Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Contrasting Clade-Specific Relative Evolutionary Rates for Site-Level Adaptation Analyses

Facundo M. Giorello

facundo.giorello@cut.edu.uy

PDU Espacio de Biología Vegetal del Noreste, CENUR, sede Tacuarembó

Joaquina Farias

PDU Espacio de Biología Vegetal del Noreste, CENUR, sede Tacuarembó

Research Article

Keywords: Relative Evolutionary Rate, Positive Selection, Purifying Selection, dN/dS, Purifying Selection space, Cetacean

Posted Date: September 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10744925/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Contrasting Clade-Specific Relative Evolutionary Rates for Site-Level Adaptation Analyses

Facundo M. Giorello[1], Joaquina Farias[1]

[1] PDU Espacio de Biología Vegetal del Noreste, CENUR, sede Tacuarembó, Udelar, Ruta 5 km 386,200, 45000 Tacuarembó, Uruguay.

Corresponding author:
Facundo M. Giorello.
facundo.giorello@cut.edu.uy

Abstract

Detecting genetic adaptations is central to evolutionary biology, yet traditional methods, such as the non-synonymous to synonymous substitution rate ratio (dN/dS), are unable to capture more nuanced evolutionary processes. While positive selection can drive a burst of non-synonymous substitutions, it can also alter the intensity and space of subsequent purifying selection within specific clades after a single adaptive amino acid substitution. Here, we test the hypothesis that clade-specific Relative Evolutionary Rate (RER) shifts can reveal a distinct footprint of purifying selection, which in turn reflects a past adaptive event. Using a dataset of over 8,000 orthologous genes across 32 artiodactyl species with high-quality genome annotations, we compared site-level RER shifts against traditional dN/dS analyses. Our results reveal a substantial overlap of genes and sites among these related but independent estimations. Furthermore, the genes detected via RER comparison and missed by dN/dS are significantly interconnected and biologically meaningful, suggesting that they hold intrinsic value for uncovering adaptive processes that lack a typical positive selection signal.

Keywords

Relative Evolutionary Rate, Positive Selection, Purifying Selection, dN/dS, Purifying Selection space, Cetacean

Statements and Declarations

Competing interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

1. Introduction

Uncovering genetic adaptations is fundamental to understanding evolution. Multiple methods have been developed to infer positive selection using phylogenetic and population approaches (Oleksyk et al. 2010). The ratio between the non-synonymous substitution and synonymous substitution rate (dN/dS) has been the most used method to identify sites or branches under positive selection (Goldman and Yang 1994; Kosakovsky Pond and Frost 2005). In spite of existing controversies about how rigorous and robust it is to uncover positive selection (Nei et al. 2010; Rahman et al. 2021), several studies have linked sites under positive selection through dN/dS with biochemical functions (e.g., Barkman et al., 2007; Farias and Giorello, 2020). However, one weakness of dN/dS, which is shared with many other selection analyses, is that they are unable to find more subtle actions of natural selection.

It is well known that different evolutionary rates vary between protein domains, since they are subject to different purifying selection intensities. Evolutionary rates are also known to change because of positive selection, through an increased non-synonymous substitution rate. However, positive selection can modify the evolutionary rate in a more subtle way, setting a new intensity and space of purifying selection after a particular substitution (Bazykin and Kondrashov 2011; Gladieux et al. 2013). This is expected because an adaptation implies a change in the protein function, which in turn implies a new space of equivalent amino acids. The new set of equivalent amino acids can be manifested by a different evolutionary or substitution rate (Bazykin and Kondrashov 2011). Thus, a clade that has undergone an adaptive event may display a differential evolutionary rate at the amino acid site of adaptation relative to other clades that did not experience the adaptive process. A direct comparison of substitution rates is inherently problematic, given that they are differentially influenced by various biological traits of the species involved. The comparison of the relative evolutionary rate (RER) (Uzzell and Corbin 1971; Echave et al. 2016; Spielman and Kosakovsky Pond 2018) is a more robust alternative to capture the shift in the purifying selection intensity caused by the adaptation event.

Relative evolutionary rates can be thought of at two levels: gene (Kowalczyk et al. 2019) and site (Spielman and Kosakovsky Pond 2018). At the gene level, by factoring out shared background variables that affect all genes—such as mutation rate, effective population size, and time since speciation—relative rates reveal exactly how much faster or slower a specific gene evolved on a given branch (Kowalczyk et al. 2019). At the site level, they indicate how rapidly each residue evolves relative to the mean protein rate (Pupko et al. 2002; Spielman and Kosakovsky Pond 2018). The latter can be used to compare if a particular residue is evolving faster or slower in one clade in regard to another while normalizing over all the factors that are affecting the protein evolution. To our knowledge, RERs have not yet been used to help reveal sites under positive selection. However, RER at the gene or “genomic region” level has been successful in identifying, for instance, accelerated regions in humans that have been shown to be key in human evolution (Cui et al. 2025). These regions are conserved genomic loci across mammals that have experienced rapid nucleotide substitutions following the divergence from chimpanzees. Relative evolutionary rates have also been used to determine genes underlying continuous or categorical traits, testing the association of RER shifts with traits across the phylogeny (Kowalczyk et al. 2019; Redlich et al. 2024).

In this study, we aimed to evaluate the potential of contrasting RER at the site level to uncover adaptive genes. The working hypothesis of this study is that differences in the RER between clades reflect a different regime of purifying selection, which in turn can reflect a different protein site functionality that could have emerged through an adaptation process. In other words, while we cannot identify the specific amino acid residue replacement responsible for the adaptation, we can analyze the RER between clades to find its distinctive footprint in the purifying selection it triggered. To empirically test our hypothesis, we evaluated the performance of RER using a dataset of over 8,000 orthologous genes across 32 artiodactyl species with high-quality genome annotations. We compared the results obtained with RER and those obtained with the dN/dS method and found a highly significant number of genes and sites shared between these related but independent approaches. Importantly, we demonstrate that genes identified exclusively through RER are biologically meaningful, showing that RER-based analyses possess inherent utility rather than merely serving as a proxy for dN/dS.

2. Materials and methods

2.1 Data collection

We downloaded the coding and protein sequences for 32 artiodactyl species from NCBI GENOMES (Table S1). All the genomes were annotated by NCBI RefSeq. Given the extensive research on morphological, physiological, and genetic adaptations in cetaceans, we chose to focus on the Artiodactyla clade, with particular emphasis on Cetacea, as this would facilitate the interpretation of our results. The 32 artiodactyl species consisted of 17 cetaceans, 5 bovids, 4 caprids, 3 cervids, and one species each of musk deer, pig, and vicuña (Table S1).

2.2 Orthology analysis, curation and phylogeny

The downloaded protein sequence for each species was used to infer orthology relationships with Orthofinder2 (Emms and Kelly 2019; Emms et al. 2025). For this purpose, only the longest protein per gene was selected using custom Python scripts. For the subsequent analysis, we used only the corresponding coding sequences (cds) of each single-copy orthologous gene obtained from Orthofinder2. Since for our evolutionary analyses we needed to work with a mirror alignment of the cds and protein sequences of each gene, we first translated the cds using MACSE v2.03 (Ranwez et al. 2018). The translated cds were then aligned using MUSCLE v3.8.1551 with default parameters (Edgar 2004). Using these alignments, we aligned and trimmed the cds using TrimAl v1.4.rev15 (Capella-Gutiérrez et al. 2009) with the options --backtrans, --strict, and --splitbystopcodon. Since we still found alignment errors through visual inspection of the sequences, we further trimmed them with TrimAl, using the options -gt 1, -st 0.5, and -sw 10. This was a very strict trim that did not allow for gaps and eliminated regions with an average similarity below 0.5 within a window of ten amino acids. In addition, we further curated the alignments using BUSTED-E (Selberg et al. 2025) from HyPhy v. 2.5.70 (Kosakovsky Pond et al. 2020) to avoid false positives in our selection analysis using FEL (see below). BUSTED-E is able to identify branches under positive selection while identifying and adjusting for sites with abnormally high dN/dS values. After running BUSTED-E, we ran error-filter under HyPhy to mask the cds and protein sequences with misalignments.

After obtaining the final mirror alignment of the trimmed cds and protein sequences, the files were split into two: one containing only the sequences of the cetacean species (CETA group) and another alignment file containing the remaining species (TERR group).

Finally, a phylogenetic tree was constructed for Cetaceans and the remaining Artiodactyls according to McGowen et al., 2020 and Chen et al., 2019, respectively. The newick file was constructed using NewickTreeModifier (Ogoniak et al. 2024) and iTOL (Letunic and Bork 2024). The trees were used as input for BUSTED-E and the subsequent analyses (see below).

2.3 Contrasting relative substitution rates

Since substitution rates are affected by generation time, mutation rate, population history, and even phylogeny node density (Lanfear et al. 2010), it is necessary to study the relative evolutionary rate (RER) per site. To analyze the relative evolutionary rate on a per-amino

acid or per-site basis, we used LEISR (Spielman and Kosakovsky Pond 2018). LEISR, which is heavily inspired by RATE4PROTEIN (Pupko et al. 2002), estimates the relative evolutionary rate to determine which regions are more or less conserved along a protein while improving speed and efficiency. LEISR's per-site rate estimation is scaled to the average rate of the entire protein. Previous to the comparison, the RER estimations were z-score transformed.

Using LEISR, the RER was estimated separately for Cetacean (CETA) (Fig. 1a) and the remaining Artiodactyla (TERR) (Fig. 1b) for each protein site. All sites that showed a non-overlapping confidence interval for the RER estimation were considered significantly different between the clades. Under our working hypothesis, RER estimates should be equal between both clades, provided that neither experiences relaxed or positive selection. Importantly, the non-overlap criterion represents a conservative approach to assessing differences between groups (Schenker and Gentleman 2001). These sites were then compared with those obtained from the dN/dS test.

2.4 The dN/dS ratio test

The coding sequences, aligned using the protein alignment, were scanned for pervasive diversifying selection using FEL (Fixed Effects Likelihood) (Kosakovsky Pond and Frost 2005). FEL uses a maximum-likelihood (ML) approach to infer nonsynonymous (dN) and synonymous (dS) substitution rates on a per-site basis for a given coding alignment and its corresponding phylogeny. This method assumes that the selection pressure for each site is constant along the entire phylogeny. Sites with a p-value equal to or lower than 0.05 were considered under pervasive diversifying selection. FEL has been extensively used (Kosakovsky Pond and Frost 2005) and produces results similar to CODEML's random effects likelihood dN/dS approach (Yang 2007). We chose FEL over alternative approaches like MEME (Murrell et al. 2012) because it tests for pervasive rather than episodic selection. This approach aligns better with our per-site and per-clade RER estimations while yielding a more conservative subset of sites.

2.5 Evaluation of Relative Evolutionary Rate to infer adaptation events

To investigate whether RER differences between clades can uncover adaptive genes, we employed three analyses: (a) evaluating the overlap in significant genes and sites identified by RER and the dN/dS approach, (b) comparing the level of protein-protein interaction (PPI) among genes identified by both methods, and (c) assessing the biological relevance of genes uniquely identified by the RER approach.

For the first analysis, we hypothesized that the RER and dN/dS approaches would show a partial overlap in the genes and sites associated with cetacean aquatic adaptation. Shared genes and sites were defined as those exhibiting a significant RER difference between clades, where the clade with the faster RER also had a dN/dS > 1 (Fig. 1c, 1d). Statistical significance of shared sites was assessed via hypergeometric-based over-representation analysis.

For the second analysis, we hypothesized that if genes identified by the RER approach are biologically meaningful and driven primarily by positive selection, they should exhibit a level of co-functionality comparable to or greater than that of genes detected via dN/dS. To test

this, we used STRING (Szklarczyk et al. 2023) to construct and analyze protein-protein interaction (PPI) networks for seven distinct gene sets: (1) all genes identified exclusively via RER (Unique RER); (2) RER-exclusive genes where CETA sites evolved faster than TERR sites (Unique RER CETA); (3) RER-exclusive genes where TERR sites evolved faster than CETA (Unique RER TERR); (4) all genes exclusively identified as being under positive selection (dN/dS > 1) by FEL (Unique FEL); (5) FEL-exclusive genes showing dN/dS > 1 specifically in CETA (Unique FEL CETA); (6) FEL-exclusive genes showing dN/dS > 1 specifically in TERR (Unique FEL TERR); and (7) the genes shared between both analyses (Shared). The p-values associated with the PPI networks were obtained at medium and high confidence levels, by setting the minimum required interaction scores to 0.4 and 0.7, respectively.

3. Results and discussion

3.1 Contrasting RER sites between clades reveal genes and sites under diversifying selection.

We propose that a difference in the Relative Evolutionary Rate (RER) between clades for a given amino acid residue could be a consequence of an adaptive event. In order to evaluate this, we studied whether protein sites under pervasive diversifying selection according to the classical dN/dS test also show evidence of different RERs between clades. The dN/dS were estimated using FEL, while the RER were estimated using LEISR. The RER estimations were z-score transformed to evaluate which clade presented a faster or slower RER by comparing their confidence intervals.

To evaluate this framework, we used artiodactyls to compare terrestrial lineages against cetaceans. The phylogenies for the cetacean (CETA) and terrestrial (TERR) clades are shown in Figures 1a and 1b. More than 1,593,063 proteins were retrieved for the 32 artiodactyl species. After selecting one protein per gene, 8,197 single-copy orthologs were inferred with OrthoFinder2. Of these, 321 were discarded because they failed to run using LEISR or FEL due to insufficient information (i.e., short alignments, a high level of masked or unknown sites).

The dN/dS test of the CETA clade yielded 357 genes with pervasive diversifying selection (dN/dS > 1). On the other hand, 559 genes showed a higher RER in CETA than in TERR (Fig. 1c). Between the analyses, we found 95 shared genes, which implies a very significant over-representation (p-value < $1x10^{-31}$). Moreover, of the 95 shared genes, both approaches pinpointed the exact same site 50% of the time (p-value < $1x10^{-88}$) (Fig. 1c). The results for the TERR clade were more modest but still significant (Fig. 1d), with 36 shared genes (p-value < $1x10^{-5}$), for which both approaches shared 15 sites (p-value < $1x10^{-20}$). These results are clearly in line with our hypothesis, which predicts that the RER approach should be able to reveal genes that underwent an adaptive process.

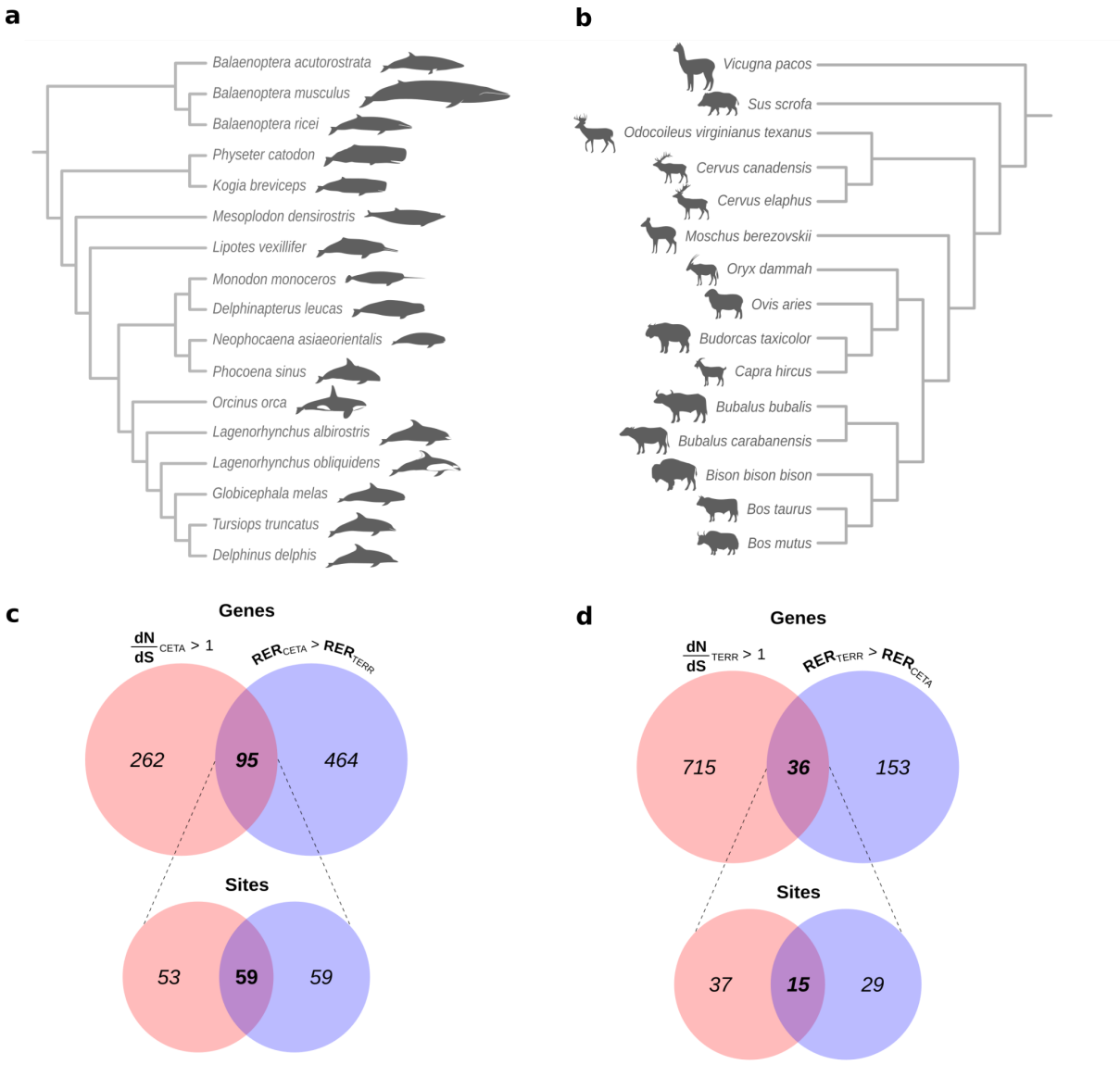


Fig 1. a. Phylogenetic tree used for the analyses of Cetaceans (CETA). **b.** Phylogenetic tree used for the analyses of Terrestrial Artiodactyls (TERR). **c.** Venn diagram of genes with a dN/dS > 1 in CETA versus genes with a higher RER at least one site in CETA compared to TERR. Lower left Venn diagram of shared sites between both approaches for the shared genes in CETA species. **d.** Venn diagram of genes with a dN/dS > 1 in TERR versus genes with a higher RER at least one site in TERR compared to CETA. Lower right: Venn diagram of shared sites between both approaches for the shared genes in TERR species. Calculations of RERs were made using LEISR while dN/dS was estimated using FEL.

The agreement between RER and dN/dS is non-trivial, as they represent independent tests. Additionally, because the RER approach does not consider synonymous substitutions, a higher dN/dS value for one clade does not imply a comparatively faster RER, and vice versa. Nevertheless, since non-synonymous nucleotide substitutions directly drive amino acid replacements, a partial concordance is to be expected. A previous study reported a moderate-to-high correlation between RER and inferred dN/dS (Sydykova and Wilke 2017). However, this does not imply that RER should differ between clades if only one clade is under positive selection (dN/dS > 1). Moreover, correlations were especially high for the inferred dN/dS under the unrealistic model of fixed dS rates across sites, and the magnitudes estimated by RER and dN/dS differed in several cases despite their high correlation (Sydykova and Wilke 2017). Thus, the agreement between these two approaches

provides new insight into how episodes of diversifying selection can drive shifts in the relative evolutionary rate of adapting clades for certain genes.

Although we found a significant over-representation of shared genes and sites for both clades, the signal for the TERR clade was less pronounced. This can be due to two main factors. First, CETA species are less divergent than TERR species; this likely causes FEL to infer more positively selected genes in TERR, while our RER framework identifies more acceleration in CETA, possibly due to skewed confidence intervals produced by few substitutions in conservative alignments. Secondly, the number of species per clade is small to intermediate, which could impact the statistical power and estimations of both RER and dN/dS. Nonetheless, our results clearly show that the RER approach is able to capture episodes of diversifying selection even with these limitations.

3.2 Unique RER have the most significant Protein-Protein Interaction

Positive selection is expected to change RER between clades for a given residual. However, differences in RER do not necessarily imply a past adaptative event because it is possible that relaxed selection could trigger RER differences between clades. Because complex phenotypic transitions—such as aquatic adaptation—typically recruit networks of functionally interacting genes, we reasoned that adaptive signals should reflect this co-functionality. Specifically, if the RER approach effectively captures aquatic adaptation, the unique set of genes it identifies (those not detected by dN/dS) should exhibit greater functional connectivity than expected by chance. To test this, we evaluated protein–protein interaction (PPI) networks across seven distinct gene sets derived from RER and dN/dS analyses (see Materials and Methods).

At medium confidence, almost all gene sets regardless of whether they were from RER or dN/dS, showed significant PPIs (Table 1). However, the two most significant networks (those with the lowest PPI associated p-values) were obtained by contrasting RER values. The most significant network was the set of Unique RER CETA, where CETA's RER values are greater than those in TERR's (p-value = 5.47×10^{-9}). This set was followed by the set Unique RER (p-value = 4.37×10^{-8}). The third most significant set is Unique FEL, composed of genes with dN/dS greater than 1 in both CETA and TERR (p-value = 4.26×10^{-5}). Only two sets were not significant: the Unique RER TERR and Unique FEL CETA. If only high-confidence interactions are considered, the same gene sets remain significant.

Under the RER framework, it is challenging to determine which clade experienced the adaptive event, making it less straightforward to define the target gene set. In other words, aquatic adaptation can trigger faster substitution rates in certain genes while promoting slower evolution in others. Consequently, using PPI networks for validation does not provide an entirely equivalent baseline, effectively penalizing RER-derived gene sets. Despite this inherent limitation, the unique gene sets identified by contrasting RERs are equally or more significant than those identified by the traditional dN/dS approach. These results strongly suggest that contrasting RERs hold intrinsic value for revealing genes involved in adaptive processes rather than merely serving as a proxy for dN/dS.

Table 1. PPI network enrichment analysis for unique and shared gene sets identified by both methods.

Gene set	No. Genes	PPI enrichment p-value medium confidence	PPI enrichment p-value high confidence
Unique RER	592	$4x10^{-8}$	$1x10^{-3}$
Unique RER CETA	464	$5x10^{-9}$	$3x10^{-3}$
Unique RER TERR	153	n.s.	n.s.
Unique FEL	928	$4x10^{-5}$	$3x10^{-2}$
Unique FEL CETA	262	n.s.	n.s.
Unique FEL TERR	715	$1x10^{-4}$	$2x10^{-2}$
Shared	95	$1x10^{-4}$	$3x10^{-2}$

A Protein-Protein Interaction (PPI) enrichment p-value indicates the statistical significance of the number of edges within a PPI network (n.s.: not significant). Although an overlap exists between the Unique RER CETA and Unique RER TERR gene sets, these shared genes contain distinct sites with elevated RER values in both cetacean and terrestrial lineages. Both datasets are mutually exclusive with respect to the genes detected by dN/dS analysis.

3.3 Biological Relevance of Unique RER Genes

To determine if the genes identified through the RER approach are biologically meaningful, we investigated whether they participate in known water-adaptation functions or if their related genes have been previously described as key to cetacean evolution.

Initially, we examined the genes shared between the FEL and RER analyses. Within the 95 shared genes, we identified several previously reported in cetacean adaptation, such as pulmonary surfactant-associated protein C (*SFTPC*), CDK5 regulatory subunit-associated protein 2 (*CDK5RAP2*), and Tumor necrosis factor receptor superfamily member 11B (*TNFRSF11B*), among other genes (Table S2). Notably, these genes had been previously detected using the dN/dS approach in CODEML (Table S2 and references therein). For instance, *SFTPC* encodes vital components of pulmonary surfactant, which is crucial for regulating alveolar surface tension and facilitating re-expansion during ascent, thereby mitigating lung squeeze injuries (Guo et al. 2024). These findings suggest that our FEL and RER results are consistent with established CODEML-based literature.

On the other hand, within the 592 unique genes identified solely by the RER approach, we found evidence suggesting their evolutionary relevance across four categories (Table 2). First, we recovered genes that had been identified in previous studies using CODEML, such as Carnitine O-palmitoyltransferase 2 (*CPT2*), Coagulation factor VII *(FVII precursor)*, Relaxin receptor 1 (*RXFP1*), and ATP-binding cassette sub-family A member 3 (*ABCA3)* (Table 2 and references therein). Second, we identified genes such as matrix metalloproteinase-14 isoform X1 (*MMP14*) and Aquaporin-5 (*AQP5*), for which paralogous genes have been previously associated with cetacean adaptation (Ramos et al. 2023; Zhang et al. 2025). In a third category, we identified the plasminogen gene (*PLG*), which belongs to

the same pathway as Plasminogen Activator, Urokinase (*PLAU*), a gene previously highlighted in cetacean evolution for its role in dissolving blood clots (Zhou et al. 2018).

Finally, a fourth category includes promising novel candidates for cetacean evolution. Among these are Endothelial PAS domain-containing protein 1 (*EPAS1)* and Neuronal PAS domain-containing protein 4 (*NPAS4*) (Fig. 2), bHLH-PAS transcription factors involved in the hypoxia response and neurogenesis (Edwards and Gorelick 2022), respectively. While *EPAS1* has shown footprints of positive selection in several species adapted to high altitudes (Wu et al. 2020), to our knowledge, this is the first time it has been associated with cetacean evolution. *NPAS4,* which is expressed only in neurons and regulates neuronal development and synaptic plasticity (Sun and Lin 2016; Edwards and Gorelick 2022) adds to the several nervous system genes that have shown evidence of adaptive evolution (McGowen et al. 2012). Other relevant genes include Thromboxane-A synthase (*TBXS1*), which is related to blood clotting (Huelsmann et al. 2019), and Cyclic nucleotide-gated cation channel beta-3 (*CNGB3*), involved in mammalian vision (Zheng et al. 2022)—a system that underwent significant transitions to function underwater. Additionally, we identified Lysocardiolipin acyltransferase 1 (*ALCAT1*), potentially involved in cardiac function (Bo et al. 2021), and Pulmonary surfactant-associated protein A (*SFTPA*), which, while not a paralog of *SFTPC*, is a collectin that forms part of the innate immune system (Crouch et al. 2000) and contributes to lung surfactant function (deMello et al. 1993).

Table 2. Characterization of unique RER genes associated with cetacean adaptation: Previously identified and candidate genes

Protein ID	Gene code	Gene name	Orthogroup name	Proposed Role	Bibliography	Reporting Method	Sites in RER analysis
NP_001233172.1	CPT2	Carnitine O-palmitoyltransferase 2	OG0006813	Fatty acid oxidation and energy metabolism	(Zhou et al. 2015)	dN/dS CODEML	625, higher on CETA
NP_001038056.1	FVII precursor	Coagulation factor VII precursor	OG0006104	Blood clotting	(Mariz and Nery 2020; Kang et al. 2024)	dN/dS CODEML and HyPhy	98, higher on CETA

XP_003356982.1	RXFP1	relaxin receptor 1 isoform X1	OG0008570	Pulmonary function	(Guo et al. 2024)	dN/dS CODEML	101, higher on CETA
XP_013843290.1	*ABCA3*	ATP-binding cassette sub-family A member 3	OG0010102	Lung development	(Guo et al. 2024)	dN/dS CODEML	965, higher on CETA
XP_020953962.1	MMP14	matrix metalloproteinase-14 isoform X1	OG0014085	Lung elasticity	Only for paralogous genes: (Zhang et al. 2025)	N/A	166, higher on CETA
NP_001103894.1	AQP5	Aquaporin-5	OG0006186	Predominantly present in lung alveolar type I cells. Regulate fluid transport; prevent pulmonary edema.	Only for paralogous genes and non marine cetaceans: (Ramos et al. 2023)	dN/dS Hyphy	210, higher on CETA 232, higher on CETA
NP_001038055.1	PLG precursor	plasminogen precursor	OG0006103	Zymogen of plasmin enzyme. Plasmin degrades fibrin clots and blood plasma proteins.	Only for related PLAU gene that acts in the same gene pathway: (Zhou et al. 2018)	N/A	281, higher on CETA 652, higher on TERR

NP_001090889.1	EPAS1 (also known as HIF2α)	endothelial PAS domain-containing protein 1	OG0006129	Hypoxia-inducible factor involved in response to oxygen concentration	N/A for cetacean adaptations. For high altitude adaptation: (Wu et al. 2020)	N/A	384, higher on CETA
XP_003122544.1	NPAS4	Neuronal PAS domain-containing protein 4	OG0007560	Immediate early basic helix-loop-helix (bHLH)-PAS transcription factor. Maximizes energy production by suppressing hypoxia-inducible factor 1α (HIF1α).	N/A	N/A	492, higher on CETA
NP_999211.1	TBXS1	thromboxane-A synthase	OG0007008	Induces vasoconstriction and promotes platelet aggregation	N/A	N/A	75, higher on CETA 181, higher on CETA
XP_003125633.3	CNG3	Cyclic nucleotide-gated cation channel beta-3	OG0007788	Ion channel. Vision.	N/A	N/A	36, higher on CETA
NP_001136317.1	ALCAT1	lysocardiolipin acyltransferase 1	OG0006324	Acyltransferase that catalyzes cardiolipin. Involved in cardiac function	N/A	N/A	117, higher on CETA 255, higher on TERR
NP_001335888.1	*SFTPA*	Pulmonary surfactant-associated protein A	OG0006928	Collectins part of the innate immune system. Found in the pulmonary surfactant in lungs.	N/A	N/A	61, higher on TERR

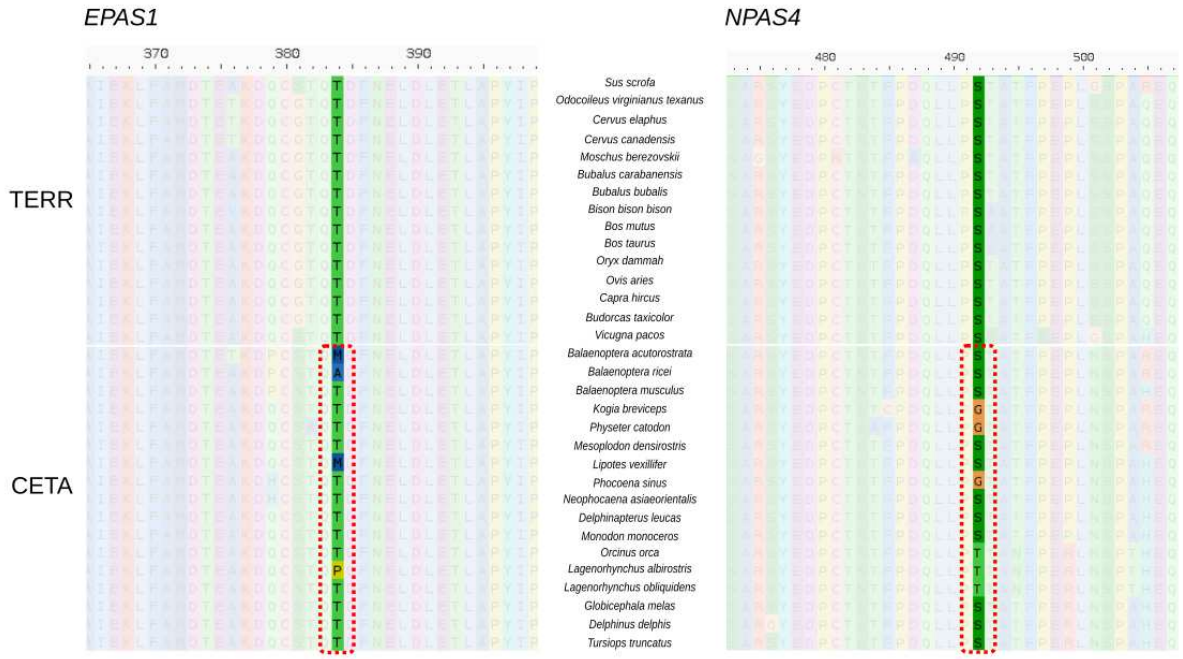


Fig. 2. Alignments of the *EPAS1* and *NPAS4* genes demonstrate a significantly higher RER for the CETA clade at sites 384 and 492, respectively. Based on z-score transformed LEISR estimations at these sites, the CETA clade exhibited higher, non-overlapping confidence intervals compared to the TERR clade.

By capturing both previously established candidates and novel genes with plausible roles in aquatic adaptation, our analysis demonstrates the potential of RER to assist in the detection of adaptive genes in cases that do not exhibit typical diversifying patterns. This highlights the potential of relative evolutionary rates to uncover a broader landscape of genetic adaptations.

Taken together, our results demonstrate that contrasting RERs across clades provides a valuable complement to dN/dS-based analyses, the traditional cornerstone for identifying sites under selection. However, we argue that conceiving the impact of selection in terms of RER offers distinct advantages over dN/dS. The definition of purifying selection—understood simply as selection that removes deleterious variants while allowing equivalent amino acids—is captured better with RER than with dN/dS. Under RER, purifying selection is conceived in a relative manner, and it is in this way that purifying selection and its consequences are comparable. Instead, when using dN/dS, the definition of purifying selection is based on both substitution rates which leads to biological inconsistencies when, for instance, dN is driven by positive selection but the dN/dS ratio is still below one.

Despite the importance of these findings, several limitations of this study should be noted. First, RER comparisons cannot detect an adaptive event if the evolutionary rate remains constant between clades, even if the preferred amino acids have changed. Second, RER comparisons cannot identify which specific clade experienced the adaptive event. Third, RER differences can arise from neutral processes like relaxed selection, such as a shift to more benign environments, rather than from adaptive positive selection. Finally, the high-throughput nature of these analyses required conservative filtering, potentially biasing evolutionary rate estimations, although such limitations are inherent to large-scale genomic datasets. Nevertheless, our results support the hypothesis that contrasting different regimes of purifying selection—here normalized using RER—can reliably identify past adaptive

events. Provided that future research focuses on addressing and ruling out relaxed selection, properly accounting for purifying selection offers a general strategy to detect departures from neutrality.

Author contributions

F.M.G.: Conceptualization, Methodology, Formal analysis, Investigation, Validation, Visualization, Writing – original draft, and Writing – review & editing. J.F.: Methodology, Visualization, Writing – original draft, and Writing – review & editing.

Data availability

The alignments (pre- and post-trimming) and the output files from the FEL and LEISR analyses generated during the current study are available in the Dryad repository, DOI: XXXX.

References

Barkman TJ, Martins TR, Sutton E, Stout JT (2007) Positive selection for single amino acid change promotes substrate discrimination of a plant volatile-producing enzyme. Mol Biol Evol 24:1320–1329. https://doi.org/10.1093/molbev/msm053

Bazykin GA, Kondrashov AS (2011) Detecting past positive selection through ongoing negative selection. Genome Biol Evol 3:1006–1013. https://doi.org/10.1093/gbe/evr086

Bo W, Ma Y, Xi Y, et al (2021) The Roles of FGF21 and ALCAT1 in Aerobic Exercise-Induced Cardioprotection of Postmyocardial Infarction Mice. Oxid Med Cell Longev 2021:8996482. https://doi.org/10.1155/2021/8996482

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Chen L, Qiu Q, Jiang Y, et al (2019) Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. Science 364:eaav6202. https://doi.org/10.1126/science.aav6202

Crouch E, Hartshorn K, Ofek I (2000) Collectins and pulmonary innate immunity. Immunol Rev 173:52–65. https://doi.org/10.1034/j.1600-065x.2000.917311.x

Cui X, Yang H, Cai C, et al (2025) Comparative characterization of human accelerated regions in neurons. Nature 640:991–999. https://doi.org/10.1038/s41586-025-08622-x

deMello DE, Heyman S, Phelps DS, Floros J (1993) Immunogold localization of SP-A in lungs of infants dying from respiratory distress syndrome. Am J Pathol 142:1631–1640

Echave J, Spielman SJ, Wilke CO (2016) Causes of evolutionary rate variation among protein sites. Nat Rev Genet 17:109–121. https://doi.org/10.1038/nrg.2015.18

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340

Edwards HE, Gorelick DA (2022) The evolution and structure/function of bHLH–PAS transcription factor family. Biochem Soc Trans 50:1227–1243. https://doi.org/10.1042/BST20211225

Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20:238. https://doi.org/10.1186/s13059-019-1832-y

Emms DM, Liu Y, Belcher L, et al (2025) OrthoFinder: scalable phylogenetic orthology inference for comparative genomics. 2025.07.15.664860

Farias J, Giorello FM (2020) Positive Selection in the Chloroplastic ATP-Synthase β-Subunit and Its Relation to Virulence Factors. J Mol Evol 88:703–713. https://doi.org/10.1007/s00239-020-09968-8

Gladieux P, Devier B, Aguileta G, et al (2013) Purifying selection after episodes of recurrent adaptive diversification in fungal pathogens. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis 17:123–131. https://doi.org/10.1016/j.meegid.2013.03.012

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11:725–736. https://doi.org/10.1093/oxfordjournals.molbev.a040153

Guo B, Sun Y, Wang Y, et al (2024) Evolutionary genetics of pulmonary anatomical adaptations in deep-diving cetaceans. BMC Genomics 25:339. https://doi.org/10.1186/s12864-024-10263-9

Huelsmann M, Hecker N, Springer MS, et al (2019) Genes lost during the transition from land to water in cetaceans highlight genomic changes associated with aquatic adaptations. Sci Adv 5:eaaw6671. https://doi.org/10.1126/sciadv.aaw6671

Kang J, Gu L, Guo B, et al (2024) Molecular evolution of wound healing-related genes during cetacean secondary aquatic adaptation. Integr Zool 19:898–912. https://doi.org/10.1111/1749-4877.12781

Kosakovsky Pond SL, Frost SDW (2005) Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. Mol Biol Evol 22:1208–1222. https://doi.org/10.1093/molbev/msi105

Kosakovsky Pond SL, Poon AFY, Velazquez R, et al (2020) HyPhy 2.5-A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. Mol Biol Evol 37:295–299. https://doi.org/10.1093/molbev/msz197

Kowalczyk A, Meyer WK, Partha R, et al (2019) RERconverge: an R package for associating evolutionary rates with convergent traits. Bioinformatics 35:4815–4817. https://doi.org/10.1093/bioinformatics/btz468

Lanfear R, Welch JJ, Bromham L (2010) Watching the clock: studying variation in rates of molecular evolution between species. Trends Ecol Evol 25:495–503. https://doi.org/10.1016/j.tree.2010.06.007

Letunic I, Bork P (2024) Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res 52:W78–W82. https://doi.org/10.1093/nar/gkae268

Mariz JPV, Nery MF (2020) Unraveling the Molecular Evolution of Blood Coagulation Genes in Fishes and Cetaceans. Front Mar Sci 7:. https://doi.org/10.3389/fmars.2020.592383

McGowen MR, Grossman LI, Wildman DE (2012) Dolphin genome provides evidence for adaptive evolution of nervous system genes and a molecular rate slowdown. Proc Biol Sci 279:3643–3651. https://doi.org/10.1098/rspb.2012.0869

McGowen MR, Tsagkogeorga G, Álvarez-Carretero S, et al (2020) Phylogenomic Resolution of the Cetacean Tree of Life Using Target Sequence Capture. Syst Biol 69:479–501. https://doi.org/10.1093/sysbio/syz068

Murrell B, Wertheim JO, Moola S, et al (2012) Detecting Individual Sites Subject to Episodic Diversifying Selection. PLOS Genet 8:e1002764. https://doi.org/10.1371/journal.pgen.1002764

Nei M, Suzuki Y, Nozawa M (2010) The neutral theory of molecular evolution in the genomic era. Annu Rev Genomics Hum Genet 11:265–289. https://doi.org/10.1146/annurev-genom-082908-150129

Ogoniak L, Steffen R, Grundmann N, et al (2024) NewickTreeModifier: A simple web tool to prune and modify Newick trees. J Hered 115:183–187. https://doi.org/10.1093/jhered/esae005

Oleksyk TK, Smith MW, O'Brien SJ (2010) Genome-wide scans for footprints of natural selection. Philos Trans R Soc B Biol Sci 365:185–205. https://doi.org/10.1098/rstb.2009.0219

Pupko T, Bell RE, Mayrose I, et al (2002) Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. Bioinformatics 18 Suppl 1:S71-77. https://doi.org/10.1093/bioinformatics/18.suppl_1.s71

Rahman S, Kosakovsky Pond SL, Webb A, Hey J (2021) Weak selection on synonymous codons substantially inflates dN/dS estimates in bacteria. Proc Natl Acad Sci U S A 118:e2023575118. https://doi.org/10.1073/pnas.2023575118

Ramos E, Selleghin-Veiga G, Magpali L, et al (2023) Molecular Footprints on Osmoregulation-Related Genes Associated with Freshwater Colonization by Cetaceans and Sirenians. J Mol Evol 91:865–881. https://doi.org/10.1007/s00239-023-10141-0

Ranwez V, Douzery EJP, Cambon C, et al (2018) MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. Mol Biol Evol 35:2582–2584. https://doi.org/10.1093/molbev/msy159

Redlich R, Kowalczyk A, Tene M, et al (2024) RERconverge Expansion: Using Relative Evolutionary Rates to Study Complex Categorical Trait Evolution. Mol Biol Evol 41:msae210. https://doi.org/10.1093/molbev/msae210

Schenker N, Gentleman JF (2001) On Judging the Significance of Differences by Examining the Overlap Between Confidence Intervals. Am Stat 55:182–186. https://doi.org/10.1198/000313001317097960

Selberg A, Clark NL, Sackton TB, et al (2025) Minus the Error: Testing for Positive Selection in the Presence of Residual Alignment Errors. BioRxiv Prepr Serv Biol 2024.11.13.620707. https://doi.org/10.1101/2024.11.13.620707

Spielman SJ, Kosakovsky Pond SL (2018) Relative evolutionary rate inference in HyPhy with LEISR. PeerJ 6:e4339. https://doi.org/10.7717/peerj.4339

Sun X, Lin Y (2016) Npas4: Linking Neuronal Activity to Memory. Trends Neurosci 39:264–275. https://doi.org/10.1016/j.tins.2016.02.003

Sydykova DK, Wilke CO (2017) Calculating site-specific evolutionary rates at the amino-acid or codon level yields similar rate estimates. PeerJ 5:e3391. https://doi.org/10.7717/peerj.3391

Szklarczyk D, Kirsch R, Koutrouli M, et al (2023) The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any

sequenced genome of interest. Nucleic Acids Res 51:D638–D646. https://doi.org/10.1093/nar/gkac1000

Uzzell T, Corbin KW (1971) Fitting discrete probability distributions to evolutionary events. Science 172:1089–1096. https://doi.org/10.1126/science.172.3988.1089

Wu D-D, Yang C-P, Wang M-S, et al (2020) Convergent genomic signatures of high-altitude adaptation among domestic mammals. Natl Sci Rev 7:952–963. https://doi.org/10.1093/nsr/nwz213

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591. https://doi.org/10.1093/molbev/msm088

Zhang Y, Lv W, Yan W, et al (2025) Molecular adaptations in MMP genes support lung elasticity and diving adaptations in cetaceans. BMC Genomics 26:562. https://doi.org/10.1186/s12864-025-11751-2

Zheng X, Hu Z, Li H, Yang J (2022) Structure of the human cone photoreceptor cyclic nucleotide-gated channel. Nat Struct Mol Biol 29:40–46. https://doi.org/10.1038/s41594-021-00699-y

Zhou X, Guang X, Sun D, et al (2018) Population genomics of finless porpoises reveal an incipient cetacean species adapted to freshwater. Nat Commun 9:1276. https://doi.org/10.1038/s41467-018-03722-x

Zhou X, Seim I, Gladyshev VN (2015) Convergent evolution of marine mammals is associated with distinct substitutions in common genes. Sci Rep 5:16550. https://doi.org/10.1038/srep16550

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1SupplementaryMaterial.xlsx
- TableS2SupplementaryMaterial.xlsx