

1 Refining Convergent Rate Analysis with Topology in Mammalian
2 Longevity and Marine Transitions

3 Stephen Treaster^{1,2*}, Jacob M. Daane^{1,2,3}, Matthew P. Harris^{1,2}

4

5 ¹ Department of Orthopaedic Research, Boston Children's Hospital, Boston, MA, 02124, USA

6 ² Department of Genetics, Harvard Medical School, Boston, MA, 02124, USA

7 ³ Department of Marine and Environmental Sciences, Northeastern University Marine Science
8 Center, Nahant, MA, 01908, USA

9

10 *communicating author, stephen_treaster@hms.harvard.edu

11

12 Key words: Mammalian longevity, marine transition, comparative genomics, convergence

13

14

1 **Abstract**

2 The quest to map the genetic foundations of phenotypes has been empowered by the modern
3 diversity, quality, and availability of genomic resources. Despite these expanding resources, the
4 abundance of variation within lineages makes the association of genetic change to specific
5 phenotypes improbable. Drawing such connections requires an *a priori* means of isolating the
6 associated changes from background genomic variation. Evolution may provide these means via
7 convergence; i.e., the shared variation that may result from replicate evolutionary experiments
8 across independent trait occurrences. To leverage these opportunities, we developed *TRACCER*:
9 *Topologically Ranked Analysis of Convergence via Comparative Evolutionary Rates*. As
10 compared to current methods, this software empowers rate convergence analysis by factoring
11 in topological relationships, because variation between phylogenetically proximate trait
12 changes is more likely to be facilitating the trait. Pairwise comparisons are performed not with
13 singular branches, but in reference to their most recent common ancestors. This ensures that
14 comparisons represent identical genetic contexts and timeframes while obviating the
15 problematic requirement of assigning ancestral states. We applied TRACCER to two case
16 studies: marine mammal transitions, an unambiguous trait which has independently evolved
17 three times, as well as the evolution of mammalian longevity, a less delineated trait but with
18 more instances to compare. TRACCER, by factoring in topology, identifies highly significant,
19 convergent genetic signals in these test cases, with important incongruities and statistical
20 resolution when compared to existing convergence approaches. These improvements in
21 sensitivity and specificity generate refined targets for downstream analysis of convergent
22 evolution and identification of genotype-phenotype relationships.

23 **Introduction**

24 When challenged with similar selective pressures, independent lineages may converge on
25 similar adaptations to those challenges (Losos, 2011). A myriad of complex mutations can be
26 sufficient to produce such adaptive traits, but the stochastic nature of evolution begets that
27 such adaptive traits will tend to manifest parsimoniously, particularly given the genetic
28 constraints and largely similar toolkits of life (Rosenblum et al., 2014). Given these
29 developmental and physiological constraints in organisms, the convergent evolution of traits

1 will tend towards similar molecular mechanisms, providing a valuable map for phenotype to
2 genotype inferences, recently coined “forward genomics” (Currie, 2013; Hiller et al., 2012).
3 However, while the parsimonious path to the same adaptive phenotype may tread upon the
4 same genetic elements or pathways, the specific molecular changes may evolve in unique ways.
5 Given a convergent trait, the molecular similarities will be shaped both by the genetic context
6 and the precision required; there are countless ways to disrupt a pathway or knockout a gene,
7 but there may only be a handful to specifically compromise a hydrophobic pocket, and only one
8 to create a disulfide bridge between two domains. A variety of changes could modulate cellular
9 activity in comparable ways and ultimately manifest as convergent phenotypes of interest.
10 Analytical tools with a more general measure are needed to capture these disparate genetic
11 signatures acting on the same system in order to make meaningful comparisons between
12 species.

13

14 Relative Evolutionary Rates and Convergence

15 The assessment of relative evolutionary rates (RERs) has arisen as a powerful and flexible way
16 to distill and compare the variable signatures acting on a system to identify convergence. RERs
17 describe the degree of change in a genomic region as compared to the background rate across
18 that genome. Not all parts of a genome evolve at the same pace, so these relative rates can
19 vary widely. While mutations are randomly distributed, the finely tuned core pathways will
20 undergo strong purifying selection, leaving them relatively unchanged, sometimes for hundreds
21 of millions of years (Katzman et al., 2007). These loci will reliably yield low RERs. Regions in the
22 wilderness of “nonfunctional junk” DNA can accumulate mutations freely, without selective
23 pressure, yielding high RERs. Duplicated and pseudogenized regions, also freed from restriction,
24 can evolve rapidly and neofunctionalize to facilitate adaptation or go to drift, and may also yield
25 high RERs. These accelerations and constraints of change on a locus can be powerfully
26 informative as to the selective pressures acting on it (Echave et al., 2016; Pollard et al., 2006).
27 Importantly, such rate analyses can scale in scope, from single positions, to genes, or even
28 discontinuous loci comprising entire pathways (Daane et al., 2020). This contrasts with analyses
29 that quantify a specific measure, like gene loss, expansion, or the detection of identical amino

1 acid substitutions. The latter, for instance, have proven informative in many systems, such as
2 echolocation (Parker et al., 2013), but their significance remains contentious (Marcovitz et al.,
3 2019; Zou and Zhang, 2015). Such specific measures alone fail to capture local, potentially
4 linked, changes that could influence the trait similarly. Single amino acid substitutions are too
5 narrow a measure to encompass the multiple dimensions molecular evolution can act upon.
6 RERs scale to the degree of specificity an experiment requires and can capture disparate yet
7 functionally comparable changes in a single measure.

8 RERs represent the substitution rate of a single element (exon, gene, CNE, etc.) relative
9 to the genome-wide average. These rates are generally represented with phylogenetic trees
10 and the relative rates can be derived by comparing individual gene trees to the species tree. For
11 example, elephant *PLCZ1* has a much shorter branch length when compared to the average
12 across its genome, resulting in a low RER, while the Elephant Shrew branch is longer than
13 expected and yields a high RER (**Figure 1A**). To identify convergence in rate, the RERs for a
14 specific element can be compared between species and intersected with trait occurrence.
15 Lineages with differing traits and differing RERs are evidence that the locus may be involved in
16 the evolution of that trait. With tens of thousands of loci and only a single species, a
17 distribution of acceleration and constraint would be expected by definition, centered on the
18 background substitution rate. By analyzing many comparisons across multiple instances of trait
19 change, consistently shared RERs for a specific element are unlikely to occur, enabling detection
20 of convergent signals from genomic noise. These probabilities and calculations are also
21 empowered by including the RER magnitude, as they are a continuous variable.

22 Until recently, tools for RER analysis were limited to individual elements or branches
23 and did not scale to the modern wealth of genomic information. The first broadscale “forward
24 genomics” approach (Hiller et al., 2012) focused solely on trait loss and was agnostic regarding
25 topology. RERConverge later efficiently analyzed RER convergence in acceleration and
26 constraint, implicating genes in marine mammal transitions (Chikina et al., 2016), mammalian
27 subterranean transitions (Partha et al., 2017), and mammalian longevity (Kowalczyk et al.,
28 2020) with fascinating insights into the genetic architecture that facilitates these dramatic
29 environmental and physiological shifts. Those studies assume each branch is an equivalent and

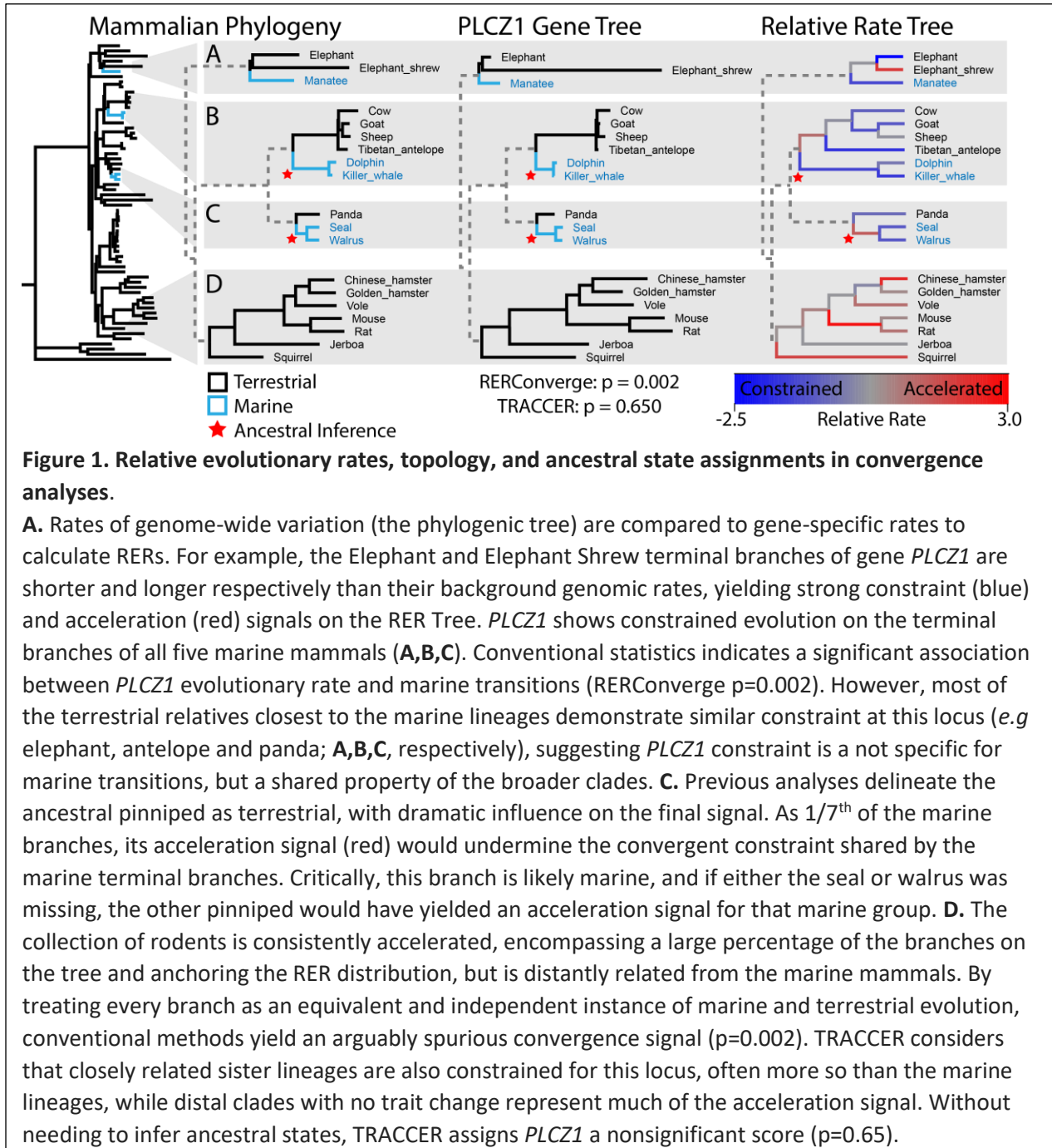
1 independent occurrence of evolution, and as such, process the signals with conventional
2 statistical tests that require independence, including the Pearson Correlation or Wilcoxon Rank-
3 Sum. However, these assumptions are incongruent with the conceptual basis of convergence
4 with consequences for both sensitivity and specificity.

5 The Value of Topology

6 Not all comparisons are equally informative, nor are they truly independent; topology and
7 genetic context must be factored in. Sister lineages with differing traits make for more valuable
8 and informative comparisons than distant ones; the changes facilitating the trait occurred
9 within a shorter-time frame and with a greater degree of shared genetic architecture. As
10 genomic variation must be facilitating the trait change, and there is less variation between
11 sister lineages, any differences revealed are thus more likely to be facilitating the trait
12 difference being investigated. The opposite is equally important; when closely related species
13 with differing traits share a genetic signature, it is strong evidence the signature is unrelated to
14 the trait. Regardless of the arguments dividing convergence and parallelism (Gould et al., 2002;
15 Pearce, 2012), neglecting topology is an oversight of many statistical approaches to associate
16 RERs with trait change. An example of these effects is highlighted in **Figure 1**. There,
17 RERConverge has assigned *PLCZ1* a significant score for convergent constraint in the evolution
18 of marine mammals (**Figure 1A,B,C**). However, the closest terrestrial relatives all share similar
19 levels of constraint, suggesting that constrained *PLCZ1* evolution is not specific for marine
20 transitions, but instead a general signature of the larger clades. In contrast, the rodents (**Figure**
21 **1D**) are enriched for accelerations and represent a substantial percentage of the branches on
22 the tree, yet are distantly related from any marine lineages. The available mammalian genomes
23 are oversampled for many clades that have little to do with marine transitions. The quantity of
24 branches in these clades can spuriously anchor conventional statistics. Despite these branches
25 representing variable genetic contexts, timeframes, and relationships to the trait being
26 analyzed, conventional statistics must treat them each *as an equivalent and independent*
27 *instance* of marine or terrestrial evolution. For instance, in the analysis of marine mammals, it is
28 problematic to include the ancestral branch leading to mice and rats -- ~100M years from the
29 closest marine mammal – as an equivalent comparison and weight as the panda terminal

1 branch, the immediate sister of pinnipeds. Individual branches on a tree are not equally
 2 meaningful; they are arbitrary fragments of evolution based on which lineages were sampled,
 3 each with their own context and varying degrees of overlap. Instead, the branches from which
 4 RERs are derived and compared should represent comparable evolutionary events.

5 To fully integrate the informative power present in the phylogeny into a convergence
 6 analysis requires delineation of which evolutionary events should be compared. Most



1 convergence analyses require flagging specific branches for analysis, with often precarious
2 assumptions of ancestral states. This is easier with well-delineated traits like marine mammals
3 that have a well-documented fossil record (Slater et al., 2012). However, misattribution of
4 these states can easily compromise an analysis. In Chikina et al 2016, the pinniped ancestral
5 branch was not assigned as marine, biasing the constraint signal for *PLCZ1* by leaving out an
6 accelerated branch (**Figure 1C**). If the genome for either the seal or walrus was unavailable, the
7 entire pinniped lineage would have been designated marine and calculated as accelerated.
8 Instead, because the branches are arbitrarily fragmented by sampling availability, both pinniped
9 terminal branches are generating a strong signal of constraint. Critically, these branches likely
10 do not represent a marine transition at all; it is likely the ancestor to the pinniped clade that
11 was undergoing the dramatic adaptation (Berta et al., 2018). By strictly defining the trait status
12 of each branch, there manifests a potential to misattribute when selection may be acting,
13 spuriously raising and lowering signals of selection. This is difficult enough with a clear trait like
14 marine transitions -- the pinniped phylogeny was contentious for some time, and many
15 originally argued it was diphyletic (Koretsky and Barnes, 2003) -- and would be nigh impossible
16 for traits like longevity or metabolism. This problematic requirement is common in convergence
17 analyses. To our knowledge, current rate convergence analyses either ignore topology, or
18 require inferring ancestral states, and/or do not scale to genome-wide data.

19 **Results**

20 **Topological Weighting and RER Statistics**

21 In light of these concerns, we developed a new analysis pipeline, *Topologically Ranked Analysis*
22 *of Convergence via Comparative Evolutionary Rates* -- *TRACCER*. We test if the analysis of
23 convergence is refined by making genomic comparisons between trait-bearing and non-trait-
24 bearing extant species in reference to their most recent common ancestor (MRCA). With this
25 approach, comparisons represent a single genetic context, diverging over the same time frame,
26 and are guaranteed to encompass a change in trait. As each value is now a comparison instead
27 of a singular branch, we can weight each comparison by the phyletic distance between them.
28 This approach allows four critical features: **I.** proximal comparisons can be given greater
29 influence, **II.** phylogenetically distal comparisons can be downplayed, **III.** it is no longer

1 necessary to infer ancestral states, and **IV.** trait distribution can be incorporated to undermine
2 sampling biases.

3 Once topology is factored in, however, we can no longer use conventional statistics to
4 determine significance as each comparison is no longer truly independent. As we are comparing
5 lineages in reference to their MRCA, many comparisons will share some ancestral branches,
6 causing varying degrees of dependence between comparisons. Instead, we use a permutation
7 strategy to repeatedly shuffle and sample branches of the element trees to simulate how likely
8 a pattern of relative rates is to occur on a given topology with the actual data. The
9 combinatorial complexity of exome-sized datasets on a broadly sampled phylogeny is sufficient
10 to sample from directly. Because RER calculations require a fixed topology, we can shuffle
11 branches between trees to yield a simulated instance of evolution using real examples of how
12 those speciation events occurred at a molecular level. Millions of these phylogenetically and
13 molecularly grounded permutations are analyzed alongside the experimental scores to
14 determine an assumption free background score distribution. The experimental results are
15 compared to this distribution to determine their significance. This contrasts with the variety of
16 evolutionary models one could draw from, which while undeniably useful, their predictions
17 often fail to describe real data (Naser-Khdour et al., 2019). Thus, TRACCER obviates the need for
18 assumptive formulae by drawing directly from the wealth of information inherent to modern
19 comparative genomic datasets.

20 TRACCER diverges in a few key details of RER calculations from previous analyses with
21 important ramifications. We use a log-transformed fold change from baseline, as opposed to a
22 residual approach used by RERConverge. Residuals fail to inversely allocate weight to variation
23 in longer branches; two SNPs in ten million years should not hold the same influence as two
24 SNPs in one million years, yet these would calculate the same residual value. The fold change
25 approach inherently corrects for branch length and embraces the “rate” concept of relative
26 rates. TRACCER also uses a median distance across all gene trees, excluding zeroes, to derive
27 the relative rates. The median calculation is robust to acceleration outliers, but would be
28 skewed by the number of zero-length branches if they were not excluded. Evolutionary models
29 and tree calculations often struggle with these regions that lack informative substitutions. Due

1 to their short-sequence and/or extreme selective pressure they are gutted to zero-length
2 branches, which does not accurately represent the forces at play genome-wide. These branches
3 should still be analyzed as undergoing constraint, and can still be compared within the same
4 tree, as each branch would be suffering from a consistent bias. The details of these calculations
5 are elaborated in the methods.

6 These analytical concepts: comparing RERs in reference to MRCAs, weighting
7 comparisons by evolutionary distance, deriving RERs as log-fold changes from zero-corrected-
8 medians, and significance as compared to permutations of the phylogenetic data, empower
9 relative rate analyses to better reveal convergent signatures. TRACCER compares relative
10 evolutionary rates in a pairwise approach that obviates the need for ancestral state
11 assumptions and allows for topologically aware comparisons; it prioritizes sister groups with
12 high comparative power and carefully weighs outgroups to have appropriate influence. RERs
13 are derived as the fold change from the corrected-median branch length, accurately describing
14 rates without branch length bias or assumptive models. Critically, TRACCER defaults include a
15 ranking transformation applied to signal magnitudes, preventing outlier lineages from
16 swamping potential signals, without leaving them ignored. The recurrent concept of
17 convergence is ensured while increasing sensitivity to subtle but meaningful changes over
18 short-time frames. Finally, by scaling each comparison by how often those branches have been
19 used, TRACCER dilutes the influence of over sampled clades and their shared ancestral
20 branches, while allowing their unique terminal branches the full influence. To benchmark the
21 impact of these strategies, we have directly compared TRACCER to a similar analysis,
22 RERConverge, which has recently demonstrated great successes analyzing both marine
23 mammal transitions and mammalian longevity.

24 **Test Case 1: Marine Mammal Convergence**

25 We obtained amino-acid level gene alignments for 62 mammals from the UCSC 100-Way
26 Vertebrate Multiz dataset (Rosenbloom et al., 2015). We chose this set to match previous
27 analyses with RERConverge (Chikina et al., 2016) and allow direct comparison of statistical
28 resolution. The species tree provided by UCSC was trimmed to maintain just the mammals and
29 used to fix the topology of 37,272 protein coding trees generated with PAML using the same

1 settings as described in RERConverge. The five marine mammals, including the cetaceans
2 (dolphin, killer whale), pinnipeds (seal, walrus) and sirenians (manatee), represent three
3 independent cases of adaptation to an aquatic environment (**Figure 1 Phylogeny**). When using
4 RERConverge, which requires ancestral branches to be classed as well, we additionally flagged
5 the cetacean ancestor as marine, but not the pinniped ancestor, again to match published
6 methodology and results. As a control group, we also copied the published RERConverge
7 selections to allow for direct comparison: aardvark, alpaca, camel, microbat, and myotis bat
8 have roughly similar topology as the marine mammals and provide a non-convergent control
9 group to assess background signal levels.

10 TRACCER has an integrated diagnostic mode to assess if a particular tree topology has
11 sufficient power to reveal convergence beyond background noise and provide an estimate of
12 how robust of a signal is required to reach that significance. It is important to note that the
13 maximum score for convergence with TRACCER is not derived from merely all trait-bearing
14 species bearing the same extremes of constraint or acceleration; their terrestrial sister lineages
15 must occupy the opposite end of the distribution for that gene. Put another way, sister lineages
16 sharing the same relative rate polarity should compromise the convergent signal, as the signal
17 is no longer specific for the trait in that genetic context. To profile these dependencies, we
18 simulated 10,000,000 gene trees with Brownian motion (Revell et al., 2008) and measured the
19 significance of their convergent signals against a backdrop of another 10,000,000 random trees
20 (**Figure 2A, 0**). This revealed that p-values less than 1/20,000 – roughly the size of a vertebrate
21 exome dataset – were possible, indicating the topology represented by the three marine
22 transitions yields a sufficiently complex score-space to parse significant hits from background
23 noise of a genomic dataset.

24 To roughly gauge the shape or robustness of the convergent pattern required to
25 produce these significant hits, the diagnostic allocates extreme rate accelerations to random
26 terminal branches on top of the Brownian backdrop. We see that with all five marine terminal
27 branches set to maximum acceleration, the vast majority of trees yield a significant score, but
28 not all of them. If the immediate sister lineages were also highly accelerated, they reduce the
29 score below the threshold (**Figure 2A, 5**). We outlined this effect by again purposefully

1 allocating extreme accelerations, this time to four marine mammals along with a number of
2 their immediate sisters (**Figure 2B**). While setting extreme accelerations to four marine terminal
3 branches can drive significant scores, even a single sister lineage sharing that extreme
4 acceleration will entirely prevent it. If the three closest terrestrial lineages share the same
5 extreme signal as four marine mammals, while all other branches are randomly distributed, it
6 will reduce the median p-value by an order of magnitude. These simulations are simplified by
7 distributing the acceleration signals only to terminal branches and randomizing all others but
8 they provide a valuable overview of the topological weight influence.

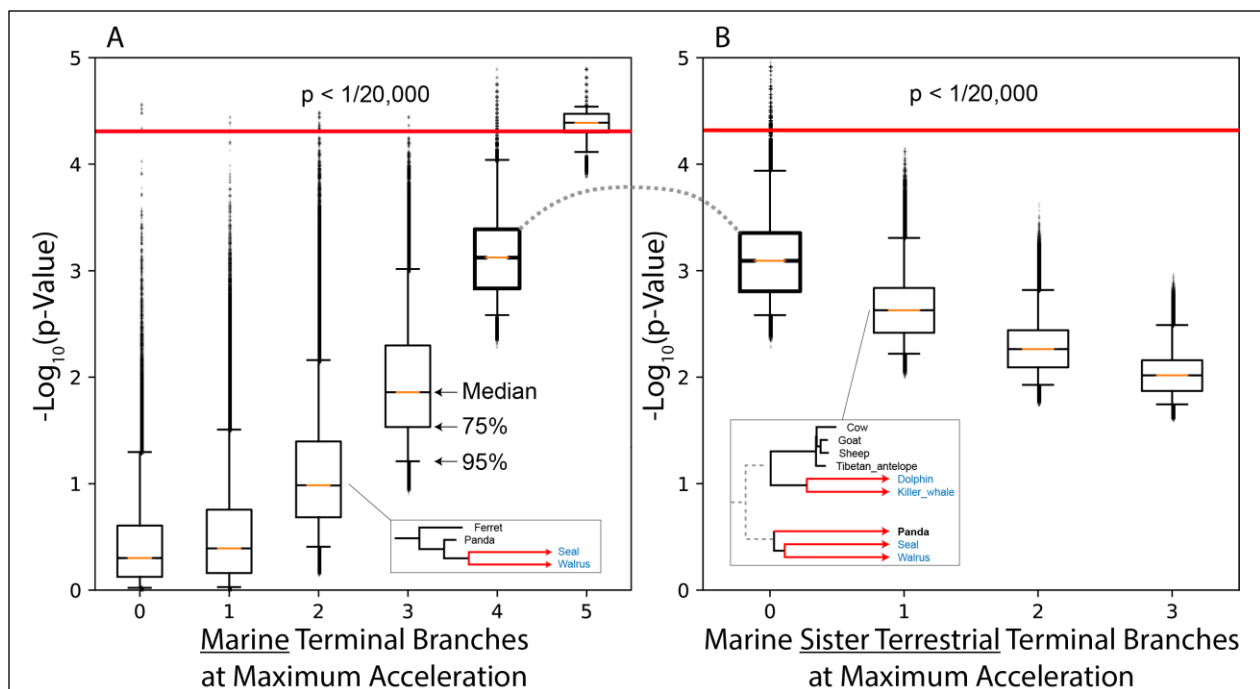
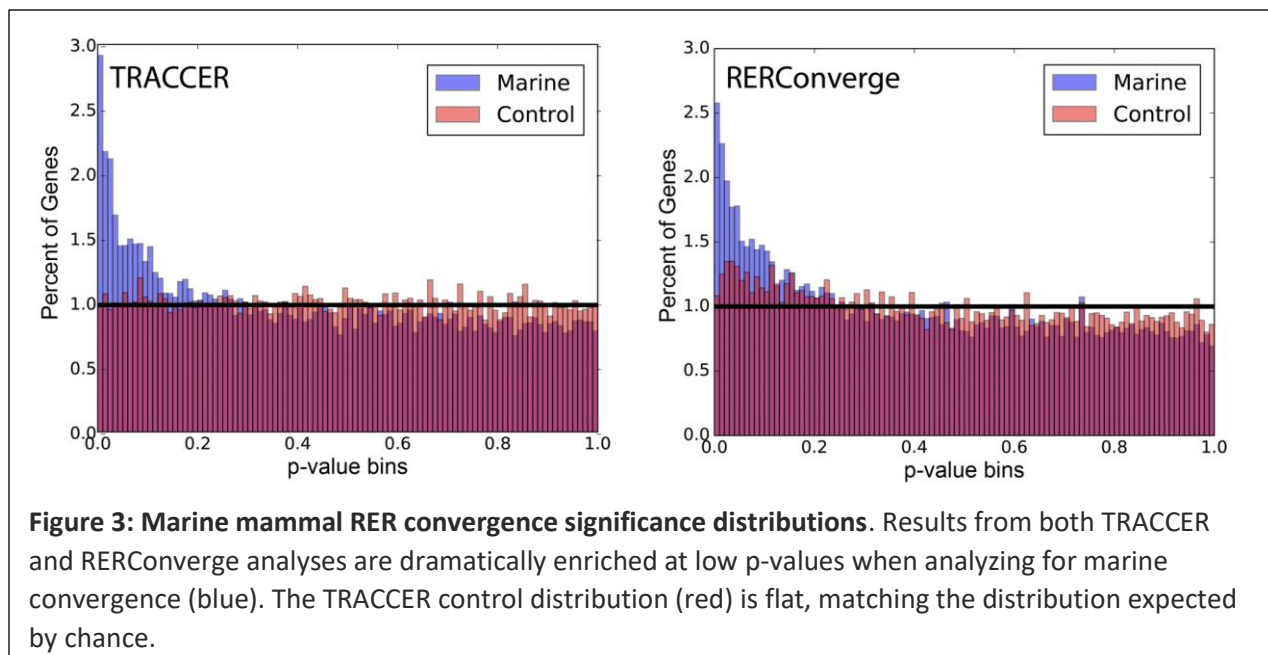
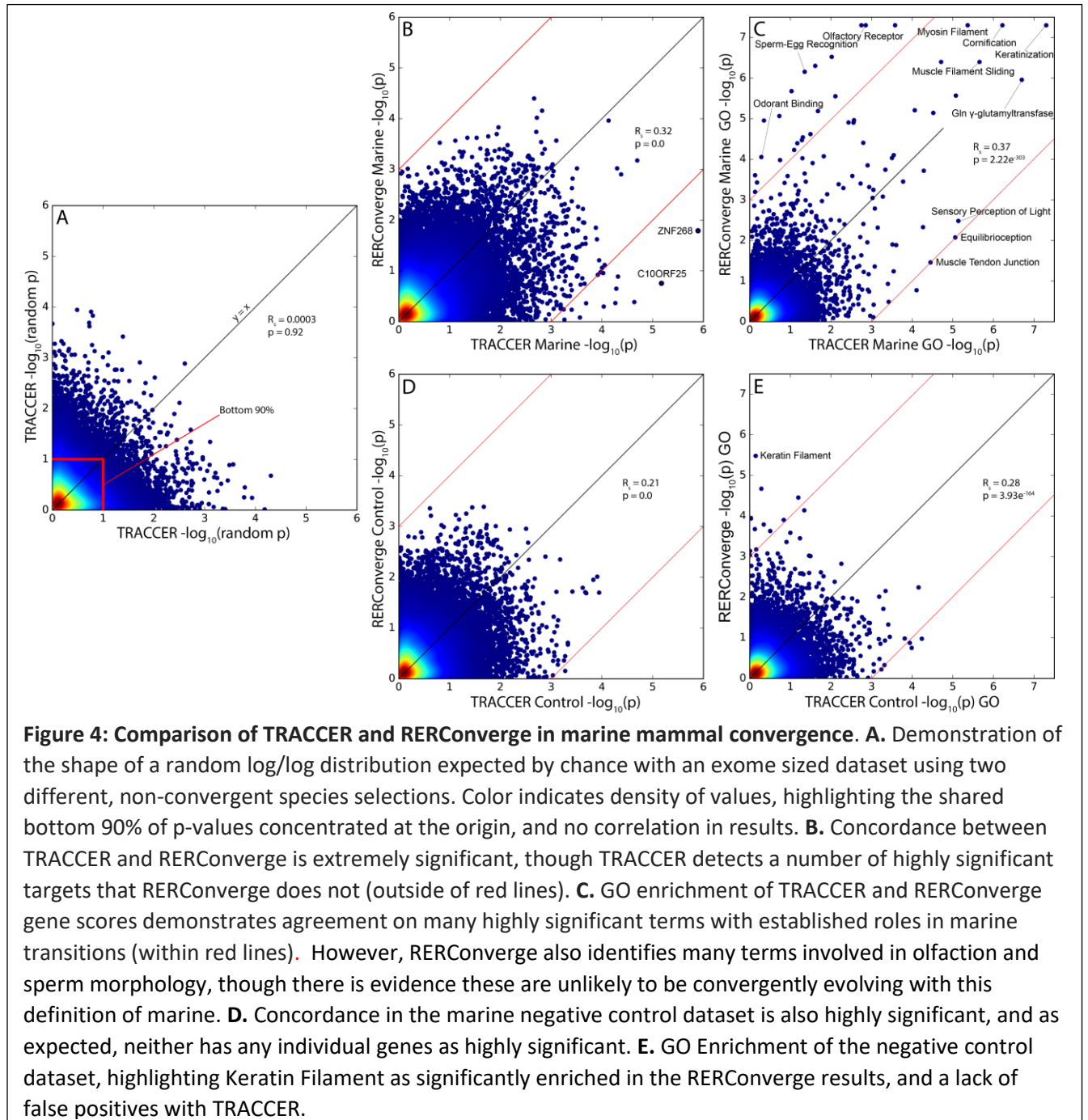


Figure 2: Boundaries of significance for marine mammal convergence. A. To highlight how topology influences significance in TRACCER, ten million randomly simulated mammalian gene trees were scored for convergence in marine lineages. Significant probability scores less than $1/20,000$ were detected, indicating the topology of the marine dataset is sufficiently complex to identify convergent signals on an exome-sized dataset (**A0**, above red line). To gauge how these p-values were influenced by the pattern of the signal, we forced random marine lineages to extreme acceleration (**A1:5**), with an example of two accelerations in the box. The range of p-values was compressed towards the significance boundary as more lineages were accelerated. However, even with all five accelerated, random variation still produced scores below significance (**A5**). **B.** The source of this variation is highlighted when immediate sister terrestrial lineages are accelerated alongside four marine lineages. While four marine lineages sharing a signal is capable of yielding significant scores, if the closest relative of any marine lineage shares the same extreme acceleration, it is sufficient to prevent those scores from reaching significance. In the example acceleration, the accelerated Panda lineage is undermining the convergent acceleration shared by four random marine lineages.

1 When applied to the real mammalian genomic dataset, TRACCER reveals substantial
2 enrichment of genes at lower p-values, indicating convergence signals well above the flat
3 background rate seen in the control group (**Figure 3**). In our use, and matching the published
4 methods, RERConverge shows a similar p-value distribution and accurately recapitulates
5 established results with that tool. Unlike those publications, in representing these results and
6 for downstream analyses, we do not segregate accelerated and constrained genes, instead
7 focusing solely on significance. The concordance between these methods is high, with a
8 Spearman correlation of 0.32 and p-value less than $1E-350$ (**Figure 4B**). However, a number of
9 genes are dramatically shifted between these analyses, towards both more and less
10 significance. Unlike RERConverge, many TRACCER gene scores remain significant after false
11 discovery correction (**Supplementary S1**) indicating an improvement in sensitivity. To expand
12 upon these differences, GO gene set enrichment was performed using the SUMSTAT approach,
13 which has demonstrated power, flexibility, and simplicity over competing gene set enrichment
14 analyses (Ackermann and Strimmer, 2009; Chen et al., 2010; Tintle et al., 2009). Both TRACCER
15 and RERConverge reveal highly significant processes and systems evolving to facilitate
16 mammalian marine transitions (**Figure 4C, Supplementary S2**). They agree on many of the most
17 significant terms, with extremely significant p-values for cornification, keratinization, myosin
18 filament, and muscle filament sliding. These each have obvious implications with the transition



- 1 to a marine environment, which required dramatic epithelial adaptations and muscular changes.
- 2 Another highly significant gene set, “protein-glutamine gamma-glutamyltransferase activity”, is
- 3 likely playing a role in the marine metabolic and dietary adaptations, particularly with glucose
- 4 regulation, and has intriguing parallels with diabetes pathology in the literature (Schermerhorn,
- 5 2013; Venn-Watson, 2014).



1 However, when we focus on the GO-terms with the greatest shift in significance
2 between the two analyses, we reveal some systems are differentially detected, including
3 sperm-egg recognition and odorant binding (**Figure 4C**). Multiple terms involving sperm
4 physiology are differentially detected as significant by RERConverge (**Supplementary S2**). As an
5 example, the ‘sperm-egg recognition’ gene set is one of the most dramatically shifted, with a p-
6 value of 7E-07 using RERConverge, but only 0.044 from TRACCER. We earlier highlighted *PLCZ1*
7 (**Figure 1**) which is the greatest contributor to the ‘sperm-head’ signal in RERConverge. With a
8 p-value shift from 0.002 to 0.65, *PLCZ1* is an excellent example of how topological weighting
9 and ancestral inferences are underscoring the observed differences in the analyses.
10 TRACCER and RERConverge also differ on the conventionally marine-associated GO terms
11 “olfactory receptor activity” and “odorant binding proteins” (ODPs). While there is plenty of
12 evidence that olfaction in general should be reduced, it is not entirely absent, and the
13 discrepancy in these terms may be important. It is certain that many olfactory receptors have
14 been pseudogenized (Hayden et al., 2010), and the olfactory bulb is reduced in most cetaceans
15 and entirely absent in toothed whales (Kishida et al., 2015). Yet, it is only modestly significant in
16 TRACCER ($p=.0014$) and extremely significant in RERConverge ($p < 1E-07$). The differential
17 detection of ODPs is driving this shift, as they are entirely a subset of “olfactory receptor
18 activity”, and while significant with RERConverge ($p = 8.80E-05$), are not with TRACCER ($p =$
19 0.52).

20 The mechanisms and functions of vertebrate ODPs have proven puzzling, but there is
21 evidence they are involved in the perception of pheromones more so than general olfaction
22 (Tegoni et al., 2000). Many marine mammals still utilize pheromones and require the molecular
23 machinery to detect them, so they are unlikely to have been universally compromised as the
24 olfactory receptors have. For instance, dolphins likely employ and detect sex pheromones in
25 their urine (Muraco and Kuczaj, 2015). Critically, the pinnipeds still maintain their vomeronasal
26 system, including strong selective pressure on *TRPC2*, an essential genetic component (Yu et al.,
27 2010), likely because they still maintain some terrestrial behavior. Strictly flagging the pinnipeds
28 as marine mammals, and not their ancestral branch, despite both spending a large fraction of
29 their life on land, may be confounding the analysis. The delineation between terrestrial and

1 marine mammals is not as cut and dry as most convergence publications suggest and echoes
2 the difficulties of inferring ancestral states. The evolution of pheromones in marine mammals
3 and the molecular systems for their detection may be more complex than originally assumed,
4 but according to TRACCER the “odorant binding” gene set is not convergently evolving with this
5 selection of marine mammals, thereby limiting the significance of the more general “olfactory
6 receptor activity” set.

7 As a negative control for the gene set analysis, we also ran GO enrichment with the
8 same settings on the marine control dataset: aardvark, alpaca, camel, microbat, and myotis bat.
9 TRACCER has no significant terms with FDRs below 0.3 and is within expected levels of
10 background noise. RERConverge, however, has 10 significantly enriched GO-Terms with FDRs
11 below 0.3, including “keratin filament” with a highly significant p-value of 3.3E-06 (**Figure 4E,**
12 **Supplemental S4**). This raises the concern that the slight enrichment in the control p-value
13 distribution produced by RERConverge (**Figure 2B**) may have a systematic trend; there are no
14 known convergent changes to the integument of the control lineages and there is no
15 immediate explanation for why the keratin gene set should be significantly enriched. In all,
16 TRACCER yielded more significant genes, yet fewer significant GO terms, while maintaining a
17 lower false positive rate in the control GO-enrichment. This suggests improvements in both
18 sensitivity and specificity.

19

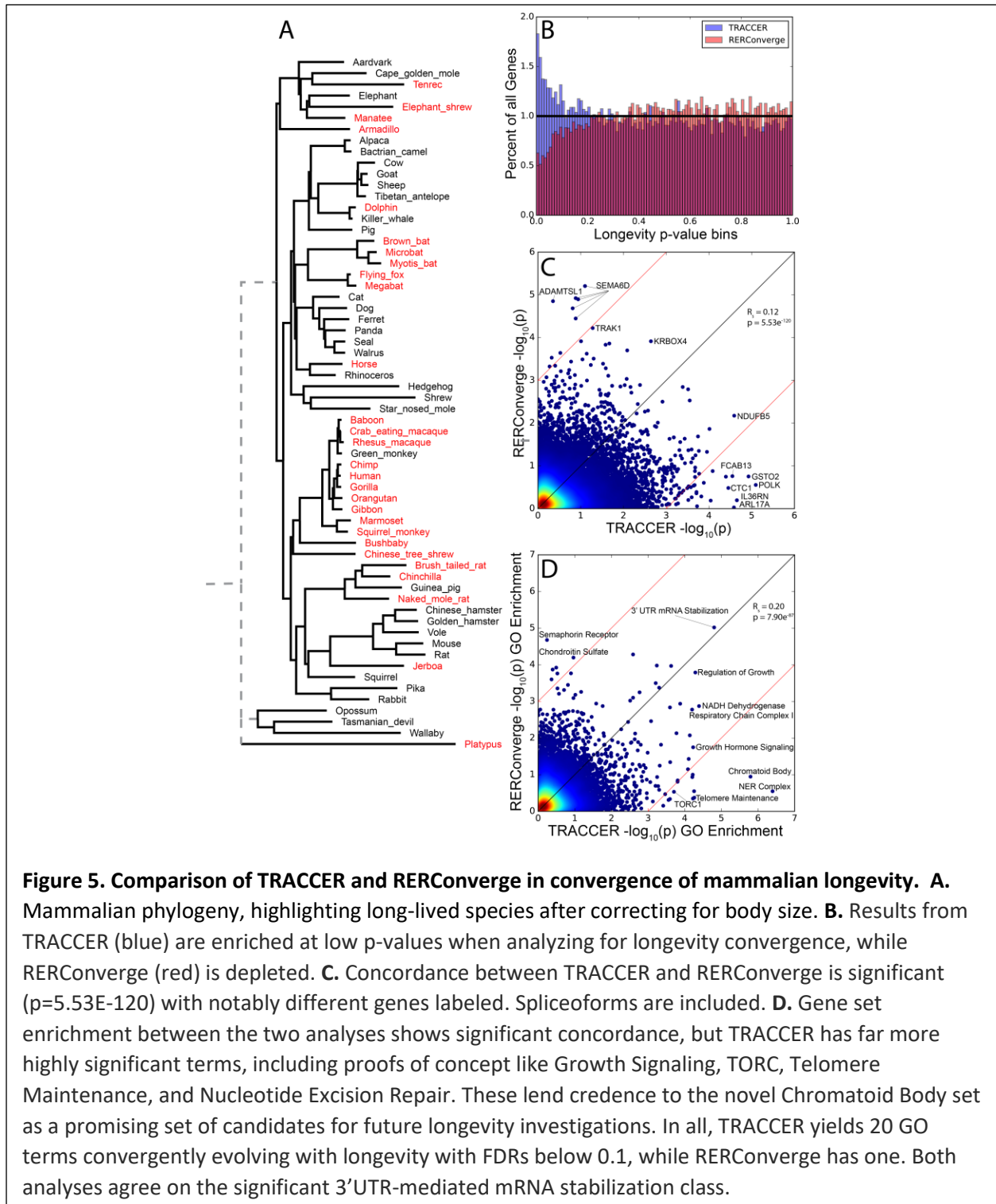
20 **Test Case 2: Mammalian Longevity Convergence**

21 To assess TRACCER’s ability to parse convergent signals on a more enigmatic, but also more
22 common trait, we applied it to longevity on the same mammalian phylogeny. Again, to compare
23 how the inclusion of topology may augment convergence analyses, we matched the design to
24 previous RERConverge studies (Kowalczyk et al., 2020). As in that study, the longevity trait was
25 calculated as the second principal component of body size and maximum lifespan. These
26 species are effectively “long-lived after correcting for the expected lifespan of their body-size”.
27 However, TRACCER currently only operates on binary traits, so all positive values were treated
28 as equivalently long-lived (**Figure 5A**). We compared these results directly to RERConverge
29 operated with similar settings. Importantly, RERConverge has the functionality to analyze

1 continuous traits; we ran it as binary nonetheless in order to keep the analyses as comparable
2 as possible and gauge the benefit of integrating topology. TRACCER demonstrated substantial
3 enrichment at low p-values, while RERConverge was depleted (**Figure 5B**). The concordance
4 between the studies was significant, with a Spearman correlation 0.12 and p-value 5.53E-120
5 (**Figure 5C**). Again, many genes were dramatically shifted between the two analytic pipelines, as
6 seen outside the red lines. Importantly, spliceforms were included, so some proteins are listed
7 multiple times to gain insight into exon specific trends. To generalize the differences observed,
8 we again performed SUMSTAT GO gene set enrichment, using only the most significant of each
9 spliceform when applicable (**Figure 5D**). TRACCER identifies “Positive Regulation of Growth”
10 ($p=5.12E-5$), “TORC1” ($p = 1.9E-4$), and “Telomere Maintenance” ($p = 5.96E-5$) as three
11 compelling proofs of concept. These terms have obvious implications in the longevity literature
12 (Foley et al., 2018; Papadopoli et al., 2019; Zhu et al., 2019) and lend credence to the others
13 identified in this analysis. Both analyses agree on the significant enrichment of “3’-UTR-
14 mediated mRNA Stabilization”, which has previously been implicated in the evolution of
15 longevity based on patterns of duplication (Doherty and de Magalhães, 2016). However,
16 TRACCER identifies Nucleotide Excision Repair two orders of magnitude more significant than
17 the other terms, and six orders greater than RERConverge scores it. This is driven in part by the
18 most significant member of the group, *POLK*, which has been previously implicated in longevity.
19 Although *POLK* deficient mice are normal in most regards, they do exhibit reduced lifespan
20 (Stancel et al., 2009). More interestingly, *POLK* is one of the greatest differentially expressed
21 repair genes when comparing human and naked mole rats – relatively long-lived mammals -- to
22 mice (MacRae et al., 2015). The second highest hit in the NER set, *ARL17A*, has its expression
23 significantly associated with human longevity past the 90th percentile (Deelen et al., 2019). NER
24 specifically may be underappreciated in the evolution of longevity.

25 These proof of concept GO classes lend credence to the highest scoring genes that are
26 not associated with any significant terms and suggests they should be investigated for
27 implications in those systems. *GSTO2*, *ILL36RN*, *NDUFB5*, *EFCAB13*, and *CTC1* all have highly
28 significant scores less than 5E-05 and FDRs less than 0.16, but are not present in the most
29 significant GO terms. They may have unappreciated roles in growth regulation or DNA repair.

- 1 The regions of variation and conservation in these genes and gene sets are exciting targets for
- 2 future longevity investigations.



1 TRACCER results are also highly enriched for the “chromatoid body” gene set ($p=1.10E-$
2 6). This functional class of genes is much less studied in longevity than sets like TORC signaling
3 and DNA repair, but has intriguing ramifications and potential. The chromatoid body is essential
4 for maintaining male germ line quality, and its structure and function is demonstrably
5 compromised with age (Santos et al., 2018). Of course, if the male germ line is compromised
6 with age, it would be inherently difficult to select for longevity or the delayed maturation that
7 reliably accompanies it. Many of these genes are validated in germ-line transposable element
8 suppression, but may be more pleiotropic than they first appear, and are potentially involved in
9 a more general transposable element suppression to maintain genomic integrity with age. For
10 instance, the most significant member of this set, *DDX25* ($p = 3.7E-4$), is an RNA helicase
11 essential for spermatogenesis, but is also broadly expressed in the brain and pituitary where it
12 is poorly characterized. Transposable element suppression may have underappreciated and
13 broad impacts on aging and longevity (Sturm et al., 2015). In all, TRACCER identifies
14 substantially more significant genes and GO terms convergently evolving than comparable
15 methods that remain agnostic to topology (**Supplemental S5,S6**).

16

17 **Discussion**

18 Convergence does not occur in isolation; it is inherently a comparative concept, and the various
19 comparisons across a phylogeny are not all equivalent. Comparisons between those with more
20 similar genetic contexts, and importantly, those encompassing trait-changes, will be the most
21 informative to unravel the regulation facilitating that trait change. TRACCER is constructed
22 around this concept in order to augment convergent relative rate analysis with the
23 phylogenetic topological relationships. It builds upon previous work using Relative Evolutionary
24 Rates to infer genomic systems facilitating instances of convergent evolution. TRACCER also
25 includes a diagnostic mode, which is particularly useful for conceptualizing the shape of a
26 convergent signal given a phylogeny of interest. The diagnostic mode is valuable when
27 designing an experiment within a particular group and can guide sampling with an estimation of
28 whether additional species will be worthwhile.

1 We compare TRACCER directly to RERConverge which uses a similar approach and was
2 the first to use RER convergence across broad datasets (Chikina et al., 2016). We found
3 RERConverge to be powerful and elegant, efficiently revealing significant genes and pathways
4 underlying convergent trait evolution. We use it as a valuable benchmark for TRACCER as it is
5 the most comparable method without factoring in the complexity of topology, though there are
6 other subtle differences in the calculation details. The differences in the results between the
7 two approaches illuminates how the inclusion of phylogenetic relationships and the bypassing
8 of ancestral state inferences may augment convergence analyses. To our knowledge, no other
9 relative rate convergence analysis includes this set of features.

10 Instead of treating each branch as an independent and equivalent event, TRACCER's
11 pairwise approach makes comparisons in reference to each pair's most recent common
12 ancestor. This ensures that each comparison represents the same genetic context evolving over
13 the same time frame, such that the calculated RERs are directly comparable. Each comparison
14 can then be weighted based on the tree topology, without assumptions about ancestral states,
15 and compensates for sampling biases. The default ranking transformations applied to both the
16 RER differences and topological weights maintain the spirit of convergence; it must be a
17 repeated signal, but some comparisons are more informative than others because of their
18 phylogenetic interrelationships. To ignore topology is to leave behind these informative
19 components of phylogenomic data that can refine the map between genotype and phenotype.

20 RERconverge implicated a number of genes in mammalian marine transitions despite
21 their closest relatives possessing similar signals and/or hinging upon the tenuous but required
22 ancestral state assumptions (**Figure 1**). Part of this is the unfortunate risk of needing to infer
23 ancestral states and flagging the pinniped ancestor as marine would be a simple change that
24 would dramatically change RERConverge results; that branch represents 1/7th of the marine
25 branches in the dataset and can easily shift a signal to and from significance. We repeated this
26 decision when using RERConverge to stay true to published results and to highlight the risks of
27 such requirements. In the case of *PLCZ1*, a gene specific for sperm heads (Ecoffier et al., 2016),
28 a significant p-value led to the enrichment of multiple sperm GO terms, none of which have an
29 *a priori* reason to be constrained when adapting to an aquatic environment. To the contrary,

1 scanning electron microscopy available for 36 ungulates and cetaceans suggests sperm-head
2 morphology should be variable in the entire group (Meisner et al., 2005). In fact, the cetacean
3 sperm are missing many of the structures present in the rest of the group, which would not
4 suggest constraint at a genetic level. The loss of such structures is generally associated with a
5 relaxation of selection (Lahti et al., 2009). Inferring ancestral states is often difficult, or even
6 impossible, and discrepancies when assigning them can propagate to databases and
7 downstream functional analyses in the field. Despite these risks, most convergence analyses
8 require it.

9 By utilizing the full wealth of information available from phylogenomics, TRACCER
10 obviates the need to infer ancestral states and can detect a more refined set of elements than
11 currently available RER approaches. Specificity was improved when applied to the marine
12 mammal dataset. This is particularly evident with the lower false discovery rates, flatter control
13 distribution, and lack of significant hits in the control gene set enrichment as compared to
14 RERConverge. With mammalian longevity, sensitivity was dramatically improved by TRACCER,
15 with many more highly significant genes and gene sets being revealed. These included
16 literature-validated controls like TORC signaling, telomere maintenance, and DNA repair as
17 aspects convergently evolving to modulate mammalian longevity. When robust “proofs of
18 concept” like these emerge, it increases confidence in the novel genes and terms arising
19 alongside, like TRACCER’s highly significant “chromatoid body” GO Term. This gene set is
20 currently only characterized in germ-line transposon suppression, but there is evidence it may
21 be more general in maintaining genomic and epigenomic integrity. The specific highly
22 significant genes that are not members of these defined gene sets are strong candidates for
23 additional investigations and classification. *ZNF268* and *C10ORF25* are the most significant hits
24 for marine transitions but have not been characterized for dermal or muscular functions.
25 *GSTO2*, *ILL36RN*, *NDUFB5*, *EFCAB13*, and *CTC1* are highly significant for convergently evolving
26 with mammalian longevity, but are not present in the most significant GO terms. These results
27 demonstrate that integrating topology into convergent relative rate analyses empowers
28 comparative genomics to discover actionable targets to modulate traits of interest.

29

1 Methods

2 TRACER operates on phylogenetic trees, the specifics of which can be decided by
 3 experimenter as appropriate for their research. The only requirements are a species tree and a
 4 set of individual element trees fixed to the species tree topology. From the species tree, a scalar
 5 for each pairwise comparison is derived from the total distance between the two species, with
 6 each branch along that path weighted by the number lineages that share it (**Figure 6**). The
 7 scalars across the phylogeny are inverted, setting the furthest comparison to the lowest weight,
 8 and the closest comparison having the greatest weight. Two options are provided for an
 9 additional transformation: the N^{th} root as decided by the user or a rank transformation across
 10 all comparisons. The analyses in this study used the rank transformation, with the closest
 11 species having the greatest rank. These transformations ensure that proximate comparisons are
 12 weighted without swamping the analysis, and distal comparisons are downplayed without
 13 being ignored.

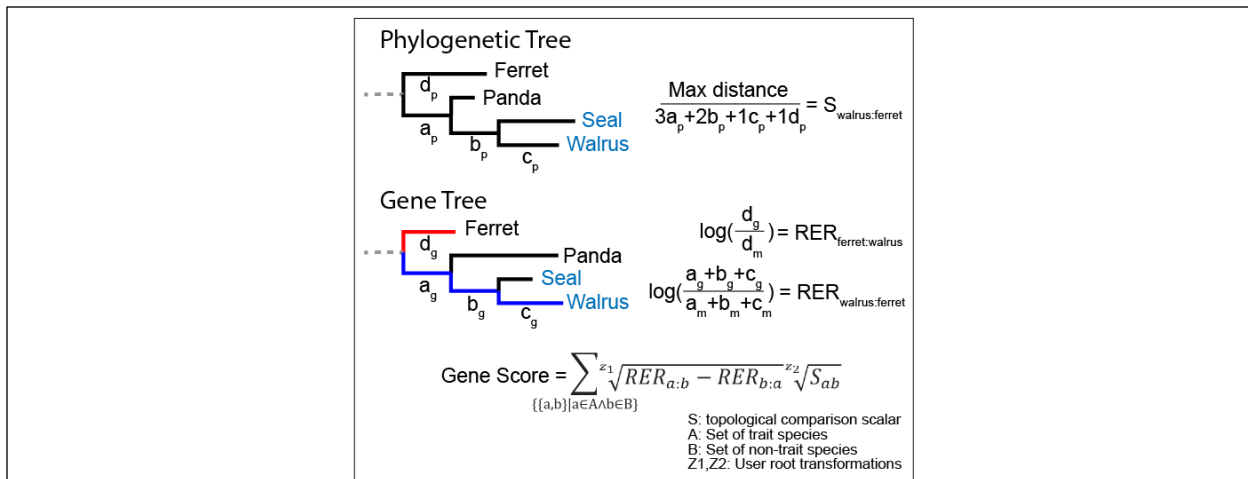


Figure 6. TRACER Example Calculations. Each pairwise comparison of trait-bearing and non-trait species has a scalar derived from the phylogenetic tree. For example, when comparing Walrus and Ferret, all branch lengths back to their MRCA on the phylogeny (a_p, b_p, \dots) are multiplied by the number of extant lineages sharing that branch, and summed. These are then inverted by taking the ratio with the maximum pairwise distance, such that the closest comparison, with the shortest distance between them, has the greatest value. Using the gene tree, RERs for each comparison are derived as the log-ratio of distance to their MRCA ($a_g + b_g, \dots$) to the median ($a_m + b_m, \dots$) of that distance across all gene trees. Finally, the difference between RERs for each comparison, and the scalar for that comparison, are z^{th} root transformed as chosen by the user, then summed across all comparisons to produce the gene score. Alternatively, as used in these analyses, they were rank transformed before summing. To calculate significance, these gene scores are then compared to scores calculated from randomly permuted combinations of branches across all gene trees.

1 The score for a given gene tree is the sum of all pairwise RER differences relative to their
2 MRCA, each scaled by proximity and branch use on the species tree. To calculate RERs for these
3 comparisons, TRACCER uses the medians of the distances across all genes for each lineage to
4 their MRCA, ignoring zeroes and discordance artifacts, to derive a fold-change. Sequence
5 divergence rates are gutted by a large number of zeroes, resulting in a non-normal
6 distribution that makes the concept of ‘relative’ rates meaningless unless they are excluded
7 from this step of the calculation. Importantly, the calculated relative rates are log-transformed.
8 The log transformation balances accelerated and decelerated fold-change relative rates, as a
9 $1/10^{\text{th}}$ deceleration from 1 to 0.1 should be the same importance as the ten-fold acceleration
10 from 1 to 10, despite an order of magnitude difference in the raw differences. As would be
11 expected, the data in our benchmarking investigations is log-normally distributed after
12 removing the zeroes and artifacts.

13 For each combination of trait- and non-trait bearing species, the log-RERs to their most
14 recent common ancestor are subtracted to quantify differences and direction from their shared
15 ancestral state. By focusing on differences, TRACCER can, for example, detect a relaxation of
16 selection on a conserved element even if the rate is still below the background rate. The
17 pairwise values across the tree are then rank transformed to correct for outliers while ensuring
18 all comparisons have an impact. Then they are multiplied by the scalar derived from the species
19 tree to weigh comparisons by proximity and branch use. All pairwise values are then summed
20 for a final gene score. An example of these calculations is in Figure 6.

21 To ascertain a given score’s significance, a permutation test is performed to determine
22 the background score distribution. As topology is fixed, branches can be shuffled between trees
23 with millions of permutations and then scored as before. Importantly, when species are missing
24 on a gene tree, it changes the shape of the possible scores and requires its own resampling to
25 compare against. Finally, gene tree scores are ranked against the permuted scores on the same
26 tree topology to determine the percentage that are greater as a p-value. Gene trees exhibiting
27 constraint are only compared to the resamples scoring below the median, and those exhibiting
28 acceleration to resamples above the median.

1 Fixing trees to the same topology is required to directly compare branches between
2 trees and derive relative rates. This can be problematic when a lineage is missing due to gene
3 loss, as it may remove key branches and undo the 1:1 mapping between trees. Fortunately,
4 because TRACCER performs comparisons in reference to their most recent common ancestor,
5 these can still be compared even when a missing lineage removes an ancestral branch.
6 However, forcing the gene trees to a specific topology will introduce artifacts when the two
7 disagree (Degnan and Rosenberg, 2009) and it would be nonsensical to derive RERs directly
8 from this discordance. Discordance over a larger distance could be detected by comparing fixed
9 and unfixed gene trees, or sequence identity directly. Ideally this information would be factored
10 as a complementary dimension of a convergence analysis, but they cannot be used for RERs
11 directly. TRACCER does compensate for one form of discordance: closely related clades and
12 slowly evolving genes often manifest short, but non-zero branch length artifacts.

13 These artifacts are due to a lack of informative sites in a slowly evolving region. When a
14 lineage has a single SNP while the rest of the clade is identical, the identical lineages will try to
15 cluster. If those are not immediate sisters, it can yield short-non-zero branch lengths. These
16 values can and should be transformed to zero as they represent no change; they are still
17 identical to the ancestral state. TRACCER has an initial curation revealing the distribution of
18 branch lengths to the user, indicating which branch lengths are likely artifacts, and allowing
19 customization of the branch lengths the user wishes to transform. In our test datasets, these
20 took the form of branches with lengths between 10^{-4} and 10^{-7} . The exact values will vary based
21 on the tree building software used, but with both PAML and IQ-Tree, in mammals, fish, and
22 birds, we found 10^{-4} to be the right cutoff. While other analyses discard these values, or use
23 them without correction, TRACCER corrects these values to zero. However, should a gene-tree
24 consist predominantly of these short values, TRACCER will discard it as evolving too slowly to be
25 informative.

26 Trait distribution and topology is specific to every experiment and dramatically
27 influences the possible score distributions. It may be impossible to differentiate noise from a
28 true convergent signal in a given dataset. To determine if one's experimental design and
29 topology is sufficient to yield a convergent signal, and how robust that signal must be, TRACCER

1 can be run in diagnostic mode. This feature uses only the species tree to generate millions of
2 gene trees with Brownian motion and forces various convergent patterns upon them. These
3 scores are compared to the strictly random trees to determine how often such scores occur by
4 chance. This can inform experimental design and sampling strategies to optimize resources
5 necessary to get a significant signal.

6

7 Gene Set Enrichment

8 GO Terms for the mammalian dataset were harvested from Ensembl. Terms with less than
9 three members, or more than five hundred, were discarded as being uninformative or
10 misleading in the context of set enrichment. The SUMSTAT approach was used on log
11 transformed p-values from each analysis, with an additional square root transformation to
12 undermine outliers. In short, the square-rooted, log transformed p-values for each gene in a set
13 were summed and compared to a distribution of randomly sampled scores of the same set size
14 to determine an enrichment p-value. Notably, we do not segregate genes into accelerated and
15 constrained bins for performing enrichment; we use the significance of their convergent signal
16 regardless of the polarity. To handle spliceoforms, only the most significant was included. FDRs
17 are calculated as the expected number of hits at that significance divided by the number of
18 actual hits at that significance or better. At the gene level, the inclusion of spliceoforms makes
19 the calculated FDRs roughly twice as conservative as they should be, but provides potential
20 insight into exon specific influence.

21

22 RERConverge Settings

23 For marine and control convergence analysis, RERConverge settings were matched to their
24 published protocols (Chikina et al., 2016). For the longevity analysis, the lineages with a positive
25 second principle component of body size and longevity were flagged as bearing the trait
26 (Kowalczyk et al., 2020). These species are effectively “long-lived after correcting for expected
27 lifespan from their body-size”. We used RERConverge’s clade-weighting flag to automatically
28 distribute their weight onto ancestral branches with clade=“all”, weighted=TRUE. When
29 calculating residuals, we used the following settings: transform = "sqrt", weighted = T, scale = T.

1 Finally, to correlate the residuals with the binary phenotype, we used min.sp=15 min.pos=5,
2 and weighted="auto". The minimum species correspond with the minimum species cutoffs
3 used with TRACCER.

4

5 TRACCER as Software

6 TRACCER's permutation strategy can be computationally intensive, particularly if variable
7 lineages are missing in many gene trees, as each tree shape will need its own set of
8 permutations to determine its unique score space. However, provided no lineages are missing
9 from any gene trees, or they can be excluded such that only a single score space is needed,
10 TRACCER is efficient and will execute mammalian exome wide datasets in a few minutes. To
11 scale the computational demand for different computing systems, TRACCER can automatically
12 scale the resampling to fit within a designated time window. Alternatively, the user can demand
13 the degree of precision and TRACCER will run until a certain number of resamples are
14 performed. The marine mammal dataset includes 37,272 protein sequences, including
15 spliceoforms, with 6,804 unique topologies. TRACCER runs on this dataset size in 578 minutes
16 with 14 CPUs to achieve a significance limit of 1.6E-7. These times are highly variable depending
17 on the size of the dataset, how many unique topologies there are, and how many unique
18 topologies are significant. The resources required are typically within the scope of any modern
19 computational center, or even a well-equipped personal desktop.

20 TRACCER was written in Python3 with minimal dependencies for easy installation and
21 execution in a variety of computing environments. Scripts, updates, and tutorials will be made
22 publicly available on github and the lab website at <http://www.fishbonelab.org/>

23

- 1 References:
- 2 Ackermann, M., and Strimmer, K. (2009). A general modular framework for gene set
- 3 enrichment analysis. *BMC Bioinformatics* *10*, 47.
- 4 Berta, A., Churchill, M., and Boessenecker, R.W. (2018). The Origin and Evolutionary Biology of
- 5 Pinnipeds: Seals, Sea Lions, and Walruses. *Annu. Rev. Earth Planet. Sci.* *46*, 203–228.
- 6 Chen, X., Wang, L., Hu, B., Guo, M., Barnard, J., and Zhu, X. (2010). PATHWAY-BASED ANALYSIS
- 7 FOR GENOME-WIDE ASSOCIATION STUDIES USING SUPERVISED PRINCIPAL COMPONENTS.
- 8 *Genet. Epidemiol.* *34*, 716–724.
- 9 Chikina, M., Robinson, J.D., and Clark, N.L. (2016). Hundreds of Genes Experienced Convergent
- 10 Shifts in Selective Pressure in Marine Mammals. *Mol. Biol. Evol.* *33*, 2182–2192.
- 11 Currie, A. (2013). Convergence as Evidence. *Br. J. Philos. Sci.* *64*, 763–786.
- 12 Daane, J.M., Auvinet, J., Stoebenau, A., Yergeau, D., Harris, M.P., and Iij, H.W.D. (2020).
- 13 Developmental constraint shaped genome evolution and erythrocyte loss in Antarctic fishes
- 14 following paleoclimate change. *PLOS Genet.* *16*, e1009173.
- 15 Deelen, J., Evans, D.S., Arking, D.E., Tesi, N., Nygaard, M., Liu, X., Wojczynski, M.K., Biggs, M.L.,
- 16 van der Spek, A., Atzmon, G., et al. (2019). A meta-analysis of genome-wide association studies
- 17 identifies multiple longevity genes. *Nat. Commun.* *10*, 3669.
- 18 Degnan, J.H., and Rosenberg, N.A. (2009). Gene tree discordance, phylogenetic inference and
- 19 the multispecies coalescent. *Trends Ecol. Evol.* *24*, 332–340.
- 20 Doherty, A., and de Magalhães, J.P. (2016). Has gene duplication impacted the evolution of
- 21 Eutherian longevity? *Aging Cell* *15*, 978–980.
- 22 Echave, J., Spielman, S.J., and Wilke, C.O. (2016). Causes of evolutionary rate variation among
- 23 protein sites. *Nat. Rev. Genet.* *17*, 109–121.

- 1 Escoffier, J., Lee, H.C., Yassine, S., Zouari, R., Martinez, G., Karaouzène, T., Coutton, C., Kherraf,
2 Z., Halouani, L., Triki, C., et al. (2016). Homozygous mutation of PLCZ1 leads to defective human
3 oocyte activation and infertility that is not rescued by the WW-binding protein PAWP. *Hum.*
4 *Mol. Genet.* 25, 878–891.
- 5 Foley, N.M., Hughes, G.M., Huang, Z., Clarke, M., Jebb, D., Whelan, C.V., Petit, E.J., Touzalin, F.,
6 Farcy, O., Jones, G., et al. (2018). Growing old, yet staying young: The role of telomeres in bats’
7 exceptional longevity. *Sci. Adv.* 4, eaao0926.
- 8 Gould, S.J., Gold, S.J., and Gould, T.A.A.P. of Z.S.J. (2002). *The Structure of Evolutionary Theory*
9 (Harvard University Press).
- 10 Hayden, S., Bekaert, M., Crider, T.A., Mariani, S., Murphy, W.J., and Teeling, E.C. (2010).
11 Ecological adaptation determines functional mammalian olfactory subgenomes. *Genome Res.*
12 20, 1–9.
- 13 Hiller, M., Schaar, B.T., Indjeian, V.B., Kingsley, D.M., Hagey, L.R., and Bejerano, G. (2012). A
14 “forward genomics” approach links genotype to phenotype using independent phenotypic
15 losses among related species. *Cell Rep.* 2, 817–823.
- 16 Katzman, S., Kern, A.D., Bejerano, G., Fewell, G., Fulton, L., Wilson, R.K., Salama, S.R., and
17 Haussler, D. (2007). Human genome ultraconserved elements are ultraselected. *Science* 317,
18 915.
- 19 Kishida, T., Thewissen, J., Hayakawa, T., Imai, H., and Agata, K. (2015). Aquatic adaptation and
20 the evolution of smell and taste in whales. *Zool. Lett.* 1, 9.
- 21 Koretsky, I., and Barnes, L. (2003). ORIGINS AND RELATIONSHIPS OF PINNIPEDS, AND THE
22 CONCEPTS OF MONOPHYLY VERSUS DIPHYLY. *J. Vertebr. Paleontol.* 23, 69A-69A.
- 23 Kowalczyk, A., Partha, R., Clark, N.L., and Chikina, M. (2020). Pan-mammalian analysis of
24 molecular constraints underlying extended lifespan. *ELife* 9, e51089.

- 1 Lahti, D.C., Johnson, N.A., Ajie, B.C., Otto, S.P., Hendry, A.P., Blumstein, D.T., Coss, R.G.,
2 Donohue, K., and Foster, S.A. (2009). Relaxed selection in the wild. *Trends Ecol. Evol.* *24*, 487–
3 496.
- 4 Losos, J.B. (2011). Convergence, Adaptation, and Constraint. *Evolution* *65*, 1827–1840.
- 5 MacRae, S.L., Croken, M.M., Calder, R.B., Aliper, A., Milholland, B., White, R.R., Zhavoronkov, A.,
6 Gladyshev, V.N., Seluanov, A., Gorbunova, V., et al. (2015). DNA repair in species with extreme
7 lifespan differences. *Aging* *7*, 1171–1182.
- 8 Marcovitz, A., Turakhia, Y., Chen, H.I., Gloudemans, M., Braun, B.A., Wang, H., and Bejerano, G.
9 (2019). A functional enrichment test for molecular convergent evolution finds a clear protein-
10 coding signal in echolocating bats and whales. *Proc. Natl. Acad. Sci.* *116*, 21094–21103.
- 11 Meisner, A.D., Klaus, A.V., and O’Leary, M.A. (2005). Sperm head morphology in 36 species of
12 artiodactylans, perissodactylans, and cetaceans (Mammalia). *J. Morphol.* *263*, 179–202.
- 13 Muraco, H., and Kuczaj, S. (2015). Conceptive Estrus Behavior in Three Bottlenose Dolphins
14 (*Tursiops truncatus*).
- 15 Naser-Khdour, S., Minh, B.Q., Zhang, W., Stone, E.A., and Lanfear, R. (2019). The Prevalence and
16 Impact of Model Violations in Phylogenetic Analysis. *Genome Biol. Evol.* *11*, 3341–3352.
- 17 Papadopoli, D., Boulay, K., Kazak, L., Pollak, M., Mallette, F., Topisirovic, I., and Hulea, L. (2019).
18 mTOR as a central regulator of lifespan and aging. *F1000Research* *8*.
- 19 Parker, J., Tsagkogeorga, G., Cotton, J.A., Liu, Y., Provero, P., Stupka, E., and Rossiter, S.J. (2013).
20 Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* *502*, 228–
21 231.
- 22 Partha, R., Chauhan, B.K., Ferreira, Z., Robinson, J.D., Lathrop, K., Nischal, K.K., Chikina, M., and
23 Clark, N.L. (2017). Subterranean mammals show convergent regression in ocular genes and
24 enhancers, along with adaptation to tunneling. *ELife* *6*, e25884.

- 1 Pearce, T. (2012). Convergence and Parallelism in Evolution: A Neo-Gouldian Account. *Br. J.*
2 *Philos. Sci.* *63*, 429–448.
- 3 Pollard, K.S., Salama, S.R., King, B., Kern, A.D., Dreszer, T., Katzman, S., Siepel, A., Pedersen, J.S.,
4 Bejerano, G., Baertsch, R., et al. (2006). Forces Shaping the Fastest Evolving Regions in the
5 Human Genome. *PLOS Genet.* *2*, e168.
- 6 Revell, L.J., Harmon, L.J., and Collar, D.C. (2008). Phylogenetic signal, evolutionary process, and
7 rate. *Syst. Biol.* *57*, 591–601.
- 8 Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer,
9 T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., et al. (2015). The UCSC Genome Browser
10 database: 2015 update. *Nucleic Acids Res.* *43*, D670–D681.
- 11 Rosenblum, E.B., Parent, C.E., and Brandt, E.E. (2014). The Molecular Basis of Phenotypic
12 Convergence. *Annu. Rev. Ecol. Evol. Syst.* *45*, 203–226.
- 13 Santos, E.G., Silva, M.A., Amorim, R.P., Giordano, L. de S., Silva, D. de S., Rasmussen, L.T., and
14 Peruquetti, R.L. (2018). Aging and chromatoid body assembly: Are these two physiological
15 events linked? *Exp. Biol. Med.* *243*, 917–925.
- 16 Schermerhorn, T. (2013). Normal Glucose Metabolism in Carnivores Overlaps with Diabetes
17 Pathology in Non-Carnivores. *Front. Endocrinol.* *4*.
- 18 Slater, G.J., Harmon, L.J., and Alfaro, M.E. (2012). Integrating fossils with molecular phylogenies
19 improves inference of trait evolution. *Evol. Int. J. Org. Evol.* *66*, 3931–3944.
- 20 Stancel, J.N.K., McDaniel, L.D., Velasco, S., Richardson, J., Guo, C., and Friedberg, E.C. (2009).
21 Polk Mutant Mice Have a Spontaneous Mutator Phenotype. *DNA Repair* *8*, 1355–1362.
- 22 Sturm, Á., Ivics, Z., and Vellai, T. (2015). The mechanism of ageing: primary role of transposable
23 elements in genome disintegration. *Cell. Mol. Life Sci.* *72*, 1839–1847.

- 1 Tegoni, M., Pelosi, P., Vincent, F., Silvia, S., Campanacci, V., Grolli, S., Ramoni, R., and Cambillau,
2 C. (2000). Mammalian Odorant Binding Proteins. *Biochim. Biophys. Acta BBA - Protein Struct.*
3 *Mol. Enzymol.* 1482, 229–240.
- 4 Tintle, N.L., Borchers, B., Brown, M., and Bekmetjev, A. (2009). Comparing gene set analysis
5 methods on single-nucleotide polymorphism data from Genetic Analysis Workshop 16. *BMC*
6 *Proc.* 3 *Suppl* 7, S96.
- 7 Venn-Watson, S. (2014). Dolphins and Diabetes: Applying One Health for Breakthrough
8 Discoveries. *Front. Endocrinol.* 5.
- 9 Yu, L., Jin, W., Wang, J., Zhang, X., Chen, M., Zhu, Z., Lee, H., Lee, M., and Zhang, Y. (2010).
10 Characterization of TRPC2, an essential genetic component of VNS chemoreception, provides
11 insights into the evolution of pheromonal olfaction in secondary-adapted marine mammals.
12 *Mol. Biol. Evol.* 27, 1467–1477.
- 13 Zhu, Y., Liu, X., Ding, X., Wang, F., and Geng, X. (2019). Telomere and its role in the aging
14 pathways: telomere shortening, cell senescence and mitochondria dysfunction. *Biogerontology*
15 20, 1–16.
- 16 Zou, Z., and Zhang, J. (2015). No Genome-Wide Protein Sequence Convergence for
17 Echolocation. *Mol. Biol. Evol.* 32, 1237–1241.
- 18