



## OPEN ACCESS

## EDITED BY

Sudhindra R. Gadagkar,  
Midwestern University, United States

## REVIEWED BY

Raghavendran Partha,  
Independent Researcher, Menlo Park, CA,  
United States  
Marian Novotny,  
Charles University, Czechia

## \*CORRESPONDENCE

Alexander G. Lucaci,  
✉ agl4001@med.cornell.edu  
Sergei L. Kosakovsky Pond,  
✉ spond@temple.edu

RECEIVED 03 February 2024

ACCEPTED 28 May 2024

PUBLISHED 24 June 2024

## CITATION

Lucaci AG, Brew WE, Lamanna J, Selberg A,  
Carnevale V, Moore AR and Kosakovsky Pond SL  
(2024), The evolution of mammalian Rem2:  
unraveling the impact of purifying selection and  
coevolution on protein function, and  
implications for human disorders.  
*Front. Bioinform.* 4:1381540.  
doi: 10.3389/fbinf.2024.1381540

## COPYRIGHT

© 2024 Lucaci, Brew, Lamanna, Selberg,  
Carnevale, Moore and Kosakovsky Pond. This is  
an open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](#). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic practice.  
No use, distribution or reproduction is  
permitted which does not comply with these  
terms.

# The evolution of mammalian Rem2: unraveling the impact of purifying selection and coevolution on protein function, and implications for human disorders

Alexander G. Lucaci<sup>1,2\*</sup>, William E. Brew<sup>3</sup>, Jason Lamanna<sup>3,4</sup>,  
Avery Selberg<sup>3,5</sup>, Vincenzo Carnevale<sup>3,4</sup>, Anna R. Moore<sup>3</sup> and  
Sergei L. Kosakovsky Pond<sup>3,5\*</sup>

<sup>1</sup>Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, United States, <sup>2</sup>Weill Cornell Medicine, The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, New York, NY, United States, <sup>3</sup>Department of Biology, Temple University, Philadelphia, PA, United States, <sup>4</sup>Institute for Computational Molecular Science, Temple University, Philadelphia, PA, United States, <sup>5</sup>Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA, United States

Rad And Gem-Like GTP-Binding Protein 2 (Rem2), a member of the RGK family of Ras-like GTPases, is implicated in Huntington's disease and Long QT Syndrome and is highly expressed in the brain and endocrine cells. We examine the evolutionary history of Rem2 identified in various mammalian species, focusing on the role of purifying selection and coevolution in shaping its sequence and protein structural constraints. Our analysis of Rem2 sequences across 175 mammalian species found evidence for strong purifying selection in 70% of non-invariant codon sites which is characteristic of essential proteins that play critical roles in biological processes and is consistent with Rem2's role in the regulation of neuronal development and function. We inferred epistatic effects in 50 pairs of codon sites in Rem2, some of which are predicted to have deleterious effects on human health. Additionally, we reconstructed the ancestral evolutionary history of mammalian Rem2 using protein structure prediction of extinct and extant sequences which revealed the dynamics of how substitutions that change the gene sequence of Rem2 can impact protein structure in variable regions while maintaining core functional mechanisms. By understanding the selective pressures, protein- and gene - interactions that have shaped the sequence and structure of the Rem2 protein, we gain a stronger understanding of its biological and functional constraints.

## KEYWORDS

molecular evolution, Rem2, Ras-like GTPases, Huntington's disease, long QT syndrome, Timothy syndrome

## Introduction

Brain development is a highly orchestrated sequence of events involving a multitude of molecular pathways. Disruptions of this sequence of events lead to impairments in brain function and may cause disease states as seen in neurological disorders such as autism spectrum disorder, Alzheimer's disease, Huntington's disease, and Parkinson's disease (Kwakowsky et al., 2023). Recent studies have highlighted the importance of understanding regulatory proteins and their roles in brain function and development. Key proteins and genes associated with neurological diseases, including those involved in synaptogenesis, neuronal architecture, and neuronal function, are of particular interest (Kwakowsky et al., 2023; Allen et al., 2016). Rem2 is a key activity-regulated gene important for synapse development and function (Ghiretti et al., 2011; 2014; Moore et al., 2013; 2018), dendritic complexity (Ghiretti et al., 2014; 2011), and intrinsic plasticity (Moore et al., 2018). Understanding the evolutionary history of Rem2, in combination with its cellular function, is critical for identifying potential therapeutic targets in disease states.

Rad and Gem-Like GTP-Binding Protein 2 (Rem2), encodes a small Ras-like GTPase within the RGK family which includes Rem, Rad, Gem and Kir proteins (Finlin et al., 2000). Since its initial discovery as a potent inhibitor of calcium channel activity (Béguin et al., 2001; Correll et al., 2008; Flynn et al., 2010), Rem2 has been identified as one of the key RGK members expressed in the brain. Rem2 is expressed within regions of the hippocampus and basal ganglia (Liput et al., 2016) and functions as an activity-dependent (Flynn et al., 2010) regulator of synapse formation, dendritic complexity, and intrinsic excitability (Ghiretti et al., 2011; Moore et al., 2013; Ghiretti et al., 2014; Moore et al., 2018). Additionally, at the cellular level, regulatory interactions have been reported between Rem2 and calmodulin (Béguin et al., 2005), L-type and N-type voltage gated calcium channels (VGCC) (Correll et al., 2008; Flynn et al., 2008), and calmodulin-dependent protein kinase II (CaMKII) (Ghiretti et al., 2013; Royer et al., 2018). Rem2 also modulates the activity of other small GTPases such as Rac1 and RhoA (Gauthier-Rouvière, 1998), which are themselves important regulators of the actin cytoskeleton and neuronal morphogenesis.

Rem2's involvement with key cellular pathways for neuronal function, its regulation on neuronal architecture, excitability, and high expression in the hippocampus (a brain region known to be associated with memory and learning) (Meck et al., 2013), implicate Rem2 as a target gene of interest for understanding neuronal health and disease states. Recently, mutations in Rem2 were identified in patients with Huntington's Disease (Nahalka, 2019) and Long QT Syndrome contributing to channelopathies [Chai et al., 2018]. Another member of the RGK family, Gem, which displays similar roles in neuronal function as Rem2 has been implicated in Timothy Syndrome [Krey et al., 2013; Boczek et al., 2015]. Thus, a further understanding of Rem2 gene evolution and function will provide deeper insight into the cellular role of the RGK subfamily and calcium channel interactions.

Codon substitution models are a workhorse of the modern field of molecular evolution and are used to assess the selective forces which have influenced the evolutionary dynamics of proteins. These analyses typically involve the estimation of the omega parameter ( $\omega$ ), which is the ratio of non-synonymous ( $\beta$  or dN) to synonymous

( $\alpha$  or dS) substitution rates, expressed as  $\omega = \beta/\alpha$  (Muse and Gaut, 1994). Non-synonymous changes may significantly affect a protein's structure and function. Conversely, synonymous changes, while not altering the amino acid at a specific site, can still subtly influence fitness through factors like codon usage bias, translation efficiency, or mRNA structural stability (Rahman et al., 2021). Despite the potential confounding of results from non-neutral synonymous changes, the rate of synonymous changes is assumed to be the rate of neutral evolution acting upon coding sequences and serves as a baseline against which non-synonymous evolutionary rates can be compared in most codon substitution models, including the kinds used in our analysis. The estimates of  $\omega$  quantify the selective pressure acting upon protein-coding genes and can be readily interpreted to discern the type of selective modality at play (e.g., positive diversifying selection with  $\omega > 1$ , negative purifying selection with  $\omega < 1$ , or neutral evolution when  $\omega$  is not significantly different from 1). Purifying selection acts to remove multiple types of variation from a population, including deleterious mutations. This process is important in maintaining the functional integrity of genes and preventing the accumulation of mutations that could lead to loss of function or disease. In the case of Rem2, this may be related to its role in neuronal development and function, such as GTP and calmodulin binding which have shown to be essential for regulating dendritic complexity (Ghiretti et al., 2011; Correll et al., 2008). Additionally, coevolution, i.e., interaction between two or more sites in a gene, may impact the evolution of a gene and play a role in constricting its functions. In the case of Rem2, coevolution could play a role in shaping the genetic interactions that regulate its function through its functional domains such as RGK.

Understanding the evolutionary history of Rem2 would offer insights into the processes that have shaped this gene and its function, including its relationship with its functional partners in biological networks. Puhl et al. (2014), investigated the evolutionary history of RGK proteins, specifically their role in modulating voltage-gated calcium channels (VGCCs). Their study revealed that the interaction between RGK proteins and VGCCs, critical for regulating calcium ion flow across cell membranes, dates to over 600 million years ago, before the divergence of protostomes and deuterostomes. The functional relationship has been conserved through Rem2's evolutionary history, despite the diverse physiological roles RGK proteins now play in various species and highlights specific structural features of RGK proteins essential for VGCC modulation and suggests that understanding these ancient protein functions can provide significant insights into their contemporary biological roles and implications for human health and disease. Our study uses a comparative genomic approach with the molecular sequences of Rem2 in different species to identify patterns of conservation and divergence. An evolutionary study of Rem2 may shed light on the broader properties and dynamics of the RGK family of small GTPases and their roles in regulating cellular functions, as Rem2 and Gem are the only two RGK family members expressed in the brain (Correll et al., 2008). Exploring potential ancestral functions of Rem2 may illuminate our current understanding of its role in regulating neuronal development and function. Additionally, the evolutionary trajectory of Rem2 may be influenced not only by the gene itself but also by its interactions with other genes. Therefore, the results from this work could be applied to

future studies to aid in the understanding of the molecular mechanisms underlying the genetic networks of neurological disorders associated with Rem2 mutations and to identify potential targets for therapeutic intervention.

## Methods

### Data retrieval and quality control

In this study, we queried the NCBI database for the *Rem2* gene using the following URL, <https://www.ncbi.nlm.nih.gov/gene/161253/ortholog?scope=1338369&term=Rem2> (last accessed December 2022). As our primary interest in this study is to explore the evolutionary history of *Rem2* in *Mammalia*, we limited our search to only include species from this taxonomic group. This query returned 186 full gene sets (transcripts and protein sequences) with one sequence per species. We downloaded all available files: Reference sequence (RefSeq) protein sequences, RefSeq transcript sequences, and Tabular data (CSV) which contains sequence metadata. In [Supplementary Table S1](#), we provide information of the species included in this analysis. We used the full protein sequence and full gene transcript files (which include 5' and 3' UTR) to retrieve coding sequences (CDS). Our process also removed low-quality protein sequences (9 sequences) from analysis as these sequences may bias or contaminate our results and analyses. For example, these sequences had invalid features such as incompletely resolved codons which result in unresolved amino acids and were subsequently excluded from our analysis.

### Sequence alignment and filtering

We generated codon-aware alignments for the filtered set of protein sequences by following the procedure available at the codon-msa GitHub repository ([github.com/veg/hyphy-analyses/tree/master/codon-msa](https://github.com/veg/hyphy-analyses/tree/master/codon-msa)). We used default parameters for the pre-msa step and used the coding sequence from Human Rem2 reference (“NM.173527.3”) sequence as part of a reference-guided alignment. We extracted and relied on the Human *Rem2* coding sequence (see [Supplementary Table S1](#) for accession) as the reference sequence for a reference-based alignment approach. In-frame nucleotide sequences were translated then the protein sequences were aligned with Multiple Alignment using Fast Fourier Transform (MAFFT) v7.505 (Kato and Standley, 2013), and then were mapped back to their respective codons. In order to raise our confidence in the quality of our MSA, we applied the “find-outliers” <https://github.com/veg/hyphy-analyses/tree/master/find-outliers> algorithm to our data, using default parameters. Briefly, the “find-outliers” script parses a Single-Linkage Ancestor Counting (SLAC) (Kosakovsky Pond and Frost, 2005) results JSON-formatted file, which contains results from ancestral sequence reconstruction and [Selberg et al., 2021] iterates over the sequences with a sliding window and examines the number of inferred multiple-nucleotide substitutions and masks these problematic codons. Masked

codon sites (“---”) improve the quality of our alignment and limit the spurious effects of molecular sequences from low-quality genome assemblies.

### Phylogenetic inference, annotation, and visualization

For our multiple sequence alignment of Rem2 sequences, we used IQ-TREE v2.2.0.3 (Minh et al., 2020) to perform maximum likelihood (ML) phylogenetic inference with 100 bootstrap replicates. To annotate our phylogenetic tree for various taxonomic groups, we employed the software found at <https://github.com/veg/hyphy-analyses/tree/master/LabelTrees>. To visualize our phylogenetic tree, we utilized the tree viewer available at <https://phyloree.hyphy.org> (Shank et al., 2018). Species tree of the Rem2 orthologs was generated using the NCBI Taxonomy database, common tree webpage at <https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi>.

### Recombination detection

Genetic Algorithm for Recombination Detection (GARD) (Kosakovsky Pond et al., 2006) is a method to screen a multiple sequence analysis for the presence of genetic recombination as a pre-processing step for evolutionary rate inference. The GARD method works by screening an alignment for putative recombination breakpoints and inferring a unique phylogenetic history for each detected non-recombinant block. GARD will search the space of all informative locations for significant changes in the tree topology across the alignment, inferring phylogenies for each putatively nonrecombinant fragment of the alignment. Goodness of fit measures are used to determine the number of optimally placed breakpoints by Akaike Information Criterion corrected for small sample size (cAIC) (Hurvich and Tsai (1989)). When recombination is not accounted for in downstream analyses, many such analyses, including selection analysis and rate estimation may become seriously biased (Fares, 2014).

### dN/dS estimation

The multiple sequence alignment and its accompanying unrooted phylogenetic tree was examined with a suite of molecular evolutionary methods, each designed to ask and answer a specific biological question, described below. We performed multiple-test correction via the false discovery rate (FDR) and reported both adjusted and non-adjusted *p*-values. All selection analyses were performed in version 2.5.40 of HyPhy (Kosakovsky Pond et al., 2020), our set of selection analyses includes the following tests (Table 1):

We applied the BUSTED model testing and averaging procedure to select the best fitting model for episodic diversifying selection, and to interpret the results of natural selection acting on Rem2. Our goal is to understand which underlying model and its parameters can detect the areas of the dataset which drive the greatest degree of evolutionary signals. Analysis is conducted as a series of tests in the

TABLE 1 An overview of the selection analyses used in this study. All methods are implemented in the HyPhy software suite.

#	Method	Biological question	Reference
1	FEL	Locates codon sites with evidence of pervasive positive diversifying or negative selection	Kosakovsky Pond and Frost (2005)
2	BUSTED (+S + MH)	Tests for gene-wide episodic selection while accounting for (optionally) synonymous rate variation (SRV) and multiple instantaneous substitutions (MH)	Murrell et al., 2015, Wisotsky et al., 2020 Lucaci et al. (2023)
3	MEME	Locates codon sites with evidence of episodic positive diversifying selection	Murrell et al. (2012)
4	SLAC	Performs substitution mapping using maximum likelihood ancestral sequence reconstruction	Kosakovsky Pond and Frost (2005)
5	BGM	Identifies groups of sites that are coevolving	Poon et al. (2008)
6	FMM	Examines model fit by permitting multiple instantaneous substitutions	Lucaci et al. (2021)

BUSTED framework of selection analysis and includes the baseline BUSTED method, and extensions to include synonymous rate variation (+S), multiple nucleotide substitutions (+MH), or both (+S + MH). A Snakemake (Mölder et al., 2021) version of the model testing procedure is available at [https://github.com/veg/BUSTED\\_ModelTest](https://github.com/veg/BUSTED_ModelTest).

## Structural assessment of proteins using ancestral sequence reconstruction (SPA)

The SPA analysis is designed to combine ancestral sequence reconstruction of a multiple sequence alignment of extant species with protein structural prediction done with AlphaFold2 (Jumper et al., 2021). It accomplishes this through the following series of steps:

First, we utilize the SLAC JSON (Kosakovsky Pond and Frost, 2005) output file from HyPhy, which generates ancestral state inference of extinct sequences. Our custom Python script “ancestralevolution.py” is used to parse and organize both extant and extinct coding sequences into a multifasta file. Subsequently, the Multifasta file is parsed into separate FASTA files, each containing a single amino acid sequence per FASTA file.

Next, we execute the ColabFold tool (Mirdita et al., 2022) available at <https://github.com/sokrypton/ColabFold>. Colabfold is a user-friendly implementation of AlphaFold2 (Jumper et al., 2021) that has been adapted to run efficiently in Google Colab and provides a way to predict protein structures based on amino-acid sequences. In batch mode, we utilize the AlphaFold2\_batch.ipynb notebook; the tool is executed with default settings, except for enabling the “Zip\_results” option.

Following the execution, PDB files are extracted, out of the five models used only the best PDB, for each species, with a rank of 1 is selected for further analysis.

To compare our results in a combinatorial fashion we ran TM-Align (Zhang and Skolnick, 2005) available at <https://anaconda.org/bioconda/tmalign> to generate TM-scores, a metric utilized to evaluate the degree of topological similarity among protein structures.

Genetic distance via the Tamura-Nei, 1993 (TN93) distance was calculated using a straightforward implementation from <https://github.com/veg/tn93> (Tamura and Nei, 1993).

The PDB structures were visualized in PyMOL version 2.5.4, the following color settings were applied: set\_color n0, [0.051, 0.341, 0.827], set\_color n1, [0.416, 0.796, 0.945], set\_color n2, [0.996, 0.851,

0.212], set\_color n3, [0.992, 0.490, 0.302], color n0, b < 100; color n1, b < 90, color n2, b < 70; color n3, b < 50.

## Software and data availability

All software, raw data, and full results, including all HyPhy selection analyzes JSON-formatted result files, used in this study is freely available via a dedicated GitHub repository at: <https://github.com/aglucaci/AOC-Rem2>.

All SPA related software and data used is available at the dedicated GitHub repository: <https://github.com/aglucaci/SPA-REM2>.

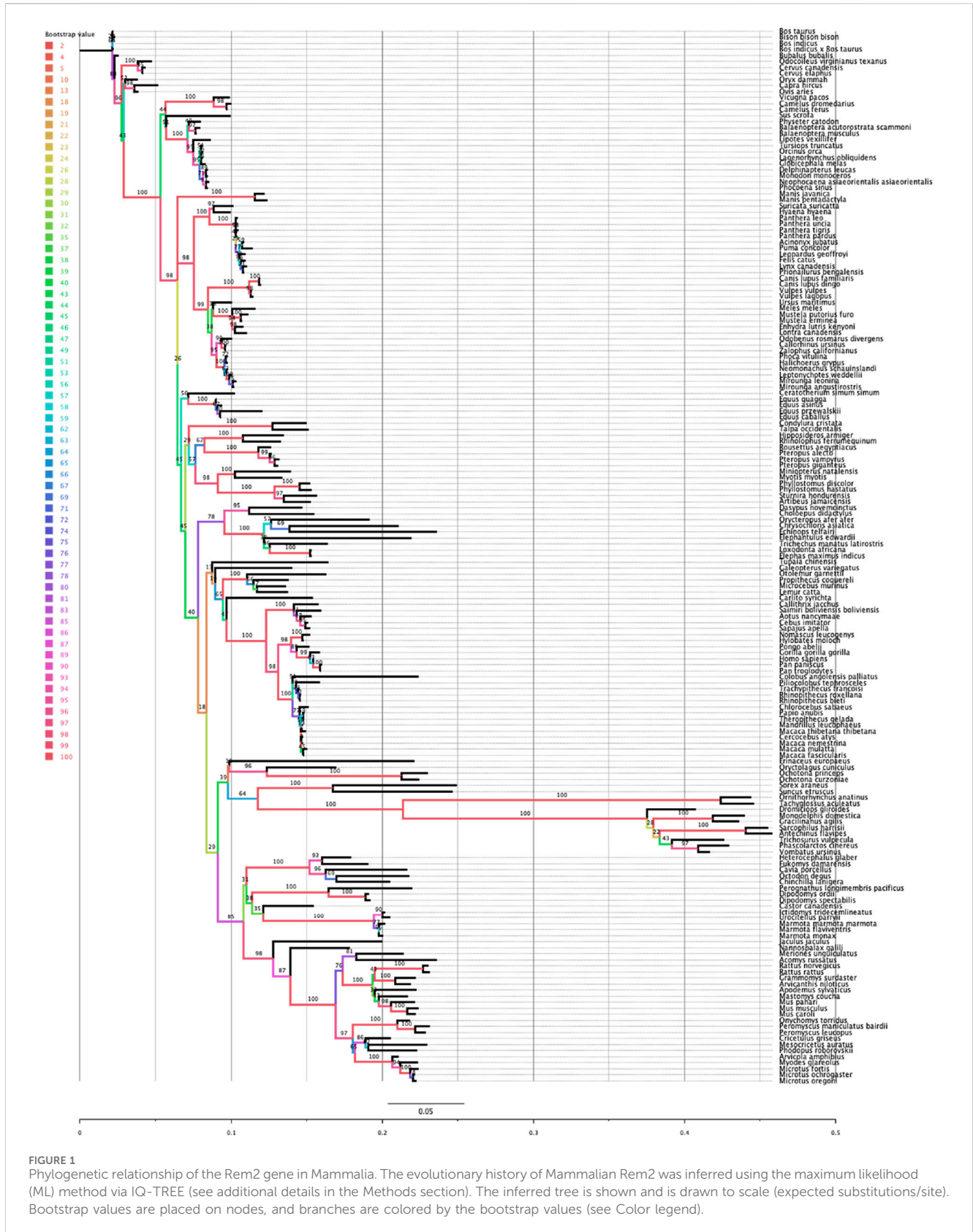
## Results

We estimated the selective pressures that have shaped the evolution of mammalian *Rem2* gene across time. Our analysis is based on 175 mammalian species used to construct a multiple sequence alignment (see Supplementary Table S1 for the full list of species and accessions used). The general structure of the Rem2 protein consists of a proximal intrinsically disordered region (IDR) and a distal main functional domain, belonging to the Rem, Rem2, Rad, Gem/Kir (RGK) subfamily of Ras-like GTPases, and is also flanked by another terminal disordered region. We find evidence in 70% of non-invariant codon sites that purifying selection has played a significant role in shaping the evolutionary history of mammalian *Rem2* using the FEL analysis (described in Methods section). These results suggest that strict regulation of *Rem2* may be important for maintaining genotypic fitness.

In addition, we also observed 251 out of 627 (or 40%) sites to be invariant at the nucleotide level, suggesting a highly conserved evolutionary importance for these loci. Evidence of coevolution was also found among 50 pairs of sites, some of which result in complex spatial relationships and potentially lend themselves to epistatic interactions within the protein. Where available, we relate several sites of interest in our results to relevant literature for important considerations for human health.

## Phylogenetic relationship of mammalian Rem2

*Rem2* is an ancient gene with molecular sequences available (at the time of this writing) for *Dipnotetrapodomorpha* which includes



tetrapods and lungfish (Hollis et al., 2012). However, the focus of this study is on Rem2’s preservation throughout the mammalian lineage, where it has been exposed to all significant evolutionary

events of the past ~250–200 million years and has played a role in a diverse and dominant terrestrial animal group (Correll et al., 2008; Edel et al., 2010; DeRocher et al., 2014). This includes all major

geological periods including earth impacts (asteroids), oxygenation level changes across time, atmospheric carbon dioxide concentrations, and changes in solar luminosity (radiation flux). A key step to frame the relative relationship of our analysis of mammalian *Rem2* is the inference of a phylogenetic tree (see Figure 1). Our results indicate that even under significant evolutionary pressures over millions of years, that *Rem2* has not been drastically modified in the mammalian lineage or in several of the taxonomic groups contained within *Mammalia* including Glires (rodents and lagomorphs), Eulipotyphla (includes hedgehogs, moles, shrews), Primates, Perissodactyla (odd-toed ungulates), Chiroptera (bats), Artiodactyla (even-toed ungulates), Carnivora. In fact, consistent with other findings (Hutter et al., 2000; Hamel et al., 2006; Anderson et al., 2016; Bennedorf et al., 2022; Lucaci et al., 2022), *Rem2* has alignment characteristics consistent with ancient genes that tend to be highly conserved and therefore are more likely to exhibit an abundance of purifying selective changes as opposed to severe adaptive evolutionary changes (Karczewski et al., 2020). Indeed, genes that undergo strong purifying selection, exhibiting minimal evolutionary changes, are more prone to being associated with severe human diseases (Lek et al., 2016). This could be explained by the similar biological roles and fitness effects that *Rem2* has in our groups under study.

Comparing the species tree of our *Rem2* orthologs and the *Rem2* gene tree using the Robinson-Foulds (RF) distance, we quantified the topological differences between the two trees to understand how well the *Rem2* gene tree reflects the evolutionary relationships represented in the species tree (Supplementary Figure S7). In our analysis, we obtained an RF distance of 122, with a maximum RF distance of 228, resulting in a normalized RF distance of 0.535. This normalized score of approximately 0.535 indicates a moderate level of dissimilarity between the species tree and the gene tree. Specifically, over half of the bipartitions present in one tree are not found in the other, suggesting that the gene tree deviates significantly from the species tree, as this divergence could be due to various biological factors such as incomplete lineage sorting, or differing rates of evolution among lineages. While the RF distance analysis helps to highlight these topological discrepancies, further analysis might be required to pinpoint the underlying causes of these differences and to assess the reliability and biological significance of the observed gene tree.

## No evidence of gene-wide evolution in *Rem2*

The conservation in the nucleotide sequences among mammalian *Rem2* and its various taxonomic subgroups is indicative of a strong purifying selective pressure acting on the *Rem2* gene. To examine evidence for gene-wide episodic diversifying selection (EDS), we used the BUSTED + S (branch-site unrestricted statistical test for episodic diversification, with synonymous rate variation) method on our multiple sequence alignment and inferred phylogenetic tree, which infers 3  $\Omega$  rate categories and their corresponding weights while integrating over their assignments to all branches and sites. Our BUSTED analysis was not statistically significant (averaged likelihood ratio test (LRT)  $p$ -value 0.23). The  $\omega_3$  rate (which is the measure of adaptive evolution in this method)

was  $\sim 0.5$  and was attributed to a miniscule proportion of the data (Table 2). This evidence suggests that a minute fraction of (branch-site) pairs in this gene are driving the signal for adaptive evolution but fail to reach statistical significance. Despite finding some evidence of EDS (using BUSTED + S) at 18 codon sites with an evidence ratio (ER) > 10 (Supplementary Figure S1), a heuristic for evolutionary signal, described in Murrell et al., 2015, the evidence disappears when accounting for MH (Table 2), as evolutionary signal is absorbed by the MH rate parameter ( $TH_{Rate}$ ).

Multi-nucleotide substitutions involve changes in multiple adjacent nucleotides in DNA and may have significant impacts on gene evolution by creating novel genetic variation, increasing the accessibility of evolutionary pathways, and altering the structure and function of proteins. Studies have shown that multi-nucleotide mutations can contribute to adaptation and evolutionary innovation in a range of organisms (Steward et al., 2022; Hensley et al., 2021; Cohen et al., 2020; Silva et al., 2023). Considering these studies, we have applied a recently developed selection analysis (Lucaci et al., 2023) that incorporates two confounding biological processes: synonymous rate variation and multi-nucleotide substitutions into the BUSTED framework. Our interpretation is that, overall, based on our model averaging approach (described in Lucaci et al., 2023), there is no evidence suggesting that mammalian *Rem2* is subject to episodic diversifying selection.

## Pervasive purifying selection in mammalian *Rem2*

Our analysis of site-level negative selection in mammalian *Rem2* was conducted using the FEL method (see Methods section for details) in HyPhy. The results (see Figure 2) provide a measure of the extent of purifying selection for each site in the human *Rem2* gene and indicate that 70% of non-invariant codon sites show a statistically significant estimate of purifying selection. To visualize the spatial organization of our results, the dN/dS estimates for the entire alignment were plotted with a confidence interval of the 95% lower and upper-bound estimates (see Figure 2 or Supplementary Table S2). Overall, mammalian *Rem2* exhibits broad evidence of purifying selection in proximal disordered regions and in distal RGGK and disordered regions. These results suggest that purifying selection has played a predominant role in shaping the functional and mature *Rem2* gene, which exhibits remarkable conservation across evolutionary epochs, leading to its rapid evolutionary adaptation in other genes and taxa.

While we searched for evidence of adaptively evolving sites using the FEL (Figure 2; Supplementary Table S2), we found a single positively selected site using FEL, after multiple tests correction, corresponding to alignment codon site 170 (human *Rem2* site 148). In comparison, using the MEME (Supplementary Table S3) method, after applying multiple tests correction, we found no sites remained statistically significant. We have included the MEME table of results for completeness in the Supplementary Material. As more information becomes available (i.e., number of mammalian genomes, or increasing the scope of the study to examine taxonomic groups outside of mammalian) we expect that some of these sites may become statistically significant, indicating a role for adaptation to unique environmental pressures.

**TABLE 2 Evidence of Gene-wide evolution in Mammalian Rem2.** This table describes summary statistics for the BUSTED model-testing approach. Briefly, in each column we applied one of four methods, each of which either accounts for a biological phenomenon of synonymous rate variation (+S) and/or multi nucleotide (+MH) substitutions or does not. Sequences and Codons correspond to the number of taxa and codon sites in the alignment, respectively. LRT *p*-value and Averaged *p*-value correspond to the unadjusted and adjusted (see Methods for details) BUSTED model-testing approach. cAIC is the small-sample Akaike information criterion. Omega rates 1-3 correspond to inferred omega estimates, while “*p*” 1-3 values are the proportion of the data that these rates were found to be the best fit. SRV rates 1-3 correspond to inferred synonymous rate variation estimates, while “SRV *p*” 1-3 values are the proportion of the data that these rates were found to be the best fit. DH and TH rates are the inferred double nucleotide and triple nucleotide substitution rate estimates.

Method	Method			
	BUSTED	BUSTED + MH	BUSTED + S	BUSTED + S + MH
Sequences	175	175	175	175
Codons	627	627	627	627
LRT <i>p</i> -value	0.0000	0.0019	0.0000	0.5000
Averaged <i>p</i> -value	0.2272	0.2272	0.2272	0.2272
cAIC	47,174.7462	47,168.7997	46,351.6387	46,352.0042
$\omega_1$	0.0000	0.0647	0.1221	0.1003
$p_1$	0.6958	0.9285	0.9959	0.6129
$\omega_2$	0.4044	1.0000	1.0000	0.1241
$p_2$	0.3013	0.0701	0.0029	0.3620
$\omega_3$	23.1578	35.2319	50.7482	1.0000
$p_3$	0.0029	0.0014	0.0012	0.0251
SRV <sub>1</sub>	-	-	0.4989	0.4939
SRV $p_1$	-	-	0.5513	0.5417
SRV <sub>2</sub>	-	-	1.3352	1.3222
SRV $p_2$	-	-	0.3799	0.3854
SRV <sub>3</sub>	-	-	3.1626	3.0585
SRV $p_3$	-	-	0.0688	0.0729
DH <sub>Rate</sub>	-	0.0448	-	0.0558
TH <sub>Rate</sub>	-	0.0652	-	0.2599

Purifying selection plays a crucial role in shaping the evolution of genes in mammals by maintaining functional genes, removing harmful mutations, and preserving genetic diversity. By analyzing orthologous sequences across species, we observe the broad distribution of negatively selected sites across the *Rem2* gene (see Figure 2; Supplementary Table S2). Protein-coding sequences with highly constrained structures are expected to fix nonsynonymous mutations at a slower rate due to the maladaptive nature of changes such as what we observe with negatively selected sites across *Rem2*.

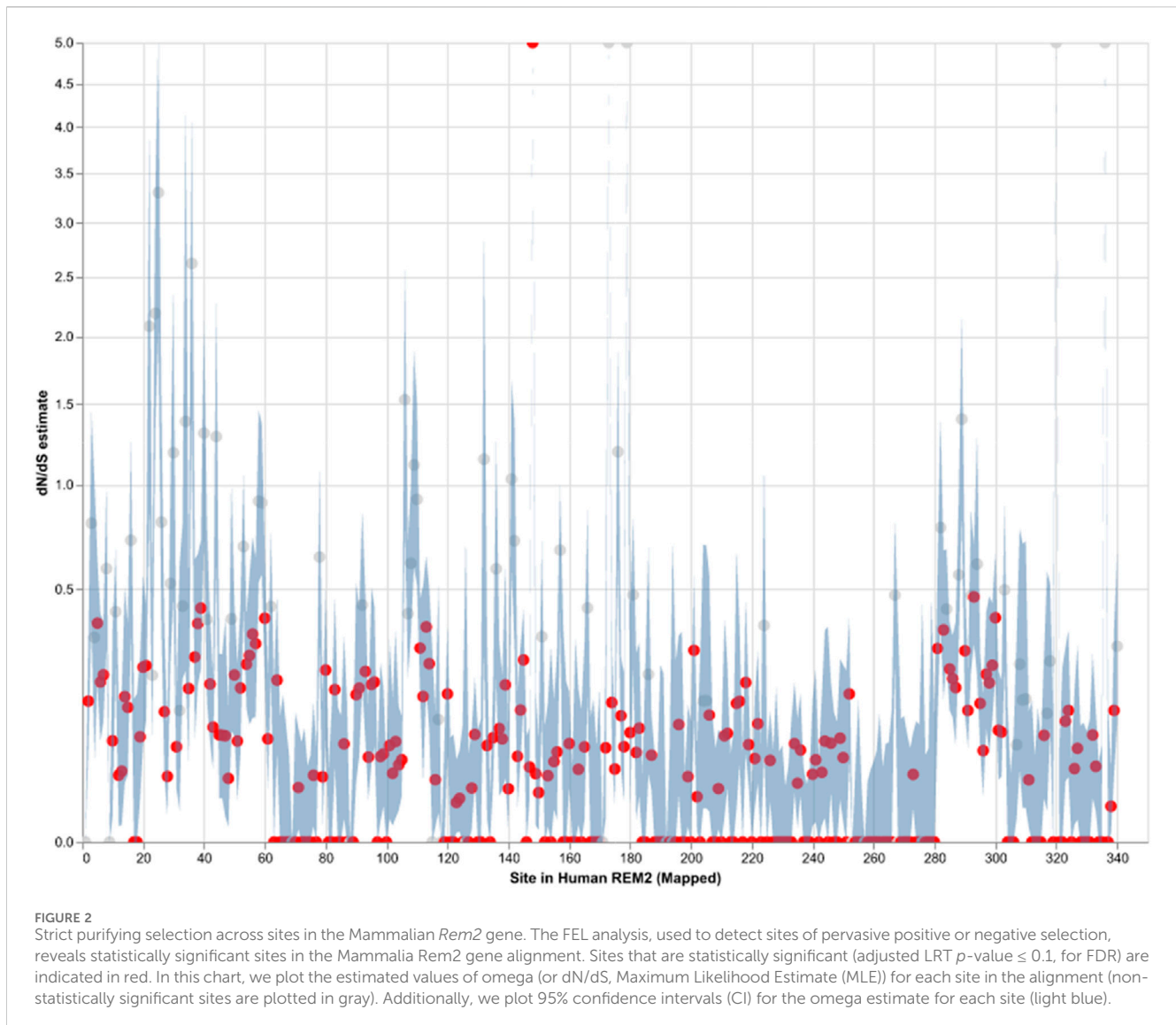
## Broad evidence of coevolutionary forces within Rem2

To examine the coevolution of codon sites in mammalian *Rem2*, (i.e., if a particular codon site was evolving in a relationship with another codon) we used the Bayesian graphical models (BGM) method (Poon et al., 2008). This method infers the inferred substitution history of an alignment with maximum-likelihood ancestral sequence reconstruction and maps these to the phylogenetic tree, which allows for the detection of correlated patterns of substitution. In our results we find evidence for pairs

and groups of 50 putatively coevolving sites in our mammalian *Rem2* alignment (see Figure 3; Supplementary Table S4). This data suggests that interaction dynamics may be impacted in tertiary space of the folded protein. Additionally, coevolving sites may be related to other fitness consequences (i.e., offering compensatory mutations) for maladaptive changes in another part of the protein sequence.

When we reviewed putatively interacting, we noticed that several pairs of coevolving sites (Figure 3) occurring only in the human *Rem2*/RGK (Supplementary Table S4), and some pairs with one site in the proximal disordered region and another within the human RGK region, linking these distant regions. Several regions display complex geometric features including a network of interactions spanning multiple sites: 86, 604, 72, 68, 120. When mapped to the human *Rem2* reference sequence all these sites fall within the RGK region, and correspond to 64, 319 (in RGK), 52, 48, 98, respectively. We examined the gnomAD database [v4.1.0, Karczewski et al., 2020] for any clinical relevance of these sites and we found the following variants at these sites:

Disruption of the delicate balance maintained in protein evolution of this type, in this network or any of the coevolving networks observed in mammalian *Rem2* could have fitness altering



clinical relevance and are worth further investigation. We appreciate that sites 48, 64, 98 in human *Rem2* are in a coevolutionary network together and have possible functional implications when mutated (Table 3).

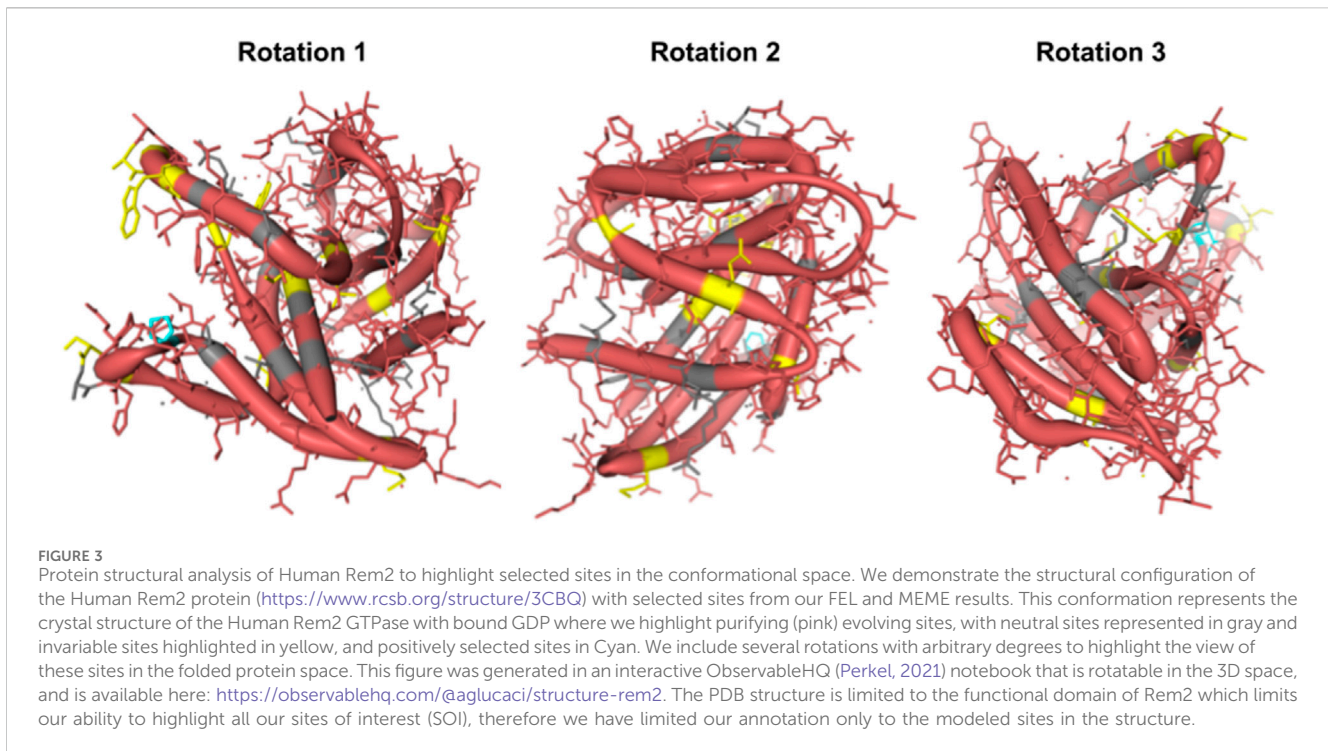
### Structural view of selected sites on the Human *Rem2* protein

To gain an appreciation of the spatial distribution of the site of interest identified by our statistical methods, we relied on a PDB structure of *Rem2* to generate a three-dimensional reconstruction. We examined the crystal structure of the human *Rem2* GTPase with bound GDP with a modeled residue count of 169. We focus on the main functional domain of the *Rem2* gene and its function as a GTPase as an example of spatial protein-ligand interactions. We believe that this additional layer of biological and physical-chemical interaction has contributed to the complex evolution of the functionality of *Rem2*. By mapping our results from the selection

analyses listed above, we examined sites with evidence to be under the effect of a purifying selective regime in the evolutionary history of *Rem2* in mammals.

Although the immediate effects of evolution on protein structure (Figure 3) are not well understood and cannot be fully verified computationally but rely on experimentally determined results, they have implications for our understanding of *Rem2* neurobiology. Our coevolutionary and selection analyses suggest that the RGK, a region in *Rem2* known for its functional effects, is implicated. Our 3D modeling (based on a known PDB structure) indicates that the patterns and underlying networks of coevolving sites are in complex relationship with each other, which may have undiscovered functional impacts. Therefore, we predict that evolutionary changes in *Rem2* are likely to reflect some form of specialization, divergence in function, or interaction with protein partners at various stages in its evolution in mammals. Further research may reveal additional connections that can provide deeper insight into *Rem2*'s diverse functions and its role in human health and disease.





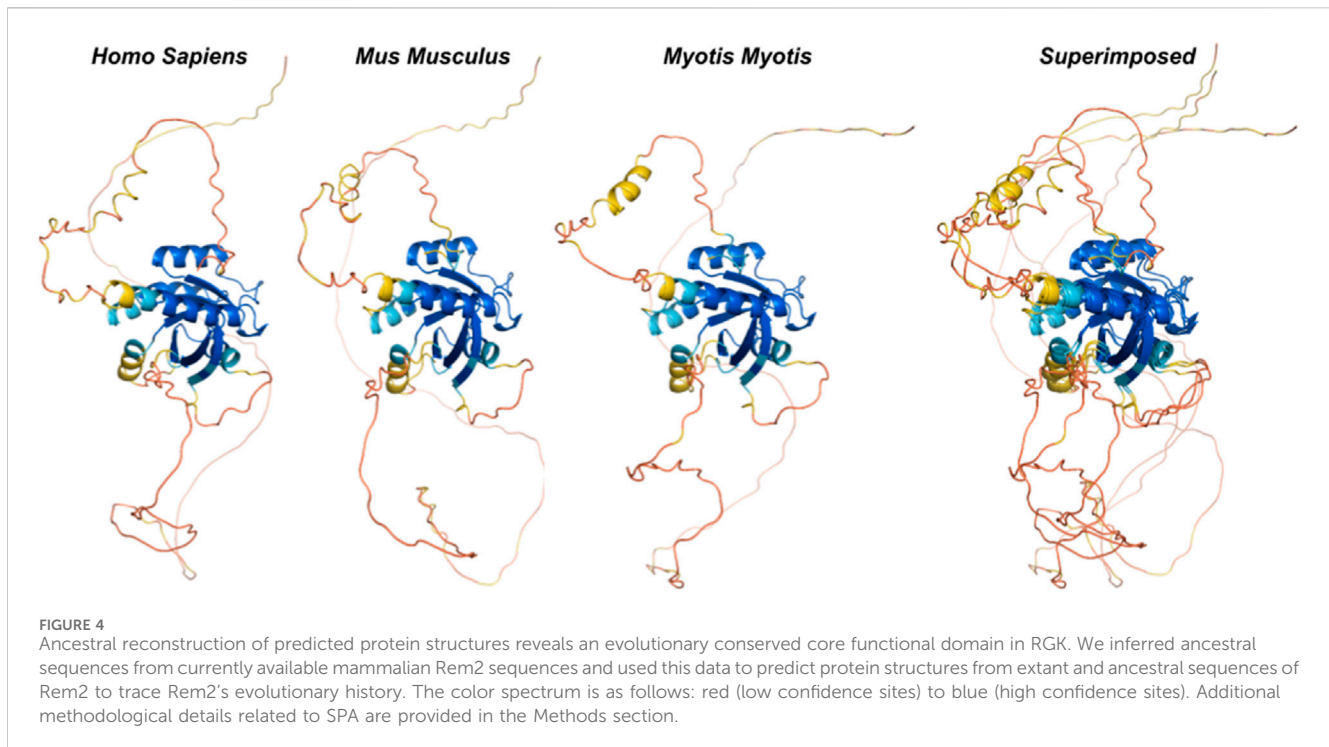
**TABLE 3** Inferred phenotypic implications of coevolutionary important codon sites in Rem2. In this table we denote the consequences of mutations found in the gnomAD database for codon sites where we found evidence of coevolutionary forces. Note that mutations at positions 23,353,970 and 23,353,971 in Rem2 fall within a potential MNV (\*) where the variant is found in phase in 1 individual, altering the amino acid sequence, the combined change is as follows: aGA to aTT becomes a R to I change.

Codon Site in Human Rem2	Genome Position (hg38: chr14)	Variant Type	Nucleotide change	Amino acid change	Mutation Consequence	PolyPhen	SIFT
48	23,353,918	Deletion	CTG- > Cdel	Leu48GlyfsTer37	Frame shift	-	-
64	23,353,969	SNV	A- > T	Arg64Ter	Stop gained	-	-
64	23,353,970	MNV*	G- > T	Arg64Ile	Missense	Possibly damaging	Deleterious
64	23,353,971	MNV*	A- > T	Arg64Ser	Missense	Possibly damaging	Deleterious
98	22,884,863	SNV	C- > T	Ser98Phe	Missense	Probably damaging	Deleterious

### Ancestral sequences of mammalian Rem2 reveal evolutionary dynamics and conservation

We used ancestral sequence reconstruction (ASR) (Selberg et al., 2021) and predicted protein structure to shed light on the evolutionary history and conservation of mammalian Rem2. By analyzing 175 extant and 159 reconstructed extinct sequences, our results suggest some degree of structural difference across Rem2 in the core functional RGK domain and in its intrinsically disordered regions, highlighting the complex interplay between purifying selection at the sequence level (noted above) and protein structural evolutionary trajectory (Figure 4). Our results (Figure 5) suggest that despite varying levels of sequence

divergence, the protein structure of Rem2 remains similar structurally, with most pairwise comparisons above a TM-Score of 0.5. Our approach also highlights the importance of examining the evolutionary history of ancient genes and pathways (Goult et al., 2023). However, how Rem2 accomplishes its activity in distinct species may be slightly different at the protein-structural level. Root Mean Square Deviation (RMSD) is a measure that quantifies the average distance between atoms, typically the backbone atoms, of proteins that have been superimposed. When examining the results for our RMSD calculations we find an average score of 3.61 (std 0.54, Supplementary Table S5), indicating a moderate level of discordance across the range of Rem2 protein structures; lower values of RMSD are better, with a score of <2 considered in good structural agreement (Maiorov and Crippen, 1994).



To circumvent some of the short-comings of RMSD we also utilized the TM-score, a metric that prioritizes smaller distance errors over larger ones, making it more sensitive to global fold similarities rather than local structural variations and has a length-dependent scale to normalize the distance errors making the magnitude of the TM-score length-independent for random structural pairs. When examining the results for our TM-score (using TM-align) calculations we find an average score of 0.62 (std 0.11, [Supplementary Table S5](#)), also indicating a high level of concordance across the range of structures; with values above 0.5 considered in about the same structural fold and values below 0.3 considered to be random ([Figure 5](#)). When filtering for strong pairwise TM-score's (above 0.5) we find that 86% of our dataset demonstrates protein structural similarity ([Supplementary Table S5](#)). However, our results are not to be overinterpreted and may be largely confounded by low confidence structural inference (low pLDDT scores), difficulty in aligning the IDR of Rem2, or alignment errors. It may also be useful to examine Rem2's protein structural domain evolution, including instances where specific domains are gained or lost across the phylogenetic tree and how these impact the comparison of structural inference. The IDR characteristic of Rem2, allows for significant flexibility and potentially enables these regions to perform various functional roles despite their lack of fixed conformation. As disordered regions are often crucial for interacting with multiple partners, facilitating dynamic responses within cellular signaling pathways, and undergoing conformational changes that activate or regulate the protein's function. Identifying and understanding these disordered segments can be vital for understanding how Rem2 operates within biological pathways and may become important regions for a new class of therapeutics ([Sader 2020](#)) designed to target IDR's.

However, we note that for even small genetic distances (<0.05) we observe a large variance TM-Score, indicating that divergence could be driven by either the disordered region of the Rem2 protein or by neutral mutations in its core functional domain, and

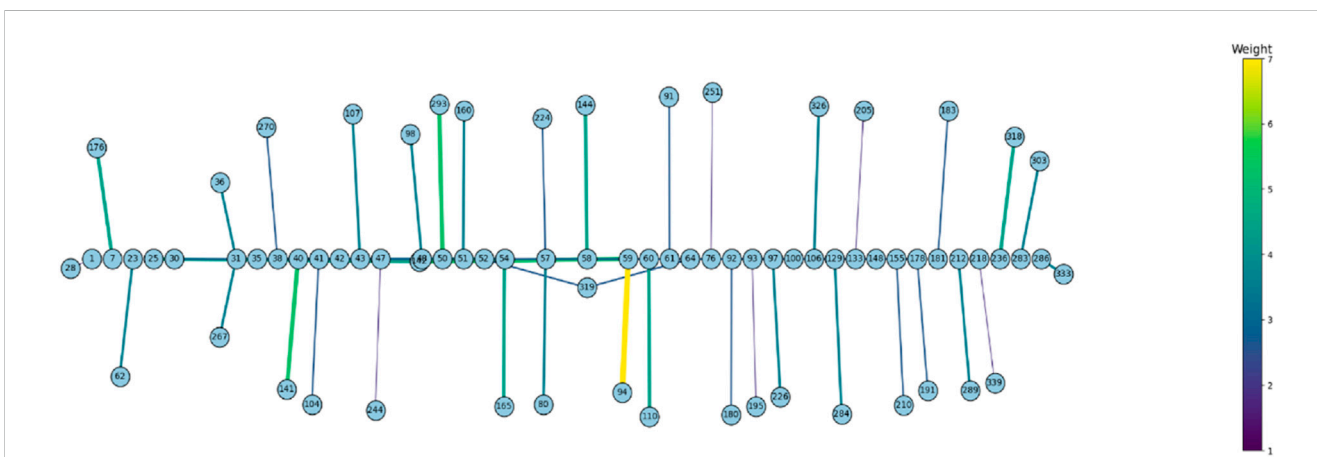
convergence explained through conservation of sequence and protein structure for closely related species. Taken together molecular sequence divergence across *Mammalia* could spur evolutionary innovation in *Rem2*.

## Discussion

We collected all sequence data available for the orthologous *Rem2* gene, gathering from a diverse set of species within the *Mammalia* taxonomic group. Our results indicate that unique evolutionary processes have shaped the evolutionary history of *Rem2*. Our study conducted several evolutionary analyses that each asked and answered specific biological questions (see Methods section for additional information) to quantify the signals of natural selection in our dataset. Our results revealed that *Rem2* is under strong purifying selection ([Figure 2](#)) acting consistently across the entire gene, suggesting its sequence and structural conservation is important in maintaining specific protein functionality. We also identified novel substitutions and areas of interest (see [Figure 2](#); [Table 3](#); [Figure 6](#); [Supplementary Tables S2, S3](#)) across regions of *Rem2* that may provide potential targets for designing molecular therapeutic approaches with significant impact on protein function. We also inferred a complex network of spatial coevolution within this gene, with evidence that 50 pairs of sites have evolved co-dependently ([Figure 6](#); [Supplementary Table S4](#)). While mutations in certain regions of Rem2 may be linked to human disorders such as Timothy Syndrome, some mutations remain unattributed to any specific condition or any observed phenotypic differences. This suggests that our comprehension of Rem2's role and biological functions is still in its preliminary stages, indicating a vast area for further exploration.



**FIGURE 5**  
 A comparison of genetic and structural differences in mammalian *Rem2* with a focus on exchanges between Primates and other clades. This chart shows scatter plots and histograms for our measures of genetic distance (via TN93) and structural distances (via TM-Score calculations), the plot shows the clade of interest, the color legend defines the taxonomic annotation of each pairwise comparison. Other clades are plotted in the (Supplementary Figures S2–S6).



**FIGURE 6**  
 The Bayesian Graphical Models (BGM) analysis, a method to detect co-evolving sites in a gene, of Mammalian *Rem2* found 50 pairs of coevolving sites out of 627 total sites, to be statistically significant with a posterior probability threshold of 0.5 (see Supplementary Table S4). In this chart, we plot the statistically significant coevolving pairs (nodes) with the number of shared substitutions between pairs of coevolving sites (edges) controlling the thickness and color of the line, with larger connections indicating a higher number of shared substitutions. The extent of coevolution tends to be distributed both proximally and distally across the *Rem2* gene, indicating interaction between the early region and its conserved core functional domain. This suggests that sequence variation incurred in this region most likely has a deleterious impact on protein fitness. Note that these sites have been mapped to correspond to sites in the Human *Rem2* protein.

## Abundant purifying selection across the Rem2 gene in Mammalia

Our results demonstrate a high degree of purifying selection observed across *Rem2*, which we hypothesize is based on the critical role of *Rem2* on its underlying network of genes governing homeostasis and normal brain development (Burkhardt et al., 2023). We believe that the conservation of function, and gene family evolution plays a role in shaping the evolutionary history of *Rem2* based on the maintenance of functional integrity. This interpretation is also consistent with the observation that *Rem2* plays a significant role in nervous system development (Ghiretti et al., 2011; 2014; Moore et al., 2013; 2018; Edel et al., 2010).

## No evidence of diversifying selection at specific sites in the Rem2 gene in Mammalia

*Rem2* has been recognized as a crucial gene for synaptogenesis and the regulation of dendritic morphology (Ghiretti and Paradis 2011; Ghiretti et al., 2014). However, due to significant evolutionary pressure shaping *Rem2* in novel environments we expected to find a small number of sites evolving adaptively. Unfortunately, new sites under positive diversifying selection were not found in the *Rem2* gene (see Supplementary Table S3) after multiple tests correction. By observing the diversification of amino acid changes which can drive species-specific adaptation and/or increase function in relation to environmental pressures, helping the scientific community understand the evolution of *Rem2* in a closer species-specific manner. However, with a limited number of species in our *Rem2* data, our results suggest that future research may yield results that hold important functional properties and may play a role in organismal disease. We expect that increasing the taxonomic range under study to all available sequences may yield different results. Additional research is needed to determine the significance of these adaptively evolving sites, including specialized or regulatory functions and their relation within the network of genes necessary for neural cell diversity and development. The 23 sites we found within our MEME results with statistically significant unadjusted *p*-values (Supplementary Table S3) may be a good starting point. While the clinical and functional significance of these sites remains a hypothesis, this study highlights their potential importance, and we look forward to further testing in controlled laboratory conditions.

## Characterization of coevolving pairs in the Rem2 gene in Mammalia

We found a network of putatively coevolving sites within the *Rem2* gene (as shown in Figure 6; Supplementary Table S4), which represents a new aspect of *Rem2* biology not reported before. Here, the coevolution of codon sites refers to a phenomenon where substitutions at one codon site co-occur along the same branches as changes occurring at another codon site. Multiple factors contribute to the coevolution of codon sites within a protein, such as functional or structural constraints, interactions at the amino acid or mRNA level, and selective pressures. For example, a scenario where specific residues hold importance for compensatory changes which may play a crucial role in maintaining a protein's proper structure and function especially in

core functional domains. Coevolving sites were not limited to a particular domain or motif and were distributed throughout the entire *Rem2* gene broadly, with some interactions in the RGK functional domain region. However, there were also coevolving sites in regulatory regions (outside of the RGK domain) linked to RGK domain sites. These relationships may play a significant role in shaping the evolution of this crucial gene and its enzymatic properties. The new evidence of coevolution suggests the importance of these sites in regulating the functional domain of *Rem2*. These residues may form crucial interactions for the functional integrity of *Rem2*, and the specific pairs that span the *Rem2* RGK and non-RGK regions suggest a new mechanism by which the functional domain may regulate the RGK or *vice versa*. Alternatively, these coevolving pairs may belong to a network of residues adapting to new environments post-speciation events and therefore serve species-specific functional requirements.

## Potential protein structural implications of evolving sites

Timothy syndrome is a rare genetic condition that is caused by mutations in the gene encoding the calcium channel protein CACNA1C. Recent studies have shown that the *Rem2* gene is involved in modulating the function of CACNA1C, and thus may play a role in the pathogenesis of Timothy syndrome, as has been observed with Gem [Krey et al., 2013; Boczek et al., 2015]. In Timothy syndrome, mutations in the CACNA1C gene result in a gain-of-function in the calcium channel, leading to an increase in intracellular calcium levels. The increased calcium levels can cause a range of symptoms, including heart problems, developmental delays, and autism. Studies have shown that the *Rem2* protein can interact with CACNA1C and modulate its activity, suggesting that the *Rem2* gene may be involved in the pathogenesis of Timothy syndrome. However, more research is needed to fully understand the relationship between *Rem2* and Timothy syndrome.

Our SPA analysis can also serve as a resource for the scientific community interested in *Rem2*'s evolutionary trajectory. The dataset provides utility for examining additional hypotheses involving the properties of molecular sequences and the protein structural dynamics. It may be particularly interesting to examine where in evolutionary time specific functional domains were lost or gained in *Rem2* in response to species-specific neurodevelopmental needs. We envision that approaches could be applied for structural-functional prediction and can be coupled with datasets from carefully designed experimental projects in future studies.

## Limitations of our computational evolutionary analysis

This study focused on *Rem2* sequences from the Mammalia taxonomic group instead of examining a more inclusive dataset for *Rem2* sequences from all Dipnotetrapodomorpha (lobe-finned fishes including tetrapods and lungfishes). Therefore, while our selection analyses are applicable to mammals, our results do not capture the entirety of *Rem2*'s evolutionary

history [Puhl et al., 2014]. We also do not explore the mutational patterns occurring outside of protein CDS (i.e., UTR's) which include complex structure and dynamics at the mRNA level of non-coding regions in the *Rem2* gene. An additional limitation of the current analysis is due to the presence of indels, spatially distributed across several regions of the *Rem2* alignment. Although there is a risk that particularly gappy regions of our multiple sequence alignment may be a computational artifact of the alignment procedure or from the quality of the genomic assembly procedures, based on all other outputs we believe that our results are interpreted and have subsequently tolerated these potential effects.

## Conclusion

Our research modeled the natural history of changes in the *Rem2* gene across 175 mammalian genomes spanning about –200 million years of evolutionary history. By analyzing a deeper phylogenetic dataset, such as additional genetic sequences contained within the Dipnotetrapodomorpha clade, one could reveal additional information on the substitutional history and lineage specific adaptations in structure and function of *Rem2*. Overall, we believe that experimental techniques in molecular biology can benefit from these types of evolutionary analysis. Specifically, where bioinformatics analysis can provide valuable information to guide experimental investigation and increase the efficiency and accuracy of biological research (Cisek and Hayden, 2021). Importantly, evolutionary history can inform both predictive modeling about the outcomes of mutations in individuals and the integration of multiple levels of biological or medical information which can lead to novel insights for disorders involving *Rem2*. In addition, both experimental design for fitness-altering regions of *Rem2* and target identification for chemical inhibitors of protein function can be informed by this style of analysis. Our discoveries may hold significant implications for advancing the development of novel treatments for targeting disease mechanisms; as they underscore the criticality of investigating the evolutionary past of genes associated with intricate neurological processes. However, further research is needed to fully understand the molecular underpinnings of *Rem2* in mammalian biology and evolution.

## Data availability statement

The data presented in the study are deposited in the <https://github.com/aglucaci/AOC-REM2> and <https://github.com/aglucaci/AMES-REM2>. Data is also archived at Zenodo with accession <https://doi.org/10.5281/zenodo.11550125>.

## Author contributions

AL: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Resources,

Software, Supervision, Validation, Visualization, Writing—original draft, Writing—review and editing. WB: Writing—original draft, Writing—review and editing. JL: Writing—original draft, Writing—review and editing. AS: Data curation, Validation, Writing—review and editing. VC: Methodology, Writing—original draft, Writing—review and editing. AM: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing—original draft, Writing—review and editing. SK: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing—original draft, Writing—review and editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported in part by grants GM151683 and HG009299 to SK. The funding bodies played no role in the design of the study, the collection, analysis, and interpretation of data, nor in writing the manuscript.

## Acknowledgments

We thank all our colleagues from the Acme Computational Molecular Evolution Group (ACME) for discussions which helped to improve this paper.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbinf.2024.1381540/full#supplementary-material>

## References

- Allen, M., Carrasquillo, M. M., Funk, C., Heavner, B. D., Zou, F., Younkin, C. S., et al. (2016). Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Sci. Data* 3, 160089. doi:10.1038/sdata.2016.89
- Anderson, D. P., Whitney, D. S., Hanson-Smith, V., Woznica, A., Campodonico-Burnett, W., Volkman, B. F., et al. (2016). Evolution of an ancient protein function involved in organized multicellularity in animals. *eLife* 5, e10147. doi:10.7554/eLife.10147
- Béguin, P., Mahalakshmi, R. N., Nagashima, K., Cher, D. H. K., Kuwamura, N., Yamada, Y., et al. (2005). Roles of 14-3-3 and calmodulin binding in subcellular localization and function of the small G-protein Rem2. *Biochem. J.* 391 (3), 712. doi:10.1042/bj3910712
- Béguin, P., Nagashima, K., Gonoï, T., Shibasaki, T., Takahashi, K., Kashima, Y., et al. (2001). Regulation of Ca<sup>2+</sup> channel expression at the cell surface by the small G-protein kir/Gem. *Nature* 411 (6838), 701–706. doi:10.1038/35079621
- Benndorf, R., Velazquez, R., Zehr, J. D., Pond, S. L. K., Martin, J. L., and Lucaci, A. G. (2022). Human HspB1, HspB3, HspB5 and HspB8: shaping these disease factors during vertebrate evolution. *Cell Stress Chaperones* 27 (4), 309–323. doi:10.1007/s12192-022-01268-y
- Burkhardt, P., Colgren, J., Medhus, A., Digel, L., Naumann, B., Soto-Angel, J. J., et al. (2023). Syncytial nerve net in a ctenophore adds insights on the evolution of nervous systems. *Science* 380 (6642), 293–297. doi:10.1126/science.ade5645
- Cisek, P., and Hayden, B. Y. (2021). Neuroscience needs evolution. *Philosophical Trans. R. Soc. B Biol. Sci.* 377 (1844), 20200518. doi:10.1098/rstb.2020.0518
- Cohen, Z. P., Brevik, K., Chen, Y. H., Hawthorne, D. J., Weibel, B. D., and Schoville, S. D. (2021). Elevated rates of positive selection drive the evolution of pestiferousness in the Colorado potato beetle (*Leptinotarsa decemlineata*, Say). *Mol. Ecol.* 30 (1), 237–254. doi:10.1111/mec.15703
- Correll, R. N., Botzet, G. J., Satin, J., Andres, D. A., and Finlin, B. S. (2008a). Analysis of the Rem2 - voltage dependant calcium channel  $\beta$  subunit interaction and Rem2 interaction with phosphorylated phosphatidylinositol lipids. *Cell Signal* 20 (2), 400–408. doi:10.1016/j.celsig.2007.10.029
- Correll, R. N., Pang, C., Niedowicz, D. M., Finlin, B. S., and Andres, D. A. (2008b). The RGK family of GTP-binding proteins: regulators of voltage-dependent calcium channels and cytoskeleton remodeling. *Cell. Signal.* 20 (2), 292–300. doi:10.1016/j.celsig.2007.10.028
- DeRocher, M. M., Armaly, F. H., Lepore, C. J., and Hollis, D. M. (2014). Rem2 in the bullfrog (*Rana catesbeiana*): patterns of expression within the central nervous system and brain expression at different ontogenetic stages. *Gene* 540 (1), 37–45. doi:10.1016/j.gene.2014.02.030
- Edel, M. J., Menchon, C., Menendez, S., Consiglio, A., Raya, A., and Izpisua Belmonte, J. C. (2010). Rem2 GTPase maintains survival of human embryonic stem cells as well as enhancing reprogramming by regulating p53 and cyclin D1. *Genes Dev.* 24 (6), 561–573. doi:10.1101/gad.1876710
- Fares, M. A. (2014). "The influence of recombination on the estimation of selection from coding sequence alignments," in *Natural selection* (United Kingdom: Taylor and Francis Group).
- Finlin, B. S., Shao, H., Kadono-Okuda, K., Guo, N., and Andres, D. A. (2000). Rem2, a new member of the Rem/Rad/Gem/Kir family of Ras-related GTPases. *Biochem. J.* 347 (1), 223–231. doi:10.1042/0264-6021:3470223
- Flynn, R., Chen, L., Hameed, S., Spafford, J. D., and Zamponi, G. W. (2008). Molecular determinants of Rem2 regulation of N-type calcium channels. *Biochem. Biophysical Res. Commun.* 368 (3), 827–831. doi:10.1016/j.bbrc.2008.02.020
- Flynn, R., Colicos, M. A., and Zamponi, G. W. (2010). Rem2 redistributes in response to neuronal stimulation. *Biophysical J.* 98 (3), 693a. doi:10.1016/j.bpj.2009.12.3805
- Flynn, R., and Zamponi, G. W. (2010). Regulation of calcium channels by RGK proteins. *Channels* 4 (6), 434–439. doi:10.4161/chan.4.6.12865
- Gauthier-Rouvière, C., Vignal, E., Mériane, M., Roux, P., Montcourier, P., and Fort, P. (1998). RhoG GTPase controls a pathway that independently activates Rac1 and Cdc42Hs. *Mol. Biol. Cell* 9 (6), 1379–1394. doi:10.1091/mbc.9.6.1379
- Ghiretti, A. E., Kenny, K., Marr, M. T., and Paradis, S. (2013). CaMKII-dependent phosphorylation of the GTPase Rem2 is required to restrict dendritic complexity. *J. Neurosci.* 33 (15), 6504–6515. doi:10.1523/jneurosci.3861-12.2013
- Ghiretti, A. E., Moore, A. R., Brenner, R. G., Chen, L. F., West, A. E., Lau, N. C., et al. (2014). Rem2 is an activity-dependent negative regulator of dendritic complexity *in vivo*. *J. Neurosci.* 34 (2), 392–407. doi:10.1523/jneurosci.1328-13.2014
- Ghiretti, A. E., and Paradis, S. (2011). The GTPase Rem2 regulates synapse development and dendritic morphology. *Dev. Neurobiol.* 71 (5), 374–389. doi:10.1002/dneu.20868
- Gouly, M., Botton-Amiot, G., Rosato, E., Sprecher, S. G., and Feuda, R. (2023). The monoaminergic system is a bilaterian innovation. *Nat. Commun.* 14 (1), 3284. doi:10.1038/s41467-023-39030-2
- Hamel, L. P., Nicole, M. C., Sritubtim, S., Morency, M. J., Ellis, M., Ehrling, J., et al. (2006). Ancient signals: comparative genomics of plant MAPK and MAPKK gene families. *Trends Plant Sci.* 11 (4), 192–198. doi:10.1016/j.tplants.2006.02.007
- Hensley, N. M., Ellis, E. A., Leung, N. Y., Coupert, J., Mikhailovsky, A., Taketa, D. A., et al. (2021). Selection, drift, and constraint in cyprinid luciferases and the diversification of bioluminescent signals in sea fireflies. *Mol. Ecol.* 30 (8), 1864–1879. doi:10.1111/mec.15673
- Hollis, D. M., Sawa, Y., Wagoner, A., Rawlings, J. S., and Goetz, F. W. (2012). Isolation and molecular characterization of Rem2 isoforms in the rainbow trout (*Oncorhynchus mykiss*): tissue and central nervous system expression. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 161 (2), 93–101. doi:10.1016/j.cbpb.2011.09.011
- Hurvich, C. M., and Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika* 76 (2), 297–307. doi:10.1093/biomet/76.2.297
- Hutter, H., Vogel, B. E., Plenefisch, J. D., Norris, C. R., Proenca, R. B., Spieth, J., et al. (2000). Conservation and novelty in the evolution of cell adhesion and extracellular matrix genes. *Science* 287 (5455), 989–994. doi:10.1126/science.287.5455.989
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596 (7873), 583–589. doi:10.1038/s41586-021-03819-2
- Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alfoldi, J., Wang, Q., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581 (7809), 434–443. doi:10.1038/s41586-020-2308-7
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30 (4), 772–780. doi:10.1093/molbev/mst010
- Kosakovsky Pond, S. L., and Frost, S. D. W. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* 22 (5), 1208–1222. doi:10.1093/molbev/msi105
- Kosakovsky Pond, S. L., Poon, A. F. Y., Velazquez, R., Weaver, S., Hepler, N. L., Murrell, B., et al. (2020). HyPhy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies. *Mol. Biol. Evol.* 37 (1), 295–299. doi:10.1093/molbev/msz197
- Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelck, C. H., and Frost, S. D. W. (2006). Automated phylogenetic detection of recombination using a genetic algorithm. *Mol. Biol. Evol.* 23 (10), 1891–1901. doi:10.1093/molbev/msl051
- Krey, J. F., Paşca, S. P., Shcheglovitov, A., Yazawa, M., Schwemberger, R., Rasmuson, R., et al. (2013). Timothy syndrome is associated with activity-dependent dendritic retraction in rodent and human neurons. *Nat. Neurosci.* 16 (2), 201–209. doi:10.1038/nn.3307
- Kwakowsky, A., Prasad, A. A., Peña-Ortega, F., and Lim, S. A. O. (2023). Editorial: neuronal network dysfunction in neurodegenerative disorders. *Front. Neurosci.* 17, 1151156. doi:10.3389/fnins.2023.1151156
- Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536 (7616), 285–291. doi:10.1038/nature19057
- Liput, D. J., Lu, V. B., Davis, M. I., Puhl, H. L., and Ikeda, S. R. (2016). Rem2, a member of the RGK family of small GTPases, is enriched in nuclei of the basal ganglia. *Sci. Rep.* 6 (1), 25137. doi:10.1038/srep25137
- Lucaci, A. G., Notaras, M. J., Kosakovsky Pond, S. L., and Colak, D. (2022). The evolution of BDNF is defined by strict purifying selection and prodomain spatial coevolution, but what does it mean for human brain disease? *Transl. Psychiatry* 12 (1), 258–317. doi:10.1038/s41398-022-02021-w
- Lucaci, A. G., Wisotsky, S. R., Shank, S. D., Weaver, S., and Pond, S. L. K. (2021). Extra base hits: widespread empirical support for instantaneous multiple-nucleotide changes. *PLOS ONE* 16 (3), e0248337. doi:10.1371/journal.pone.0248337
- Lucaci, A. G., Zehr, J. D., Enard, D., Thornton, J. W., and Kosakovsky Pond, S. L. (2023). Evolutionary shortcuts via multinucleotide substitutions and their impact on natural selection analyses. *Mol. Biol. Evol.* 40 (7), msad150. doi:10.1093/molbev/msad150
- Maiorov, V. N., and Crippen, G. M. (1994). Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *J. Mol. Biol.* 235 (2), 625–634. doi:10.1006/jmbi.1994.1017
- Meck, W. H., Church, R. M., and Olton, D. S. (2013). Hippocampus, time, and memory. *Behav. Neurosci.* 127 (5), 655–668. doi:10.1037/a0034188
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37 (5), 1530–1534. doi:10.1093/molbev/msaa015
- Mirdita, M., Schütze, K., Moriawaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nat. Methods* 19 (6), 679–682. doi:10.1038/s41592-022-01488-1

- Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., et al. (2021). Sustainable data analysis with Snakemake. *F1000Res* 10, 33. doi:10.12688/f1000research.29032.1
- Moore, A. R., Ghiretti, A. E., and Paradis, S. (2013). A loss-of-function analysis reveals that endogenous Rem2 promotes functional glutamatergic synapse formation and restricts dendritic complexity. *PLOS ONE* 8 (8), e74751. doi:10.1371/journal.pone.0074751
- Moore, A. R., Richards, S. E., Kenny, K., Royer, L., Chan, U., Flavahan, K., et al. (2018). Rem2 stabilizes intrinsic excitability and spontaneous firing in visual circuits. *Elife* 7, e33092. doi:10.7554/elife.33092
- Murrell, B., Weaver, S., Smith, M. D., Wertheim, J. O., Murrell, S., Aylward, A., et al. (2015). Gene-wide identification of episodic selection. *Mol. Biol. Evol.* 32 (5), 1365–1371. doi:10.1093/molbev/msv035
- Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., and Pond, S. L. K. (2012). Detecting individual sites subject to episodic diversifying selection. *PLOS Genet.* 8 (7), e1002764. doi:10.1371/journal.pgen.1002764
- Muse, S. V., and Gaut, B. S. (1994). A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol. Biol. Evol.* 11 (5), 715–724. doi:10.1093/oxfordjournals.molbev.a040152
- Nahalka, J. (2019). The role of the protein-RNA recognition code in neurodegeneration. *Cell Mol. Life Sci.* 76 (11), 2043–2058. doi:10.1007/s00018-019-03096-3
- Perkel, J. M. (2021). Reactive, reproducible, collaborative: computational notebooks evolve. *Nature* 593 (7857), 156–157. doi:10.1038/d41586-021-01174-w
- Poon, A. F. Y., Lewis, F. I., Frost, S. D. W., and Kosakovsky Pond, S. L. (2008). Spidermonkey: rapid detection of co-evolving sites using Bayesian graphical models. *Bioinformatics* 24 (17), 1949–1950. doi:10.1093/bioinformatics/btn313
- Puhl, H. L., Lu, V. B., Won, Y. J., Sasson, Y., Hirsch, J. A., Ono, F., et al. (2014). Ancient origins of RGK protein function: modulation of voltage-gated calcium channels preceded the protostome and deuterostome split. *PLoS One* 9 (7), e100694. doi:10.1371/journal.pone.0100694
- Rahman, S., Kosakovsky, P. S. L., Webb, A., and Hey, J. (2021). Weak selection on synonymous codons substantially inflates dN/dS estimates in bacteria. *Proc. Natl. Acad. Sci.* 118 (20), e2023575118. doi:10.1073/pnas.2023575118
- Royer, L., Herzog, J. J., Kenny, K., Tzvetkova, B., Cochrane, J. C., Marr, M. T., et al. (2018). The Ras-like GTPase Rem2 is a potent inhibitor of calcium/calmodulin-dependent kinase II activity. *J. Biol. Chem.* 293 (38), 14798–14811. doi:10.1074/jbc.ra118.003560
- Sadar, M. D. (2020). Discovery of drugs that directly target the intrinsically disordered region of the androgen receptor. *Expert Opin. Drug Discov.* 15 (5), 551–560. doi:10.1080/17460441.2020.1732920
- Selberg, A. G. A., Gaucher, E. A., and Liberles, D. A. (2021). Ancestral sequence reconstruction: from chemical paleogenetics to maximum likelihood algorithms and beyond. *J. Mol. Evol.* 89 (3), 157–164. doi:10.1007/s00239-021-09993-1
- Shank, S. D., Weaver, S., and Kosakovsky Pond, S. L. (2018). phylotree.js - a JavaScript library for application development and interactive data visualization in phylogenetics. *BMC Bioinforma.* 19 (1), 276. doi:10.1186/s12859-018-2283-2
- Silva, S. R., Miranda, V. F. O., Michael, T. P., Plachno, B. J., Matos, R. G., Adamec, L., et al. (2023). The phylogenomics and evolutionary dynamics of the organellar genomes in carnivorous *Utricularia* and *Genlisea* species (Lentibulariaceae). *Mol. Phylogenetics Evol.* 181, 107711. doi:10.1016/j.ympev.2023.107711
- Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., and Kosakovsky Pond, S. L. (2015). Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* 32 (5), 1342–1353. doi:10.1093/molbev/msv022
- Steward, R. A., de Jong, M. A., Oostra, V., and Wheat, C. W. (2022). Alternative splicing in seasonal plasticity and the potential for adaptation to environmental change. *Nat. Commun.* 13 (1), 755. doi:10.1038/s41467-022-28306-8
- Tamura, K., and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10 (3), 512–526. doi:10.1093/oxfordjournals.molbev.a040023
- Wisotsky, S. R., Kosakovsky Pond, S. L., Shank, S. D., and Muse, S. V. (2020). Synonymous site-to-site substitution rate variation dramatically inflates false positive rates of selection analyses: ignore at your own peril. *Mol. Biol. Evol.* 37 (8), 2430–2439. doi:10.1093/molbev/msaa037
- Zhang, Y., and Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* 33 (7), 2302–2309. doi:10.1093/nar/gki524