. Example control files are available at https://github.com/nclark-lab/bsmodels-dropout.

2.3.4.2 Real Data

The mammal phylogeny was pruned from a previously reported phylogeny (Wynn K Meyer et al., 2018) to remove closely-related species. Such pruning reduces branch-site and sites model runtime, helps the models converge without reducing phylogenetic information, and remedies oversampling in some clades (such as primates). Mammal mass values were from the

Anage Longevity Database (Tacutu et al., 2018), and mammals above 100,000g were defined as "large" foreground species. Gene alignments were from the UCSC 100-way alignment (Blanchette et al., 2004; Harris, 2007; Kent et al., 2002). Selected genes were pseudorandom – all genes tested had alignment file sizes equal to 15 kilobytes (a proxy for sequence length) and at least 50 mammal species of the 62 included in the full alignment. Selection of short genes allowed for fast runtimes and therefore testing of numerous genes (133 total genes) and helped ensure model convergence. Selecting highly conserved genes (present in most species) mitigated genome quality concerns and ensured that genes had similar foreground/background species counts. The four genes shown are the only genes that demonstrated foreground positive selection from branch-site models.

2.3.4.3 Branch-site and sites models

Branch-site models were conducted using BS1 and BS2 templates with tree topologies and foreground species as shown in Figure 11 and Figure 12. Briefly, both models specify model = 2 for two or more *dN/dS* ratios for branches and NSsites = 2 for selection. BS1 specifies fix_omega = 1 and omega = 1 to fix foreground omega at 1 while BS2 specifies fix_omega = 0 and omega = 0.4 to allow foreground sites to have omega greater than 1.

Sites models were conducted using M1 and M2 templates topologies as show in Figure 11 and Figure 12 with foreground species removed. Briefly, both models specify fix_omega = 0 and omega = 0.4 to estimate omega values. M1 specifies NSsites = 1 for negative selection and neutral evolution with no codon class allowed to have omega values greater than 1. M2 specifies NSsites = 2 to build on model M1 by allowing an additional codon class with omega values greater than or equal to 1.

All parameter values are available in control files in the GitHub repository (https://github.com/nclark-lab/bsmodels-dropout).

3.0 Convergent evolutionary rate shifts in coding sequence underlie the evolution of longevity in mammals

I used the methods described in the previous chapter with two phenotypes, the first of which is longevity in mammals. How species age and evolve to live longer is of great interest to the scientific community because of its implications for healthy aging in humans. If we can fully understand the genomics of aging, we can help people live longer, healthier lives. This work was of further interest because it represented the first successful use of RERconverge with a continuous phenotype, and one of the first ever uses of any similar methods with continuous phenotypes.

I completed the entirety of work described in this chapter individually, with the exception of assistance with running branch-site models for positive selection and gene tree generation that were conducted by Raghavendran Partha. This work was originally published in *eLife* at (Amanda Kowalczyk et al., 2020).

3.1 Introduction

Humans age in the sense that an individual's probability of dying increases as a function of time lived. Interestingly, this trend is not true of all species but is true of mammals generally (Jones et al., 2014). Numerous hypotheses of mammal-specific aging exist, including the antagonistic pleiotropy hypothesis (Williams, 1957) and the mutation accumulation hypothesis (Medawar, 1952), both of which refer to changes within an individual throughout its lifetime that

result in aging. The antagonistic pleiotropy hypothesis postulates that genes that are beneficial to early life become detrimental later in life. Such genes are retained because they increase early-life reproductive output and thus increase fitness. The mutation accumulation hypothesis predicts that a gradual accumulation of errors in DNA sequence as a result of repeated replication during a lifetime's worth of cell divisions will lead to a gradual breakdown of functionality. Support for both hypotheses has been found in individual species. Recent work has identified disease-related SNPs in age-related genes that are beneficial in early life and detrimental in later life in humans, thus indicating selective pressures associated with gene evolution related to aging and supporting the antagonistic pleiotropy hypothesis (Rodríguez et al., 2017). A study of SNVs in the human brain found that number of mutations was positively correlated with age and that mutations were at loci associated with age-related disease, thus supporting the mutation accumulation hypothesis (Lodato et al., 2018). However, in human populations, variability in the aging phenotype is limited and many confounding biological changes are correlated with aging, making it difficult to pinpoint specific biological processes that are causal and thus amenable to manipulation.

On the other hand, lifespan varies dramatically (>100-fold) across mammals (Nowak, 1999), making comparative genomics a fruitful avenue for aging research. Numerous studies have investigated the genomic features of mammals with extreme lifespan such as bats (Foley et al., 2018; Seim et al., 2013), naked mole-rats (Kim et al., 2011), whales (Keane et al., 2015), and elephants (Sulak et al., 2016) to identify potential causative genetic changes. In *Myotis*, the longest-lived bat genus, species show lack of telomere shortening and corresponding expression changes in telomere maintenance and DNA repair genes (Foley et al., 2018). Comparative genomics studies have also suggested that changes in the insulin growth factor 1 pathway may enable increased lifespan and cancer resistance in *Myotis brandtii* (Brandt's bat) (Seim et al.,

2013). Similarly, naked mole-rats show differential regulation of genes associated with macromolecule degradation, mitochondrial function, and TERT, a gene associated with telomere maintenance, as well as changes to genes related to tumor suppression (Kim et al., 2011). In primates, genes associated with cardiovascular function, coagulation, and healing have been demonstrated to show evolutionary correlations with lifespan (Muntané et al., 2018). Sequencing of the bowhead whale genome revealed species-specific changes in DNA repair, cell cycle, and aging genes (Keane et al., 2015). In elephants, a striking increase in TP53 copy number has been linked to increased cancer resistance enabling longer lifespan (Sulak et al., 2016). Despite compelling results, these single- and limited-species studies have limitations. The species studied differ from their nearest sequenced relatives in multiple physiological traits as well as millions of nucleotides. Thus, while single-species studies have yielded some credible candidates for genes associated with increased lifespan, it is difficult to know to what extent these represent insights into the universal mechanisms of lifespan regulation rather than species-specific adaptation or coincidental neutral changes. In this study, we develop new methodology to evaluate the relationship between the evolutionary constraint of genes and pathways and quantitative lifespan traits in an unbiased, genome-wide, pan-mammalian analysis.

The wide range of lifespans across the mammalian phylogeny (Figure 13A) provides the ideal dataset to investigate lifespan from a comparative genomics perspective. Because independent changes in lifespan occurred repeatedly in the mammalian species tree, lifespan can be viewed as a convergent trait. Molecular features that correlate with convergent changes in lifespan therefore may also occur repeatedly across a variety of organisms. In our study we use protein evolutionary rates quantified as the number of amino acid substitutions on a phylogenetic

branch to infer convergent rate shifts associated with lifespan traits across the mammalian phylogeny.

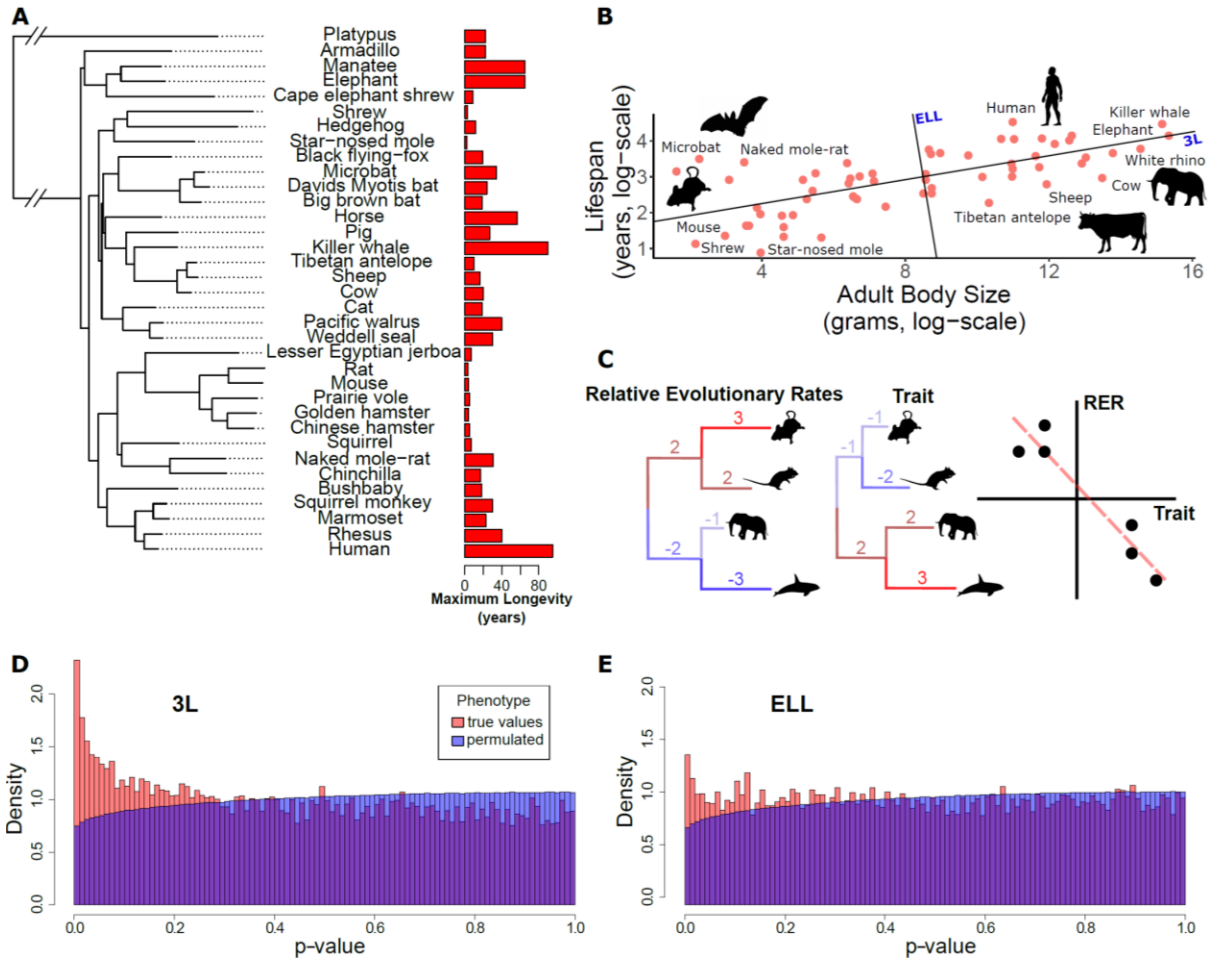


Figure 13 RERconverge and Longevity. Many genes have evolutionary rates correlated with longevity phenotypes as demonstrated by analysis with RERconverge. A) A subset of species used for this analysis alongside their maximum longevity values. Lifespan varies widely across mammals independent of phylogeny. B) Mammal body size and maximum lifespan values for 61 species. Lines represent the 3L phenotype and the ELL phenotype (also see Figure 14). C) RERconverge pipeline to find correlation between relative evolutionary rates of genes and change in lifespan phenotypes. D) and E) Distribution of p-values from correlations between evolutionary rates of genes and change in the 3L and ELL phenotypes indicate an enrichment of significant correlations (also see Figure 29).

Evolutionary rates are useful for linking phenotypes to genes because they reflect evolutionary constraint experienced by a genetic element (Jianzhi Zhang & Yang, 2015). In the absence of diversifying selection, genetic elements that support a specific trait are expected to be more constrained in species where the trait has a larger contribution to fitness. In agreement with this expectation, multiple studies have shown that a genetic element providing a function less important for a given species is under less constraint and hence exhibits a faster evolutionary rate (Clark et al., 2013; Janiak, Chaney, & Tosi, 2018; Roscito et al., 2018; Wertheim et al., 2015). Reciprocally, when an element becomes relatively more important, its rate is expected to slow. Thus, in cases of phenotypic convergence, rates can be exploited to reveal important genes associated with the phenotype, such as changes to muscle and skin genes associated with the mammalian transition to a marine environment (Chikina et al., 2016) and loss of constraint of vision-related genetic elements in subterranean mammals (Partha et al., 2017a; Prudent et al., 2016). Rate shifts can thus provide an evolutionary perspective on the contribution of genes, non-coding elements, or pathways to phenotypes of interest (Hiller et al., 2012b). Here we report the genome-wide, pan-mammalian correlations between evolutionary rates of genes and lifespan phenotypes.

3.2 Results

In mammals, lifespan is strongly positively correlated with adult body size such that the largest mammals (whales) are longest-lived and the smallest mammals (small rodents) are shortest lived (Figure 13B). However, if lifespan is corrected for body size, species including bats, the

naked mole-rat, and some primates are clearly exceptionally long-lived given their body sizes. The term "longevity" is applied to both phenotypes and previous studies have focused on small numbers of species at both phenotypic extremes to find genetic and physiological explanations for the extended phenotypes (Foley et al., 2018; Keane et al., 2015; Kim et al., 2011; Seim et al., 2013; Sulak et al., 2016). In our study we explicitly distinguish two different extended longevity traits: the "long-lived large-bodied" trait (3L) and the "exceptionally long-lived given body size" trait (ELL). Using maximum lifespan and body size data (Tacutu et al., 2018) we define the 3L and ELL phenotypes (Supp. File 1 in (A. Kowalczyk, Partha, Clark, & Chikina, 2020)) to be the first and second principal components of body size and maximum lifespan (Figure 13B, Figure 14). The resulting trait values are orthogonal with respect to each other, ensuring that they can be analyzed independently.

Having defined the 3L and ELL phenotypes, we compute the association between these phenotypes and protein-specific relative evolutionary rates (RERs) using the RERconverge method (Amanda Kowalczyk et al., 2019; Partha et al., 2019) (Figure 13C and Methods). Relative evolutionary rates quantify the deviation in evolutionary rate of a protein along a specific phylogenetic branch from proteome-wide expectations. Negative RERs indicate fewer substitutions than expected due to increased constraint. Positive RERs correspond to more substitutions than expected, which could arise due to relaxation of constraint or positive selection.

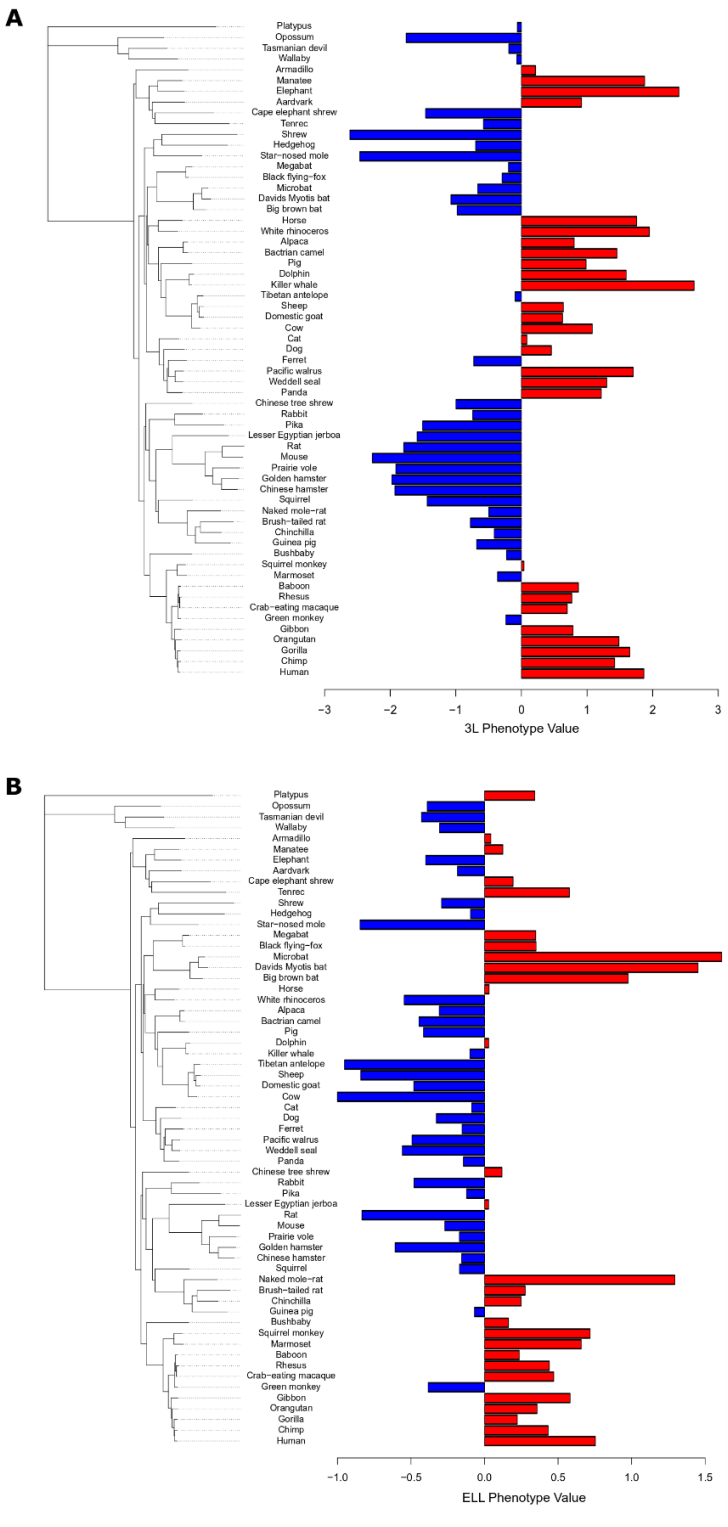


Figure 14 Longevity Phenotypes. 3L phenotype values (A) and ELL phenotype values (B) for 61 mammal species alongside mammalian phylogenetic tree.

After computing correlations between all protein relative evolutionary rates and the 3L and ELL phenotypes, we find an excess of low p-values (Figure 13D and Figure 13E). In order to evaluate how much of the signal is due to true association with the phenotypes, we used a phylogenetically restricted permutation strategy (termed "permulations", see Permulation Analysis (Phylogenetically-Restricted Permutations)) to generate null correlation statistics. We find that the real and permuted p-value distributions are indeed different, indicating that a large fraction of genes are correlated with lifespan phenotypes. We quantify the fraction of non-null genes using the π_1 method (John D. Storey, 2003). Using our permulation p-values as the null distribution, the fraction of true positives was inferred to be ~15% (π_1=0.153) for the 3L phenotype and ~7.5% (π_1=0.075) for the ELL phenotype. Using the theoretical uniform null distribution, the corresponding values are π_1=0.108 and π_1=0.021, respectively. Regardless of the null distribution choice, our analysis clearly demonstrates a significant molecular signal for gene evolutionary rates correlated with lifespan phenotypes, with a considerably higher number of associations for the 3L phenotype.

Our analysis investigates both positive and negative correlations between evolutionary rates of genes and changes in lifespan phenotypes. Positive correlations represent genes with faster evolutionary rates in species with high 3L and ELL phenotype values relative to species with low phenotype values. Conversely, negative correlations represent genes with slower evolutionary rates in species with high 3L and ELL phenotype values relative to species with low phenotype values. Note that correlation directionality is relative – faster evolution in low phenotype values corresponds to slower evolution in high phenotype values and vice versa. However, the choice has important consequences for biological interpretation and necessitates an in-depth discussion.

RERs reflect the amount of constraint on the genetic element. A decrease in RER implies greater evolutionary constraint and a greater contribution of that gene to fitness. An increase in RER, on the other hand, has two nearly opposite interpretations. An increased RER could arise because of a relaxation of evolutionary constraint driven by a reduced contribution to fitness. Alternatively, an increased RER could arise due to positive (also termed directional) selection, which implies that that the gene is actively undergoing directed evolution and thus could be contributing to trait-related innovation. However, since there is no default evolutionary rate for protein coding sequence, genes evolving *slower* in *long-lived* species could just as easily be interpreted as evolving *faster* in *short-lived* species. In our study, we have chosen to interpret the rate changes with reference to the effect in the long-lived species. The ancestral mammal is believed to have been small and short-lived, and thus large values of 3L and ELL are derived traits. Consequently, our interpretation reduces to assuming that change in phenotype and shifts in evolutionary rate coincide.

With this interpretation, negative correlations imply that a gene is more important in species with large values of 3L and ELL, while positive correlations imply either a relaxation of constraint or positive selection in species with large values of 3L and ELL. While such positively correlated genes would be of great interest if they indeed represent molecular innovations underlying evolution of extended lifespan, we find relatively fewer such genes, no evidence of positive selection among them (see Supp. File 6 in (A. Kowalczyk et al., 2020)), and no enriched pathways associated with them.

We thus focus our analysis on the negatively correlated genes, which we interpret as being under increased purifying selection in species with high longevity (3L or ELL) values. It is theoretically possible for negative correlations to be caused by accelerated evolution in small and

short-lived species because of directional selection associated with development of low longevity values. However, this possibility can be ruled out via branch-site models for positive selection using the low longevity species as foreground, as these show little evidence for positive selection (see Supp. File 6 in (A. Kowalczyk et al., 2020)). Together, these analyses support a single interpretation of the main rate convergence signal as a *decrease* in evolutionary rate, and thus an *increase* in purifying selection, experienced by species with large 3L and ELL values. It is important to emphasize that these genes are unlikely to have contributed to molecular innovation that lead to the establishment of 3L and ELL traits, but rather these represent existing biological systems that become especially important after the traits are established (see Discussion).

While we observe a clear excess of genes at low p-values, we focus on pathway enrichment analysis which both demonstrates a stronger signal and facilitates interpreting our results in the context of existing knowledge. We investigate enriched pathways for both 3L and ELL phenotypes using a rank-based method (see Supp. File 2 in (A. Kowalczyk et al., 2020)). After performing standard multiple-hypothesis testing corrections on the empirical p-values from permulations, there remains considerable pathway-level signal underlying the 3L and ELL traits. Both the gene-level and pathway-level results were highly robust to species removal, which indicates the biological pathways revealed here are important for longevity across mammals and are not restricted to specific species (see Species Robustness through Subtree Analysis, Figure 15 and, Figure 16, and Supp. File 4 and Supp. File 5 in (A. Kowalczyk et al., 2020)).

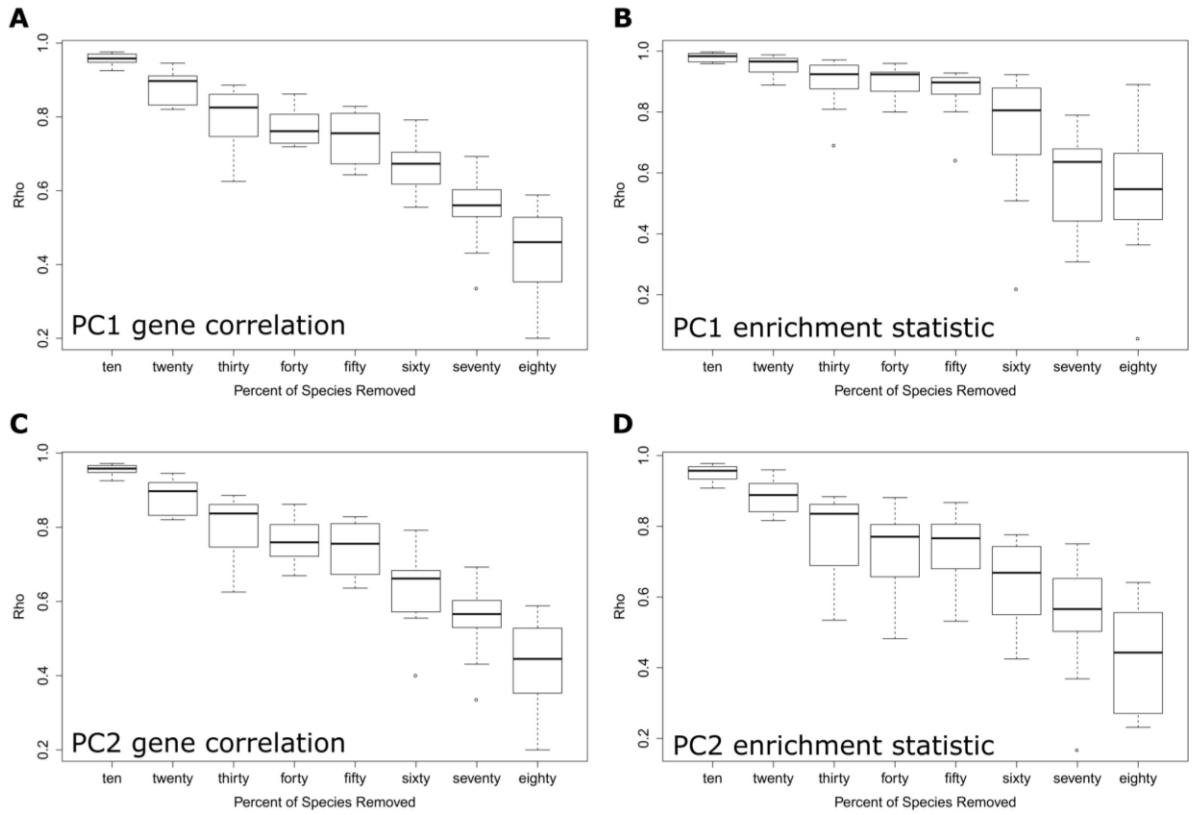


Figure 15 Random Subtree Analysis. Each panel demonstrates the correlation between results using all species and results with ten to eighty percent of species removed. Panels A and B show results from the PC1 phenotype 3L and panels C and D show results from the PC2 phenotype ELL. Panels A and C show correlations between gene results as quantified by the sign of the correlation statistic times the negative log p-value of the gene correlation. Panels B and D show correlations between enrichment results as quantified by the sign of the pathway enrichment statistic times the negative log p-value of the pathway enrichment statistic. There is a strong relationship between results from the full dataset and results from partial datasets, which suggest that the results are generally robust to species presence and absence.

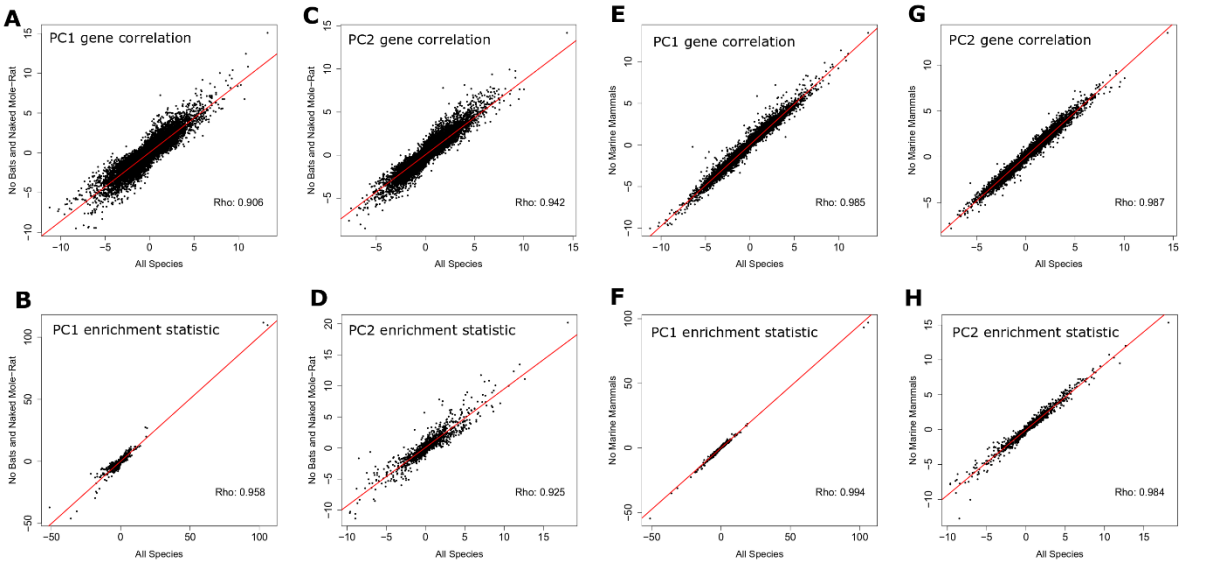


Figure 16 Targeted Subtree Analysis. Correlation statistics for genes and enrichment statistics for canonical pathways plotted with statistics calculated from data with bat and naked mole-rat removed (A, B, C, and D) and with marine species removed (E, F, G, and H). Points plotted are the sign of the correlation statistic times the negative log p-value of the correlation for genes (A, C, E, and G) and the sign of the enrichment statistic times the negative log p-value of canonical pathway enrichment (B, D, F, and H). These data indicate a strong relationship between results from the full species set and the partial species sets, which suggests the bats, naked mole-rat, and marine mammals do not strongly influence the results.

Among pathways under increased constraint in 3L species, we find a striking abundance of pathways related to cancer control. Those pathways can be organized into the broad categories of "cell cycle control", "cell death", and "innate and adaptive immunity", and they also include other cancer-related pathways such as p53 regulation and telomere maintenance (Figure 17A). We likewise see a significant enrichment in cancer-related genes more broadly. We compared correlation statistics between known cancer genes from Bailey et al. (2018) and found that tumor suppressor genes had significantly lower correlation statistics (rho) than all genes (Wilcoxon rank-

sum p-value=4.114e-8) while oncogenes had no significant difference in correlation statistics compared to all genes (Wilcoxon rank-sum p-value=0.3745) (Figure 18). This may indicate preferential purifying selection on tumor suppressor genes in large, long-lived species. Considering these findings, our 3L results can be naturally interpreted in the context of Peto's Paradox (Peto, 2016). The paradox reasons as follows: if all cells have a similar probability of undergoing a malignant transformation, organisms with more cells should have a greater risk of developing cancer. However, empirical cancer rates do not vary with body size (Peto, 2016), which implies that larger animals harbor mechanisms to suppress cancer rates. Top 3L constrained pathways are associated with multiple cancer control mechanisms, including DNA repair, cell cycle control, cell death, and immune function (Figure 17B and Figure 17C). A normal cell's transformation to malignancy involves failure of all these processes, and our analysis suggests that 3L animals are invested in the maintenance of each of their associated pathways through increased purifying selection. Based on enrichment and permulation results, we can infer that cell cycle fidelity, an early step in cancer development, is most important over evolutionary time scales for 3L species. Further, there is no evidence for enrichment of pathways associated with metastasis and angiogenesis, later steps in cancer development. This finding suggests that large, long-lived species have experienced increased selective pressure to protect pathways involved in early cancer stages but not later stages, perhaps because the most severely negative fitness impacts of cancer are felt earlier in its development. Species-specific cancer control mechanisms have been identified in individual species, such as increased *TP53* copy number in elephants (Sulak et al., 2016), but we show here that investment in cancer control is key to longevity across the entire dmammalian phylogeny because top enriched pathways for the 3L phenotype do not depend on a handful of species (Figure 15 and Figure 16).

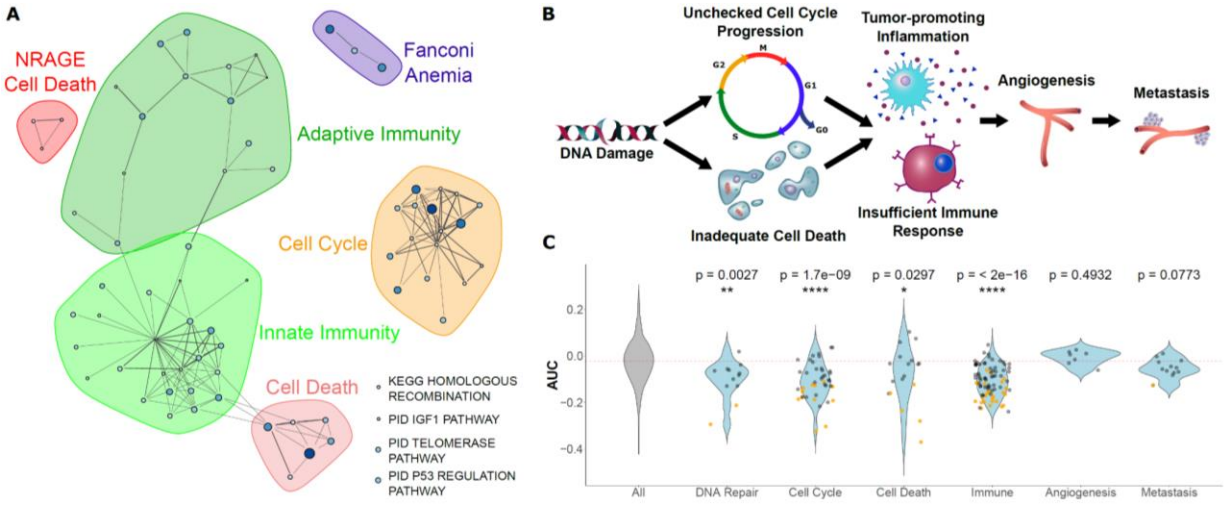


Figure 17 Cancer Control Mechanisms Underlying Mammalian Longevity. Pathways that evolve slower in long-lived, large-bodied mammals are related to control of cancer. A) Significantly enriched pathways under increased constraint in species with larger values of the 3L phenotype. Each dot represents a pathway, and the size and color of the dot represents the negative log of the rank-sum enrichment statistic. Width of lines connecting pathways represent the number of genes the pathways have in common. B) and C) Pathways under increased constraint in 3L species play various roles in cancer control. Pathways associated with early stages of cancer development (DNA repair, cell cycle control, cell death, and immune functions) are significantly enriched, while pathways for later stages of cancer development (angiogenesis and metastasis) are not enriched. In C), each dot represents a pathway. Yellow dots have significant permulation p-values while black dots do not. Note that dots for "All" pathways are excluded for the sake of clarity.

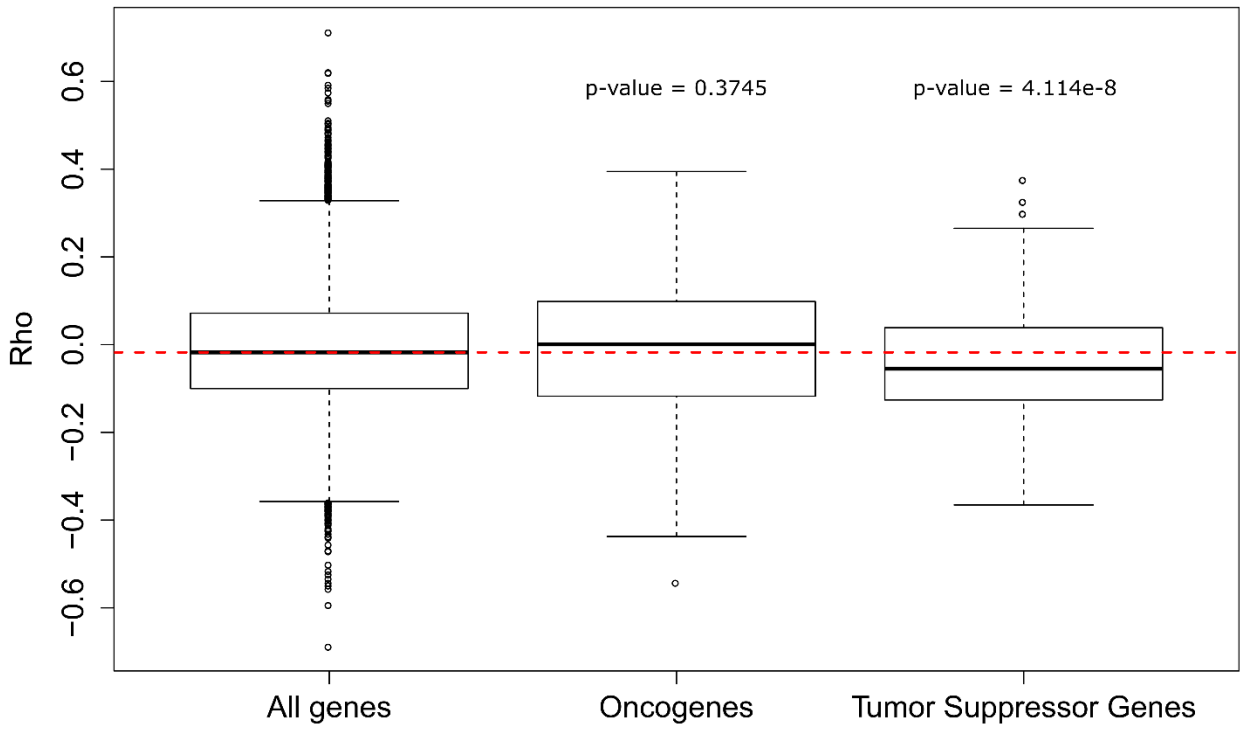


Figure 18 Oncogenes and Tumor Suppressor Genes. Correlation statistics (Rho) for gene evolutionary rate with PC1 3L phenotype. Oncogenes show no significant difference in Rho compared to all genes, which tumor suppressor genes show significantly more negative correlations (Wilcoxon Rank-Sum).

An additional pathway that shows a strong signal of increased constraint with the 3L phenotype is the insulin-like growth factor (IGF1) signaling pathway (Figure 19), which deserves special consideration. Perturbations of IGF1 signaling result in changes in lifespan and body size in diverse organisms (S. Johnson, Rabinovitch, & Kaeberlein, 2013; Kimura et al., 1997; Stout et al., 2013), which suggests that the IGF1 pathway may be a source of innovation underlying the evolution of the 3L trait. However, we find that IGF1 pathway genes in fact evolve more slowly and are thus under increased *purifying* selection in large, long-lived species. The magnitude of this signal is quite striking; IGF1 and IGF1R are ranked 1014 and 94 respectively for increased constraint and several other pathway members are near the top (see Figure 17A). A possible

explanation is that the IGF1 pathway plays an important role in cancer control (Larsson, Girnita, & Girnita, 2005), and this cancer-related signal of constraint dominates any adaptive signal related to lifespan. There are also further reasons to believe that the IGF1 pathway is not the main source of the 3L trait. Across the mammalian phylogeny, lifespan is strongly correlated with body size, but genetic perturbations in the IGF pathway result in longer-lived individuals that are of smaller size (Holzenberger et al., 2003; Sutter et al., 2007) thus decoupling the two traits. This strongly suggests that changes in the IGF1 pathway are unlikely to drive the natural evolution of the large, long-lived phenotype, which is established through a different, yet-unknown mechanism.

For the ELL phenotype, we find a smaller, more focused set of enriched genes and pathways. Of those significantly enriched constrained pathways for ELL, we see some overlap with functional groups represented in results from the 3L phenotype, notably immune-related and DNA repair pathways (Figure 20A). However, although the functional groups are the same, the pathways contained within them differ between the two phenotypes (see File S2, File S3 in (A. Kowalczyk et al., 2020)). In particular, the only significantly constrained DNA repair pathways for the 3L phenotype involve Fanconi's anemia, while the ELL phenotype shows significantly constrained DNA repair pathways for a variety of repair functions (Figure 20). Such pathways stand out not only because of the connection between DNA repair and cancer control, but also because of the observed relationship between DNA repair and aging independent of cancer incidence. This relationship can be demonstrated experimentally by creating double-stranded DNA breaks in laboratory mice to induce an aging phenotype (White et al., 2015). There is also evidence that DNA damage causes dysregulation of the cellular chromatin state and thus can contribute to aging even in post-mitotic cells (Oberdoerffer et al., 2008; Shi & Oberdoerffer, 2012).

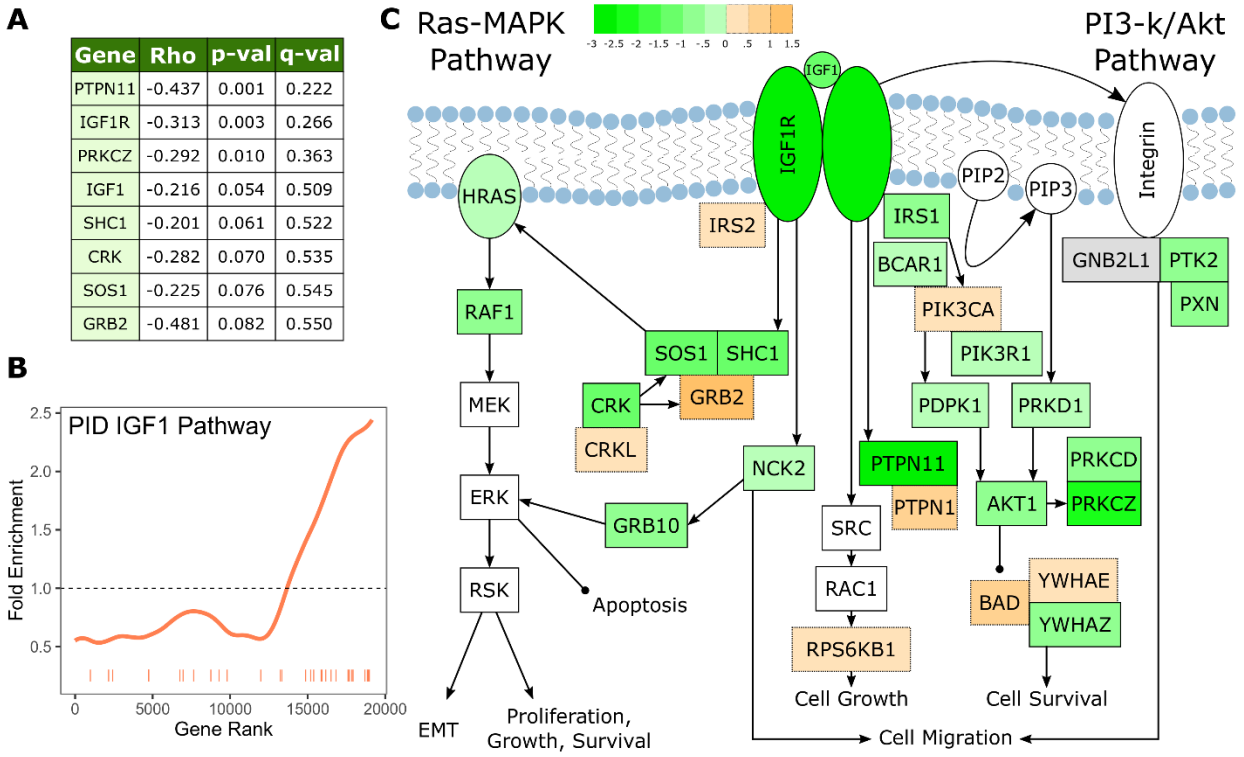


Figure 19 IGF Pathway. The IGF1 signaling pathway is under increased evolutionary constraint in long-lived, large-bodied mammals. A) IGF1 signaling pathway genes are significantly correlated with change in the 3L phenotype. B) The IGF1 signaling pathway is significantly enriched for increased evolutionary constraint in large, long-lived species. The barcode indicates ranks of genes in the pathway within the list of all pathway-annotated genes. The worm indicates enrichment as calculated by a tricube moving average, a type of moving average in which values near the end of the sliding window are down weighted to reduce the effect of extreme values in any given window. The dashed horizontal line marks the null value indicating no enrichment. C) The IGF1 signaling pathway contains many genes whose evolutionary rates are negatively correlated with the 3L phenotype. Shading indicates the Rho-signed negative log p-value for the correlation. Genes in white are not included in the IGF1 pathway annotation used to calculate pathway enrichment statistics, but they are included in the diagram for sake of completeness. The GNB2L1 gene (gray) is in the IGF1 pathway annotation, but correlation statistics were not calculated for that gene because too few branches in the gene tree met the minimum branch length cut-off.

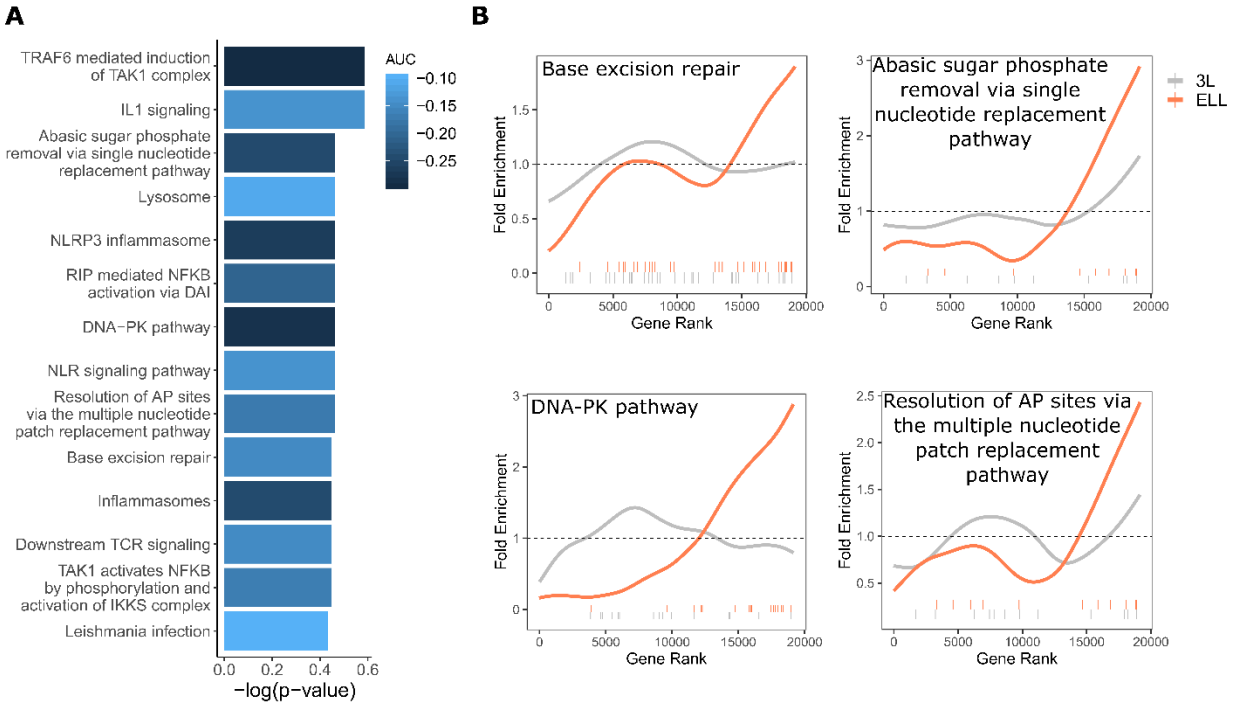


Figure 20 DNA Repair Pathways. DNA repair pathways are under increased evolutionary constraint in mammals that are exceptionally long-lived given their body size. A) Significantly enriched pathways under increased constraint in species with larger values of the ELL phenotype. Bar height indicates the negative log permulation p-value for each pathway, and the color of bars indicates the pathway enrichment statistic. B) DNA repair pathways are more significantly enriched for increased evolutionary constraint in species with large values of the ELL phenotype than species with large values of the 3L phenotype. The barcodes indicate ranks of genes in the pathways within the list of all pathway-annotated genes. The worms indicate enrichment as calculated by a tricube moving average, a type of moving average in which values near the end of the sliding window are down weighted to reduce the effect of extreme values in any given window. The dashed horizontal lines mark the null value indicating no enrichment.

In addition to DNA repair-related pathways, we also noted pathways related to NFKB signaling for which overexpression in downstream targets has been associated with aging. Experimental evidence suggests a connection between NFKB signaling and DNA repair through

sirtuins, a chromatin regulator family that has already been implicated in lifespan control (Howitz et al., 2003; Mao et al., 2012). Sirtuins mediate DNA damage-induced dysregulation and are also responsible for silencing NFKB-regulated genes, thus connecting the two processes (Salminen et al., 2008). Overall, our analysis strongly suggests that fidelity in DNA repair and NFKB signaling contributes to the fitness of ELL species, indicating that these pathways may be a fruitful avenue for aging research and intervention.

3.3 Discussion

We employed an evolutionary rates-based method as an unbiased, genome-wide, pan-mammalian scan to identify genes and pathways that evolve significantly slower in long-lived large-bodied species (3L) and species that are exceptionally long-lived given their body size (ELL). Pathways related to cancer control, including cell cycle, DNA repair, cell death, and immunity, evolve significantly slower in 3L species, which suggests that cancer resistance is an important functionality to enable to evolution of large and long-lived species. Alternatively, a broader set of DNA repair genes and a more focused set of immune genes related to NFKB signaling evolve significantly slower in ELL species, both of which may be linked to effective DNA repair in order to preserve chromatin state.

Our analysis differs from previous efforts in both methodology and results. Firstly, we do not consider individual amino acid changes within a protein as the unit of convergence, but rather we calculate the overall evolutionary rate of a protein on each branch of a phylogeny and consider these evolutionary rates as the unit of convergence. Secondly, unlike previous studies that focused

on lineage- or species-specific changes, we look for correlations between evolutionary rates and quantitative life history traits across the entire phylogeny. This pan-mammalian approach allows us to generalize our findings to describe evolutionary trends throughout all mammals. We also draw a careful distinction between absolute lifespan and relative lifespan, which allows us to make district conclusions about the 3L and the ELL phenotypes.

The most important distinction between our work and previous work is that prior studies began with the assumption that some genes must be under positive selection in association with evolution of extended lifespan. From our unbiased analysis, we find that in fact the strongest signal of evolutionary convergence is that of *increased* constraint on certain genes and gene families in long-lived species (see Figure 17). While some of the pathways have been reported previously (such as cell cycle control, DNA repair, telomerase repair, and IGF1 signaling) our finding is actually the *opposite* of (though not necessarily contradictory to) the positive selection signal that has been emphasized in previous work, which from the perspective of evolutionary rate is decreased constraint. In fact, we find comparatively few genes and no significant pathway enrichments for the opposite trend – faster relative rates in longer lived species – which would correspond to genes potentially under positive selection in longer-lived species. We therefore focus our discussion on genes and pathways evolving slower in species with large values for the 3L and ELL traits. We considered these genetic elements to be important to the evolution of the 3L and ELL traits because they have been protected from accumulating deleterious mutations and therefore evolve slower.

Despite considerable previous work on genes under positive selection in species with extreme lifespan traits, our analysis did not recover any positively selected pathways in the pan-mammalian analysis. However, for any new trait to arise, some corresponding genetic changes

must occur. It is possible that many of the molecular innovations that lead to increased lifespan are species-specific and thus would not be detected by our method. However, it is also true that our relative evolutionary rates method is much more suited to detecting the kind of long-term convergent increases in purifying selection that we observe. Figure 21 represents a simple schematic history of the evolution of a trait and corresponding rate changes. During the establishment of a new trait, some genes experience a brief period of positive selection which generates the molecular innovation to create the new trait, after which the derived sequence will again be subject to purifying selection. Using extant species as a genomic reference for evolutionary history, we can only infer the average rate over the entire history and the brief period of positive selection results in a small, potentially undetectable signal. On the other hand, we can hypothesize another class of genes which become more important as a trait is established and thus experience continuously increased purifying selection after the trait exists. These are the trait "enabling" as opposed to the trait "establishing" genes. Because the rate change in the case of trait-enabling genes is permanent, it has a greater impact on the inferred average rate observable from extant data. Our analyses strongly support that in the case of extended longevity, such "enabling" genes not only exist but are also convergent across independent trait change events.

We have interpreted the rate correlation from the prospective of species with large phenotype-values thereby assuming phenotype changes and rate changes coincide. The implicit choice of default rate does not affect the final gene ranking on which our analysis is based, it does have one important consequence. While decreases in rate are easily interpretable as increased constraint, rate increase can be due to relaxation of constraint or positive selection. We address this asymmetry simply by testing both possibilities. We test the relative accelerated genes for positive selection in the long-lived species and the relatively decelerated genes for positive

selection in the short-lived species. We find no evidence of systematic positive selection in either direction.

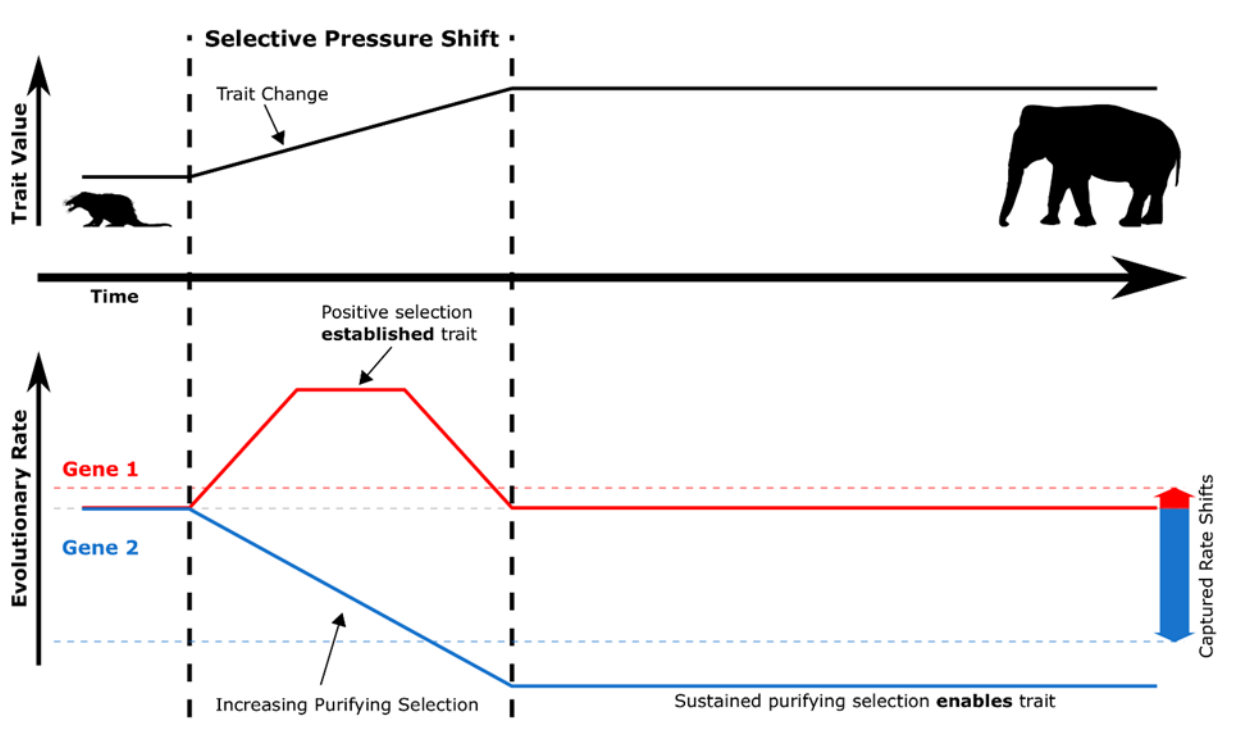


Figure 21 Diagram of Selective Pressure Shifts. During selective pressure shifts that drive phenotypic changes, the genetic evolutionary rate landscape shifts for many genes. Imagine a simplified case where Gene 1 and Gene 2 start at the same evolutionary rate (solid red and blue lines, respectively). Trait-driving genes like Gene 1 enter a transient state of increased evolutionary rates to drive trait change. Complementary genes that support the trait like Gene 2 experience increased purifying selection to allow the trait to persist as it is established. True rates (solid red and blue lines) are not measurable because they represent rates at inaccessible ancestral evolutionary timepoints – only accessible extant sequences can be used to access average rates over time (dashed red and blue lines, which represent positive and negative RERs). Therefore, transient periods of positive selection are less readily able to be captured by RERconverge than sustained purifying selection due to the greater magnitude of their captured rate shifts.

We interpret our strongest pathway enrichment signals as indicating that cancer control is important for enabling evolution of the 3L phenotype, which is in agreement with previous work that has found similar changes in expression levels of cell cycle and immune function genes in both cancerous and aging cells (Chatsirisupachai, Palmer, Ferreira, & Magalhães, 2019). However, there may be alternative explanations. Specifically, differential pathogen pressure is expected to affect some of the same pathways. The connection to immune pathways is clear, and viruses are known to co-opt the cell cycle. TP53 and cell death are likewise important mechanisms by which cells control viral infection. All of these pathways may thus be under increased constraint in species more likely to experience viral infections. However, we believe these explanations are unlikely for several reasons. Firstly, we do not see an enrichment of genes that are specialized for virus control, such as MX1, OAS1, DHX58, and genes activated by type I interferon. Secondly, the interaction between pathogen pressure and evolutionary rates has been studied extensively and it is found that it typically drives adaptive changes in pathogen interacting protein as pathogens are often unique to a species (Kosiol et al., 2008; Sackton et al., 2007; Shultz & Sackton, 2019). This is at odds with the increased constraint of immune genes that we observe. Finally, the probability of cancer increases with age and there is theoretical and empirical evidence that importance of cancer resistance increases with body-size and lifespan. On the other hand, for humans the probability of dying from many infections decreases sharply in early life (though increases again at post-reproductive ages) (Palmer, Albergante, Blackburn, & Newman, 2018) suggesting that infection resistance should not be preferentially selected for in long-lived species.

Overall, the genes and pathways we identified whose constraint is negatively correlated with the 3L and ELL traits encompass functionalities important for the evolution of extended lifespan, and they therefore represent candidate genes and pathways for further experimental

exploration. Importantly, such genes were uncovered using an unbiased, genome-wide pan-mammalian scan. As such, these results point to keys to exceptional longevity that are not specific to one or a handful of species, but that are universal across mammals.

3.4 Methods

RERconverge measures relationships between relative evolutionary rates of orthologous genes and phenotype values based on a set of gene trees whose branch lengths represent protein evolutionary rates and a set of phenotype values, either binary or continuous. The RERconverge package is freely-available for use at https://github.com/nclark-lab/RERconverge along with walkthroughs for beginner users (Amanda Kowalczyk et al., 2019). The original RERconverge method was successfully used to find genes undergoing convergent evolutionary rate shifts in marine mammals (Chikina et al., 2016) and subterranean mammals (Partha et al., 2017a), and recent statistical improvements have made RERconverge more robust and have given it more power to detect such rate shifts (Partha et al., 2019). Here, the RERconverge methods were extended to use on continuous traits.

To perform the RERconverge analysis, we first used PAML (Yang, 2007) to create maximum likelihood gene trees whose branch lengths represent evolutionary rates by means of number of amino acid substitutions. Gene trees were generated for 19,149 amino acid alignments of orthologous genes from the 62 mammal species from the UCSC 100-way alignment (Blanchette et al., 2004; Harris, 2007; Kent et al., 2002) using the hg19 genome, all of which are available at http://genome.ucsc.edu/. Trees were read into R using the readTrees() function from

RERconverge. This step also estimated a master tree with branch lengths that represented the average branch lengths across many gene trees. For all further steps, only 61 species were used as listed in Supp. File 1; the cape golden mole was excluded because longevity data were not available for it. Relative evolutionary rates corrected for genome-wide evolutionary rates for each species as well as corrected for branch length heteroskedasticity using a weighted regression approach (Partha et al., 2019) were calculated using getAllResiduals() with weight=T, scale=T, cutoff=.001, and useSpecies=species names for the 61 mammals species.

Relative evolutionary rates (RERs) quantify the deviation in evolutionary rate of a protein along a specific phylogenetic branch from proteome-wide expectations. Negative RERs indicate fewer substitutions than expected due to increased constraint. Positive RERs correspond to more substitutions than expected, which could arise due to relaxation of constraint or positive selection. To correlate RERs with lifespan phenotypes, we use phenotypic change along phylogenetic branches computed from maximum likelihood ancestral state reconstruction (Revell, 2012). This transformation is equivalent to phylogenetically independent contrasts (Felsenstein, 1985) and thus removes phylogenetic dependence from the phenotype values.

We used the char2Paths() function on each phenotype with trees read in using readTrees() to create a phenotype vector that represented the predicted difference in phenotype values between each species and its ancestor. The char2Paths() function first uses maximum likelihood estimation through the fastAnc() function from phytools (Revell, 2012) and then subtracts the values between connected pairs of species to calculate a phylogenetically-independent measure of the change in the phenotype over evolutionary time. Finally, the getAllCor() function was run with each phenotype and the RER matrix with method = "p", min.pos = 0, and winsorizeRER = 3 to perform Pearson correlations with RERs winsorized such that the two most extreme RER values are set to

the value of the third most extreme RER. See Figure 22 for evolutionary rate-phenotype scatterplots with high correlation values.

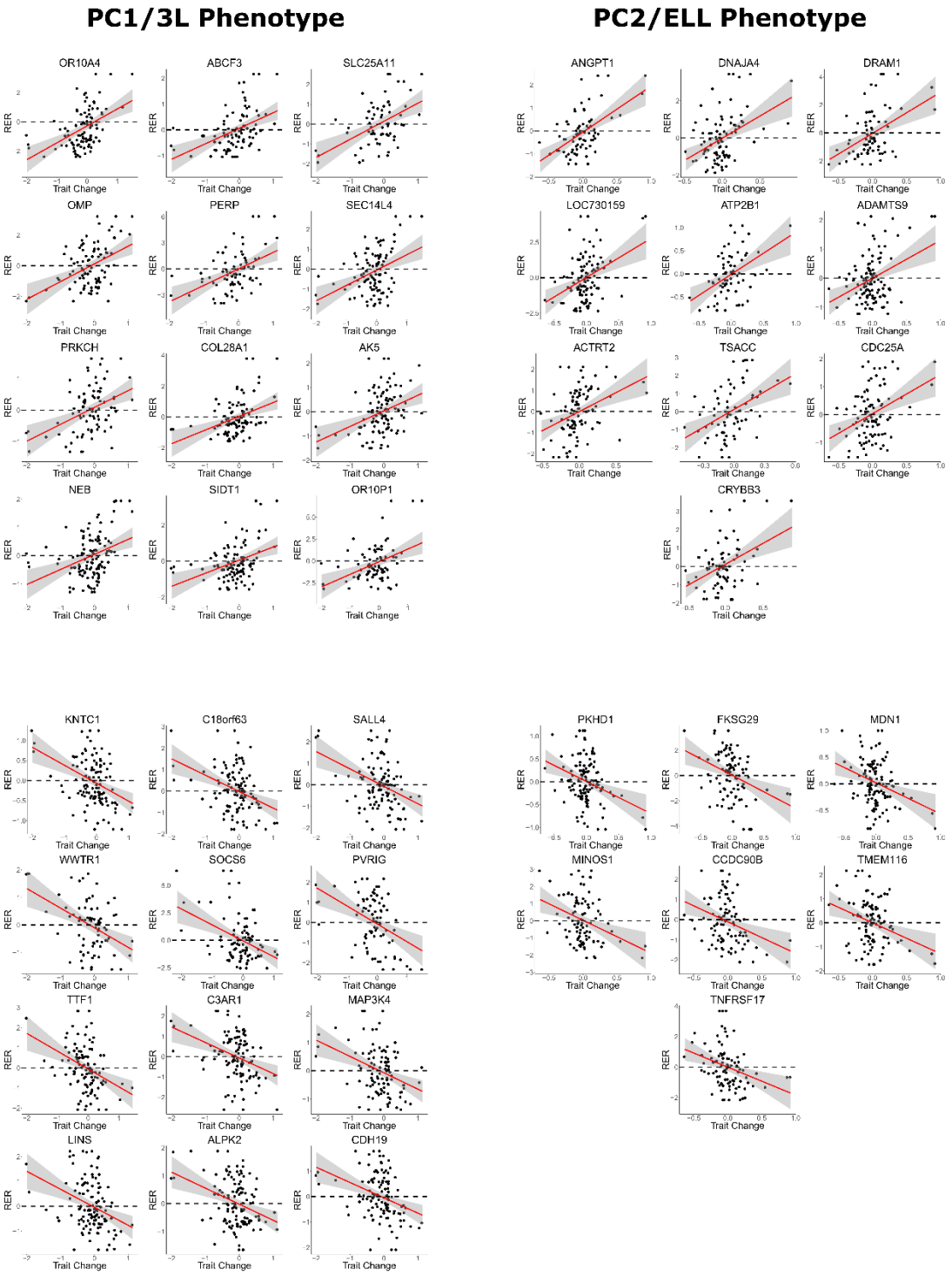


Figure 22 RERconverge Scatterplots. Scatterplots for both 3L and ELL phenotype trait change versus relative evolutionary rate. Genes depicted for the 3L phenotype have adjusted correlation p-values below

0.15, genes depicted for ELL positive correlations have adjusted correlation p-values below 0.4, and genes depicted for ELL negative correlations have adjusted correlation p-values below 0.5.

The gene lists produced from the RERconverge correlation analyses were used to calculate pathway enrichments for all canonical pathways from mSigDB (Liberzon et al., 2011) and Mouse Genome Informatics (MGI) functional annotations (Blake et al., 2003; Eppig et al., 2015). For each pathway, a Wilcoxon Rank-Sum statistic was calculated to compare the sign of Rho times the negative log p-value for correlations of genes in the pathway to the same measure for all genes included in a pathway annotation using getStat() from RERconverge to calculate the sign of Rho times the negative log p-value for correlations for each gene and fastwilcoxGMTall() to quickly calculate an approximation of the Wilcoxon Rank-Sum statistic.

3.4.1 Permulation Analysis (Phylogenetically-Restricted Permutations)

In addition to calculating theoretical enrichment statistics, we developed a novel phylogenetically-restricted permutation strategy dubbed "permulations" to calculate empirical p-values for each pathway enrichment statistic. Permulations are a combination of permutations and simulations in reference to the strategy for generating null phenotype values to use to generate null statistics. To perform permulations, first phenotype values (3L and ELL) were simulated for each species using phylogenetic simulations, and then original phenotype values were reassigned to the species based on the rank of the simulated values (see Figure 23). Simulations were performed using the geiger library in R using a Brownian motion approach and the average evolutionary rate tree created by RERconverge (Harmon et al., 2008). After creating new phenotype values,

RERconverge analyses were performed to calculate correlations between evolutionary rates of genes and phenotype values, and enrichment statistics were calculated using these gene results. The process of creating new phenotype values and calculating enrichment statistics was repeated 1,000 times for each phenotype and the empirical p-value for pathway enrichment was calculated as the proportion of times the permulation enrichment statistic was greater than the enrichment statistic calculated using real phenotype values.

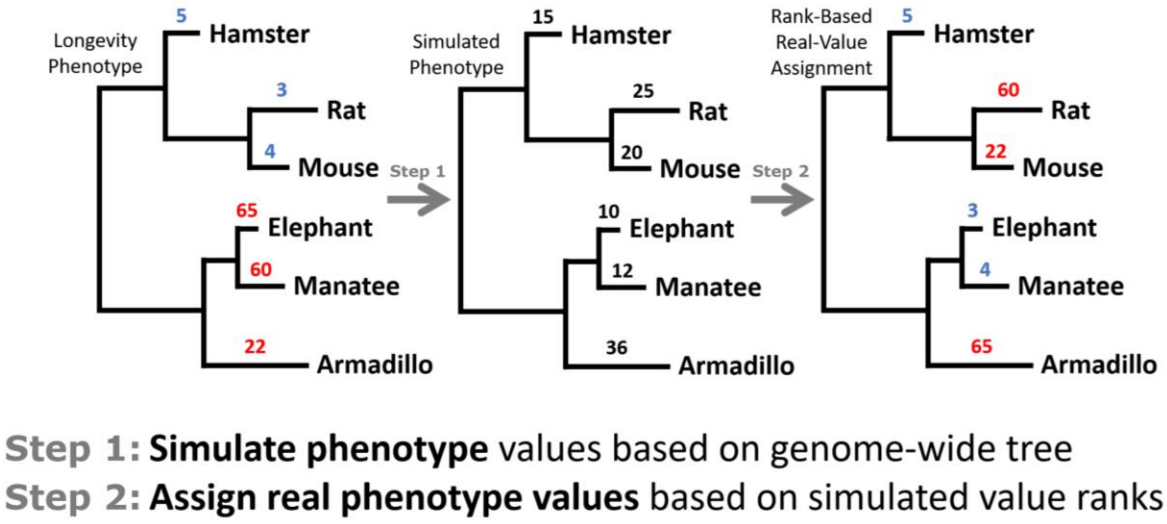


Figure 23 Permulation Diagram. Diagram of a toy example of permulation calculations. Starting with 3L or ELL phenotypes, new phenotype values are simulated based on average genome-wide evolutionary rate. Original phenotype values are then reassigned to species based on the ranks of simulated values and used as the new permulated phenotype values.

Permulations are an attractive option to calculate empirical p-values because they use fabricated phenotypes that are independent of RERs but respect the underlying phylogenetic relationships between species (due to use of phylogenetic simulations) while maintaining the original range of the data (due to rank-based assignment of true phenotype values). These types

of analyses are a common tool in genomics because they allow for control of the dependence across tests caused by intercorrelations among genes. Permulations are highly analogous to permutations commonly performed in differential expression analysis, and we use them for the same purpose, namely to calculate empirical pathway-level statistics (Aravind Subramanian et al., 2005). This step is critical for evaluating pathway enrichment results because many pathways have elevated RER correlations even when conditioned on the phenotype of interest. A pathway with high RER correlation among its genes is more likely to show up as enriched when using a test that assumes gene independence, such as the Wilcoxon Rank-Sum test. Permulations allow us to generate an empirical null distribution for pathway enrichment statistics to correct for interdependence among gene ranks. Indeed, we find multiple pathways that show significant enrichment using a Wilcoxon Rank-Sum test but insignificant empirical p-values using permulations. We also show that permulation p-values, which take phylogeny into account, are more conservative than permutation p-values, which ignore phylogenetic dependence, and equally as conservative as the phylogenetic simulation p-values (see Figure 24).

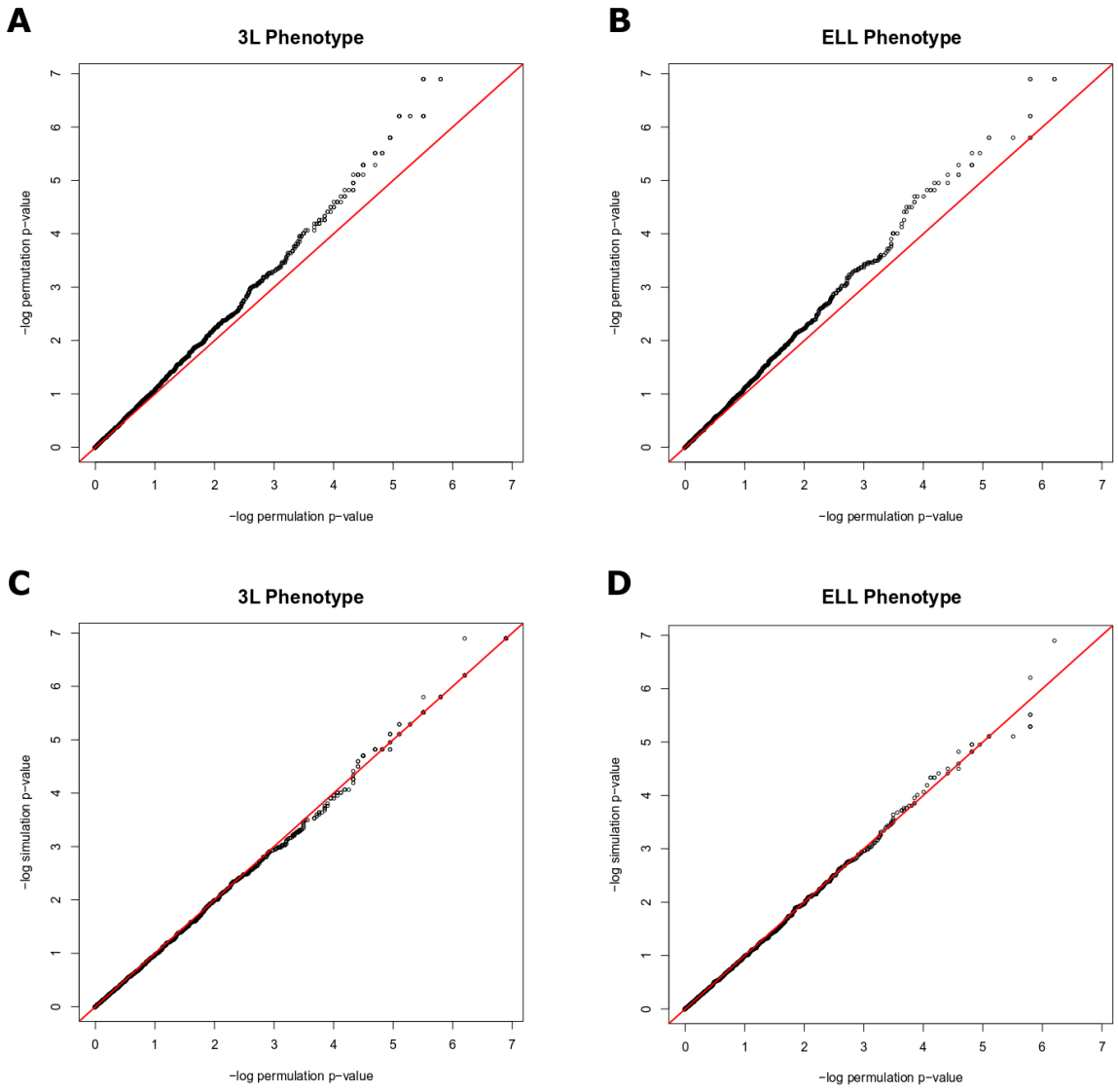


Figure 24 Compare Permulations, Permutations, and Simulations. Quantile-quantile plots demonstrating that permulation p-values are more conservative than permutation p-values for both 3L and ELL phenotypes (A and B) and permulation p-values are equally as conservative as simulation p-values for both 3L and ELL phenotypes (C and D).

3.4.2 Positive Selection Tests

In genes significantly correlated with the 3L and the ELL phenotypes, we investigated evidence for relaxation of constraint and positive selection on trait-defining foreground branches using phylogenetic models of codon evolution. We did so using a representative subset of the full mammalian phylogeny (see Figure 25). Trait-defining foreground branches were specified independently for the two phenotypes along both axes of trait values – positive and negative. In total we have four sets of trait-defining foreground branches, namely positive 3L, negative 3L, positive ELL, and negative ELL. Figure 26 shows the phylogeny of species used for this analysis and the four sets of trait-defining foreground branches. Species were selected for the representative subset based on three criteria: 1) all species were placental mammals (the monotreme and marsupials were excluded), 2) species with highest and lowest 3L and ELL phenotype values were included and used as foreground species in their respective tests, and 3) closely-related non-foreground species were included. Only placental mammals were used because the inference of positive selection can be confounded in the non-placental clade due to the long divergence time between the clades and hence saturation of synonymous sites for many genes. Foreground species were selected based on 3L and ELL values. A subset of non-foreground species was selected to correct for over- and under-sampling in particular clades while maintaining valid "outgroups" for foreground species. This subset also eliminated species that were essentially duplicates in terms of tests for positive selection. For example, the primate clade is sampled much more than other clades, i.e. more primate genomes are available. Additionally, because most of these species are non-foreground species, their presence has little impact on positive selection tests. To correct for the oversampling, we used five out of twelve primates in our subtree, three of which

were foreground species (human, marmoset, and squirrel monkey) and two of which were outgroups (rhesus as an outgroup to human and bushbaby as an outgroup to the whole clade). We used similar logic to select outgroup species throughout the rest of the phylogeny.

We inferred the significance of relaxation of constraint on each foreground branch set using likelihood ratio tests (LRT) between Branch-site Neutral (BS Neutral) and its nested null model M1 (sites neutral model) in PAML (Yang, 2007). Similarly, we performed LRTs between branch-site selection model (BS Alt Mod) and its null BS Neutral were used to infer positive selection on the foreground branches. Probabilities for each of these two LRTs were estimated using the chi-square distribution with 1 degree of freedom. We additionally inferred significance of mammal-wide relaxation of constraint and positive selection using the LRTs between M8A (Neutral model) vs M0 (null model) and M8 (positive selection model) vs M8A respectively. Prior to performing the mammal-wide tests in genes corresponding to each of the four foreground branch sets, we removed the corresponding foreground branches, allowing us to obtain unbiased estimates for significance of relaxation of constraint and positive selection from only the background mammalian branches.

Figure 25 Full Phylogenetic Tree. Phylogenetic tree with all 61 mammal species used for RERconverge analysis. Branch lengths represent average evolutionary rates for each species as estimated by RERconverge. This tree topology was first reported in (Wynn K Meyer et al., 2018) and was created based on two widely-accepted phylogenies (Bininda-Emonds et al., 2007; Meredith et al., 2011) and extensive literature review.

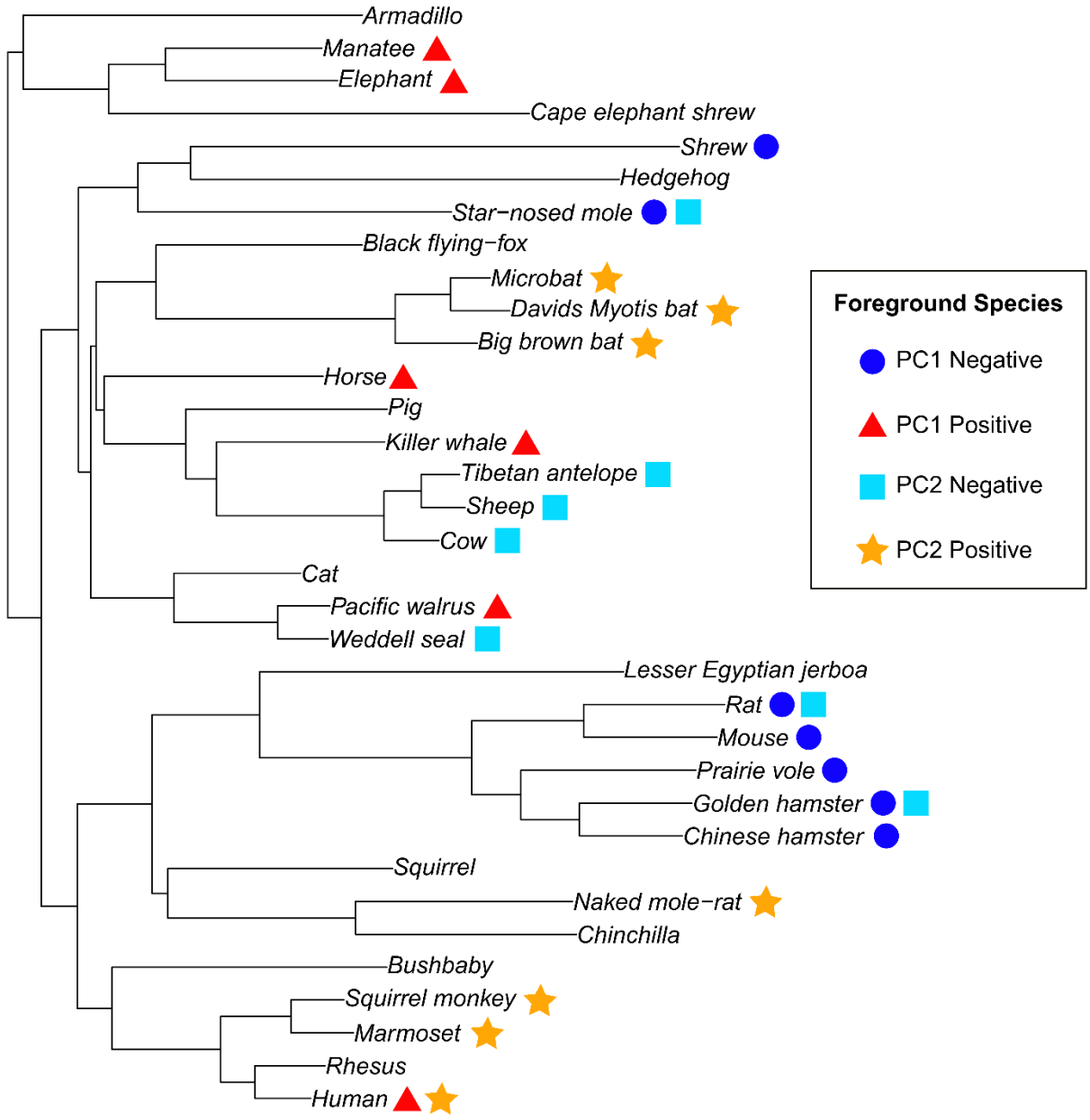


Figure 26 Subtree for Branch-Site Models. Phylogenetic tree with 34 placental mammal species used for branch-site tests for positive selection. Species used as foreground for each of the four phenotypes (PC1=3L, PC2=ELL) are represented by symbols and shapes.

3.4.3 Species Robustness through Subtree Analysis

We were interested in assessing the sensitivity of our results to the choice of mammalian species used in our analyses. The mammalian genomes available represent a subset of not only all extant mammals, but all mammal species that have ever existed. Therefore, the genomes used in this analysis represent an incomplete, and perhaps even a biased representation of mammal species. Since we would like to extend our conclusions to pertain to mammals in general, we sought to quantify the effects of our incomplete data on our gene and pathway results.

To do this, we created subsets of our data that contained fewer species. These subsets had 10, 20, 30, 40, 50, 60, 70, and 80 percent of species randomly removed (6, 12, 18, 24, 30, 36, 42, and 48 species removed out of 61 total, respectively) with ten random subsets created for each species removal level. We then ran the standard RERconverge analysis on these subsets to acquire correlation statistics representing the relationship between evolutionary rate of each gene across species and longevity phenotypes in those species. We also ran enrichment analyses on the subset gene results to acquire pathway enrichment statistics. After performing these analyses, we calculated correlations between results from our full dataset and results from the subsets, where results were quantified as the negative log p-value times the sign of the statistic.

We further tested the sensitivity of our results to specific species presence/absence by performing targeted species removal of species groups with which we expected to have other phenotypes confounded with body size, namely marine mammals, and species that may have non-convergent genetic mechanisms for lifespan extension, namely bats and the naked mole rat. We created two new data subsets that only contained non-marine species (dolphin, manatee, killer whale, walrus, and Weddell seal were removed) and only non-bat and non-naked-mole-rat species

(megabat, black flying-fox, microbat, David's myotis bat, big brown bat, and naked mole-rat were removed). We then performed the standard RERconverge analysis to find correlations between the evolutionary rates of genes across species and longevity phenotypes in those species. We also performed pathway enrichment analyses on the gene results.

We found good correlation of both enrichment and correlation statistics between subset results and full dataset results based on negative log p-values times the sign of the statistic (Figure 15 and Figure 16). For data subsets in which a proportion of species were randomly eliminated, there was upwards of a median 60% correlation between results even with half all species removed from our analyses, which indicates that randomly removing species did not significantly change our results. These findings are quantified numerically as an among-subtree variance in Supp. File 4 and Supp. File 5 in (A. Kowalczyk et al., 2020). We also compared full dataset results to results from targeted data subsets without marine mammals and without bats and the naked mole-rat by quantifying correlations between their negative log p-values times the sign of the statistic for gene correlations and pathway enrichment statistics (Figure 16). From these comparisons, we found a strong relationship between full dataset results and targeted subset results, which indicates that presence or absence of potentially problematic species such as marine mammals, bats, and the naked mole-rat is not strongly impacting our results. Quantitative representations of these findings are available in Supp. File 4 and Supp. File 5 in (A. Kowalczyk et al., 2020). Together, these findings indicate that our results are not species, clade, or species subgroup-specific, but instead represent trends across all mammals.

3.4.4 Alternate Tree Topology Analysis

In addition to determining whether individual or groups of species were disproportionately driving our conclusions, we were also interested in verifying that the fixed tree topology used for analyses was not affecting results. To do so, we reran analyses using alternate plausible topologies representing different ancestral relationships among species (trees shown in Figure 27). Differences in the alternate tree topologies represent points of potential incomplete lineage sorting that lead to uncertainty, they and are thus reasonable alternatives compared to the original tree used for analyses.

RERconverge was run using gene trees generated using the alternate topologies and results were compared to results using the original tree. As shown in Figure 28, there is a strong correlation between all sets of results, which indicates that uncertainties in tree topology do not strongly affect results. This is true for very similar topologies (Robinson-Foulds distance 6) and fairly different topologies (Robinson-Foulds distance 22).

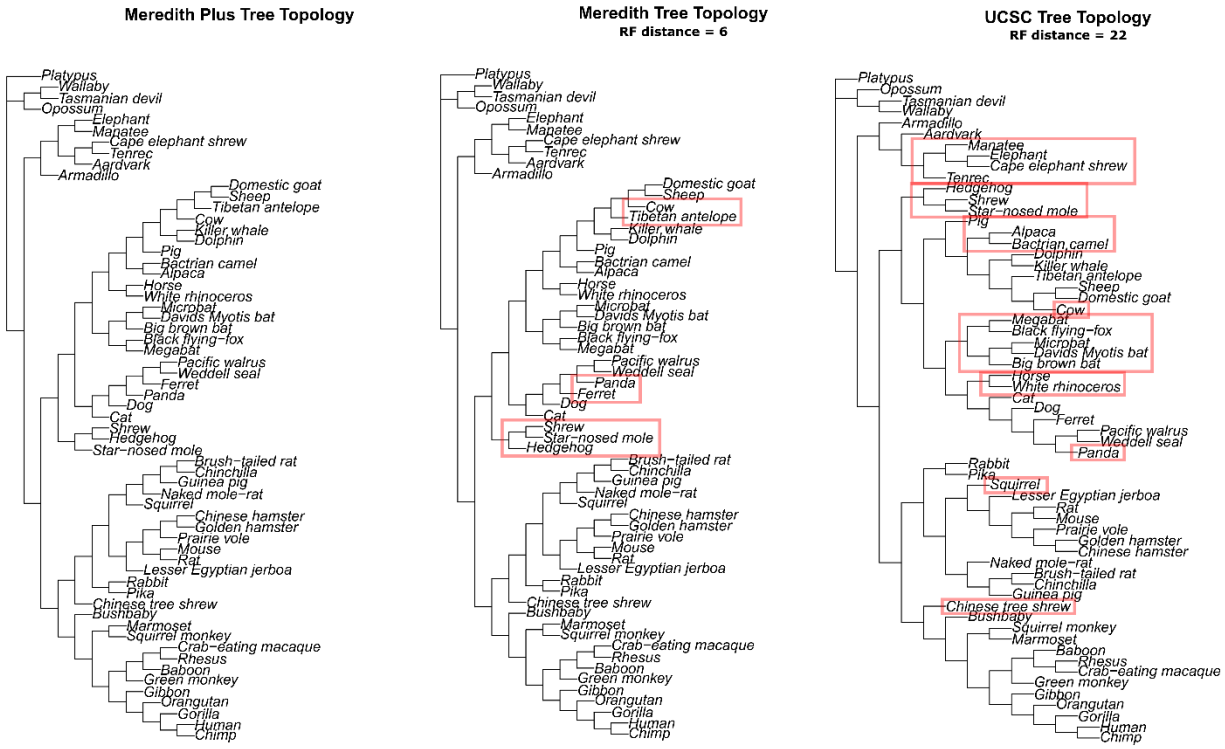


Figure 27 Mammalian Dendrogram Topologies. Alternative tree topologies used to test for robustness to phylogeny topology errors and incomplete lineage sorting.

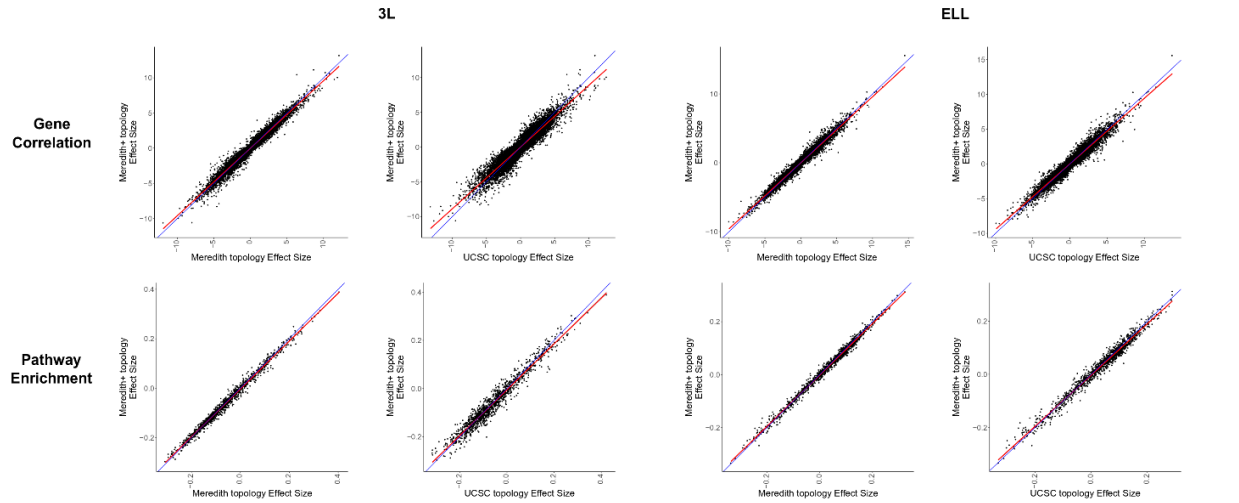


Figure 28 RERconverge Results from Alternate Tree Topologies. Correlations between gene correlation and pathway enrichment statistics between alternative tree topologies and the Meredith+ tree topology used for all other analyses. Correlations between results are very strong, indicating that the results are robust to differences in tree topology.

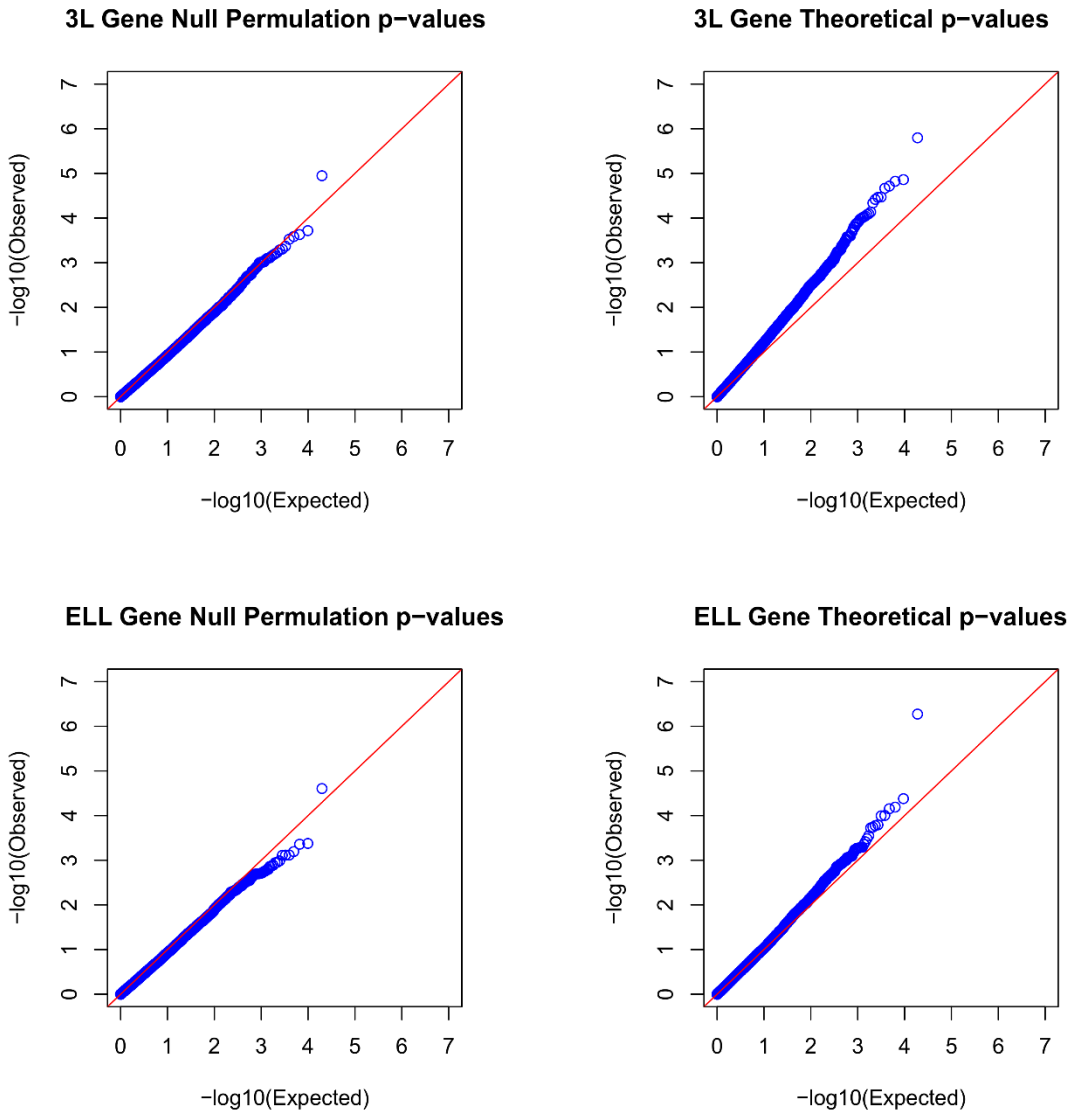


Figure 29 Q-Q Plots of RERconverge Results. Q-Q plots demonstrating the relationship between null gene permulation p-values and a standard uniform distribution and theoretical gene p-values and a standard uniform distribution for both 3L and ELL phenotypes.

3.4.5 Phylogenetic Trees

Full Tree (61 mammals):

((((((((((Human:0.005957477577,Chimp:0.006721826689):0.001382639829,Gorilla:0.007765177171):0.005572327638,Orangutan:0.0164503644):0.002187630666,Gibbon:0.01770384793):0.007043113559,(Green_monkey:0.007693724903,((Crab-eating_macaque:0.001292320552,Rhesus:0.00713015786):0.002951690224,Baboon:0.005199240711):0.002049749893):0.01566263562):0.0135408115,(Marmoset:0.02474184521,Squirrel_monkey:0.02096868307):0.02784675729):0.04299750653,Bushbaby:0.108738222):0.01379370868,((((((Guinea_pig:0.09048639907,(Chinchilla:0.05332953299,Brush-tailed_rat:0.08476954109):0.01287861561):0.02118937782,Naked_mole-rat:0.08588673524):0.07432515556,Squirrel:0.08896424642):0.006291577528,((((Chinese_hamster:0.04084640027,Golden_hamster:0.04456203524):0.02314125062,Prairie_vole:0.06932402649):0.01947113467,(Mouse:0.05273642272,Rat:0.05576007402):0.04435347588):0.08380065137,Lesser_Egyptian_jerboa:0.1438649666):0.04270536633):0.01663675397,(Pika:0.125654444 5,Rabbit:0.07131655591):0.06535533418):0.009050428462,Chinese_tree_shrew:0.1191189141):0.003894252213):0.01425600689,(((((Panda:0.03854019703,((Weddell_seal:0.02002160645,Pacific_walrus:0.02064385875):0.01734764946,Ferret:0.04613997497):0.002879093616):0.009005888384,Dog:0.05339127565):0.01185166857,Cat:0.05020331605):0.03285617057,((((((Cow:0.02168740723,((Domestic_goat:0.01157093136,Sheep:0.01246322594):0.0049716126,Tibetan_antelope:0.01522587482):0.01465511149):0.0662523666,(Killer_whale:0.006371664911,Dolphin:0.01086552617):0.06014682602):0.01216198069,Pig:0.0796745271):0.006785823323,(Bactrian_camel:0.01240650215,Alpaca:0.01096629635):0.06374554586):0.02551888691,(White_rhin

oceros:0.04977357056,Horse:0.061454379):0.02510111297):0.00331214686,((Big_brown_bat:0.03248546656,(Davids_Myotis_bat:0.02344332842,Microbat:0.01567729315):0.02193849809):0.09455328094,(Black_flying-fox:0.005833353548,Megabat:0.01611220178):0.07567400302):0.02385546003):0.002057771224):0.004845253848,(Star-nosed_mole:0.1239823369,(Hedgehog:0.1696142244,Shrew:0.1934205791):0.02079474546):0.0235875333):0.01477733374):0.01316436193,(((((Cape_golden_mole:0.1017903453,Tenrec:0.1749615473):0.01592632003,Cape_elephant_shrew:0.1516860647):0.006610995228,Aardvark:0.08326528894):0.008243787904,(Elephant:0.06812658238,Manatee:0.06198982615):0.0224994529):0.03384011363,Armadillo:0.1342602666):0.005989703247):0.2206952867,((Wallaby:0.1270943532,Tasmanian_devil:0.09944141622):0.02717055443,Opossum:0.1181200712):0.1802966572):0,Platypus:0.4322118716);

Branch Site Tree (34 placental mammals):

((((((Human:0.02214318927,Rhesus:0.0277942336):0.0135408115,(Marmoset:0.02474184521,Squirrel_monkey:0.02096868307):0.02784675729):0.04299750653,Bushbaby:0.108738222):0.01379370868,(((Chinchilla:0.08739752642,Naked_mole-rat:0.08588673524):0.07432515556,Squirrel:0.08896424642):0.006291577528,((((Chinese_hamster:0.04084640027,Golden_hamster:0.04456203524):0.02314125062,Prairie_vole:0.06932402649):0.01947113467,(Mouse:0.05273642272,Rat:0.05576007402):0.04435347588):0.08380065137,Lesser_Egyptian_jerboa:0.1438649666):0.04270536633):0.02958143464):0.01425600689,(((Weddell_seal:0.02002160645,Pacific_walrus:0.02064385875):0.04108430003,Cat:0.05020331605):0.03285617057,(((((Cow:0.02168740723,(Sheep:0.01743483854,Tibetan_antelope:0.01522

587482):0.01465511149):0.0662523666,Killer_whale:0.06651849094):0.01216198069,Pig:0.07
96745271):0.03230471024,Horse:0.08655549197):0.00331214686,((Big_brown_bat:0.0324854
6656,(Davids_Myotis_bat:0.02344332842,Microbat:0.01567729315):0.02193849809):0.094553
28094,Black_flying-
fox:0.08150735657):0.02385546003):0.002057771224):0.004845253848,(Star-
nosed_mole:0.1239823369,(Hedgehog:0.1696142244,Shrew:0.1934205791):0.02079474546):0.
0235875333):0.01477733374):0.01316436193,((Cape_elephant_shrew:0.1665408478,(Elephant:
0.06812658238,Manatee:0.06198982615):0.0224994529):0.03384011363,Armadillo:0.1342602
666):0.005989703247);

4.0 Complementary evolution of coding and noncoding sequence underlies mammalian hairlessness

This final chapter detailing my thesis research covers the use of RERconverge for a different phenotype: mammalian hairlessness. Studying this phenotype allowed for two key benefits. First, better understanding the genomics of hair growth may help us find new genomic regions to tackle conditions that cause undesired hair loss in humans. Second, I developed several new computational strategies to study hairlessness, including using Bayes factors to disentangle confounded phenotypes, using weight-regressed RERs to remove confounding with body size, and developing binary permulations that were first used in this work. Such strategies will be invaluable for further use in studying other phenotypes.

I independently completed all of the work described in this chapter. This work is currently available as a preprint on *bioRxiv* at (Amanda Kowalczyk et al., 2021).

4.1 Introduction

Hair is a defining mammalian characteristic with a variety of functions, from sensory perception to heat retention to skin protection (Pough, Heiser, & McFarland, 1989). Although the mammalian ancestor is believed to have had hair, and in fact the development of hair is a key evolutionary innovation along the mammalian lineage (Eckhart et al., 2008), numerous mammals subsequently lost much of their hair. Many marine mammals, including whales, dolphins,

porpoises, manatees, dugongs, and walruses, have sparse hair coverage likely related to hydrodynamic adaptations to allow those species to thrive in a marine environment (Z. Chen, Wang, Xu, Zhou, & Yang, 2013; Nery, Arroyo, & Opazo, 2014). Large terrestrial mammals such as elephants, rhinoceroses, and hippopotamuses also have little hair, likely to enable heat dissipation diminished by the species' large sizes (Fuller, Mitchell, Maloney, & Hetem, 2016). Notably, humans are also relatively hairless, a phenotypic characteristic that, while stark, has long been of mysterious origin (Kushlan, 1980). Just as hair coverage varies across mammal species, coverage for an individual organism can change over time in response to environmental factors. For example, Arctic mammals such as foxes and hares famously demonstrate dramatic coat changes in different seasons (E. Johnson, 1981).

Hair follicles are established during embryonic development as a result of interactions between epithelial and mesenchymal cells in the skin, and such interactions also drive follicle movement in adults (Zhou et al., 2018). Hair follicles consist of a complex set of structures under the skin that support the hair shaft itself, which protrudes above the skin. The hair shaft contains an outer layer called the cuticle, an inner cortex later, and sometimes a central medulla core (Plowman, Harland, & Deb-Choudhury, 2018). Structures under the skin support the growth and formation of the hair follicle. Of particular interest are the dermal papilla and matrix region, both located at the base of the hair follicle. The dermal papilla is a key controller of regulation of hair growth and follicle morphogenesis (Veraitch et al., 2017). In fact, transplantation of dermal papilla cells has been repeatedly demonstrated to result in hair growth in previously hairless tissue (C. A.B. Jahoda, Horne, & Oliver, 1984; Colin A.B. Jahoda, Reynolds, & Oliver, 1993; Reynolds & Jahoda, 1992). Just above the dermal papilla, the matrix generates stem cells to the growing hair shaft and the root sheath (Plowman et al., 2018). The two regions work together to regulate and

carry out hair growth – the dermal papilla is the master controller that instructs the hair-growing engine of the matrix region.

During hair growth, a hair follicle goes through three stages of growth called anagen, catagen, and telogen phases. During the anagen phase, the hair shaft is generated and grows out through the skin, while catagen phase ends hair growth and telogen phase causes the follicle to become dormant (Alonso & Fuchs, 2006).

Changes to several hair-related genes are known to result in hairlessness in specific species. The *Hr* gene in mice, so named because of its role in the hair phenotype, results in hairless mice when knocked-out (Benavides, Oberyszyn, VanBuskirk, Reeve, & Kusewitt, 2009). In Mexican dogs, the *FOXI3* gene has been found to be associated not only with hairlessness, but also associated dental abnormalities (Drögemüller et al., 2008). In the American Hairless Terrier, mutation in a different gene, *SGK3*, is responsible for relative hairlessness (Parker, Harris, Dreger, Davis, & Ostrander, 2017). Fibroblast growth factor genes such as *FGF5* and *FGF7* are also heavily implicated in hair growth because their absence causes drastic changes to coat length and appearance in mice (Ahmad et al., 1998). Such genes are associated with keratinocyte growth in which keratins and keratin-associated proteins play a key role. Unsurprisingly, specific structural proteins that comprise hair shafts and their associated genes, known as *KRTAP* genes or hair-specific keratins, are also heavily implicated in hair-related functions (Plowman et al., 2018). They also appear to be unique to mammals, although some *KRTAP*-like genes have been found in reptiles (Eckhart et al., 2008),

Although genetic changes associated with induced hairlessness in specific domesticated species are useful, it is unclear if such changes reflect evolutionary changes that result in spontaneous hairlessness and how much such changes are convergent across all or many naturally

hairless species. By taking advantage of natural biological replicates of independent evolution of hairlessness in mammals, we can learn about global genetic mechanisms underlying the hairless phenotype.

Mammalian hairlessness is a convergent trait since it independently evolved multiple times across the mammalian phylogeny. We can therefore characterize the nature of its convergence at the molecular level to provide insights into the mechanisms underlying the trait. For example, if a gene is evolving quickly in hairless species and slowly in non-hairless species, that implies that the gene may be associated with hairlessness. We focus on the relative evolutionary rate of genomic sequence, which is a measure of how fast the sequence is evolving relative to its expected rate. Unlike seeking sequence convergence to a specific amino acid or nucleotide, using an evolutionary-rates-based method detects convergent shifts in evolutionary rates across an entire region of interest (such as a gene or putative regulatory element). Evolutionary rate shifts reflect the amount of evolutionary pressure acting on genomic elements, and multiple studies investigating diverse phenotypes have found that phenotypic convergence is indeed associated with convergent changes in evolutionary rates (Chikina et al., 2016; Hiller et al., 2012b; Hu et al., 2019; Kapheim et al., 2015; Amanda Kowalczyk et al., 2020; Partha et al., 2017a, 2019; Prudent et al., 2016; Wertheim et al., 2015). We used RERconverge, an established computational pipeline, to link convergently evolving genes and noncoding regions to convergent evolution of mammalian hairlessness. Previous work using RERconverge (Amanda Kowalczyk et al., 2019) to detect convergent evolutionary rate shifts in genes and noncoding elements associated with convergently evolving traits has identified the putative genetic basis of the marine phenotype in mammals (Chikina et al., 2016), the fossorial phenotype in subterranean mammals (Partha et al., 2017a, 2019), and extreme longevity in mammals (Amanda Kowalczyk et al., 2020). Those

studies revealed trends that are not species-specific, but instead represent relevant genetic changes that occurred phylogeny-wide.

Here, we further explored the genetic basis of hairlessness across the mammalian phylogeny by finding genes and noncoding regions under relaxation of evolutionary constraint (i.e. evolving faster) in hairless species. Such genetic elements likely have reduced selective constraint in species with less hair and thus accumulate substitutions at a more rapid rate. To find genetic elements under accelerated evolution in hairless species, we performed an unbiased, genome-wide scan across 62 mammal species using RERconverge on 19,149 orthologous genes and 343,598 conserved noncoding elements. In addition to recapturing known hair-related elements, we also identified novel putative hair-related genetic elements previously overlooked by targeted studies. Importantly, newly uncovered genes and noncoding regions were not only related to keratins, but they also represented a suite of genetic functionality underlying hair growth. Such findings represent strong candidates for future experimental testing related to the hair phenotype.

4.2 Results

4.2.1 Phenotype Assignment

The hairless phenotype in mammals arose at least nine independent times along the mammalian phylogeny (Figure 30A). Genomic regions that experienced evolutionary rate shifts in tandem with mammalian loss of hair were considered potentially associated with phenotype loss. Ten extant and one ancestral hairless species were identified based on species hair density

(Figure 30A). Broadly, species with skin visible through hair were classified as hairless, namely rhinoceros, elephant, naked mole-rat, human, pig, armadillo, walrus, manatee, dolphin, and orca. The cetacean (dolphin-orca) ancestor was also included because it was likely a hairless marine mammal.

An ancestral point of phenotypic ambiguity existed at the ancestor of manatee and elephant. Considerable uncertainty exists as to whether the ancestral species had hair and independent trait losses occurred on the manatee and elephant lineages or, alternatively, whether the ancestral species lost hair prior to manatee-elephant divergence and regained hair along mammoth lineages post-divergence (Roca et al., 2009). Since foreground assignment of the manatee-elephant ancestor had little impact on skin-specific signal, we retained the parsimonious assignment of the ancestral species as haired with inferred independent losses in the manatee and post-mammoth elephant lineages (Figure 31B). Similarly, assigning foreground branches based on the state of being hairless or the transition from haired to hairless – i.e. assigning the entire cetacean clade as foreground versus only assigning the cetacean ancestor as foreground – had little impact on skin-specific signal (Figure 31A). In the case of cetaceans, we retained all three branches (orca, dolphin, and the orca-dolphin ancestor) as foreground to maximize statistical power.

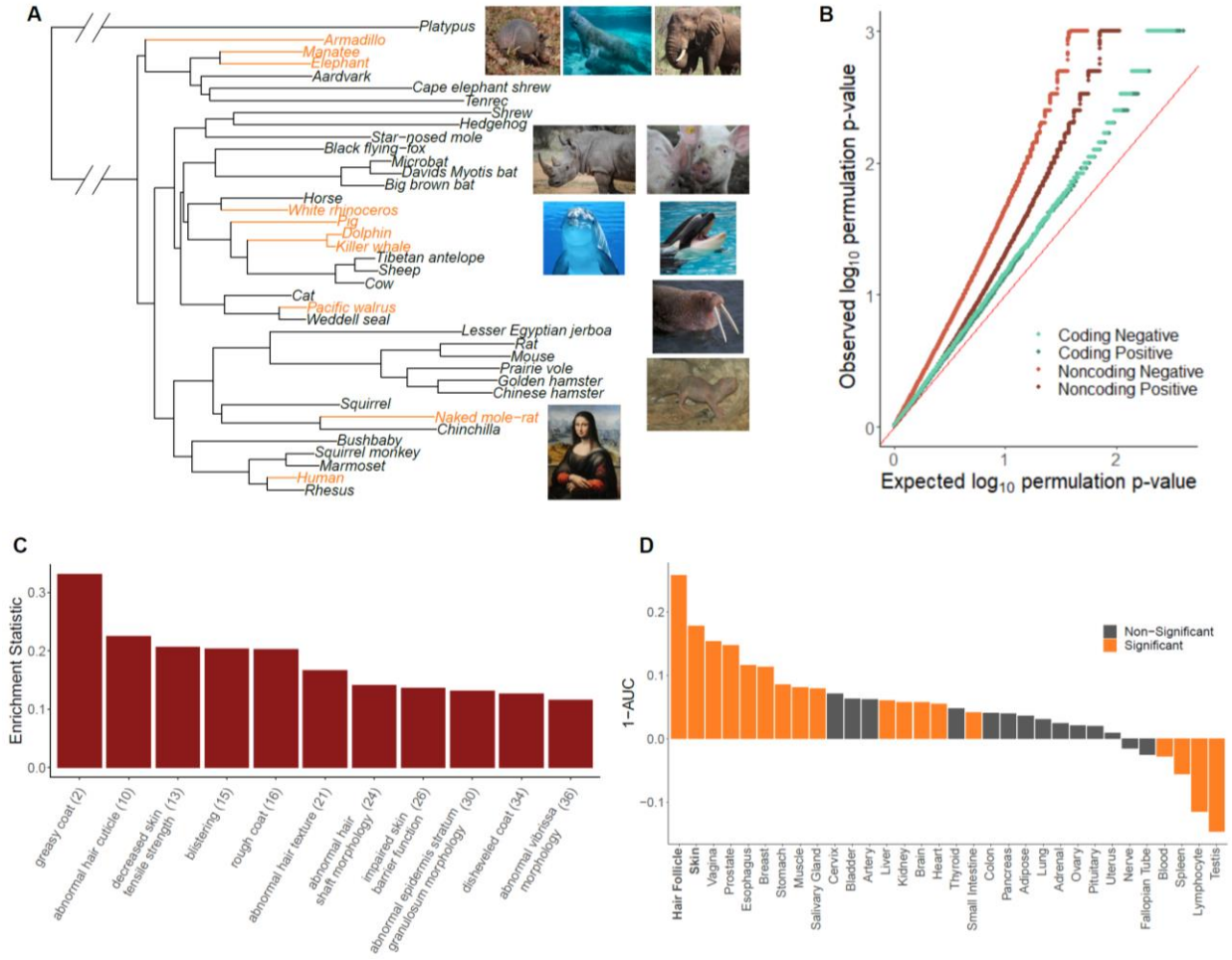


Figure 30 RERconverge and Hairlessness. Hairless species show an enrichment of hair-related genes and noncoding elements whose evolutionary rates are significantly associated with phenotype evolution. A) Phylogenetic tree showing a subset of the 62 mammal species used for analyses. Foreground branches representing the hairless phenotype are depicted in orange alongside photographs of the species. B) Q-Q plots for uniformity of permulation p-values for association tests per genetic element for coding and noncoding elements. Shown are both positive associations that indicate accelerated evolution in hairless species and negative associations that indicate decelerated evolution in hairless species. The deviation from the red line (the identity) indicates an enrichment of low permulation p-values – there are more significant permulation p-values than we would observe under the uniform null expectation. This indicates significant evolutionary rate shifts for many genes and noncoding elements in hairless mammals. C) Hair-related MGI category genes are under significantly accelerated evolution in hairless species. Shown are one minus the

AUC values (maximum enrichment statistic = 0.5; statistic = 0 indicates no enrichment) for each hair- or skin-related pathway with a permulation p-value less than or equal to 0.01. In parentheses are the statistic-based ranks of those pathways among all pathways under accelerated evolution in hairless mammals with permulation p-values less than or equal to 0.01. D) Skin- and hair-expressed genes are under significant evolutionary rate acceleration in hairless species. All genesets except hair follicle are from the GTEx tissue expression database. Hair follicle genes are the top 70 most highly expressed genes from (Jing Zhang et al., 2017) hair follicle RNA sequencing that are not ubiquitously expressed across GTEx tissue types.

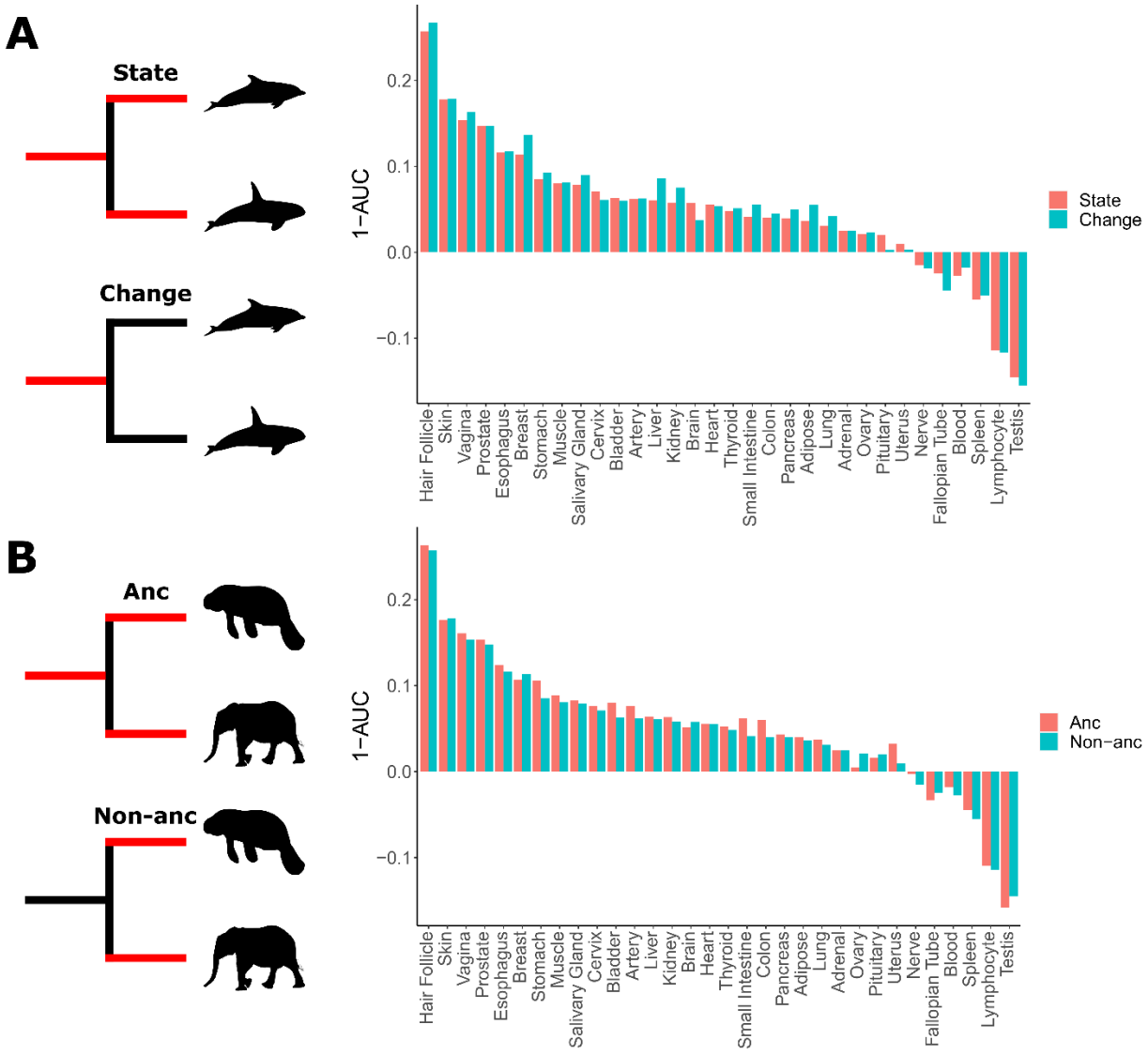


Figure 31 Ancestral Hair Phenotypes. Skin-related genes evolve faster in hairless species. A) When considering either the state of being hairless as the foreground or the process of changing from haired to hairless as the foreground, enrichment of skin-related genes shows little difference. B) When considering the elephant/manatee ancestor as haired or hairless, enrichment of skin-related genes shows little difference.

4.2.2 Phenotypic Confounders

Hairless species share other convergent characteristics that could confound associations between the hairless phenotype and evolutionary rate shifts. In particular, several hairless species are large and many are marine mammals. Therefore, any signal related to hairless species could be driven instead by confounders. Problems with these two confounders were handled in two different ways.

To handle large body size as a confounder, body size was regressed from relative evolutionary rates on an element-by-element basis. In other words, the residuals from the linear relationship between body size and relative evolutionary rates were retained to eliminate the effect of body size on relative evolutionary rate trend. In doing so, any effects related to the relationship between body size and hairlessness were mitigated.

Marine status, on the other hand, is a trait of potential interest because marine mammals experienced unique hair and skin changes during the transition from a terrestrial to a marine environment. However, it is also of interest how much signal is driven by the marine phenotype versus the hairless phenotype. Therefore, Bayes factors were used to quantify the amount of support for the marine phenotype versus the hairless phenotype. A larger Bayes factor indicated more contribution from one model versus another. A ratio of five or greater for the hairless phenotype versus the marine phenotype indicated strongly more support for signal driven by hairlessness. Many hair-related pathways evolving faster in hairless species according to RERconverge also indicated that signal was indeed driven by the hairless phenotype as opposed to its heavy confounder, the marine phenotype, according to Bayes factor analyses (Figure 32).

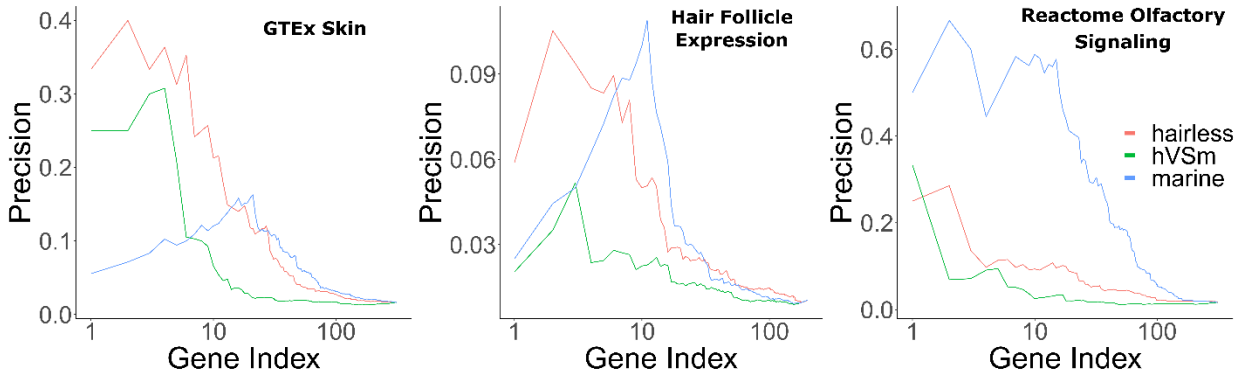


Figure 32 Bayes Factors for Detangling Confounded Phenotypes. Bayes factors reveal the proportion of signal driven by the marine phenotype versus the hairless phenotype. Depicted are precision-recall curves demonstrating how Bayes factors of the contrasting hairless and marine phenotypes rank genes related to skin, hair, and olfaction. Also plotted is a ranking based on the ratio of hairlessness and marine Bayes factors (hVSm = hairlessness Bayes factor/marine Bayes factor). The ratio of the Bayes factors quantifies the amount of support for the hairless phenotype beyond the support for the marine phenotype per gene. In other words, a high Bayes factor ratio indicates a signal of evolutionary convergence associated with hairlessness that is not only driven by signals of convergence in hairless marine mammals. The hairless phenotype had much greater power to enrich for genes expressed in skin (GTEx data) compared to the marine phenotype, indicating that accelerated evolution is driven more strongly by hairlessness. Both the marine and hairless phenotypes enriched for genes in hair follicle expression genes, indicating that both contribute to accelerated evolution of those genes. Olfactory genes on the other hand are expected to show acceleration only related to the marine phenotype. As expected, the marine phenotype much more strongly enriched for olfactory genes than the hairless phenotype.

4.2.3 Known Hair-Related Genomic Elements Evolve Faster in Hairless Species

We used RERconverge to identify genes and noncoding elements evolving at significantly faster or slower rates in hairless species compared to haired species (see Methods). Briefly, the

evolutionary rates of genetic elements were compared in hairless versus haired species using a rank-based hypothesis test, and we generated p-values empirically with a newly developed method, termed permulations, that uses phylogenetically constrained phenotype permutations (Saputra, Kowalczyk, Cusick, Clark, & Chikina, 2020). The permulation method compares the correlation statistics from the true phenotype to correlation statistics that arise from randomized phenotypes that preserve the relative species relationships. Thus, small p-values indicate a specific association with the hairless phenotype.

We find that quantile-quantile (QQ) plots of permulation p-values from hypothesis tests for all genetic elements indicate a large deviation from the expected uniform distribution and thus an enrichment of significant permulation p-values (Figure 30B). We show enrichment of significant p-values for both positive and negative evolutionary rate shifts, and the direction of the rate shifts is critical to interpretation. Positive rate shifts imply rate acceleration, which we interpret as a relaxation of evolutionary constraint. While positive rate shifts could theoretically be driven by positive selection, we demonstrate that this is not the case for our top accelerated genes. Branch-site models to test for positive selection were performed using PAML (Yang, 2007) on top accelerated genes. Tests showed little evidence for foreground-specific positive selection; out of 199 genes tested, 27 genes demonstrated hairless acceleration, but all such genes also showed evidence for tree-wide positive selection, suggesting that positive selection was not specific to hairless species although perhaps stronger. Thus, regions with positive rate shifts evolve faster in hairless species due to relaxation of evolutionary constraint, perhaps because of reduced functionality driving or in conjunction with the hairlessness phenotype. Negative rate shifts indicate increased evolutionary constraint in hairless species, which implies increased functional importance of a genomic region. While negative shifts are more difficult to interpret in

the context of trait loss, they may represent compensatory phenotypic evolution in response to trait loss.

To demonstrate that the statistical signal from individual genes and noncoding regions is meaningful, we evaluated to what extent those RERconverge results enrich for known hair-related elements. We calculated pathway enrichment statistics using a rank-based test and statistics from element-specific results to evaluate if genes or noncoding elements that are part of a predefined biologically coherent set are enriched in our ranked list of accelerated regions. Using numerous gene sets associated with hair growth, such as KRTs, KRTAPs, hair follicle expressed genes (Jing Zhang et al., 2017), skin-expressed genes (Papatheodorou et al., 2018), and Gene Ontology (GO) (Ashburner et al., 2000), Mouse Genome Informatics (MGI) (Eppig et al., 2015), and canonical hair-annotated genes (Liberzon et al., 2011), we indeed find that our results are highly enriched for hair-related functions. As shown in Figure 1C, many of the top-enriched MGI phenotypes are hair-related. Likewise, enrichment analyses using the GTEx tissue expression database (Papatheodorou et al., 2018) supplemented with hair follicle expressed genes (Jing Zhang et al., 2017) show strong enrichment for both skin and hair follicle genes, as well as signal for other epithelial tissues such as vagina and esophagus (Figure 30D).

Hair-related pathways remained enriched among rapidly evolving genes even when KRTs and KRTAPs were removed. This implies that hairless-related genetic changes are not merely structural, but instead they are broadly driven by many genes related to the hair cycle. Similarly, no individual hairless species had an undue impact on enrichment of known hair-related pathways as indicated by consistent findings when individual hairless species were removed from analyses (Figure 33).

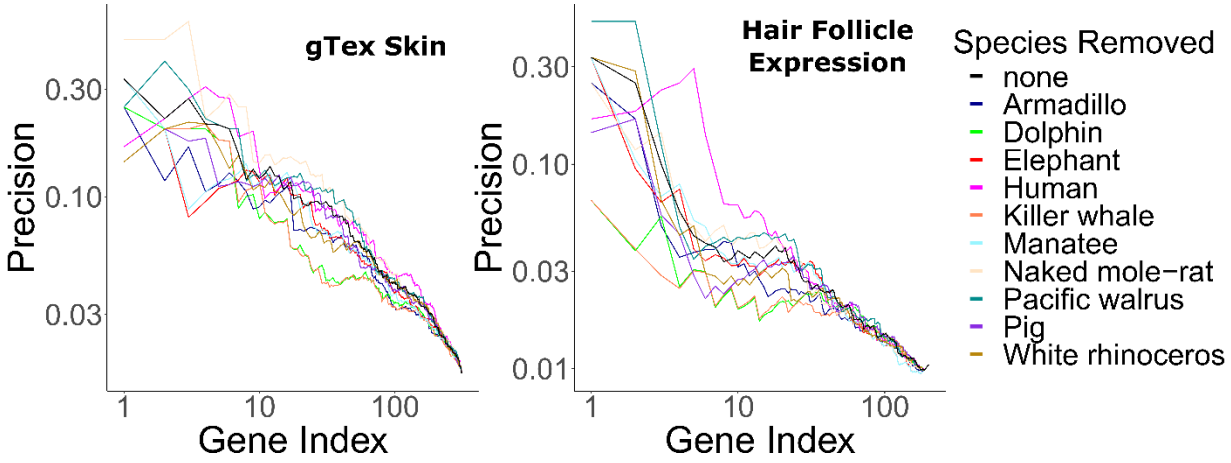
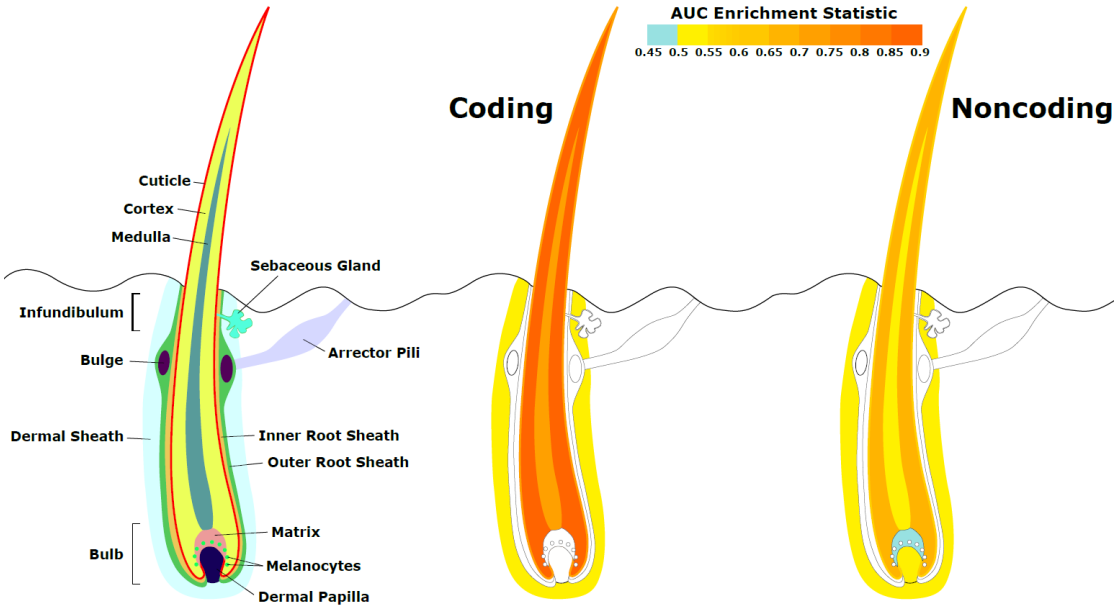


Figure 33 RERconverge Results are Robust to Species Removal. Hair-related pathways are enriched for genes with evolutionary rates significantly accelerated in hairless species. Enrichment is consistent even when individual hairless species are removed.

Investigating a focused list of gene sets associated with specific structures of the hair follicle revealed an interesting contrast between coding and noncoding sequence (Figure 34). Significantly accelerated genes were primarily within the hair shaft itself for coding sequence. Noncoding regions near genes related to the hair shaft were also under accelerated evolution, and additionally, noncoding regions near genes for the matrix and dermal papilla also showed patterns of decelerated and accelerated evolution, respectively, in hairless species. Since the matrix and dermal papilla play key roles in hair follicle localization, development, and cycling, evolutionary rate shifts in those compartments' noncoding regions suggests that regulatory sequence evolution rather than coding sequence evolution may drive changes in hair follicle formation.



	Coding				Noncoding			
	p-value	adjusted p-value	AUC	Number of Genes	p-value	adjusted p-value	AUC	Number of CNEs
Cortex	0.00042	0.0050	0.86	8	4.20e-7	2.94e-6	0.65	93
Cuticle	0.010	0.035	0.72	11	1.58e-7	2.21e-6	0.59	287
Medulla	0.0059	0.028	0.72	13	5.80e-3	2.71e-2	0.55	263
Inner Root Sheath	0.15	0.41	0.60	19	0.54	0.66	0.51	626
Bulge	0.34	0.52	0.57	15	0.25	0.50	0.48	397
Outer Root Sheath	0.30	0.52	0.56	28	0.55	0.66	0.49	695
Matrix	0.71	0.90	0.55	4	0.078	0.18	0.46	191
Dermal Sheath	0.00072	0.0050	0.55	436	0.059	0.16	0.50	13482
Arrector Pili	0.30	0.52	0.54	45	0.39	0.66	0.51	2421
Sebaceous Gland	0.21	0.48	0.54	94	0.75	0.75	0.50	2709
Bulb	0.96	0.96	0.50	12	0.57	0.66	0.49	470
Infundibulum	0.87	0.94	0.48	7	0.73	0.75	0.51	218
Dermal Papilla	0.88	0.94	0.48	5	0.049	0.16	0.55	139
Melanocytes	0.52	0.72	0.46	19	0.48	0.66	0.49	663

Figure 34 RERconverge Results in Hair Compartments. Diagram of hair shaft and follicle with shading representing region-specific enrichment for coding and noncoding sequence. Both coding and noncoding sequence demonstrate accelerated evolution of elements related to hair shaft (cortex, cuticle, and medulla). Noncoding regions demonstrate accelerated evolution of matrix and dermal papilla elements not observed in coding sequence. All compartment genesets were compiled from MGI annotations that contained the name of the compartment except arrector pili (Santos et al., 2015) and dermal sheath (Heitman et al., 2020) genesets.

Overall, these results indicate strong enrichment for hair-related function in both protein coding genes and non-coding regions that are convergently accelerated in hairless species.

4.2.4 Analyses Reveal Novel Putative Hair-Related Genetic Elements

After extensive filtering using RERconverge statistics, Bayes Factors, and permulation statistics, several novel putatively hair-related genes were uncovered. Shown in Table 2, the top accelerated gene associated with hairlessness with strong support for hairless-related signal as opposed to marine-related signal was *FGF11*. While *FGF11* has no known role in hair growth, its expression is highly enriched in the skin and other fibroblast growth factor genes are known to be related to hair growth (Kawano et al., 2005; Lee et al., 2019; Nakatake, Hoshikawa, Asaki, Kassai, & Itoh, 2001; Rosenquist & Martin, 1996; Suzuki, Ota, Ozawa, & Imamura, 2000). Together these observations support *FGF11* as another strong candidate for hair-related function.

The second-ranked gene, *GLRA4*, a glycine receptor subunit, is more difficult to interpret because while generally conserved across mammals, it is a pseudogene in humans, so it has been relatively less studied. Glycine receptors are often involved in motor reflex circuits (Callister, 2010), and thus with respect to any functional relevance to hair we hypothesize that GLRA4 may contribute to regulating the reflexive piloerection response (hairs "standing on end") observed in many mammals.

Table 2 Top RERconverge Hair Results. Genes whose evolutionary rates are significantly associated with the hairless phenotype with significant parametric p-values, significant permulation p-values, positive statistic, and hairless versus marine Bayes factors (BF) greater than five. BF Marine and BF Hairless are bayes factors for those phenotypes individually, while BF HvM is the ratio of the two (BF Hairless/BF Marine). The ratio of the Bayes factors quantifies the amount of support for the hairless phenotype beyond the support for the marine phenotype per gene. In other words, a high Bayes factor ratio indicates a signal of evolutionary convergence associated with hairlessness that is not only driven by signals of convergence in hairless marine mammals. Also shown are enrichment statistics for noncoding regions near top genes. Adjusted p-values are Benjamini-Hochberg corrected. Note that permulation p-values observed as 0 were adjusted to 0.001 (the smallest observable permulation p-value) prior to multiple hypothesis testing correction. Cells with missing values (for "Noncoding Enrich Stat" and "Noncoding Enrich p-adj") do not have enough observations to calculate enrichment statistics because too few conserved noncoding elements were detected in the vicinity of those genes.

Gene	Statistic	p-adj	BF HvM	BF Hairless	BF Marine	Perm p-adj	Noncoding Enrich Stat	Noncoding Enrich p-adj
FGF11	0.403	0.205	116.4	6354.7	54.6	0.201	-0.115	0.051
GLRA4	0.332	0.179	22.6	1908.3	84.3	0.201	-0.159	0.068
ANXA11	0.328	0.179	25.5	45.2	1.8	0.201		
PTPRM	0.326	0.179	51.7	4393.6	85.0	0.201	0.146	1.19e-9
PKP1	0.323	0.179	5.6	2669.0	478.9	0.201	0.117	0.410
KRT2	0.304	0.205	2235.7	27034.4	12.1	0.201	0.175	0.181
MYH4	0.297	0.205	28.0	11447.2	409.3	0.201	0.147	0.311
KRT35	0.293	0.205	8.6	1954.5	227.3	0.201	0.142	0.211

Other top-accelerated genes are *KRT2*, *KRT35*, *PKP1*, and *PTPRM*, all of which are known hair-related genes. *KRT2* protein product localizes in the hair follicle and may play a role in hair and skin coloration (Cui et al., 2016), and *KRT35* is a known target of *HOXC13* and is essential for hair differentiation (Lin et al., 2012). *PKP1* mutations lead to ectodermal dysplasia/skin

fragility syndrome, which includes abnormalities of both skin and hair development (Sprecher et al., 2004). *PTPRM* regulates cell-cell communication in keratinocytes (Peng et al., 2015).

The remaining accelerated genes are also plausibly connected to skin- and hair-related functions. *ANXA11* has been strongly linked to sarcoidosis in humans (Hofmann et al., 2008), an inflammatory disease in epithelial tissue. *MYH4*, a myosin heavy-chain protein, has surprisingly also been implicated in skin and hair growth, both through upregulation during hair follicle cycling and skin healing (Carrasco et al., 2015) and upregulation in response to overexpression glucocorticoid receptors that drive hair follicle morphogenesis (Donet, Bayo, Calvo, Labrie, & Pérez, 2008). Note that in both cases of *MYH4* upregulation, it was the only myosin with significantly different expression in the tissues studied, suggesting a unique role for the protein in skin and hair growth.

In addition to identifying genes with significant evolutionary rate shifts in coding sequence, we have also found many other protein-coding genes with significant enrichment of hairless-accelerated non-coding elements in their vicinity (https://pitt.box.com/s/b8ozkcwzile4znq8tw9ri8s160zjb3uw and Figure 35). There is a global trend in correlation between evolutionary rate shift statistics for protein-coding regions and enrichment statistics for their nearby non-coding regions (Pearson's Rho = 0.177). Concordance between accelerated evolutionary rates in genes and their nearby noncoding regions is particularly strong for *KRT*s and *KRTAP*s, which are known to be skin- and hair-related (Figure 35B) – out of 69 *KRT*s and *KRTAP*s for which noncoding enrichment could be calculated, 66 showed accelerated evolution in both protein-coding sequence and non-coding regions. However, across all genes with strong signals for nearby noncoding regions under accelerated evolution in hairless mammals (permulation p-value less than or equal to 0.03), acceleration in the coding sequence itself spans a

wide range of values (Figure 35A), and in many cases there is little evidence of evolutionary rate shifts in the coding sequence. This range likely reflects the requirement that some protein-coding sequences remain under strong evolutionary constraint because of their continued importance in non-hair-related tissues.

Top-ranked genes with accelerated nearby noncoding regions include several known hair-related regulator genes (*ELF3*, *FOXC1*, and others) (Figure 35C). *FOXC1* is a transcription factor involved in maintaining the hair follicle stem cell niche (Lay et al., 2016; L. Wang, Siegenthaler, Dowell, & Yi, 2016) and *ELF3* is known to regulate transcription of keratin genes (Aldinger et al., 2009). These genes showed no coding region acceleration, which is expected since they are highly pleiotropic. Regulatory proteins tend to have many functions – for example, in addition to their hair-related functions, *FOXC1* regulates embryonic development (Brembeck, Opitz, Libermann, & Rustgi, 2000; Seo et al., 2006) and *ELF3* is involved in the epithelial to mesenchymal transition (Sengez et al., 2019) – so we expected to observe no loss of constraint in the coding sequence for those proteins. Instead, changes to regions that regulate expression of those regulatory proteins appear to be driving the convergent evolution of hairlessness. While regulation of transcription factor expression is highly complex, our analysis pinpoints regions that are candidates for hair-specific regulation.

The global analysis of noncoding regions also revealed under-characterized regions (*CCDC162-SOHLH2*, *FAM178B*), and regions that may plausibly be connected to hair or skin (*UVSSA* (Sarasin, 2012), *OLFM4* (Jaks et al., 2008; Muñoz et al., 2012), *ADRA1D* (Rezza et al., 2016)). These noncoding regions are excellent candidates for further experimental analyses to explore their role in regulating hair and skin growth, development, and cycling.

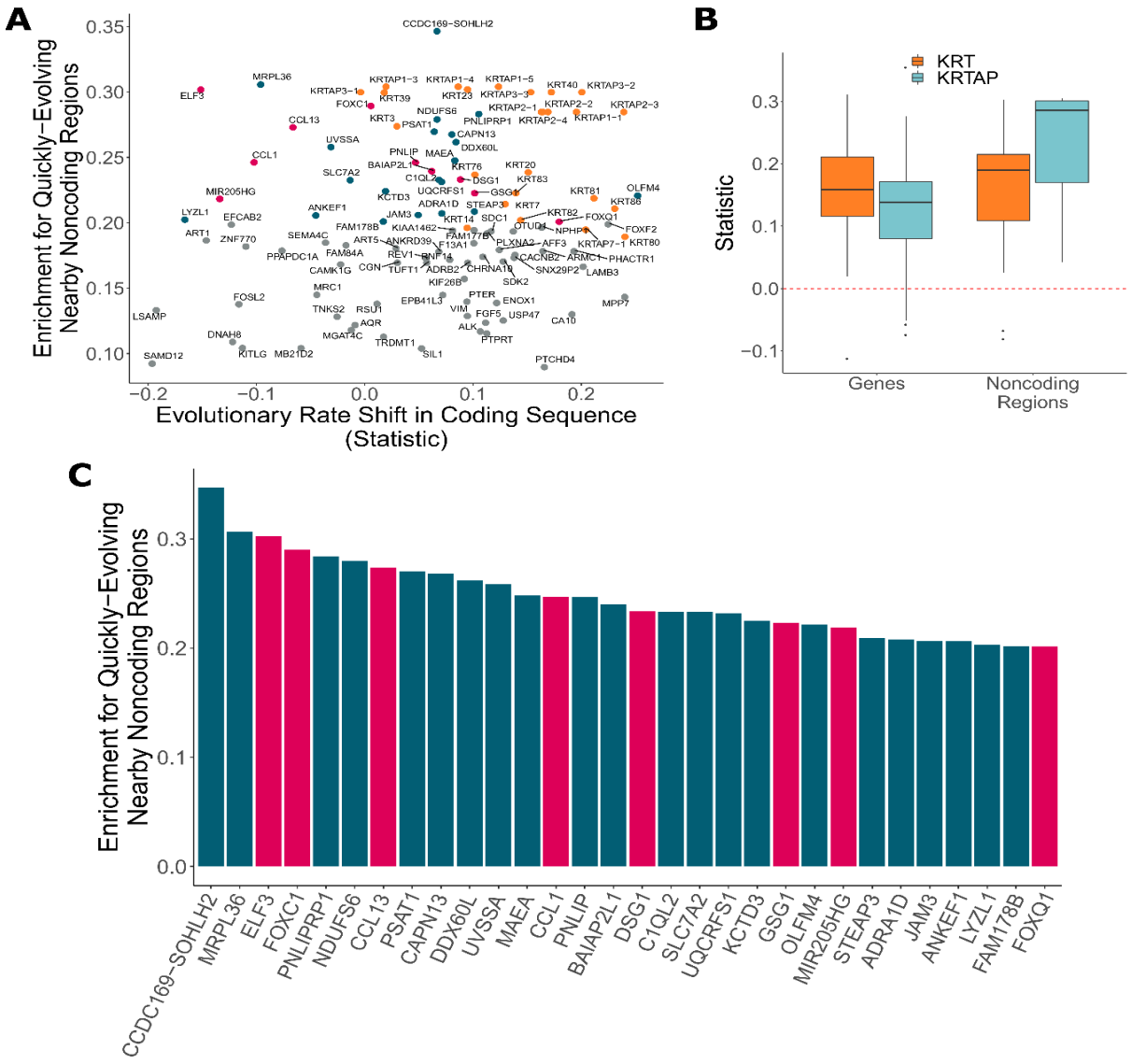


Figure 35 Hairless Coding and Noncoding Results. Noncoding regions near hair-related genes evolve faster in hairless species. A) Genes with a significant enrichment for quickly evolving nearby noncoding regions (permulation p-value of 0.03 or less) only sometimes demonstrate evolutionary rate shifts in their protein-coding sequences. In orange are keratins and keratin-associated proteins, which tend to show accelerated evolutionary rates in both genes and nearby non-coding regions. In pink are top genes also in pink in panel C. In blue are all other genes in panel C. B) Keratin (*KRT*) and keratin-associated protein (*KRTAP*) genes and nearby noncoding sequence show enrichment for accelerated evolutionary rates. Show are rate shift statistics for genes and enrichment statistics for noncoding regions. C) Many top-ranked genes for nearby

quickly evolving noncoding regions are hair-related. Depicted are the top 30 genes (*KRT*s and *KRTAP*s excluded) based on enrichment statistic with enrichment permulation p-value of 0.03 or less. No genes had significant evolutionary rate shifts in coding sequence except *OLFM4*, which evolves faster in hairless species. In pink are genes with hair-related functions in the literature (citations: ELF3 (Blumenberg, 2013), FOXC1 (Lay, Kume, & Fuchs, 2016), CCL13 (Michel et al., 2017; Suárez-Fariñas et al., 2015), CCL1 (Nagao et al., 2012), DSG1 (Jing Zhang et al., 2017), GSG1 (Aya, Shimokawa, & Doi, 2009), *MIR205HG* (D. Wang et al., 2013), *FOXQ1* (Ashburner et al., 2000; Carbon et al., 2019)).

Perhaps even more so than genes and their regulatory regions, microRNAs are strong candidates for hair-related functions. A key component of hair follicle cycling is persistence of stem cells, and microRNAs are known to be important players in stem cell regulation (Peng et al., 2015). Too small to be analyzed via their sequence alone using our analysis strategy, we mapped noncoding regions to nearby microRNAs and performed enrichment analyses to identify groups of microRNA-associated noncoding regions enriched for significant association with the hairlessness phenotype (Figure 36A). The top enriched microRNA with rapidly evolving nearby noncoding regions was mir205, a microRNA known to be associated with skin and hair development (D. Wang et al., 2013). *Mir205* is readily studied because its host gene (*mir205hg*) is long enough to be captured using standard methods, including bulk RNA sequencing. Reanalyzed data from a previous study (Jing Zhang et al., 2017) focused on coding sequence revealed read pileups at *mir205hg* even without microRNA-specific capture methods (Figure 36C and Figure 36D). Through our study, we now know which noncoding regions in the gene desert around *mir205* potentially control its expression in hair follicles as opposed to other tissues. Through this scan for associated noncoding elements, we similarly identified several poorly characterized microRNAs with significant hair-related signal that are less studied and are also

strong candidates for hair-related functions (Figure 36A). Furthermore, we have identified the precise noncoding regions that likely control their expression in the context of hair and hair follicles.

A

	Statistic	p-adj
MIR205	0.238	8.6e-10
MIR1305	0.162	1.1e-6
MIR924	0.178	6.8e-6
MIR124-1	0.218	2.4e-5
MIR346	0.283	2.4e-5
MIR759	0.152	8.3e-5
MIR3167	0.150	2.1e-4
MIR4255	0.189	4.4e-4

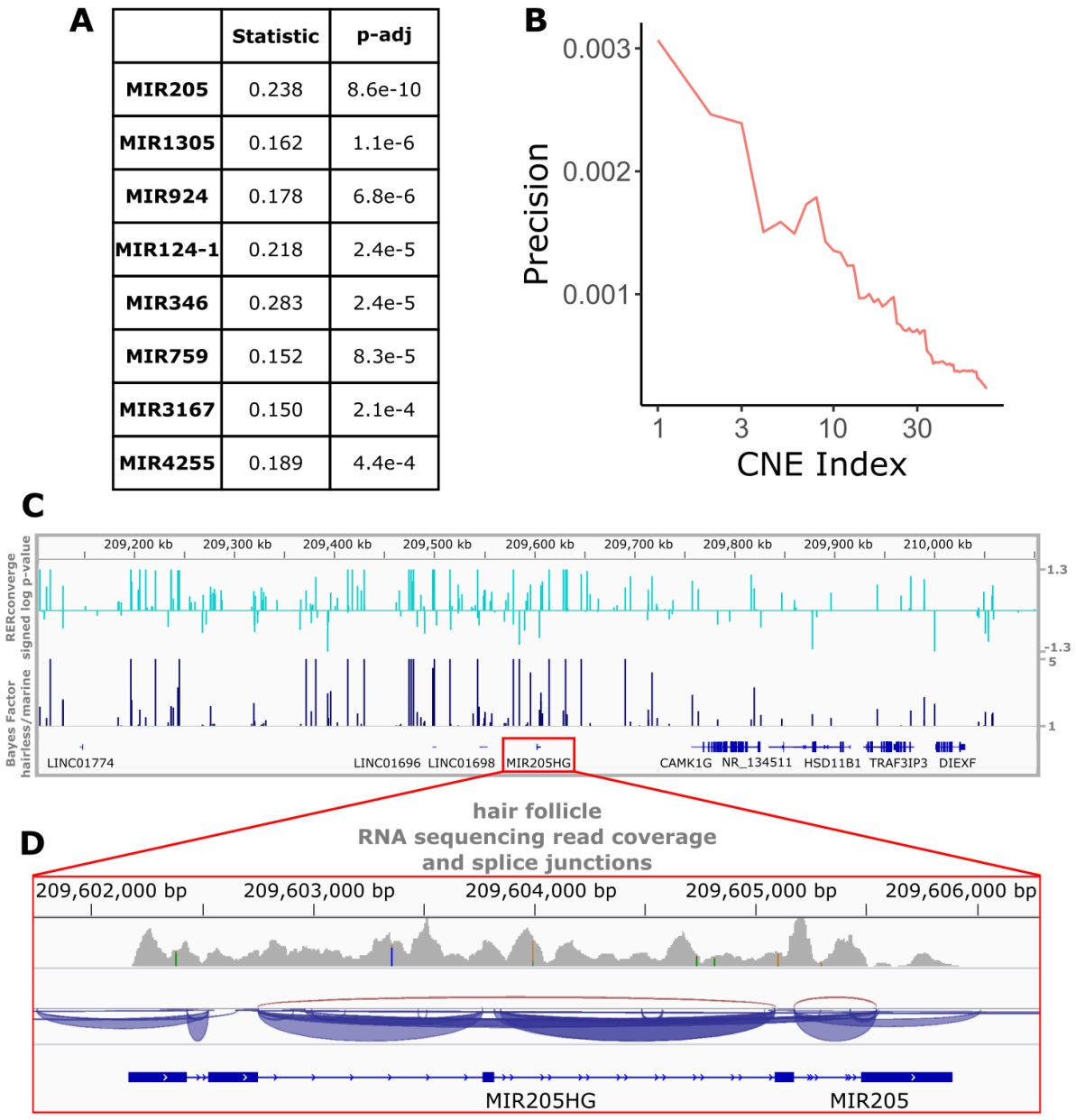


Figure 36 microRNA Results. Top miRNAs with nearby noncoding regions with evolutionary rates significantly associated with the hairless phenotype. A) Wilcoxon rank-sum enrichment statistics and Benjamini-Hochberg corrected p-values for top-ranked miRNAs. B) Precision recall curve of statistic ranks for CNEs near *mir205* demonstrates an enrichment of CNEs with accelerated evolution near *mir205* compared to all noncoding regions near microRNAs. C) The chromosomal region around *mir205* shows a large number of CNEs accelerated in hairless species, as seen for RERconverge and Bayes factor scores. Note

the relative decline of peaks in the vicinity of nearby protein-coding genes such as *CAMK1G* to the right. D) *mir205* is well-known to be associated with hair and skin growth and structure. Its transcriptional unit on chromosome 1 shows clear read pileups from hair follicle RNAseq data (Jing Zhang et al., 2017). Gray peaks represent the number of RNAseq read coverage and blue curves represent splice junctions.

4.3 Discussion

These analyses successfully used RERconverge, a method to link convergent evolutionary rates of genetic elements with convergent phenotypes, to identify known hair-related genes in mammals. In addition to identifying known genes, other understudied genes and microRNAs were also identified as key plausible targets for further inquiry into the genetic basis of hairlessness, and a suite of putative regulatory elements associated with hair and skin were uncovered.

The top-ranked gene was *FGF11*, a fibroblast growth factor gene. It evolved faster in hairless species due to relaxation of evolutionary constraint, indicating that it has reduced functionality in hairless species. Fibroblast growth factors are readily studied for a variety of functions, but the precise functionality of *FGF11* is unknown. The gene may be associated with cancer development through interaction with T-cells (Ye et al., 2016), and it has also been implicated in tooth development in mice (Kettunen, Furmanek, Chaulagain, Hals Kvinnsland, & Luukko, 2011). Interestingly, the gene related to hairlessness in Mexican Hairless dogs is also related to dentition (Drögemüller et al., 2008), implying plausibility for a hair-related gene to also be tooth-related. Furthermore, numerous fibroblast growth factor genes have been studied in relation to hair growth (Rosenquist & Martin, 1996), including work that found errors in *FGF5* resulted in longer hair in goats (G. Li et al., 2019), *FGF5* and *FGF7* regulation controlled hair

anagen phase in mice (Lee et al., 2019), and *FGF2* stimulated hair growth when applied to mouse skin (Xu, Chen, Wang, Xue, & Fu, 2018). *FGF11* is an excellent candidate to perform similar tests for hair-related functions, among other high-scoring genes in the list such as *MYH4* and *ANXA11*.

Compared to coding sequence, study of noncoding regions is more challenging for several reasons. First, identifying such regions genome-wide is difficult because they lack the defining characteristics that genes share, such as start and stop codons, and thus finding putative regulatory elements using sequence alone is an ongoing area of study. Our strategy of using conserved regions as putative regulatory elements likely misses many real regulatory sequences while simultaneously capturing conserved elements with no regulatory function. However, our method is also unbiased and provides a robust set of sequences to analyze, many of which likely do have regulatory functions. Second, validating our findings from noncoding regions is difficult because few CNEs have known functions. Therefore, to validate our noncoding results, we mapped noncoding regions to nearby genes and inferred CNE functions based on the functions of those genes. Such proximity-based mapping has known flaws because enhancers can have distal effects and chromatin state controls enhancers' access to genes regardless of distance. However, despite all of the potential sources of error, we identify global signal for noncoding regions under accelerated evolution in hairless species (Figure 30B) and signal for hair-related acceleration of noncoding regions (Figure 34, Figure 35, and Figure 36).

Further analyses of noncoding regions revealed an interesting deviation from signals of accelerated evolution in coding regions. Namely, coding regions primarily showed acceleration in genes related to texture and the structure of the hair shaft itself. Noncoding regions, on the other hand, showed accelerated evolution near genes related to the dermal papilla and the matrix. Both

regions are essential for hair growth. The dermal papilla is the master controller of hair follicle development and hair growth, and it has in fact been repeatedly shown to be sufficient to cause hair growth. Dermal papilla cells, when transplanted to hairless skin such as footpads, has consistently been shown to result in development of hair follicles (C. A.B. Jahoda et al., 1984; Colin A.B. Jahoda et al., 1993; Reynolds & Jahoda, 1992). Since all mammals are capable of growing hair and do have at least some hair at some point in their life cycles, these findings imply that function of genes related to the dermal papilla must be preserved, and spatial and temporal changes in hair growth may be driven by noncoding regions. Much like the dermal papilla, the hair follicle matrix is essential for hair growth – mitotically-active matrix cells give rise to all other inner hair structures, including the hair shaft and the root sheath. Early-stage matrix differentiation can even progress without dermal papilla signaling (Mesler, Veniaminova, Lull, & Wong, 2017). Hair cannot exist without the dermal papilla and matrix, and alterations to their related noncoding regions could plausibly have a large impact on hair growth capabilities. Changes to their associated regulatory regions, on the other hand, may be more flexible and allow for the changes in hair localization, texture, and density that we observe in near-hairless mammals.

Other genes with nearby accelerated noncoding regions likewise demonstrate conservation in protein-coding sequence, possibly because of strong pleiotropy of hair- and skin-related genes. In fact, among the top-ranked non-keratin genes with quickly evolving nearby noncoding regions, only one gene showed a significant evolutionary rate shift in protein-coding sequence (Figure 35). *FOXC1* and *ELF3*, among the top-ranked genes, are strongly linked to hair and skin development (Brembeck et al., 2000; Lay et al., 2016; L. Wang et al., 2016) but also have other essential functions (Aldinger et al., 2009; Sengez et al., 2019; Seo et al., 2006). Our findings imply that many hair-related genes may have similar pleiotropy preventing accelerated evolution of coding

sequence in hairless species. Instead, plasticity of gene regulation through accelerated evolution of noncoding regions may allow for the evolution of hairlessness.

Additionally based on noncoding sequences near regions of interest, mir205 was found to be the top-ranked microRNA with nearby noncoding sequences under accelerated evolution. Mir205 is well-established microRNA related to hair and skin development (D. Wang et al., 2013), and it thereby serves as a strong validation that signals of convergent evolution are successfully identifying hair-related elements. The second-ranked mir1305 has been implicated in skin functionality with significantly different expression levels in damaged versus healthy skin (Liang et al., 2012). Numerous microRNAs have been implicated in hair- and skin-related functions (Andl & Botchkareva, 2015; Fu, Zhao, Zheng, Li, & Zhang, 2014), likely a subset of the total elements involved in hair growth. In general, microRNAs are likely key players in hair follicle cycling because of their importance in stem cell regulation (Peng et al., 2015), and the microRNAs and their associated noncoding regions identified in this work serve as a valuable list of candidates for further inquiry. Likewise, noncoding regions near other hair-related genes are also under accelerated evolution in hairless species and may regulate hair- and skin-related functions. Further, some under-characterized and plausibly hair- and skin-related genes, such as *CCDC169-SOHLH2* and *FAM178B*, have nearby accelerated noncoding regions and thus identify those genes and their regulatory regions as candidates for further experimental testing.

This study has revealed a slew of fresh candidate genes, noncoding regions, and microRNAs putatively associated with hair growth. As an unbiased, genome-wide scan across a large swath of the mammalian phylogeny, it represents not only a step toward fully understanding hair growth, but also understanding the evolution of hair across all mammals.

4.4 Methods

4.4.1 Calculating body size-regressed relative evolutionary rates

The RERconverge package in R was used to generate phylogenetic trees for each gene and noncoding region in which branch length represented the amount of evolutionary change, or the number of nonsynonymous substitutions, that occurred along that branch as described in several previous publications (Amanda Kowalczyk et al., 2019, 2020; Partha et al., 2019). Alignments for 19,149 genes in 62 mammal species were obtained from the UCSC 100-way alignment (Blanchette et al., 2004; Harris, 2007; Kent et al., 2002). The topology used to generate element-specific trees is included below under Phylogenetic trees.

Likewise, alignments for 343,598 conserved noncoding elements were extracted based on phastCons conservations scores across the 62 mammal species and the blind mole-rat (*Nanospalax galili*) (Siepel et al., 2005). Briefly, the full set of conserved elements across 46 placental mammals and their respective phastCons scores were downloaded from the UCSC genome browser (Kent et al., 2002) from the hg19 (human genome) "Cons 46-way" track (phastConsElements46wayPlacental). Regions that overlapped coding regions were removed using the UCSC genome browser "Intersection" utility and the "Genes and Gene Predictions" annotations from the "GENCODE V28lift37" track. Elements with phastCons scores greater than 350 were maintained, and elements less than 10 base-pairs apart were merged. Finally, elements with fewer than 40 base-pairs were discarded to result in the final 343,598 regions. Orthologs for all 63 mammals were downloaded from the UCSC 100-way alignment. Blind mole-rat elements were added based on the pairwise alignment between hg38 (human genome) and *Nannospalax*

galili genome (Zerbino, Wilder, Johnson, Juettemann, & Flicek, 2015) by first mapping hg19 coordinates to hg38 coordinates (Supp. File 7 and 8). Orthologs were added to the 62 mammal species alignments using MUSCLE (Edgar, 2004). Data generation steps and code are available here: https://pitt.box.com/s/m6gineor6ergnjk07s8515o27wm37lyx.

Alignments were used to generate evolutionary rate trees based on a well-established topology of the mammalian phylogeny in the Phylogenetic Analysis by Maximum Likelihood (PAML) program (Wynn K Meyer et al., 2018; Yang, 2007). Briefly, RERconverge was used to convert evolutionary rate information from each gene- or noncoding element-specific tree by correcting for the mean-variance relationship among branch lengths and normalizing each branch for the average evolutionary rate along that branch such that the final branch length was relative to the expectation for that branch (Partha et al., 2019).

The resulting relative evolutionary rates were used to calculate body size-regressed relative evolutionary rates. Using adult weight information for the 62 mammal species obtained from the Anage Animal Aging and Longevity Database (Tacutu et al., 2018), RERconverge functions were used to predict body size phenotype values throughout the mammalian phylogeny. Residuals from a linear model fitted to the phenotype values and the relative evolutionary rates for each gene and conserved noncoding element were extracted and used as the body size-regressed relative evolutionary rates for that element.

RER matrices and phylogenetic trees are available at: https://pitt.box.com/s/newoupek4nlfnx1y5xri7xnx9i96ir9s

4.4.2 Defining hairless species

Since all mammals have some hair during at least one stage of life, no species are truly hairless. Therefore, classification of species as "hairless" versus "haired" was qualitatively based on density of hair covering and quantitively based on the impact of removing species on the hair-related signal detected during analyses. Tendency was the err on the side of leniency when assigning species as hairless – any species with reduced hair quantity was classified as hairless.

Extant species classified as hairless were armadillo, elephant, white rhinoceros, pig, naked mole-rat, human, and marine mammals (manatee, pacific walrus, dolphin, and orca). The hairless set comprised all but one marine mammal in the 62 mammal species (the furry Weddell seal is not included in the hairless set). The only non-extant species classified as hairless was the orca-dolphin ancestor (the cetacean ancestor) because that species was likely also a hairless marine species (Z. Chen et al., 2013; Nery et al., 2014). The elephant-manatee ancestor was not classified as hairless because modern elephants have known extinct hairy sister species (wooly mammoths) that diverged after the elephant-manatee divergence (Roca et al., 2009). Thus, classifying the elephant-manatee ancestor as hairy was the most parsimonious phenotype assignment for the afrotherian clade. The classification was also supported by the data, which indicated a stronger signal for skin-related genes when the elephant-manatee ancestor was classified as hairy (Figure 31).

Although some species are undeniably hairy (dog, cat, sheep, etc.) and some are undeniably relatively hairless (orca, dolphin, elephant, etc.), some species are borderline cases. For example, the tenrec and hedgehog appear to have "spikes" rather than hair. However, tenrec and hedgehog spikes (as well as porcupine quills), are modified hairs (Leon Augustus, 1920), so we classified

tenrec and hedgehog as hairy. Armadillo, pig, and human are likewise classified as hairless species but have relatively greater hair quantity than the other hairless species. The armadillo, like the tenrec and hedgehog, has a unique external modification, but unlike the tenrec and hedgehog, the armadillo's shell is made of bone, not hair (I. H. Chen et al., 2011), so we classified the armadillo as hairless. Pig and human, on the other hand, have non-modified skin that is nearly completely covered in hair (and in the case of humans, the hair is quite dense in some body areas), but both species have large swaths of body area where hair is so sparse that sun-exposed skin is clearly visible. Both species were classified as hairless due to this pervasive low hair density. To assess the impact of species assignment on skin- and hair-related signal, hairless species were systematically removed and relevant enrichment statistics were recalculated. No specific species has a consistently detrimental impact on enrichment for gene sets of interest (Figure 33).

4.4.3 Calculating element-specific association statistics

For each genetic element, evolutionary rates for haired species versus hairless species were compared using Kendall's Tau. Haired species included ancestral species inferred to be haired in addition to extant haired species. Resulting p-values were multiple hypothesis testing corrected using a standard Benjamini-Hochberg correction (Benjamini & Hochberg, 1995).

In addition to calculating parametric p-values, empirical p-values were calculated using a novel permulation strategy modified from a similar strategy developed for continuous phenotypes (Amanda Kowalczyk et al., 2020). First, 1,000 null phenotypes were generated by using Brownian motion phylogenetic simulations and assigning the top ten values as hairless species. Resulting phenotypes were backpropagated along the phylogeny to ensure that final null phenotypes

contained a total of eleven foreground species with only a single ancestral species classified as hairless. Such a procedure matched the organization of null phenotype values to true phenotype values. Hypothesis testing was repeating using all null phenotypes, and the empirical p-values were calculated as the proportion of permulations with statistics as extreme or more extreme than the parametric statistic for the real phenotype values.

4.4.4 Calculating element-specific Bayes Factors

In addition to calculating element-specific association statistics, Bayes factors were calculated for each gene using the marine and hairless phenotypes. Briefly, Bayes factors quantify the support for a linear model predicting phenotype using evolutionary rate information from each gene, with a higher Bayes factor indicating greater support. The ratio of Bayes factors between the hairless and marine phenotypes quantifies the level of support of one phenotype over the other and thus can be used to tease apart intricacies of the two heavily-confounded phenotypes.

4.4.5 Calculating enrichment statistics

Enrichment statistics were calculated using Mouse Genome Informatics (MGI) gene sets (Blake et al., 2003), GTEx tissue annotations (Papatheodorou et al., 2018), GO annotations (Ashburner et al., 2000; Carbon et al., 2019), and genes highly expressed in hair follicles (Jing Zhang et al., 2017). The 70 hair-follicle-specific genes were obtained by selecting the top 200 hair-follicle-expressed genes and removing genes that were included in the top 10,000 genes with the highest minimum median expression across GTEx tissues, i.e. ubiquitously expressed genes.

Noncoding regions were mapped to annotations via distance from relevant genes – regions within 10,000 bases of a gene were assigned to that gene and its pathways. Noncoding regions were also mapped to microRNA coordinates using the same distance-based metric. All annotations are available at: https://pitt.box.com/s/b8ozkcwzile4znq8tw9ri8s160zjb3uw

Pathway enrichment statistics were calculated using the Wilcoxon Rank-Sum Test, which compares ranks of foreground values for elements in a pathway to background values for non-pathway elements. For each gene or noncoding element, the sign of the statistic times the log of the p-value were used to generate ranks. Empirical p-values from permulations were also generated using the same null phenotypes used for individual elements and detailed in previous work (Amanda Kowalczyk et al., 2020).

4.4.6 Permulations

In addition to computing parametric statistics directly from standard statistical tests, empirical p-values were also calculated using a permulation strategy. Permulations were used to generate null phenotype values, and the empirical p-value was calculated as the proportion of null statistics as extreme or more extreme than the observed parametric statistics. Such a strategy corrects for a non-uniform empirical null distribution at the gene level (Figure 30) and non-independence among genetic elements at the pathway level (Saputra et al., 2020).

4.4.7 Positive selection tests

For top-ranked genes under accelerated evolution in hairless species, all KRT and KRTAP genes, and various genes in top-ranked pathways under accelerated evolution in hairless species, branch-site models to test for positive selection were performed identify if rapidly-evolving genes were undergoing positive selection or merely under relaxation of constraint. Such models were performed using a subset of the full 62 species mammalian phylogeny as shown in the Phylogenetic trees section below.

Significance of relaxation of constraint for hairless species was assessed using likelihood ratio tests (LRT) between Branch-site Neutral (BS Neutral) and its nested null model M1 (sites neutral model) in PAML (Yang, 2007). Similarly, LRTs between branch-site selection model (BS Alt Mod) and its null BS Neutral were used to infer positive selection in hairless species. For each test, p-values were estimated using the chi-square distribution with one degree of freedom. Phylogeny-wide relaxation of constraint was additionally quantified using the LRTs between M2 (sites selection model) vs M1 (sites neutral model) and M8 (sites selection model) vs M8A (sites neutral model) respectively. Prior to performing the mammal-wide tests, hairless foreground species were removed to allow for unbiased estimates of significance of relaxation of constraint and positive selection from only the background mammalian branches. Genes with significant signals of positive selection and non-significant signals of phylogeny-wide acceleration were inferred to be under positive selection.

4.4.8 Phylogenetic trees

Master tree topology with average branch lengths: ((((((((((((ailMel1:0.03854019703,((lepWed1:0.02002160645,odoRosDi:0.02064385875):0.01734764946,musFur1:0.04613997497):0.002879093616):0.009005888384,canFam3:0.05339127565):0.01185166857,felCat5:0.05020331605):0.03285617057,((((((bosTau7:0.02168740723,((capHir1:0.01157093136,oviAri3:0.01246322594):0.0049716126,panHod1:0.01522587482):0.01465511149):0.0662523666,(orcOrc1:0.006371664911,turTru2:0.01086552617):0.06014682602):0.01216198069,susScr3:0.0796745271):0.006785823323,(camFer1:0.01240650215,vicPac2:0.01096629635):0.06374554586):0.02551888691,(cerSim1:0.04977357056,equCab2:0.061454379):0.02510111297):0.00331214686,((eptFus1:0.03248546656,(myoDav1:0.02344332842,myoLuc2:0.01567729315):0.02193849809):0.09455328094,(pteAle1:0.005833353548,pteVam1:0.01611220178):0.07567400302):0.02385546003):0.002057771224):0.004845253848,(conCri1:0.1239823369,(eriEur2:0.1696142244,sorAra2:0.1934205791):0.02079474546):0.0235875333):0.01477733374,((((((chrAsi1:0.1017903453,echTel2:0.1749615473):0.01592632003,eleEdw1:0.1516860647):0.006610995228,oryAfe1:0.08326528894):0.008243787904,(loxAfr3:0.06812658238,triMan1:0.06198982615):0.0224994529):0.03384011363,dasNov3:0.1342602666):0.005989703247,(((macEug2:0.1270943532,sarHar1:0.09944141622):0.02717055443,monDom5:0.1181200712):0.1802966572,ornAna1:0.4322118716):0.2206952867):0.01316436193):0.01425600689,((((((cavPor3:0.09048639907,(chiLan1:0.05332953299,octDeg1:0.08476954109):0.01287861561):0.02118937782,hetGla2:0.08588673524):0.07432515556,speTri2:0.08896424642):0.006291577528,(((criGri1:0.04084640027,mesAur1:0.04456203524):0.02314125062,micOch1:0.06932402649):0.01947113467,(mm10:0.05273642272,rn5:0.05576007402):0.04435347588):0.08380065137,jac

Jac1:0.1438649666):0.04270536633):0.01663675397,(ochPri3:0.1256544445,oryCun2:0.07131655591):0.06535533418):0.009050428462,tupChi1:0.1191189141):0.003894252213):0.01379370868,otoGar3:0.108738222):0.04299750653,(calJac3:0.02474184521,saiBol1:0.02096868307):0.02784675729):0.0135408115,(chlSab1:0.007693724903,((macFas5:0.001292320552,rheMac3:0.00713015786):0.002951690224,papHam1:0.005199240711):0.002049749893):0.015662635620):0.007043113559,nomLeu3:0.01770384793):0.002187630666,ponAbe2:0.0164503644):0.005572327638,gorGor3:0.007765177171):0.001382639829,hg19:0.005957477577,panTro4:0.006721826689);

Subset of master tree used for branch-site models for positive selection:

((((((((((lepWed1:0.02002160645,odoRosDi:0.02064385875):0.01734764946,musFur1:0.04613997497):0.02373665057,felCat5:0.05020331605):0.03285617057,((((((bosTau7:0.02168740723,oviAri3:0.03208995002):0.0662523666,(orcOrc1:0.006371664911,turTru2:0.01086552617):0.06014682602):0.01216198069,susScr3:0.0796745271):0.006785823323,vicPac2:0.0747118422):0.02551888691,(cerSim1:0.04977357056,equCab2:0.061454379):0.02510111297):0.00331214686,(myoDav1:0.1399351075,pteAle1:0.08150735657):0.02385546003):0.002057771224):0.004845253848,(conCri1:0.1239823369,sorAra2:0.2142153246):0.0235875333):0.0147773337 4,(((eleEdw1:0.1582970599,oryAfe1:0.08326528894):0.008243787904,(loxAfr3:0.0681265823 8,triMan1:0.06198982615):0.0224994529):0.03384011363,dasNov3:0.1342602666):0.01915406518):0.01425600689,(((((cavPor3:0.1116757769,hetGla2:0.08588673524):0.07432515556,speTri2:0.08896424642):0.006291577528,((criGri1:0.08345878555,mm10:0.09708989861):0.08380065137,jacJac1:0.1438649666):0.04270536633):0.01663675397,oryCun2:0.1366718901):0.009050428462,tupChi1:0.1191189141):0.003894252213):0.01379370868,otoGar3:0.108738222):0.0

4299750653,calJac3:0.0525886025):0.0135408115,(chlSab1:0.007693724903,rheMac3:0.01213159798):0.01566263562):0.01618571169,hg19:0.005957477577,panTro4:0.006721826689);

5.0 Broader Impacts: Greensburg Area Science Program

While completing my dissertation research, I launched and organized an outreach effort called the Greensburg Area Science Program (GASP). Originally started as a program through a few classrooms at Greensburg Salem High School, the program has since grown to include students from any school districts in Westmoreland County, an area immediately east of Pittsburgh. The program includes field trips to allow students to tour research facilities, guest lectures at classrooms, opportunities for mentorship, and a yearly science fair.

I chose the Greensburg area as the target for my outreach efforts because I grew up in the area and attended the Greensburg Salem School District, so I know firsthand that the community is underserved by science professionals. Since there are no large research universities and few scientific industry hubs near Greensburg, students never meet scientists or learn what science careers look like. When students are choosing a career path, in order for them to believe that the career is actually possible for them, it is essential that they 1) see people in general following that career path and 2) see people like them following that career path. Since I fit both of those criteria for students in the Greensburg area, it seemed like a natural choice for a region to enrich with scientific outreach.

Beyond being underserved, the Greensburg area is also home to a wide array of individuals from a huge span of socioeconomic backgrounds. While a small number of students are from wealthy backgrounds, many students are from relatively lower income homes. Many such students would be first-generation college students, and many more would be first-generation graduate

school students. Without a parent with an advanced education to guide their paths, many students have no idea what it means to get a PhD, let alone how to get one.

My outreach efforts in the Greensburg area started with yearly field trips to bring students to visit research facilities at the University of Pittsburgh. In the 2017-2018 and 2018-2019 school years, students from one or two classrooms visited the University of Pittsburgh to attend a seminar about how to get a PhD, hear lightning talks about computational biology, tour laboratory facilities, and be paired with a graduate student, postdoctoral fellow, or a professor for a mini one-on-one job shadow to learn what a job in scientific research looks like. Unfortunately, the pandemic eliminated the possibility of field trips during the 2019-2020 and 2020-2021 school years, but the two earlier years were huge successes. They sparked longer term job shadows for multiple students, and at least one student remained in contact with their mini job shadow graduate student to receive mentorship for a research project.

In addition to field trips and in order to serve more students, I also visited Greensburg Salem High School to give several guest lectures during the 2017-2018 and 2018-2019 school years. Some topics were more exploratory, such as "What is computational biology?" and "What is a PhD?", but other lectures were more focused on scientific topics the students were learning during the course. One such lecture focused on phylogenetics and how to use BLAST to find sequences and build a phylogenetic tree. Another focused on reading scientific literature and asked students to read and be prepared to discuss a scientific paper. During every lecture, students started out unsure and hesitant, but always ended visibly more confident in their ability to understand "real science".

After conducting lectures and field trips, my outreach program grew to include a yearly science fair for students in the Spring of 2019, 2020 (virtual), and 2021 (virtual). The first science

fair included only five groups with six total students, and this year's event has grown to 12 groups with 25 total students. Originally targeted to Greensburg Salem High School students for its first two year, this year's science fair is open to middle and high school students throughout Westmoreland County. Unlike traditional science fairs in which students conduct projects primarily independently, the GASP science fair includes mentorship and numerous workshops to guide students through the research process. Each group is paired with a graduate student mentor as a direct point of contact to help them organize their research project, which not only gives the student participants valuable knowledge and networking, but also helps graduate students gain mentoring skills. Numerous workshops help students pick their project topics, design their experiments, perform basic statistical analyses, and prepare their poster presentations for the science fair event. The science fair is open to the public and is largely supported by community organizations, such as the Delmont Lions Club and the Kiwanis Club of Greensburg. Additional support comes from the University of Pittsburgh Biomedical Graduate Student Association, the Department of Computational and Systems Biology, and the Integrative Systems Biology PhD Program. The 2019 science fair was also supported by an American Society for Cell Biology COMPASS grant.

Together, these efforts represent a substantial push to introduce Greensburg area students to careers in science. I hope to see participation and impact continue to grow over the years to continue showing even more students that careers in science are possible for them.

6.0 Conclusions

In this dissertation, I have demonstrated that computational methods to connect genetic elements to convergently evolving phenotypes are powerful and statistically robust tools. As genome sequencing technology continues to improve and more genomes are available, these tools will become even more relevant to study a diverse array of phenotypes and to perform preliminary functional annotation tasks for genome regions such as enhancers and microRNAs whose functions can be difficult to identify experimentally. My results demonstrate the utility of RERconverge, permulations, and branch-site models for positive selection, all of which are versatile and scalable to large datasets.

In chapter 1, I discuss my three methods, RERconverge, permulations, and branch-site models for positive selection, and demonstrated their effectiveness. Both new methods, RERconverge and permulations, are already used by numerous external labs despite only being recently published, which further highlights their utility in a huge array of research designs. Tests of RERconverge alone indicate that it is fast, easy to use, and scalable to large genomic datasets for use with both binary and continuous phenotypes. Permulations enhance RERconverge function by generating an empirical null under a phylogenetic simulation framework. For individual genetic elements, the empirical null corrects for non-independence and other unknown confounders when seeking correlations between each genetic element and a convergent phenotype. For pathway enrichment analyses, the non-independence correction is even more dramatic because genes are known to coevolve, and thus permulations provide a more valid pathway enrichment analysis than RERconverge alone. Finally, a note about proper use of branch-site models for

positive selection to test for convergent positive selection highlights the necessity of proper model formulation to distinguish phylogeny-wide positive selection from convergent branch-specific positive selection. Although branch-site models are a foundational and widely used method, in published manuscripts, they are often used incorrectly to test for positive selection. My strategy shows that the "drop-out method", a test in which species with a convergent phenotype are removed prior to running a second set of tests for selection, helps distinguish phylogeny-wide positive selection from branch-specific positive selection.

In chapter 2, I demonstrate work to use the methods from chapter 1 to identify genetic elements associated with the evolution of extreme lifespan in mammals. Using my new longevity phenotypes, the first and second principal components of body size and lifespan, I identify a suite of genes and pathways that evolve slower in long-lived species. Little signal is found for genes evolving faster in species with extreme lifespan, perhaps because the period of positive selection to drive lifespan change was very short in terms of evolutionary time. On the other hand, as some genes were theoretically positively selected to increase lifespan and body size, the entire evolutionary landscape of genes shifted and genes whose function was essential to enable long lifespan became under increased constraint to protect them from mutations. Such increased evolutionary constraint persists to the current day, and thus those signals of convergence are stronger and can be detected using RERconverge. For large and long-lived species, we find that cancer control functionalities, and in particular cell cycle control, evolve slower and thus are likely essential to enable the phenotype. This finding helps to resolve Peto's Paradox, which states that large (and long-lived) species should get cancer more often than small (and short-lived) species simply because they have more cells than duplicate more times and thus have greater opportunity to acquire deleterious mutations. However, cancer rates across species of different sizes are

actually fairly similar. My findings suggest that robust cancer control mechanisms across large and long-lived species may be the key to protect them from cancer. For species that are long-lived independent of body size, I find a unique set of DNA repair mechanisms under increased evolutionary constraint. This is consistent with findings from previous work in individual species that found efficient DNA repair was important to slow senescence independent of the size of the organism.

In chapter 3, I show results using RERconverge with a binary phenotype, mammalian hairlessness. Unlike the longevity phenotypes, the hairlessness phenotype shows many genes that are under accelerated evolution. Since branch-site models for positive selection show little evidence for positive selection, it appears that many genes are under reduced evolutionary constraint in association with hairlessness, perhaps because of reduced functionality. Since many such genes were known to be hair-related, such as keratins and genes associated with hair follicle compartments, gene-based analyses served as validation that the method was working correctly. Analysis of noncoding regions showed regions with both increased under decreased evolutionary rates, and distance-based mapping of noncoding regions to nearby genes showed that genes and their putative regulatory elements often have very different evolutionary rates landscapes in association with hairlessness. Notably, genes associated with the matrix and dermal papilla of the hair follicle showed little evolutionary rate shifts in genes but significant shifts in noncoding regions, suggesting a regulatory component of natural mammalian hairlessness associated with key hair follicle regions. Noncoding analysis further allowed me to identify putative enhancers and microRNAs that are potentially associated with hairlessness and are strong candidates for further experimental validation.

Finally, my broader impacts section highlights my work with a local underserved community to help high school and middle school students see that a career in science is possible for them. I describe the initiatives that I piloted, including organizing field trips and guest lectures, facilitating mentorship and job shadows, and planning a yearly science fair. Since its conception at the beginning of my doctoral research, the program has grown to include numerous high schools and include many local students, and I plan to continue growing the program in the future. Diversity of individuals conducting scientific research is essential to foster innovation and ensure equity across all populations of people. I hope that my outreach serves as a model for future outreach initiatives to help students pursue careers in the sciences.

As we move into the future of science, technologies to study genomic sequences will continue to improve. Soon we will have genome sequences available for every species on Earth, and we will need methods to analyze those genomes. Continuing to implement and improve methods like RERconverge will allow us to continue performing rigorous analyses of convergent sequence evolution like those described in this dissertation. In particular, study of regulatory elements and their networks remains in its infancy, and the field is in dire need of gold standard strategies to computationally identify regulatory sequences, map regulatory sequences to their associated genes, and build networks of transcriptional networks at the whole genome scale. Measuring sequence changes alone is insufficient to fully understand how genomic sequence functions – we must strive for a multi-dimensional view of not only what the sequence looks like, but when and where it is expressed, how its expression affects the expression of other sequence, how it is co-expressed with other sequences, and how it and other sequences ultimately work together to drive phenotypes. A full understanding will require multi-faceted approaches and innovation that I hope to be a part of in my future work.

A fuller understanding of whole genome function and interactivity would provide innumerable benefits to other genomic technologies. Fully understanding regulatory function in particular would allow us to design gene therapies that target specific cell types or preferentially target tumor cells over healthy cells. It could also help guide modifications using the CRISPR-Cas9 gene editing system by allowing us to accurately and fully predict how edits will affect the genome. As it currently stands, CRISPR-Cas9 is like the queen on a chess board – incredibly powerful, but it is impossible to know the best place to move her without being able to see the rest of the pieces. The rest of the pieces is a full understanding of the genome, and that would allow us to make the most out of gene editing technology.

Improvements in technology in the world at large are also continually making science more accessible to all, and outreach efforts that adapt to the modern world will be essential to continuing to diversify scientific research. Particularly for computational research, many resources are completely free and available to anyone with a computer and internet access. Such resource availability makes it possible for anyone, anywhere to conduct research. In conjunction with technological advancements, the movement toward all-open-access journals will allow anyone to access scientific research information and make science more equitable and accessible to all. Scientific innovation cannot continue without welcoming a diverse set of individuals with a wide range of ideas to apply to difficult research questions. My future work will tackle those difficult questions, and I plan to continue engaging in outreach to attract diverse minds to help me answer them.

Appendix Permissions to Reuse Copyright Content

Text from Chapter 1 about RERconverge originally published in (A. Kowalczyk et al., 2019) was reused from *Bioinformatics* under policies that allow full reuse in author theses: https://global.oup.com/academic/rights/permissions/autperm/?cc=us&lang=en&

Text from Chapter 1 about permulations originally published in (Saputra et al., 2021) was reused from *Molecular Biology and Evolution* under https://global.oup.com/academic/rights/permissions/autperm/?cc=us&lang=en&

I retain copyright for the text from Chapter 1 about proper use of branch-site models to detect convergent positive selection, which is not yet available as a preprint or a peer-reviewed manuscript.

Text from Chapter 2 about implementation of RERconverge to analyze the longevity phenotype originally published in (A. Kowalczyk et al., 2020) was reused from *eLife* under https://creativecommons.org/licenses/by/4.0/legalcode

I retain copyright for the text from Chapter 3 about the implementation of RERconverge to analyze the hairlessness phenotype, currently available as a preprint on *bioRxiv* at (Amanda Kowalczyk et al., 2021).

Bibliography

Ahmad, W., Ul Haque, M. F., Brancolini, V., Tsou, H. C., Ul Haque, S., Lam, H., … Christiano, A. M. (1998). Alopecia universalis associated with a mutation in the human hairless gene. *Science*, *279*(5351), 720–724. https://doi.org/10.1126/science.279.5351.720

Aldinger, K. A., Lehmann, O. J., Hudgins, L., Chizhikov, V. V, Bassuk, A. G., Ades, L. C., … Millen, K. J. (2009). FOXC1 is required for normal cerebellar development and is a major contributor to chromosome 6p25.3 Dandy-Walker malformation. *Nature Genetics*, *41*(9), 1037–1042. https://doi.org/10.1038/ng.422

Allison, D. B., Gadbury, G. L., Heo, M., FernÃ andez, J. R., Lee, C.-K., Prolla, T. A., & Weindruch, R. (2002). A mixture model approach for the analysis of microarray gene expression data. In *Computational Statistics & Data Analysis* (Vol. 39). Retrieved from www.elsevier.com/locate/csda

Alonso, L., & Fuchs, E. (2006). The hair cycle. *Journal of Cell Science*, *119*(3), 391–393. https://doi.org/10.1242/jcs.02793

Andl, T., & Botchkareva, N. V. (2015). MicroRNAs (miRNAs) in the control of HF development and cycling: The next frontiers in hair research. *Experimental Dermatology*, *24*(11), 821–826. https://doi.org/10.1111/exd.12785

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Sherlock, G. (2000, May). Gene ontology: Tool for the unification of biology. *Nature Genetics*, Vol. 25, pp. 25–29. https://doi.org/10.1038/75556

Aya, U. I., Shimokawa, I., & Doi, K. (2009). Time-course expression profiles of hair cycle-associated genes in male mini rats after depilation of telogen-phase hairs. *International Journal of Molecular Sciences*, *10*(5), 1967–1977. https://doi.org/10.3390/ijms10051967

Bailey, M. H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., … Ding, L. (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell*, *173*(2), 371-385.e18. https://doi.org/10.1016/J.CELL.2018.02.060

Benavides, F., Oberyszyn, T. M., VanBuskirk, A. M., Reeve, V. E., & Kusewitt, D. F. (2009, January 1). The hairless mouse in skin research. *Journal of Dermatological Science*, Vol. 53, pp. 10–18. https://doi.org/10.1016/j.jdermsci.2008.08.012

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple. In *Source: Journal of the Royal Statistical Society. Series B (Methodological)* (Vol. 57). Retrieved from https://www-jstor-org.pitt.idm.oclc.org/stable/pdf/2346101.pdf?refreqid=excelsior%3A56a1f0a45f63fb9afb02

67135a742b13

Bielawski, J. P., & Yang, Z. (2004). A Maximum Likelihood Method for Detecting Functional Divergence at Individual Codon Sites, with Application to Gene Family Evolution. *Journal of Molecular Evolution*, *59*, 121–132.

Bininda-Emonds, O. R. P., Cardillo, M., Jones, K. E., MacPhee, R. D. E., Beck, R. M. D., Grenyer, R., … Purvis, A. (2007). The delayed rise of present-day mammals. *Nature*, *446*(7135), 507–512. https://doi.org/10.1038/nature05634

Blake, J. A., Richardson, J. E., Bult, C. J., Kadin, J. A., Eppig, J. T., & Mouse Genome Database Group. (2003). MGD: the Mouse Genome Database. *Nucleic Acids Research*, *31*(1), 193–195. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12519980

Blanchette, M., Kent, W. J., Riemer, C., Elnitski, L., Smit, A. F. A., Roskin, K. M., … Miller, W. (2004). Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Research*, *14*(4), 708–715. https://doi.org/10.1101/gr.1933104

Blumenberg, M. (2013). *Transcriptional Regulation of Keratin Gene Expression*. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK6213/

Boyle, E. I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J. M., & Sherlock, G. (2004). GO::TermFinder - open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics*, *20*(18), 3710–3715.

Brembeck, F. H., Opitz, O. G., Libermann, T. A., & Rustgi, A. K. (2000). Dual function of the epithelial specific ets transcription factor, ELF3, in modulating differentiation. *Oncogene*, *19*(15), 1941–1949. https://doi.org/10.1038/sj.onc.1203441

Callister, R. J. (2010). Early history of glycine receptor biology in mammalian spinal cord circuits. *Frontiers in Molecular Neuroscience*, *3*, 13. https://doi.org/10.3389/fnmol.2010.00013

Carbon, S., Douglass, E., Dunn, N., Good, B., Harris, N. L., Lewis, S. E., … Westerfield, M. (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research*, *47*(D1), D330–D338. https://doi.org/10.1093/nar/gky1055

Carrasco, E., Calvo, M. I., Blázquez-Castro, A., Vecchio, D., Zamarrón, A., De Almeida, I. J. D., … Espada, J. (2015). Photoactivation of ROS Production in Situ Transiently Activates Cell Proliferation in Mouse Skin and in the Hair Follicle Stem Cell Niche Promoting Hair Growth and Wound Healing. *Journal of Investigative Dermatology*, *135*(11), 2611–2622. https://doi.org/10.1038/jid.2015.248

Chatsirisupachai, K., Palmer, D., Ferreira, S., & Magalhães, J. P. (2019). A human tissue-specific transcriptomic analysis reveals a complex relationship between aging, cancer, and cellular senescence. *Aging Cell*, *18*(6). https://doi.org/10.1111/acel.13041

Chen, I. H., Kiang, J. H., Correa, V., Lopez, M. I., Chen, P. Y., McKittrick, J., & Meyers, M. A. (2011). Armadillo armor: Mechanical testing and micro-structural evaluation. *Journal of the Mechanical Behavior of Biomedical Materials*, *4*(5), 713–722. https://doi.org/10.1016/j.jmbbm.2010.12.013

Chen, Z., Wang, Z., Xu, S., Zhou, K., & Yang, G. (2013). Characterization of hairless (Hr) and FGF5 genes provides insights into the molecular basis of hair loss in cetaceans. *BMC Evolutionary Biology*, *13*(1), 1–11. https://doi.org/10.1186/1471-2148-13-34

Cheverud, J. M., & Dow, M. M. (1985). An autocorrelation analysis of genetic variation due to lineal fission in social groups of rhesus macaques. *American Journal of Physical Anthropology*, *67*(2), 113–121. https://doi.org/10.1002/ajpa.1330670206

Chikina, M., Robinson, J. D., & Clark, N. L. (2016). Hundreds of Genes Experienced Convergent Shifts in Selective Pressure in Marine Mammals. *Molecular Biology and Evolution*, *33*(9), 2182–2192. https://doi.org/10.1093/molbev/msw112

Clark, N. L., Alani, E., & Aquadro, C. F. (2012). Evolutionary rate covariation reveals shared functionality and coexpression of genes. *Genome Research*, *22*(4), 714–720. https://doi.org/10.1101/gr.132647.111

Clark, N. L., Alani, E., & Aquadro, C. F. (2013). Evolutionary rate covariation in meiotic proteins results from fluctuating evolutionary pressure in yeasts and mammals. *Genetics*, *193*(2), 529–538. https://doi.org/10.1534/genetics.112.145979

Cui, Y., Song, Y., Geng, Q., Ding, Z., Qin, Y., Fan, R., … Geng, J. (2016). The expression of KRT2 and its effect on melanogenesis in alpaca skins. *Acta Histochemica*, *118*(5), 505–512. https://doi.org/10.1016/j.acthis.2016.05.004

Davies, K. T. J., Bennett, N. C., Tsagkogeorga, G., Rossiter, S. J., & Faulkes, C. G. (2015). Family Wide Molecular Adaptations to Underground Life in African Mole-Rats Revealed by Phylogenomic Analysis. *Molecular Biology and Evolution*, *32*(12), msv175. https://doi.org/10.1093/molbev/msv175

Debian, D. E., & François, R. (2011). *Rcpp: Seamless R and C++ Integration*. Retrieved from http://cran.r-project.org/package=Rcpp.

Donet, E., Bayo, P., Calvo, E., Labrie, F., & Pérez, P. (2008). Identification of novel glucocorticoid receptor-regulated genes involved in epidermal homeostasis and hair follicle differentiation. *Journal of Steroid Biochemistry and Molecular Biology*, *108*(1–2), 8–16. https://doi.org/10.1016/j.jsbmb.2007.05.033

Drögemüller, C., Karlsson, E. K., Hytönen, M. K., Perloski, M., Dolf, G., Sainio, K., … Leeb, T. (2008). A mutation in hairless dogs implicates FOXI3 in ectodermal development. *Science*, *321*(5895), 1462. https://doi.org/10.1126/science.1162525

Eckhart, L., Dalla Valle, L., Jaeger, K., Ballaun, C., Szabo, S., Nardi, A., … Tschachler, E. (2008).

Identification of reptilian genes encoding hair keratin-like proteins suggests a new scenario for the evolutionary origin of hair. Retrieved from www.pnas.orgcgidoi10.1073pnas.0805154105

Eden, E., Lipson, D., Yogev, S., & Yakhini, Z. (2007). Discovering Motifs in Ranked Lists of DNA Sequences. *PLoS Computational Biology*, *3*(3), e39. https://doi.org/10.1371/journal.pcbi.0030039

Eden, E., Navon, R., Steinfeld, I., Lipson, D., & Yakhini, Z. (2009). GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, *10*(1), 48. https://doi.org/10.1186/1471-2105-10-48

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Eppig, J. T., Blake, J. A., Bult, C. J., Kadin, J. A., Richardson, J. E., & Mouse Genome Database Group, T. M. G. D. (2015). The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. *Nucleic Acids Research*, *43*(Database issue), D726-36. https://doi.org/10.1093/nar/gku967

Eyre-Walker, A., & Keightley, P. D. (1999). High genomic deleterious mutation rates in hominids. *Nature*, *397*(6717), 344–347. https://doi.org/10.1038/16915

Eyre-Walker, A., Keightley, P. D., Smith, N. G. C., & Gaffney, D. (2002). Quantifying the slightly deleterious mutation model of molecular evolution. *Molecular Biology and Evolution*, *19*(12), 2142–2149. https://doi.org/10.1093/oxfordjournals.molbev.a004039

Eyre-Walker, A., Woolfit, M., & Phelps, T. (2006). The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics*, *173*(2), 891–900. https://doi.org/10.1534/genetics.106.057570

Fang, X., Nevo, E., Han, L., Levanon, E. Y., Zhao, J., Avivi, A., … Wang, J. (2014). Genome-wide adaptive complexes to underground stresses in blind mole rats Spalax. *Nature Communications*, *5*(1), 1–11. https://doi.org/10.1038/ncomms4966

Feldman, C. R., Brodie, E. D., Brodie, E. D., & Pfrender, M. E. (2012). Constraint shapes convergence in tetrodotoxinresistant sodium channels of snakes. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(12), 4556–4561. https://doi.org/10.1073/pnas.1113468109

Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American Naturalist*, *125*(1), 1–15. https://doi.org/10.1086/284325

Foley, N. M., Hughes, G. M., Huang, Z., Clarke, M., Jebb, D., Whelan, C. V., … Teeling, E. C. (2018). Growing old, yet staying young: The role of telomeres in bats' exceptional longevity. *Science Advances*, *4*(2), eaao0926. https://doi.org/10.1126/sciadv.aao0926

Foote, A. D., Liu, Y., Thomas, G. W. C., Vinař, T., Alföldi, J., Deng, J., … Gibbs, R. A. (2015). Convergent evolution of the genomes of marine mammals. *Nature Genetics*, *47*(3), 272–275. https://doi.org/10.1038/ng.3198

Fu, S., Zhao, H., Zheng, Z., Li, J., & Zhang, W. (2014). Melatonin regulating the expression of miRNAs involved in hair follicle cycle of cashmere goats skin. *Yi Chuan = Hereditas / Zhongguo Yi Chuan Xue Hui Bian Ji*, *36*(12), 1235–1242. https://doi.org/10.3724/SP.J.1005.2014.1235

Fuller, A., Mitchell, D., Maloney, S. K., & Hetem, R. S. (2016). Towards a mechanistic understanding of the responses of large terrestrial mammals to heat and aridity associated with climate change. *Climate Change Responses*, *3*(1), 1–19. https://doi.org/10.1186/s40665-016-0024-1

Gharib, W. H., & Robinson-Rechavi, M. (2013). The branch-site test of positive selection is surprisingly robust but lacks power under synonymous substitution saturation and variation in GC. *Molecular Biology and Evolution*, *30*(7), 1675–1686. https://doi.org/10.1093/molbev/mst062

Gittleman, J. L., & Kot, M. (1990). Adaptation: Statistics and a Null Model for Estimating Phylogenetic Effects. *Systematic Zoology*, *39*(3), 227. https://doi.org/10.2307/2992183

Grafen, A. (1989). The phylogenetic regression. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *326*(1233), 119–157. https://doi.org/10.1098/rstb.1989.0106

Hadfield, J. D., & Nakagawa, S. (2010). General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of Evolutionary Biology*, *23*(3), 494–508. https://doi.org/10.1111/j.1420-9101.2009.01915.x

Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., & Challenger, W. (2008). GEIGER: investigating evolutionary radiations. *Bioinformatics*, *24*(1), 129–131. https://doi.org/10.1093/bioinformatics/btm538

Harris, R. S. (2007). *Improved Pairwise Alignment of Genomic DNA*. Retrieved from https://etda.libraries.psu.edu/catalog/7971

Heitman, N., Sennett, R., Mok, K. W., Saxena, N., Srivastava, D., Martino, P., … Rendl, M. (2020). Dermal sheath contraction powers stem cell niche relocation during hair cycle regression. *Science*, *367*(6474), 161–166. https://doi.org/10.1126/science.aax9131

Hiller, M., Schaar, B. T., Indjeian, V. B., Kingsley, D. M., Hagey, L. R., & Bejerano, G. (2012a). A "forward genomics" approach links genotype to phenotype using independent phenotypic losses among related species. *Cell Reports*, *2*(4), 817–823. https://doi.org/10.1016/j.celrep.2012.08.032

Hiller, M., Schaar, B. T., Indjeian, V. B., Kingsley, D. M., Hagey, L. R., & Bejerano, G. (2012b). A "Forward Genomics" Approach Links Genotype to Phenotype using Independent Phenotypic Losses among Related Species. *Cell Reports*, *2*(4), 817–823. https://doi.org/10.1016/J.CELREP.2012.08.032

Hofmann, S., Franke, A., Fischer, A., Jacobs, G., Nothnagel, M., Gaede, K. I., … Schreiber, S. (2008). *Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis*. https://doi.org/10.1038/ng.198

Holzenberger, M., Dupont, J., Ducos, B., Leneuve, P., Géloën, A., Even, P. C., … Le Bouc, Y. (2003). IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. *Nature*, *421*(6919), 182–187. https://doi.org/10.1038/nature01298

Hosner, P. A., Faircloth, B. C., Glenn, T. C., Braun, E. L., & Kimball, R. T. (2016). Avoiding Missing Data Biases in Phylogenomic Inference: An Empirical Study in the Landfowl (Aves:Galliformes). *Molecular Biology and Evolution*, *33*(4), 1110–1125.

Housworth, E. A., Martins, E. P., & Lynch, M. (2004). The Phylogenetic Mixed Model. *The American Naturalist*, *163*(1), 84–96. https://doi.org/10.1086/380570

Howitz, K. T., Bitterman, K. J., Cohen, H. Y., Lamming, D. W., Lavu, S., Wood, J. G., … Sinclair, D. A. (2003). Small molecule activators of sirtuins extend Saccharomyces cerevisiae lifespan. *Nature*, *425*(6954), 191–196. https://doi.org/10.1038/nature01960

Hu, Z., Sackton, T. B., Edwards, S. V., & Liu, J. S. (2019). Bayesian Detection of Convergent Rate Changes of Conserved Noncoding Elements on Phylogenetic Trees. *Molecular Biology and Evolution*, *36*(5), 1086–1100. https://doi.org/10.1093/molbev/msz049

Hughes, A. L., & Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*, *335*(6186), 167–170. https://doi.org/10.1038/335167a0

Jahoda, C. A.B., Horne, K. A., & Oliver, R. F. (1984). Induction of hair growth by implantation of cultured dermal papilla cells. *Nature*, *311*(5986), 560–562. https://doi.org/10.1038/311560a0

Jahoda, Colin A.B., Reynolds, A. J., & Oliver, R. F. (1993). Induction of hair growth in ear wounds by cultured dermal papilla cells. *Journal of Investigative Dermatology*, *101*(4), 584–590. https://doi.org/10.1111/1523-1747.ep12366039

Jaks, V., Barker, N., Kasper, M., Van Es, J. H., Snippert, H. J., Clevers, H., & Toftgård, R. (2008). Lgr5 marks cycling, yet long-lived, hair follicle stem cells. *Nature Genetics*, *40*(11), 1291–1299. https://doi.org/10.1038/ng.239

Janiak, M. C., Chaney, M. E., & Tosi, A. J. (2018). Evolution of Acidic Mammalian Chitinase Genes (CHIA) Is Related to Body Mass and Insectivory in Primates. *Molecular Biology and Evolution*, *35*(3), 607–622. https://doi.org/10.1093/molbev/msx312

Johnson, E. (1981). Environmental Influences on the Hair Follicle. In *Hair Research* (pp. 183–194). https://doi.org/10.1007/978-3-642-81650-5_27

Johnson, S., Rabinovitch, P., & Kaeberlein, M. (2013). mTOR is a key modulator of aging and age-related disease. *Nature*, *493*(7432), 338–345. https://doi.org/10.1038/nature11861

Jones, O. R., Scheuerlein, A., Salguero-Gómez, R., Camarda, C. G., Schaible, R., Casper, B. B., … Vaupel, J. W. (2014). Diversity of ageing across the tree of life. *Nature*, *505*(7482), 169–173. https://doi.org/10.1038/nature12789

Juan, D., Pazos, F., & Valencia, A. (2008). High-confidence prediction of global interactomes based on genome-wide coevolutionary networks. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(3), 934–939. https://doi.org/10.1073/pnas.0709671105

Kapheim, K. M., Pan, H., Li, C., Salzberg, S. L., Puiu, D., Magoc, T., … Zhang, G. (2015). Genomic signatures of evolutionary transitions from solitary to group living. *Science*, *348*(6239), 25. https://doi.org/10.1126/science.aaa4788

Kawano, M., Komi-Kuramochi, A., Asada, M., Suzuki, M., Oki, J., Jiang, J., & Imamura, T. (2005). Comprehensive analysis of FGF and FGFR expression in skin: FGF18 is highly expressed in hair follicles and capable of inducing anagen from telogen stage hair follicles. *Journal of Investigative Dermatology*, *124*(5), 877–885. https://doi.org/10.1111/j.0022-202X.2005.23693.x

Keane, M., Semeiks, J., Webb, A. E., Li, Y. I., Quesada, V., Craig, T., … de Magalhães, J. P. (2015). Insights into the Evolution of Longevity from the Bowhead Whale Genome. *Cell Reports*, *10*(1), 112–122. https://doi.org/10.1016/J.CELREP.2014.12.008

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The human genome browser at UCSC. *Genome Research*, *12*(6), 996–1006. https://doi.org/10.1101/gr.229102

Kettunen, P., Furmanek, T., Chaulagain, R., Hals Kvinnsland, I., & Luukko, K. (2011). Developmentally regulated expression of intracellular Fgf11-13, hormone-like Fgf15 and canonical Fgf16, -17 and -20 mRNAs in the developing mouse molar tooth. *Acta Odontologica Scandinavica*, *69*(6), 360–366. https://doi.org/10.3109/00016357.2011.568968

Kim, E. B., Fang, X., Fushan, A. A., Huang, Z., Lobanov, A. V., Han, L., … Gladyshev, V. N. (2011). Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature*, *479*(7372), 223–227. https://doi.org/10.1038/nature10533

Kimura, K. D., Tissenbaum, H. A., Liu, Y., Ruvkun, G., Yin, C.-M., & Garofalo, R. S. (1997). daf-2, an insulin receptor-like gene that regulates longevity and diapause in Caenorhabditis elegans. *Science (New York, N.Y.)*, *277*(5328), 942–946. https://doi.org/10.1126/science.277.5328.942

Kosiol, C., Vinař, T., da Fonseca, R. R., Hubisz, M. J., Bustamante, C. D., Nielsen, R., & Siepel, A. (2008). Patterns of Positive Selection in Six Mammalian Genomes. *PLoS Genetics*, *4*(8), e1000144. https://doi.org/10.1371/journal.pgen.1000144

Kowalczyk, A., Meyer, W. K., Partha, R., Mao, W., Clark, N. L., & Chikina, M. (2019). RERconverge: An R package for associating evolutionary rates with convergent traits. *Bioinformatics*, *35*(22). https://doi.org/10.1093/bioinformatics/btz468

Kowalczyk, A., Partha, R., Clark, N. L., & Chikina, M. (2020). Pan-mammalian analysis of molecular constraints underlying extended lifespan. *ELife*, *9*. https://doi.org/10.7554/eLife.51089

Kowalczyk, Amanda, Clark, N., & Chikina, M. (2021). Complementary evolution of coding and noncoding sequence underlies mammalian hairlessness. *BioRxiv*.

Kowalczyk, Amanda, Meyer, W., Partha, R., Mao, W., Clark, N., & Chikina, M. (2019). RERconverge: an R package for associating evolutionary rates with convergent traits. *Bioinformatics*, *35*(22), 4815–4817. https://doi.org/10.1101/451138

Kowalczyk, Amanda, Partha, R., Clark, N. L., & Chikina, M. (2020). Pan-mammalian analysis of molecular constraints underlying extended lifespan. *ELife*, *9*. https://doi.org/10.7554/eLife.51089

Kronforst, M. R., Barsh, G. S., Kopp, A., Mallet, J., Monteiro, A., Mullen, S. P., … Hoekstra, H. E. (2012). Unraveling the thread of nature's tapestry: the genetics of diversity and convergence in animal pigmentation. *Pigment Cell & Melanoma Research*, *25*(4), 411–433. https://doi.org/10.1111/j.1755-148X.2012.01014.x

Kryukov, G. V., Pennacchio, L. A., & Sunyaev, S. R. (2007). Most rare missense alleles are deleterious in humans: Implications for complex disease and association studies. *American Journal of Human Genetics*, *80*(4), 727–739. https://doi.org/10.1086/513473

Kushlan, J. A. (1980). The Evolution of Hairlessness in Man. *The American Naturalist*, *116*(5), 727–729. https://doi.org/10.1086/283663

Larsson, O., Girnita, A., & Girnita, L. (2005). Role of insulin-like growth factor 1 receptor signalling in cancer. *British Journal of Cancer*, *92*(12), 2097–2101. https://doi.org/10.1038/sj.bjc.6602627

Lartillot, N., & Poujol, R. (2011). A Phylogenetic Model for Investigating Correlated Evolution of Substitution Rates and Continuous Phenotypic Characters. *Molecular Biology and Evolution*, *28*(1), 729–744. https://doi.org/10.1093/molbev/msq244

Lay, K., Kume, T., & Fuchs, E. (2016). FOXC1 maintains the hair follicle stem cell niche and governs stem cell quiescence to preserve long-term tissue-regenerating potential. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(11), E1506–E1515. https://doi.org/10.1073/pnas.1601569113

Lee, K. H., Choi, D., Jeong, S. Il, Kim, S. J., Lee, C. H., Seo, H. S., & Jeong, H. S. (2019). Eclipta prostrata promotes the induction of anagen, sustains the anagen phase through regulation of FGF-7 and FGF-5. *Pharmaceutical Biology*, *57*(1), 105–111. https://doi.org/10.1080/13880209.2018.1561729

Leon Augustus, H. (1920). Structural Characteristics of the Hair of Mammals. *The American Naturalist*, *54*(635), 496–523. https://doi.org/10.2307/2456345

Li, G., Zhou, S., Li, C., Cai, B., Yu, H., Ma, B., … Wang, X. (2019). Base pair editing in goat: nonsense codon introgression into FGF5 results in longer hair. *FEBS Journal*, *286*(23), 4675–4692. https://doi.org/10.1111/febs.14983

Li, Y., Liu, Z., Shi, P., & Zhang, J. (2010, January 26). The hearing gene Prestin unites echolocating bats and whales. *Current Biology*, Vol. 20, pp. R55–R56. https://doi.org/10.1016/j.cub.2009.11.042

Liang, P., Lv, C., Jiang, B., Long, X., Zhang, P., Zhang, M., … Huang, X. (2012). MicroRNA profiling in denatured dermis of deep burn patients. *Burns*, *38*(4), 534–540. https://doi.org/10.1016/j.burns.2011.10.014

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., & Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics (Oxford, England)*, *27*(12), 1739–1740. https://doi.org/10.1093/bioinformatics/btr260

Lin, Z., Chen, Q., Shi, L., Lee, M., Giehl, K. A., Tang, Z., … Yang, Y. (2012). Loss-of-function mutations in HOXC13 cause pure hair and nail ectodermal dysplasia. *American Journal of Human Genetics*, *91*(5), 906–911. https://doi.org/10.1016/j.ajhg.2012.08.029

Lodato, M. A., Rodin, R. E., Bohrson, C. L., Coulter, M. E., Barton, A. R., Kwon, M., … Walsh, C. A. (2018). Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science*, *359*(6375), 555–559. https://doi.org/10.1126/science.aao4426

Lynch, M. (1991). Methods for the Analysis of Comparative Data in Evolutionary Biology. *Evolution*, *45*(5), 1065. https://doi.org/10.2307/2409716

Majewski, I. J., Ritchie, M. E., Phipson, B., Corbin, J., Pakusch, M., Ebert, A., … Blewitt, M. E. (2010). Opposing roles of polycomb repressive complexes in hematopoietic stem and progenitor cells. *Blood*, *116*(5), 731–739. https://doi.org/10.1182/blood-2009-12-260760

Mao, Z., Tian, X., Van Meter, M., Ke, Z., Gorbunova, V., & Seluanov, A. (2012). Sirtuin 6 (SIRT6) rescues the decline of homologous recombination repair during replicative senescence. *Proceedings of the National Academy of Sciences*, *109*(29), 11800–11805. https://doi.org/10.1073/pnas.1200583109

Martins, E. P. (2000). Adaptation and the comparative method. *Trends in Ecology & Evolution*, *15*(7), 296–299. https://doi.org/10.1016/S0169-5347(00)01880-2

McGowen, M. R., Tsagkogeorga, G., Williamson, J., Morin, P. A., & Rossiter, and S. J. (2020). Positive Selection and Inactivation in the Vision and Hearing Genes of Cetaceans. *Molecular Biology and Evolution*, *37*(7), 2069–2083. https://doi.org/10.1093/molbev/msaa070

Medawar, P. B. (1952). An unsolved problem in biology. In *H.K.* Lewis, London.

Meredith, R. W., Janecka, J. E., Gatesy, J., Ryder, O. A., Fisher, C. A., Teeling, E. C., … Murphy, W. J. (2011). Impacts of the Cretaceous Terrestrial Revolution and KPg Extinction on Mammal Diversification. *Science*, *334*(6055), 521–524. https://doi.org/10.1126/science.1211028

Mesler, A. L., Veniaminova, N. A., Lull, M. V., & Wong, S. Y. (2017). Hair Follicle Terminal Differentiation Is Orchestrated by Distinct Early and Late Matrix Progenitors. *Cell Reports*, *19*(4), 809–821. https://doi.org/10.1016/j.celrep.2017.03.077

Meyer, W.K., Jamison, J., Richter, R., Woods, S. E., Partha, R., Kowalczyk, A., … Clark, N. L. (2018). Ancient convergent losses of Paraoxonase 1 yield potential risks for modern marine mammals. *Science*, *361*(6402). https://doi.org/10.1126/science.aap7714

Meyer, Wynn K, Jamison, J., Richter, R., Woods, S. E., Partha, R., Kowalczyk, A., … Clark, N. L. (2018). Ancient convergent losses of Paraoxonase 1 yield potential risks for modern marine mammals. *Science (New York, N.Y.)*, *361*(6402), 591–594. https://doi.org/10.1126/science.aap7714

Michel, L., Reygagne, P., Benech, P., Jean-Louis, F., Scalvino, S., Ly Ka So, S., … Hocquaux, M. (2017). Study of gene expression alteration in male androgenetic alopecia: evidence of predominant molecular signalling pathways. *British Journal of Dermatology*, *177*(5), 1322–1336. https://doi.org/10.1111/bjd.15577

Muñoz, J., Stange, D. E., Schepers, A. G., Van De Wetering, M., Koo, B. K., Itzkovitz, S., … Clevers, H. (2012). The Lgr5 intestinal stem cell signature: Robust expression of proposed quiescent ′ +4′ cell markers. *EMBO Journal*, *31*(14), 3079–3091. https://doi.org/10.1038/emboj.2012.166

Muntané, G., Farré, X., Rodríguez, J. A., Pegueroles, C., Hughes, D. A., de Magalhães, J. P., … Navarro, A. (2018). Biological Processes Modulating Longevity across Primates: A Phylogenetic Genome-Phenome Analysis. *Molecular Biology and Evolution*, *35*(8), 1990–2004. https://doi.org/10.1093/molbev/msy105

Nagao, K., Kobayashi, T., Moro, K., Ohyama, M., Adachi, T., Kitashima, D. Y., … Amagai, M. (2012). Stress-induced production of chemokines by hair follicles regulates the trafficking of dendritic cells in skin. *Nature Immunology*, *13*(8), 744–752. https://doi.org/10.1038/ni.2353

Nakatake, Y., Hoshikawa, M., Asaki, T., Kassai, Y., & Itoh, N. (2001). Identification of a novel fibroblast growth factor, FGF-22, preferentially expressed in the inner root sheath of the hair follicle. *Biochimica et Biophysica Acta - Gene Structure and Expression*, *1517*(3), 460–463.

https://doi.org/10.1016/S0167-4781(00)00302-X

Nam, K., Lee, K. W., Chung, O., Yim, H. S., Cha, S. S., Lee, S. W., … Jeong, J. Y. (2017). Analysis of the FGF gene family provides insights into aquatic adaptation in cetaceans. *Scientific Reports*, *7*(1), 1–13. https://doi.org/10.1038/srep40233

Nery, M. F., Arroyo, J. I., & Opazo, J. C. (2014). Increased rate of hair keratin gene loss in the cetacean lineage. *BMC Genomics*, *15*(1), 1–9. https://doi.org/10.1186/1471-2164-15-869

Nowak, R. M. (1999). *Walker's Mammals of the World* (6th ed.). Baltimore: The Johns Hopkins University Press.

Oberdoerffer, P., Michan, S., McVay, M., Mostoslavsky, R., Vann, J., Park, S.-K., … Sinclair, D. A. (2008). SIRT1 Redistribution on Chromatin Promotes Genomic Stability but Alters Gene Expression during Aging. *Cell*, *135*(5), 907–918. https://doi.org/10.1016/J.CELL.2008.10.025

Ochoa, D., & Pazos, F. (2014). Practical aspects of protein co-evolution. *Frontiers in Cell and Developmental Biology*, *2*(APR), 14. https://doi.org/10.3389/fcell.2014.00014

Pagel, M., & Meade, A. (2006). Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *American Naturalist*, *167*(6), 808–825. https://doi.org/10.1086/503444

Palmer, S., Albergante, L., Blackburn, C. C., & Newman, T. J. (2018). Thymic involution and rising disease incidence with age. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(8), 1883–1888. https://doi.org/10.1073/pnas.1714478115

Papatheodorou, I., Fonseca, N. A., Keays, M., Tang, Y. A., Barrera, E., Bazant, W., … Petryszak, R. (2018). Expression Atlas: Gene and protein expression across multiple studies and organisms. *Nucleic Acids Research*, *46*(D1), D246–D251. https://doi.org/10.1093/nar/gkx1158

Parker, H. G., Harris, A., Dreger, D. L., Davis, B. W., & Ostrander, E. A. (2017). The bald and the beautiful: hairlessness in domestic dog breeds. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1713), 20150488. https://doi.org/10.1098/rstb.2015.0488

Partha, R., Chauhan, B. K., Ferreira, Z., Robinson, J. D., Lathrop, K., Nischal, K. K., … Clark, N. L. (2017a). Subterranean mammals show convergent regression in ocular genes and enhancers, along with adaptation to tunneling. *ELife*, *6*. https://doi.org/10.7554/eLife.25884

Partha, R., Chauhan, B. K., Ferreira, Z., Robinson, J. D., Lathrop, K., Nischal, K. K., … Clark, N. L. (2017b). Subterranean mammals show convergent regression in ocular genes and enhancers, along with adaptation to tunneling. *ELife*, *6*, e25884. https://doi.org/10.7554/eLife.25884

Partha, R., Kowalczyk, A., Clark, N. L., & Chikina, M. (2019). Robust method for detecting

convergent shifts in evolutionary rates. *Molecular Biology and Evolution*, *36*(8), 1817–1830. https://doi.org/10.1101/457309

Peng, H., Park, J. K., Katsnelson, J., Kaplan, N., Yang, W., Getsios, S., & Lavker, R. M. (2015). microRNA-103/107 Family Regulates Multiple Epithelial Stem Cell Characteristics. *STEM CELLS*, *33*(5), 1642–1656. https://doi.org/10.1002/STEM.1962@10.1002/(ISSN)1549-4918.THEBESTPAPERSFROMOUR2016YOUNGINVESTIGATORS

Peto, R. (2016). Epidemiology, multistage models, and short-term mutagenicity tests. *International Journal of Epidemiology*, *45*(3), 621–637. https://doi.org/10.1093/ije/dyv199

Plowman, J., Harland, D., & Deb-Choudhury, S. (2018). *The Hair Fibre: Proteins, Structure and Development* (I. Cohen, A. Lajtha, J. Lambris, R. Paoletti, & N. Rezaei, Eds.). Singapore.

Pough, F., Heiser, J., & McFarland, W. (1989). *Vertebrate Life*. New York: Macmillan Publishing Company.

Prudent, X., Parra, G., Schwede, P., Roscito, J. G., & Hiller, M. (2016). Controlling for Phylogenetic Relatedness and Evolutionary Rates Improves the Discovery of Associations Between Species' Phenotypic and Genomic Differences. *Molecular Biology and Evolution*, *33*(8), 2135–2150. https://doi.org/10.1093/molbev/msw098

Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Reynolds, A. J., & Jahoda, C. A. (1992). Cultured dermal papilla cells induce follicle formation and hair growth by transdifferentiation of an adult epidermis. *Development*, *115*(2).

Rezza, A., Wang, Z., Sennett, R., Qiao, W., Wang, D., Heitman, N., … Rendl, M. (2016). Signaling Networks among Stem Cell Precursors, Transit-Amplifying Progenitors, and their Niche in Developing Hair Follicles. *Cell Reports*, *14*(12), 3001–3018. https://doi.org/10.1016/j.celrep.2016.02.078

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47. https://doi.org/10.1093/nar/gkv007

Roca, A. L., Ishida, Y., Nikolaidis, N., Kolokotronis, S. O., Fratpietro, S., Stewardson, K., … Greenwood, A. D. (2009). Genetic variation at hair length candidate genes in elephants and the extinct woolly mammoth. *BMC Evolutionary Biology*, *9*(1), 232. https://doi.org/10.1186/1471-2148-9-232

Rodríguez, J. A., Marigorta, U. M., Hughes, D. A., Spataro, N., Bosch, E., & Navarro, A. (2017). Antagonistic pleiotropy and mutation accumulation influence human senescence and disease. *Nature Ecology and Evolution*, *1*(3). https://doi.org/10.1038/s41559-016-0055

Romiguier, J., & Roux, C. (2017). Analytical biases associated with GC-content in molecular evolution. *Frontiers in Genetics*, *8*(FEB), 16. https://doi.org/10.3389/fgene.2017.00016

Roscito, J. G., Sameith, K., Parra, G., Langer, B. E., Petzold, A., Moebius, C., … Hiller, M. (2018). Phenotype loss is associated with widespread divergence of the gene regulatory landscape in evolution. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-07122-z

Rosenquist, T. A., & Martin, G. R. (1996). Fibroblast growth factor signalling in the hair growth cycle: expression of the fibroblast growth factor receptor and ligand genes in the murine hair follicle. *Developmental Dynamics : An Official Publication of the American Association of Anatomists*, *205*(4), 379–386. https://doi.org/10.1002/(SICI)1097-0177(199604)205:4<379::AID-AJA2>3.0.CO;2-F

Sackton, T. B., Lazzaro, B. P., Schlenke, T. A., Evans, J. D., Hultmark, D., & Clark, A. G. (2007). Dynamic evolution of the innate immune system in Drosophila. *Nature Genetics*, *39*(12), 1461–1468. https://doi.org/10.1038/ng.2007.60

Sahm, A., Bens, M., Szafranski, K., Holtze, S., Groth, M., Görlach, M., … Platzer, M. (2018). Long-lived rodents reveal signatures of positive selection in genes associated with lifespan. *PLoS Genetics*, *14*(3), e1007272. https://doi.org/10.1371/journal.pgen.1007272

Sakamoto, M., & Venditti, C. (2018). Phylogenetic non-independence in rates of trait evolution. *Biology Letters*, *14*(10), 20180502. https://doi.org/10.1098/rsbl.2018.0502

Salminen, A., Huuskonen, J., Ojala, J., Kauppinen, A., Kaarniranta, K., & Suuronen, T. (2008). Activation of innate immunity system during aging: NF-kB signaling is the molecular culprit of inflamm-aging. *Ageing Research Reviews*, *7*(2), 83–105. https://doi.org/10.1016/J.ARR.2007.09.002

Santos, A., Tsafou, K., Stolte, C., Pletscher-Frankild, S., O'Donoghue, S. I., & Jensen, L. J. (2015). Comprehensive comparison of large-scale tissue expression datasets. *PeerJ*, *2015*(6). https://doi.org/10.7717/peerj.1054

Saputra, E., Kowalczyk, A., Cusick, L., Clark, N., & Chikina, M. (2020). Phylogenetic Permulations: a statistically rigorous approach to measure confidence in associations between phenotypes and genetic elements in a phylogenetic context. *BioRxiv*, 2020.10.14.338608. https://doi.org/10.1101/2020.10.14.338608

Saputra, E., Kowalczyk, A., Cusick, L., Clark, N., & Chikina, M. (2021). Phylogenetic Permulations: A Statistically Rigorous Approach to Measure Confidence in Associations in a Phylogenetic Context. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msab068

Sarasin, A. (2012). UVSSA and USP7: New players regulating transcription-coupled nucleotide excision repair in human cells. *Genome Medicine*, *4*(5), 44. https://doi.org/10.1186/gm343

Seim, I., Fang, X., Xiong, Z., Lobanov, A. V., Huang, Z., Ma, S., … Gladyshev, V. N. (2013).

Genome analysis reveals insights into physiology and longevity of the Brandt's bat Myotis brandtii. *Nature Communications*, *4*(1), 2212. https://doi.org/10.1038/ncomms3212

Sengez, B., Aygün, I., Shehwana, H., Toyran, N., Tercan Avci, S., Konu, O., … Alotaibi, H. (2019). The Transcription Factor Elf3 Is Essential for a Successful Mesenchymal to Epithelial Transition. *Cells*, *8*(8), 858. https://doi.org/10.3390/cells8080858

Seo, S., Fujita, H., Nakano, A., Kang, M., Duarte, A., & Kume, T. (2006). The forkhead transcription factors, Foxc1 and Foxc2, are required for arterial specification and lymphatic sprouting during vascular development. *Developmental Biology*, *294*(2), 458–470. https://doi.org/10.1016/j.ydbio.2006.03.035

Shi, L., & Oberdoerffer, P. (2012). Chromatin dynamics in DNA double-strand break repair. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1819*(7), 811–819. https://doi.org/10.1016/J.BBAGRM.2012.01.002

Shultz, A. J., & Sackton, T. B. (2019). Immune genes are hotspots of shared positive selection across birds and mammals. *ELife*, *8*. https://doi.org/10.7554/eLife.41815

Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., … Haussler, D. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, *15*(8), 1034–1050. https://doi.org/10.1101/gr.3715005

Sprecher, E., Molho-Pessach, V., Ingber, A., Sagi, E., Indelman, M., & Bergman, R. (2004). Homozygous splice site mutations in PKP1 result in loss of epidermal plakophilin 1 expression and underlie ectodermal dysplasia/skin fragility syndrome in two consanguineous families. *Journal of Investigative Dermatology*, *122*(3), 647–651. https://doi.org/10.1111/j.0022-202X.2004.22335.x

Stone, G. N., Nee, S., & Felsenstein, J. (2011). Controlling for non-independence in comparative analysis of patterns across populations within species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, Vol. 366, pp. 1410–1424. https://doi.org/10.1098/rstb.2010.0311

Storey, J. D., Bass, A. J., Dabney, A., & Robinson, D. (2020). *qvalue: Q-value estimation for false discovery rate control*. R package version 2.20.0.

Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, *100*(16), 9440–9445. https://doi.org/10.1073/pnas.1530509100

Storey, John D. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *Annals of Statistics*, *31*(6), 2013–2035. https://doi.org/10.1214/aos/1074290335

Stout, G. J., Stigter, E. C. A., Essers, P. B., Mulder, K. W., Kolkman, A., Snijders, D. S., … Brenkman, A. B. (2013). Insulin/IGF-1-mediated longevity is marked by reduced protein metabolism. *Molecular Systems Biology*, *9*(1), 679. https://doi.org/10.1038/msb.2013.35

Suárez-Fariñas, M., Ungar, B., Noda, S., Shroff, A., Mansouri, Y., Fuentes-Duculan, J., … Guttman-Yassky, E. (2015). Alopecia areata profiling shows TH1, TH2, and IL-23 cytokine activation without parallel TH17/TH22 skewing. *Journal of Allergy and Clinical Immunology*, *136*(5), 1277–1287. https://doi.org/10.1016/j.jaci.2015.06.032

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., … Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

Subramanian, Aravind, Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., … Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

Sulak, M., Fong, L., Mika, K., Chigurupati, S., Yon, L., Mongan, N. P., … Lynch, V. J. (2016). TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. *ELife*, *5*. https://doi.org/10.7554/eLife.11994

Sutter, N. B., Bustamante, C. D., Chase, K., Gray, M. M., Zhao, K., Zhu, L., … Ostrander, E. A. (2007). A single IGF1 allele is a major determinant of small size in dogs. *Science (New York, N.Y.)*, *316*(5821), 112–115. https://doi.org/10.1126/science.1137045

Suzuki, S., Ota, Y., Ozawa, K., & Imamura, T. (2000). Dual-mode regulation of hair growth cycle by two Fgf-5 gene products. *Journal of Investigative Dermatology*, *114*(3), 456–463. https://doi.org/10.1046/j.1523-1747.2000.00912.x

Tacutu, R., Thornton, D., Johnson, E., Budovsky, A., Barardo, D., Craig, T., … de Magalhães, J. P. (2018). Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Research*, *46*(D1), D1083–D1090. https://doi.org/10.1093/nar/gkx1042

Team, R. C. (2018). *R: A Language and Environment for Statistical Computing*. Retrieved from https://www.r-project.org/

Veraitch, O., Mabuchi, Y., Matsuzaki, Y., Sasaki, T., Okuno, H., Tsukashima, A., … Ohyama, M. (2017). Induction of hair follicle dermal papilla cell properties in human induced pluripotent stem cell-derived multipotent LNGFR(+)THY-1(+) mesenchymal cells. *Scientific Reports*, *7*(1), 1–13. https://doi.org/10.1038/srep42777

Wang, D., Zhang, Z., O'Loughlin, E., Wang, L., Fan, X., Lai, E. C., & Yi, R. (2013). MicroRNA-205 controls neonatal expansion of skin stem cells by modulating the PI(3)K pathway. *Nature Cell Biology*, *15*(10), 1153–1163. https://doi.org/10.1038/ncb2827

Wang, L., Siegenthaler, J. A., Dowell, R. D., & Yi, R. (2016). Stem cells: Foxc1 reinforces quiescence in self-renewing hair follicle stem cells. *Science*, *351*(6273), 613–617.

https://doi.org/10.1126/science.aad5440

Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky Pond, S. L., & Scheffler, K. (n.d.). *RELAX: Detecting Relaxed Selection in a Phylogenetic Framework*. https://doi.org/10.1093/molbev/msu400

Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky Pond, S. L., & Scheffler, K. (2015). RELAX: Detecting Relaxed Selection in a Phylogenetic Framework. *Molecular Biology and Evolution*, *32*(3), 820–832. https://doi.org/10.1093/molbev/msu400

White, R. R., Milholland, B., de Bruin, A., Curran, S., Laberge, R.-M., van Steeg, H., … Vijg, J. (2015). Controlled induction of DNA double-strand breaks in the mouse liver induces features of tissue ageing. *Nature Communications*, *6*(1), 6790. https://doi.org/10.1038/ncomms7790

Williams, G. C. (1957). Pleiotropy, Natural Selection, and the Evolution of Senescence. *Evolution*, *11*(4), 398. https://doi.org/10.2307/2406060

Wu, D., & Smyth, G. K. (2012). Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Research*, *40*(17), e133–e133. https://doi.org/10.1093/nar/gks461

Xu, H. L., Chen, P. P., Wang, L. fen, Xue, W., & Fu, T. L. (2018). Hair regenerative effect of silk fibroin hydrogel with incorporation of FGF-2-liposome and its potential mechanism in mice with testosterone-induced alopecia areata. *Journal of Drug Delivery Science and Technology*, *48*, 128–136. https://doi.org/10.1016/j.jddst.2018.09.006

Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution*, *24*(8), 1586–1591. https://doi.org/10.1093/molbev/msm088

Ye, S. B., Zhang, H., Cai, T. T., Liu, Y. N., Ni, J. J., He, J., … Li, J. (2016). Exosomal miR-24-3p impedes T-cell function by targeting FGF11 and serves as a potential prognostic biomarker for nasopharyngeal carcinoma. *Journal of Pathology*, *240*(3), 329–340. https://doi.org/10.1002/path.4781

Zerbino, D. R., Wilder, S. P., Johnson, N., Juettemann, T., & Flicek, P. R. (2015). The Ensembl Regulatory Build. *Genome Biology*, *16*(1). https://doi.org/10.1186/s13059-015-0621-5

Zhang, Jianzhi, & Yang, J. R. (2015, July 19). Determinants of the rate of protein sequence evolution. *Nature Reviews Genetics*, Vol. 16, pp. 409–420. https://doi.org/10.1038/nrg3950

Zhang, Jing, Wallace, S. J., Shiu, M. Y., Smith, I., Rhind, S. G., & Langlois, V. S. (2017). Human hair follicle transcriptome profiling: A minimally invasive tool to assess molecular adaptations upon low-volume, high-intensity interval training. *Physiological Reports*, *5*(23), e13534. https://doi.org/10.14814/phy2.13534

Zhou, L., Wang, H., Jing, J., Yu, L., Wu, X., & Lu, Z. (2018). Regulation of hair follicle development by exosomes derived from dermal papilla cells. *Biochemical and Biophysical Research Communications*, *500*(2), 325–332. https://doi.org/10.1016/j.bbrc.2018.04.067