STUDIA UNIVERSITATIS BABEŞ-BOLYAI BIOLOGIA, 71, 1, 2026 (p. 125-159)
doi:10.24193/subbbiol.2026.1.07

Phylogenetic structuring of intrinsic disorder and charge patterning in vertebrate Ermin

Ilka Koszorus[1]✉ and Ferencz Kósa[1,2]

[1]Babeş-Bolyai University, Hungarian Department of Biology and Ecology, Cluj-Napoca, Romania;
[2]Babeş-Bolyai University, Center for Systems Biology, Biodiversity and Bioresources, Sociobiology and Insect Ecology Lab, Cluj-Napoca, Romania;
✉Corresponding author, E-mail: koszorusilka@gmail.com.

Article history: Received 27 February 2026; Revised 17 June 2026; Accepted 17 June 2026; Available online 25 June 2026

Abstract. Ermin (ERMN) is an oligodendrocyte-enriched cytoskeletal protein implicated in myelin sheath formation and stability. Although previously described as intrinsically disordered, its evolutionary diversification across vertebrates has not been systematically examined. Here, we performed a comprehensive phylogenetic and comparative analysis of 159 vertebrate Ermin orthologs to characterize the evolution of intrinsic disorder and physicochemical sequence properties. Across all major vertebrate clades, Ermin consistently fulfilled criteria for highly intrinsically disordered proteins based on Average Disorder Score (ADS) and Percent of Predicted Disordered Residues (PPDR), demonstrating that extensive disorder represents a deeply conserved structural feature. Despite this conserved disorder abundance, the architectural organization of intrinsically disordered regions (IDRs) varied among lineages. Mammalian Ermin proteins exhibited increased fragmentation of long IDRs compared to birds, which retained largely continuous long-IDR segments, indicating that IDR partitioning evolves independently of total disorder content. Amino-acid composition analyses revealed conserved depletion of aromatic and bulky hydrophobic residues across vertebrates, combined with clade-structured variation in charged and disorder-promoting residues.

Phylogenetic generalized least squares (PGLS) analyses detected strong phylogenetic signal across disorder and electrostatic traits. Root-to-tip regressions revealed a significant directional decrease in mean hydrophobicity across vertebrate divergence, whereas disorder propensity and sequence charge decoration (SCD) did not exhibit consistent monotonic trends. Fraction of charged residues (FCR) showed more modest evolutionary variation. Multivariate models further demonstrated that hydrophobicity, global charge density, and sequence charge decoration independently contribute to evolutionary variation in Ermin sequence architecture.

Together, these results indicate that Ermin evolution is characterized by architectural remodeling and compositional retuning of an already disordered scaffold rather than by progressive increases in intrinsic disorder. This study highlights IDR fragmentation and electrostatic organization as key evolutionary dimensions shaping the diversification of intrinsically disordered proteins across vertebrate lineages.

Keywords: Ermin (ERMN), Intrinsically disordered proteins (IDPs), Intrinsic disorder evolution, Sequence charge decoration (SCD), Phylogenetic comparative analysis (PGLS)

Introduction

Myelination is a defining feature of vertebrate nervous systems, enabling the rapid saltatory propagation of action potentials along axons and thereby ensuring efficient neural signalling (Nave and Werner, 2014). Proper myelin formation and maintenance rely on a limited set of highly specialized dosage-sensitive proteins that coordinate membrane organization with the underlying cytoskeleton (Boggs, 2006; Nawaz *et al.*, 2015). Several myelin-associated proteins deviate from classical globular architectures and are enriched in intrinsically disordered regions, a property linked to dynamic regulatory roles rather than static structural support (Jahn *et al.*, 2009; Harauz and Boggs, 2013). Despite their functional importance, comparative and evolutionary analyses of such proteins remain limited, particularly beyond mammalian systems (Baumann and Pham-Dinh, 2001; Aggarwal *et al.*, 2011; Nave and Werner, 2014). Notably, myelin is unusually enriched in proteins with extensive intrinsic disorder, many of which are involved in myelin regulation and are affected in demyelinating disease (Raasakka and Kursula, 2020).

Intrinsically disordered proteins (IDPs) and regions (IDRs) are widespread in eukaryotic proteomes and are particularly enriched among proteins involved in transcriptional regulation, signalling, and cytoskeleton-associated functions (Wright and Dyson, 1999; Uversky *et al.*, 2000; Bellay *et al.*, 2011; van der Lee *et al.*, 2014; Peng *et al.*, 2015; Holehouse and Kragelund, 2024). Large-scale analyses of disease-associated variants have shown that intrinsically disordered regions are frequent targets of pathogenic mutations, despite lacking stable tertiary structure. Approximately 20–25% of disease-related mutations map to predicted disordered regions, indicating that intrinsic disorder is neither evolutionarily neutral nor functionally dispensable (Vacic and Iakoucheva, 2012). Rather than disrupting global folding, disease-associated mutations in disordered regions often perturb short functional elements, including molecular recognition motifs, binding interfaces, or charge-patterned segments that are critical for interaction specificity and regulatory function (Vacic *et al.,* 2012). Together, these findings establish intrinsic disorder as a functionally constrained and disease-relevant sequence feature, motivating comparative analyses of disorder, amino-acid composition, and charge-related properties in proteins such as Ermin.

Ermin (also known as Juxtanodin, Fig. 1) was first described in 2005 as a previously uncharacterized protein enriched in the central nervous system (Zhang *et al.*, 2005). Early sequence analyses revealed the presence of a C-terminal actin-binding region with similarity to ezrin–radixin–moesin (ERM) family proteins, despite the absence of a canonical ERM FERM domain (Zhang *et al.*, 2005; Brockschnieder *et al.,* 2006). Subsequent experimental studies demonstrated that Ermin associates with F-actin and modulates cytoskeleton-dependent cell morphology, including effects on cellular process formation and arborization (Zhang *et al.*, 2005; Meng *et al.*, 2010; Wang *et al.*, 2011; Ruskamo *et al.*, 2012). Although Ermin is enriched in oligodendrocytes and localizes to non-compact myelin regions during late stages of myelination, its precise mechanistic role in myelin formation and maintenance remains unclear (Brockschnieder *et al.*, 2006; Wang *et al.*, 2020). Importantly, Ermin expression is not restricted to oligodendrocytes, as it is also detected in retinal pigment epithelial cells, where it localizes to actin-rich subcellular regions and influences actin organization (Liang *et al.*, 2018). Taken together, existing evidence supports viewing Ermin primarily as an actin-associated cytoskeletal regulator with cell-type-specific functional consequences rather than as a structural myelin protein per se, despite its enrichment in oligodendrocytes (Zhang *et al.*, 2005; Brockschnieder *et al.*, 2006; Wang *et al.*, 2020).

Altered ERMN expression or genetic variation has been reported in multiple neurological and neurodevelopmental disorders, including autism spectrum disorder, epilepsy, schizophrenia, and multiple sclerosis (Salek Esfahani *et al.*,

2019; Wang *et al.*, 2020; Shiva *et al.*, 2021, Ziaei *et al.*, 2022). The ERMN gene is located within the AUTS5 locus, a genomic region that has been repeatedly implicated in autism, language impairment, and cognitive traits (Shiva *et al.*, 2021). Both reduced expression and dosage imbalance of ERMN have been associated with disease phenotypes, suggesting that tight regulation of Ermin sequence and expression is biologically important (Homs *et al.*, 2016; Shiva *et al.*, 2021). Together, these disease associations provide strong motivation for investigating Ermin's sequence properties and evolutionary constraints.

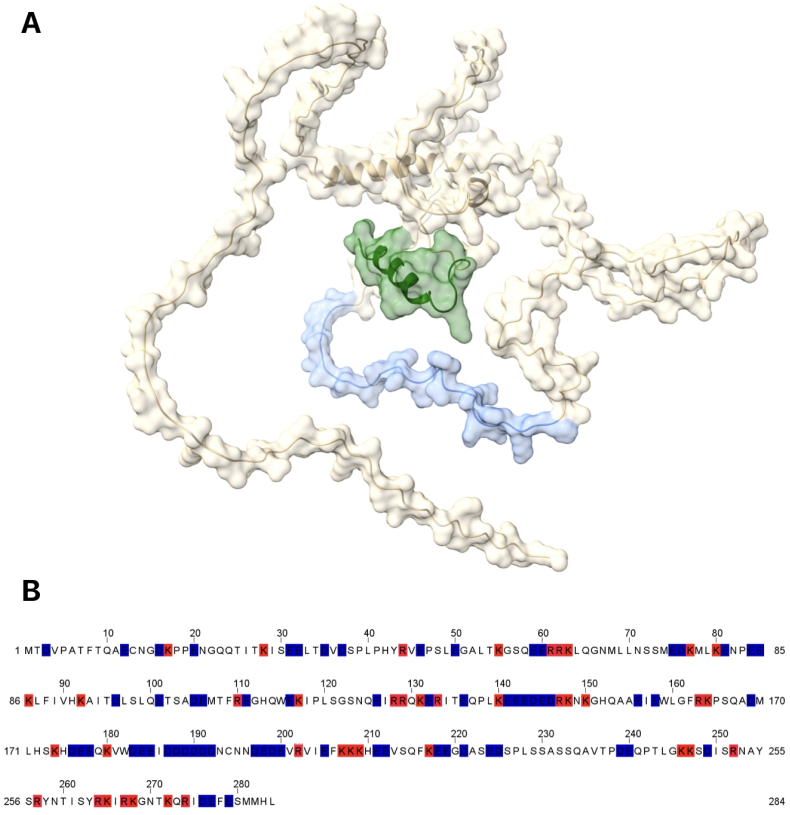

Figure 1. Structural and functional features of human Ermin (UniProt ID: Q8TAM6). AlphaFold-predicted three-dimensional structure of Ermin (AF-Q8TAM6-F1-v6; https://alphafold.ebi.ac.uk/entry/AF-Q8TAM6-F1) visualized using ChimeraX (Pettersen *et al.*, 2021). (A) The experimentally characterized F-actin binding site (Zhang *et al.*, 2005; Ruskamo *et al.*, 2012) is highlighted in green, and the acidic region (Ruskamo *et al.*, 2012) is shown in blue. (B) Amino acid sequence of human Ermin (UniProt ID: Q8TAM6), with positively charged residues (Arg, Lys) indicated in red and negatively charged residues (Asp, Glu) in blue.

Ermin lacks extensive stable secondary or tertiary structure and is predicted to contain large IDRs, consistent with sequence-based analyses and functional studies of the protein (Ruskamo *et al*., 2012). Intrinsically disordered regions are well suited to signaling, scaffolding, cytoskeletal regulation, and sensitivity to amino-acid composition and charge patterning, features that are particularly relevant for cytoskeleton-associated proteins (Wright and Dyson, 1999; Uversky *et al.*, 2000).

Despite growing interest in Ermin's cellular and pathological relevance, its evolutionary history across vertebrates has not been systematically examined. It remains unclear how conserved Ermin sequence features are across major vertebrate clades, whether intrinsic disorder and charge-related properties are evolutionarily stable or exhibit clade-specific variation, and how amino-acid compositional biases within disordered regions differ among lineages. Moreover, most existing studies of Ermin are restricted to mammalian systems, which limits broader evolutionary inference.

The primary objective of this study was to investigate the evolutionary diversification of intrinsic disorder in vertebrate Ermin using a comprehensive phylogenetic and sequence-based framework. Specifically, we aimed to (i) determine whether Ermin is consistently classified as a highly intrinsically disordered protein across major vertebrate lineages; (ii) characterize the architectural organization of intrinsically disordered regions (IDRs), with particular emphasis on the distribution, continuity, and fragmentation of long IDRs; (iii) quantify lineage-specific variation in amino-acid composition, hydrophobicity, and electrostatic features, including fraction of charged residues (FCR), net charge per residue (NCPR), and sequence charge decoration (SCD); and (iv) assess evolutionary trends and phylogenetic structuring of these properties using phylogenetically informed comparative methods.

By integrating disorder prediction, polymer-inspired electrostatic metrics, and phylogenetic generalized least squares (PGLS) analyses, we sought to test whether Ermin evolution is characterized by monotonic changes in disorder content or instead by lineage-specific remodelling of disorder architecture and sequence composition. Through this approach, we aim to establish a quantitative framework for understanding how intrinsically disordered proteins diversify over deep evolutionary timescales while maintaining conserved functional disorder.

Materials and methods

Sequence retrieval

Ermin amino acid sequences were collected using the same NCBI BLAST-based workflow (https://blast.ncbi.nlm.nih.gov) described in our previously published myelin basic protein study (Koszorus and Kósa, 2025). Sequence collection was restricted to species with available, well-annotated whole-genome assemblies. BLASTp searches were conducted separately for each major vertebrate clade using clade-specific Ermin reference sequences.

For cartilaginous fishes (*Chondrichthyes*), *Callorhinchus milii* (XP_007888107.2) was used as the reference sequence. In ray-finned fishes (*Actinopterygii*), two reference sequences were employed to account for the pronounced length variability observed among Ermin homologs in this clade: *Danio rerio* (XP_073761994.1), and *Perca fluviatilis* (XP_039674551.1). Single representative reference sequences were used for amphibians (*Xenopus tropicalis*, XP_002933329.2), reptiles (*Gekko japonicus*, XP_015267048.1), birds (*Gallus gallus*, XP_040531879.1), and mammals (*Homo sapiens*, Q8TAM6).

All BLAST parameters, filtering criteria, and orthology assignment steps were identical to those described previously (Koszorus and Kósa, 2025). In total, 159 Ermin sequences were curated, with clade-specific distribution (cartilaginous fishes: 13, ray-finned fishes: 50, amphibians: 16, reptiles: 17, birds: 19, and mammals: 44).

Prediction of protein disorder

The intrinsic disorder of Ermin proteins was predicted using the Rapid Intrinsic Disorder Analysis Online (RIDAO) platform (https://ridao.app, Dayhoff and Uversky, 2022), which incorporates six per residue predictors of disorder: PONDR VLXT (Li *et al.*, 1999), PONDR VL3 (Radivojac *et al.*, 2003), PONDR VSL2B (Peng *et al.*, 2006), PONDR FIT (Xue *et al.*, 2010), IUPred (short and long modes) (Dosztányi *et al.*, 2005, Erdős and Dosztányi, 2020). All predictors provide a score that characterizes the disorder propensity of each position along the sequence. Residues with scores above 0.5 are predicted as disordered, while scores below 0.5 indicate order. Predictor outputs were aggregated and exported as mean disorder profiles (MDP). The mean disorder profile was used to identify the intrinsically disordered regions (IDRs) in the protein, and allowed us to calculate the disorder properties related to the whole protein, such as the average disorder score (ADS), percentages of predicted disordered residues (PPDR), and the length and positions of the individual disordered regions.

Identification, classification, and evolutionary comparison of Ermin IDR architecture

Residues were classified as disordered (IDR) or ordered using a disorder threshold of 0.5, such that residues with disorder scores ≥0.5 were assigned to IDRs, while residues below this threshold were considered ordered. According to the length, intrinsically disordered regions were classified in short IDRs (SDR, 5-29 consecutive residues with disorder scores ≥0.5) (Monzon *et al.*, 2020) and long IDRs (LDR, ≥ 30 residues) (van der Lee *et al.*, 2014; Monzon *et al.*, 2020).

Each IDR was further assigned a positional class (N-terminal, internal, or C-terminal) based on its location relative to the full ungapped protein sequence. For each Ermin ortholog, summary descriptors were calculated, including the number of long and short IDRs, their lengths, and the fraction of amino acids residing in each IDR class.

To visualize the phylogenetic distribution of IDR architecture, IDR maps were plotted alongside a Neighbour-Joining phylogenetic tree inferred from aligned Ermin sequences (Fig. 2). IDRs were shown as horizontal bars scaled to sequence length, enabling direct comparison of disorder organization across clades.

Quantification of IDR fragmentation

To decouple overall disorder abundance from architectural organization, new IDR fragmentation indices were defined. Fragmentation indices normalize the number of disordered segments by the fraction of disordered amino acids, thereby capturing how disorder is partitioned rather than how much disorder is present. Two indices were calculated per sequence: Total-IDR fragmentation index (FI_{total}), $FI_{total} = N_{total\ IDRs}$ / *(fraction of amino acids in long + short IDRs),* where $N_{total\ IDRs}$ is the total number of IDRs; and Long-IDR fragmentation index (FI_{long}), $FI_{long} = N_{long\ IDRs}$ / *fraction of amino acids in long IDRs*, where $N_{long\ IDRs}$ is the number of long IDRs. Fragmentation indices were summarized by major vertebrate clades and compared statistically using non-parametric tests.

Amino-acid compositional bias of intrinsically disordered regions

To further characterize intrinsically disordered regions, we compared the amino-acid composition of residues classified as disordered with that of residues located in ordered regions. Amino-acid compositional bias between intrinsically disordered regions (IDRs) and ordered regions was quantified using per-residue disorder predictions and protein sequence data (Vacic *et al.*, 2007, Djulbegovic and Uversky, 2022).

For each amino acid, pooled frequencies were calculated separately for IDR and ordered regions within the same protein sequences, and aggregated across all sequences. Relative amino-acid composition was then computed as *Relative composition* = $(f_{IDR} - f_{ordered})/f_{ordered}$, where f_{IDR} and $f_{ordered}$ denote the fractional abundance of a given amino acid in IDR and ordered regions, respectively. Positive values indicate enrichment in IDRs, whereas negative values indicate depletion relative to ordered regions. Confidence intervals for relative composition estimates were obtained by non-parametric bootstrap resampling of protein sequences (10,000 iterations). In each bootstrap replicate, sequences were sampled with replacement, residue counts were pooled, and relative composition values were recalculated. The 2.5th and 97.5th percentiles of the bootstrap distributions were taken as the 95% confidence intervals. Amino acids whose confidence intervals overlapped zero were considered not significantly enriched or depleted.

Phylogenetic comparative framework

To investigate the evolutionary dynamics of Ermin sequence properties across vertebrates, we employed phylogenetic comparative analyses that explicitly account for non-independence among species due to shared evolutionary history. Phylogenetic generalized least squares (PGLS) regression was used as the primary statistical framework, following the classical formulation of phylogenetic regression and its subsequent extensions for comparative data analysis (Grafen, 1989; Martins and Hansen, 1997; Freckleton *et al.*, 2002). This approach incorporates the expected covariance structure among species derived from a phylogenetic tree and allows hypothesis testing on trait evolution while controlling for phylogenetic relatedness.

Phylogenetic tree reconstruction

A species phylogeny was constructed using amino-acid sequences aligned with ClustalW and inferred using the Neighbor-Joining (NJ) method implemented in MEGA12. The NJ algorithm provides an efficient distance-based approximation of phylogenetic relationships and is widely used for comparative and exploratory evolutionary analyses when the primary goal is to capture relative divergence among taxa rather than time-calibrated histories (Saitou and Nei, 1987; Kumar *et al.*, 2018). Branch lengths correspond to sequence divergence. For PGLS analyses, the exact topology and relative branch length structure are more critical than time calibration, and distance-based Neighbor-Joining trees are sufficient for estimating phylogenetic covariance among taxa when the objective is to control for shared ancestry rather than to reconstruct time-calibrated evolutionary histories (Grafen, 1989; Freckleton *et al.*, 2002).

Prior to PGLS analyses, the tree was midpoint-rooted, an established rooting approach when a reliable outgroup is unavailable, which places the root at the midpoint of the longest path between taxa and preserves relative branch length structure (Farris, 1972; Hess and De Moraes Russo, 2007). Because branch lengths represent evolutionary divergence rather than absolute time, root-to-tip distances were interpreted as relative evolutionary distances along the phylogeny rather than chronological age, and were used to assess directional trends in trait evolution rather than absolute evolutionary rates. Phylogenetic signal was quantified using Pagel's λ, which rescales internal branch lengths to best fit the observed trait covariance under a Brownian motion model (Pagel, 1999). λ was estimated by maximum likelihood for each fitted model and interpreted as a measure of the degree to which trait variation reflects phylogenetic structure, with values approaching 1 indicating strong phylogenetic signal and values near 0 indicating phylogenetic independence (Blomberg *et al.*, 2003).

Sequence-derived trait calculation

For each Ermin sequence, multiple physicochemical and disorder-related traits were computed. Disorder scores were obtained from per-residue disorder predictions (RIDAO MDP) using the mean disorder probability (MDP) score across residues.

Fraction of charged residues (FCR) was defined as the fraction of positively and negatively charged residues (FCR = $f^+ + f^-$, where $f^+ = N^+/N$ and $f^- = N^-/N$, N is total of residues in the sequence, and $N^+$and N^- the number of positive, negatively charged residues) (Das and Pappu, 2013; Das *et al*., 2015). Net charge per residue (NCPR) was calculated as the difference between positively and negatively charged residues (NCPR = $f^+ - f^-$), preserving the sign of the net charge (Das and Pappu, 2013; Das *et al.*, 2015). Sequence charge decoration (SCD) was computed following established definitions that quantify charge patterning along the sequence, with higher values indicating greater segregation of like charges (Sawle and Ghosh, 2015; Ghosh *et al*., 2022). The mean hydrophobicity was calculated using the Kyte–Doolittle hydrophobicity scale (Kyte and Doolittle, 1982).

Clade-wise PGLS analyses and pairwise contrasts

Univariate PGLS models were fitted separately for each sequence property to test for differences among major vertebrate clades (*Chondrichthyes, Actinopterygii, Amphibia, Reptilia, Aves,* and *Mammalia*), with clade identity treated as a categorical predictor. To formally assess whether clade identity improved model fit beyond phylogenetic structure alone, each full model

(including clade identity) was compared to a reduced, intercept-only model using likelihood-ratio tests (Burnham and Anderson, 2002). The resulting likelihood-ratio statistics and associated p-values were used to evaluate whether inclusion of clade identity improved model fit relative to phylogenetic covariance alone. Pairwise contrasts between clade means were derived from the fitted PGLS models to identify which clade pairs contributed most strongly to overall differences. To control for multiple testing across contrasts, p-values were adjusted using the Holm–Bonferroni procedure, which controls the family-wise error rate (Holm, 1979).

Root-to-tip trend analyses

To test for directional evolutionary trends, additional PGLS models were fitted in which each trait was regressed against root-to-tip phylogenetic distance, calculated as the sum of branch lengths from the root to each species. Under this framework, a significant regression slope, if detected, indicates a directional trend in trait evolution, whereas non-significant slopes are consistent with clade-structured or stationary evolution (Hunt, 2006). Root-to-tip analyses were conducted within a Brownian/PGLS framework rather than Ornstein–Uhlenbeck models to avoid over-parameterization and misinterpretation in the presence of strong phylogenetic signal (Cooper *et al.*, 2016).

Phylogenetic multiple regression

To assess the independent contributions of multiple sequence properties, a phylogenetic multiple regression was fitted including sequence length, mean hydrophobicity, FCR, NCPR, and SCD as predictors. Regression coefficients were standardized to allow direct comparison of effect sizes under the phylogenetic covariance structure. This analysis complements univariate and clade-wise comparisons by identifying which properties explain evolutionary variation in Ermin sequence architecture when other correlated traits are held constant.

Results

Vertebrate Ermin Proteins Are Highly Disordered

Proteins can be classified into structural disorder categories based on the Average Disorder Score (ADS) and the Percentage of Predicted Disordered Residues (PPDR), typically as highly ordered, moderately disordered, or highly disordered (Gsponer *et al.* 2008; Rajagopalan *et al.* 2011; Djulbegovic and Uversky, 2022). In our dataset, all 159 analyzed vertebrate Ermin sequences exhibited ADS values above 0.5 and PPDR values exceeding 30% (Table 1.).

According to commonly accepted classification criteria, these parameters place Ermin in the highly disordered category. This classification is consistent with previous observations for Myelin Basic Protein, which similarly falls within the highly disordered class (Koszorus and Kósa, 2025).

Table 1. Intrinsic disorder metrics and IDR architectural parameters of vertebrate Ermin across major clades. Descriptive summary of intrinsic disorder–related properties for 159 vertebrate Ermin sequences grouped by major lineages. Parameters include Average Disorder Score (ADS), Percent of Predicted Disordered Residues (PPDR), total number of intrinsically disordered regions (IDRs), number and mean length of long (LDRs; ≥30 residues) and short (SDRs; 5–29 residues) disordered regions, mean number of LDRs and SDRs per sequence, and the percentage of residues within LDRs and SDRs. Values are reported as clade-wise means with observed minimum–maximum ranges and are presented for descriptive purposes only.

Parameter	*Chondrichthyes*	*Actinopterygii*	*Amphibia*	*Reptilia*	*Aves*	*Mammalia*
Mean sequence length (aa)	419	268.2	245.25	255.76	252.42	279.14
PPDR (%)	92.7	88.19	87.67	92.93	97.46	86.45
ADS	0.72	0.72	0.713	0.724	0.81	0.69
Nr of IDRs/sequence (min-max)	1-6	1-6	1-4	1-3	1-2	2-4
Mean nr of IDRs/sequence	2.92	2.34	2.25	1.82	1.53	2.77
Percent of residues in IDR (%)	91.03	86.23	87.59	90.03	96.97	85.80
Nr of SDRs/sequence (min-max)	0-2	0-4	0-1	0-2	0-1	0-1
Mean nr of SDRs/sequence	0.61	0.96	0.81	0.65	0.32	0.09
Mean length of SDR (aa)	9.33	17.18	16.61	26.77	23.66	17
Percent of residues in SDR (%)	1.41	5.34	5.53	6.77	2.93	0.55
Nr of LDRs/sequence (min-max)	1-4	0-3	1-3	1-2	1-2	2-3
Mean nr of LDRs/sequence	2.30	1.38	1.44	1.18	1.21	2.68
Mean length of LDR (aa)	206.57	175.72	161.80	193.11	211.39	92.20
Percent of residues in LDR (%)	89.61	80.89	82.05	83.26	94.04	85.25

Conserved long-IDR–rich architecture across vertebrate Ermin proteins

Across all examined vertebrate Ermin orthologs (n = 159), protein architecture was dominated by long intrinsically disordered regions. In every major clade, long IDRs accounted for the majority of disordered residues, indicating that extensive disorder is a deeply conserved feature of Ermin.

Phylogeny-aligned visualization of IDR architecture revealed that this conservation masks substantial variation in how long disordered residues are organized within individual proteins (Fig. 2). While overall disorder content remained high across clades (Table 1), the number, continuity, and positional arrangement of IDRs differed markedly.

Clade-specific organization and localization of IDRs

C-terminal long IDRs were nearly universal across tetrapods, suggesting strong functional constraint on this region. In contrast, internal long IDRs displayed pronounced lineage specificity: they were frequent in mammals but rare or absent in birds and reptiles (Fig. 2).

Short IDRs were common in actinopterygians and amphibians, often located at N-terminal or internal positions, but were strongly reduced in birds and mammals. Notably, mammalian Ermin sequences exhibited an almost complete absence of short IDRs despite maintaining extensive long-IDR coverage, indicating a shift in how disorder is partitioned rather than a reduction in overall disorder per se.

Mammalian Ermin shows increased fragmentation of long IDRs

Quantitative analysis of fragmentation indices revealed clear clade-specific differences. Birds (*Aves*) exhibited the lowest total-IDR fragmentation index (FI_{total}), with a narrow distribution centered below the global median (Fig. 3A). In contrast, mammals showed significantly elevated FI_{total} values relative to birds, reflecting increased segmentation of disordered segments.

This pattern was even more pronounced for long-IDR fragmentation (FI_{long}). Birds exhibited the lowest FI_{long} values, consistent with largely continuous long IDRs, whereas mammals displayed a strong and statistically significant increase in FI_{long} (Fig. 3B). Intermediate fragmentation levels were observed in actinopterygians, amphibians, and reptiles, while chondrichthyans showed high variance, indicating heterogeneous IDR architectures.

These results demonstrate that mammalian Ermin proteins are characterized by increased partitioning of long disordered regions, rather than by an increase in total disorder content.

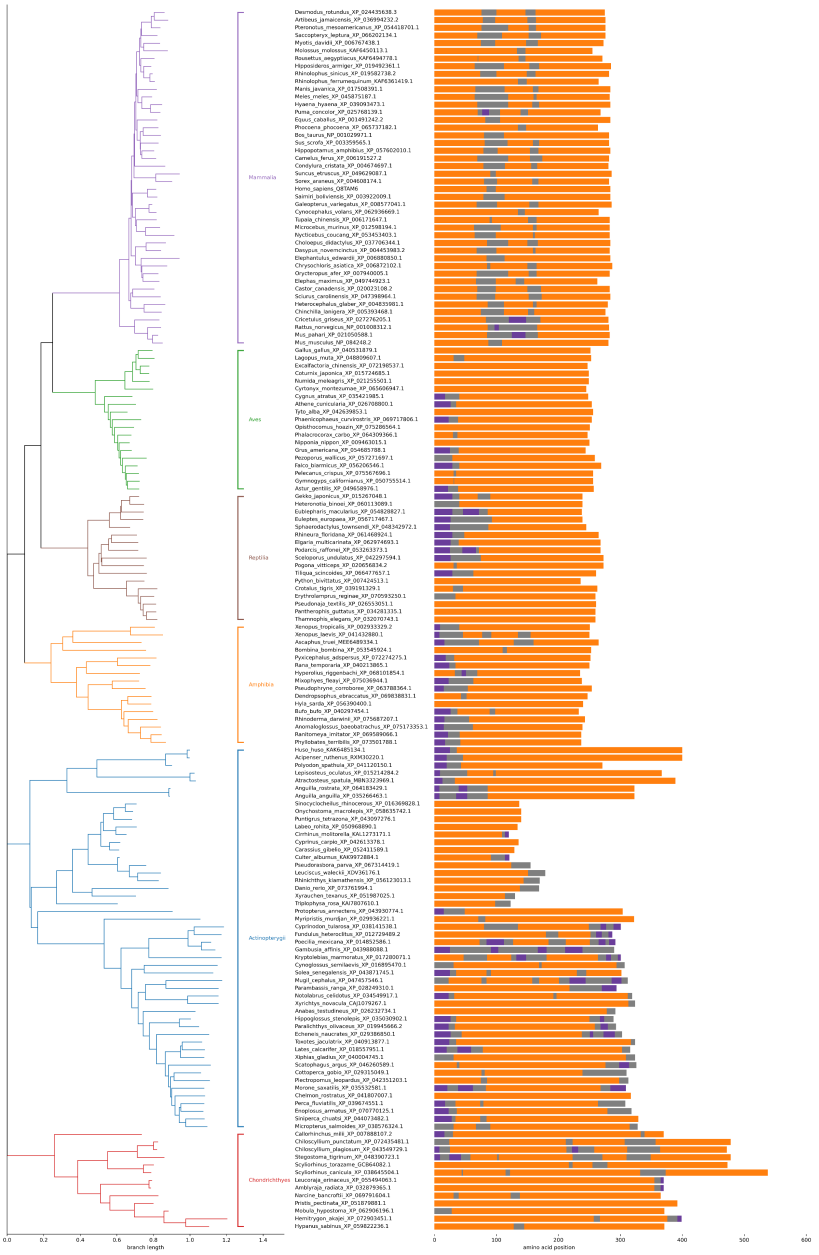

Figure 2. Phylogeny-linked IDR architecture of vertebrate Ermin. A phylogenetic tree of 159 Ermin sequences is shown with branches colored by major vertebrate clade. Horizontal bars depict per-sequence disorder architecture in ungapped sequence coordinates: ordered regions (grey), long IDRs (orange), and short IDRs (purple). Bars are aligned at the N-terminus and scaled by each sequence's ungapped length.

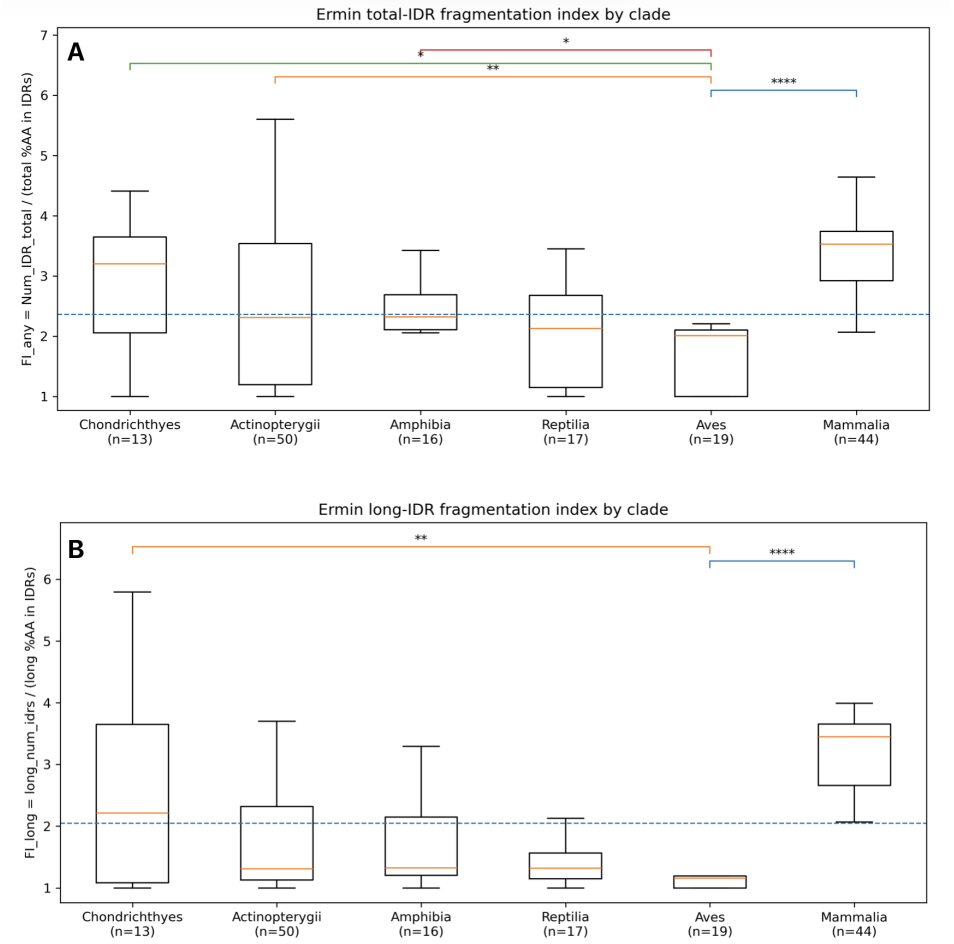

Figure 3. Clade-specific fragmentation of Ermin intrinsically disordered regions. Boxplots show the distribution of intrinsically disordered region (IDR) fragmentation indices across major vertebrate clades. (A) Total-IDR fragmentation index (FI_{total}), defined as the number of all IDRs normalized by the total fraction of amino acids classified as disordered. (B) Long-IDR fragmentation index (FI_{long}), defined as the number of long IDRs normalized by the fraction of amino acids classified as long-IDR. Boxes indicate the interquartile range with medians shown as horizontal lines; whiskers extend to 1.5× IQR. Numbers in parentheses denote the number of sequences per clade. The blue dashed line marks the global median fragmentation index across all Ermin sequences, serving as a vertebrate-wide reference. Statistical significance was assessed using Kruskal–Wallis tests followed by Dunn's post-hoc comparisons with Holm correction. Brackets indicate pairwise comparisons relative to the avian clade (***Aves***), which exhibits the lowest fragmentation and serves as a baseline for continuous IDR architecture (* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$).

Global features of vertebrate IDR composition

Analysis of amino-acid composition revealed pronounced differences between intrinsically disordered regions (IDRs) and ordered regions of Ermin proteins (Fig. 4). IDRs were significantly depleted in aromatic (tryptophan, tyrosine, phenylalanine, histidine) and large hydrophobic residues (isoleucine, leucine, valine, methionine), consistent with reduced hydrophobic packing requirements in disordered conformations. In contrast, IDRs exhibited strong enrichment in proline, as well as increased representation of polar and charged residues, including serine, glutamine, arginine, aspartate, and glutamate.

Proline showed the strongest and most consistent enrichment among all amino acids, reflecting its well-established role in disrupting secondary structure and promoting conformational flexibility. Charged residues displayed moderate but significant enrichment, supporting an IDR architecture dominated by electrostatic interactions rather than hydrophobic collapse. Cysteine was markedly depleted in IDRs, consistent with its preferential involvement in structured domains and disulfide-bond formation.

Bootstrap confidence interval analysis indicated that most observed compositional biases were statistically robust across sequences. Only a small subset of residues exhibited confidence intervals overlapping zero, indicating limited or variable enrichment across species. Across vertebrate clades, IDRs display both conserved disorder signatures and striking lineage-specific compositional specializations. Chondrichthyans exhibit predominantly polar, uncharged IDRs enriched in serine and threonine, with strong depletion of aromatic and hydrophobic residues and minimal charge enrichment, yielding a low-aromatic, low-hydrophobic architecture (Fig. S1A). In contrast, *Actinopterygii* show the most pronounced acidic signature, with strong enrichment of glutamate and aspartate and exceptional relative enrichment of proline, indicating a highly charged, electrostatically driven IDR composition (Fig. S1B). Amphibian IDRs retain acidic dominance but display increased compositional heterogeneity, with marked enrichment of asparagine and glutamate alongside variable contributions from basic and aromatic residues (Fig. S1C). Reptilian IDRs are distinguished by extreme and heterogeneous enrichment of histidine, threonine, and lysine, superimposed on conserved depletion of cysteine and bulky hydrophobics, suggesting clade-specific modulation of charge and potential pH sensitivity (Fig. S1D).

Aves exhibit the strongest low-complexity profile, characterized by enrichment of methionine, leucine, cysteine, proline, threonine, arginine and glutamate and profound depletion of branched hydrophobic residues (isoleucine, valine), as well as near absence of aromatics, reflecting highly flexible, compositionally simplified

IDRs (Fig. S1E). Finally, Mammalian IDRs combine strong enrichment of disorder-promoting residues — particularly arginine, and proline — with marked depletion of bulky aromatics, yielding a composition that integrates charge enrichment with maintained proline-driven structural disruption (Fig. S1F). Collectively, these patterns indicate progressive diversification of IDR electrostatic and low-complexity features across vertebrate evolution, with distinct clade-specific biases superimposed on a conserved depletion of aromatic and hydrophobic residues.

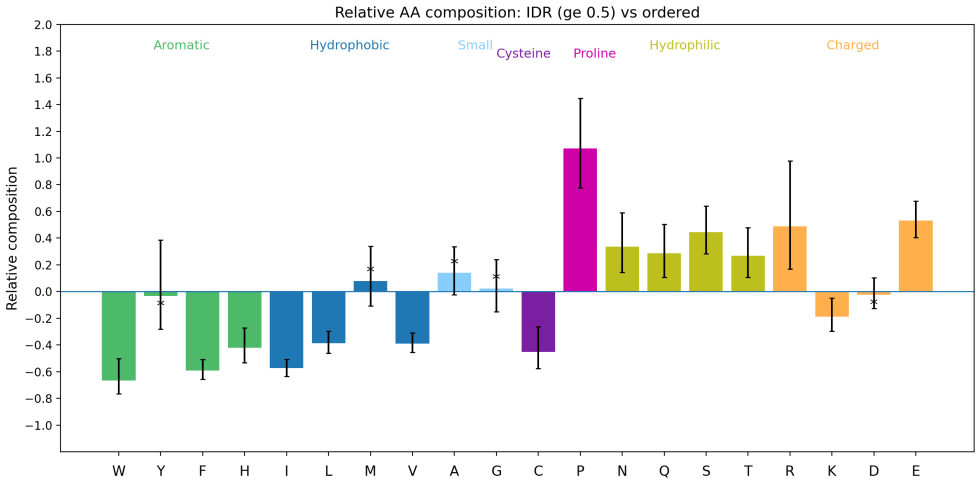

Figure 4. Relative amino-acid composition of intrinsically disordered regions (IDRs) versus ordered regions in Ermin proteins. Relative composition was calculated as $(f_{IDR} - f_{ordered}) / f_{ordered}$; positive values indicate enrichment in IDRs and negative values indicate depletion. Amino acids are grouped by physicochemical properties, with cysteine and proline shown separately. Error bars represent 95% bootstrap confidence intervals (10,000 sequence resamplings). Asterisks denote residues whose confidence intervals overlap zero (non-significant enrichment or depletion).

Phylogenetic distribution of Ermin sequence properties

Mapping z-score–normalized sequence properties onto the Ermin phylogeny revealed pronounced structuring across disorder, hydrophobicity, and electrostatic traits (Fig. 5). Closely related species clustered together in trait space, and distinct clade-level signatures were evident for multiple properties. Disorder propensity showed a non-monotonic distribution across vertebrates: early-diverging lineages, particularly chondrichthyans, already exhibited high disorder values; birds displayed the highest disorder overall; and mammals showed

comparatively lower disorder than birds and several fish lineages. Thus, high intrinsic disorder is not restricted to derived vertebrate groups and does not increase monotonically across the phylogeny.

In contrast, mean Kyte–Doolittle hydrophobicity showed a clearer lineage-dependent pattern, with mammals and reptiles exhibiting more negative values than most fish lineages, consistent with reduced hydrophobicity. Electrostatic properties also displayed strong lineage dependence: fraction of charged residues (FCR), net charge per residue (NCPR), and sequence charge decoration (SCD) each showed clade-specific patterns, with SCD in particular separating vertebrate groups in the phylogeny-aligned heatmap. Pagel's λ values were consistently high across traits ($\lambda \approx 0.85$–1.0), indicating substantial phylogenetic signal and strong covariance among species.

Clade-wise distributions and pairwise PGLS contrasts

Clade-wise comparisons of raw trait distributions confirmed the lineage-specific patterns observed in the phylogenetic mapping (Fig. 6). Disorder propensity exhibited substantial variation across clades but did not follow a simple monotonic trajectory. Birds exhibited the highest disorder distributions, whereas mammals showed lower disorder relative to birds and some early-diverging lineages. Mean hydrophobicity differed among clades, with mammals and reptiles tending toward more negative values, consistent with reduced hydrophobicity. Electrostatic traits also varied among clades. FCR and SCD showed broader interclade separation than NCPR, which remained within a relatively narrow range across vertebrates.

To formally evaluate clade effects while accounting for phylogenetic non-independence, univariate PGLS models including clade identity were compared with reduced intercept-only models using likelihood-ratio tests. In all cases, inclusion of clade identity did not significantly improve model fit after accounting for phylogenetic covariance (Table S2). These results indicate that the apparent clade-level differences observed in raw trait distributions are largely explained by shared evolutionary history rather than by statistically discrete shifts among categorical clades.

Consistent with this finding, pairwise PGLS contrasts among clade means revealed directional differences in several traits; however, none of the individual pairwise comparisons remained significant after Holm–Bonferroni correction for multiple testing (Table S1). The absence of statistically significant pairwise contrasts reflects the combined effects of strong phylogenetic signal (high Pagel's λ values), moderate within-clade variance, and conservative multiple-testing correction.

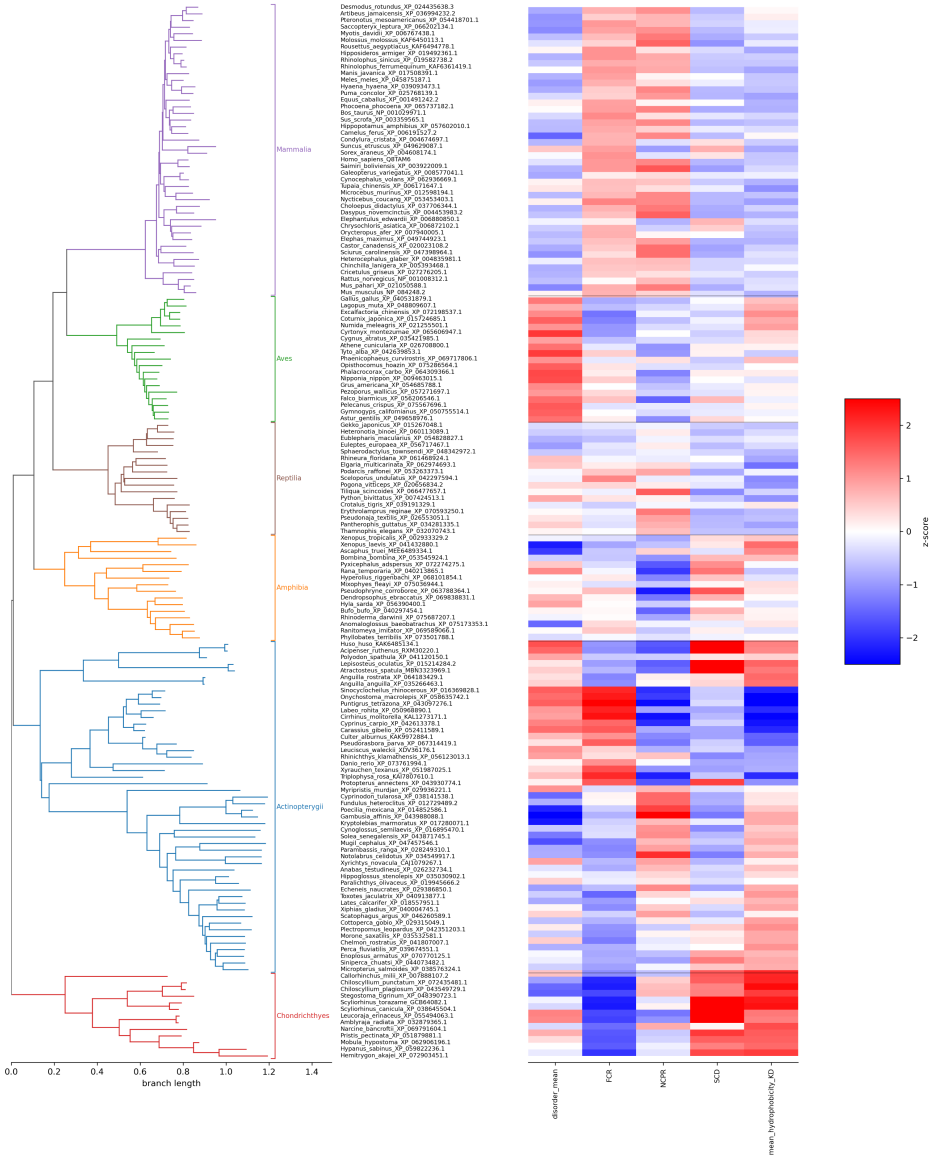

Figure 5. Phylogenetic distribution of disorder, mean hydrophobicity and electrostatic properties of Ermin across vertebrates. A species phylogeny of vertebrate Ermin sequences is shown on the left, with branches colored by major clades (*Chondrichthyes, Actinopterygii, Amphibia, Reptilia, Aves, Mammalia*). Brackets indicate clade boundaries. To the right, a heatmap displays z-score–normalized values of mean disorder, fraction of charged residues (FCR), net charge per residue (NCPR), sequence charge decoration (SCD) and mean hydrophobicity for each species, aligned to the corresponding tree tip. Red and blue indicate values above and below the overall mean, respectively.

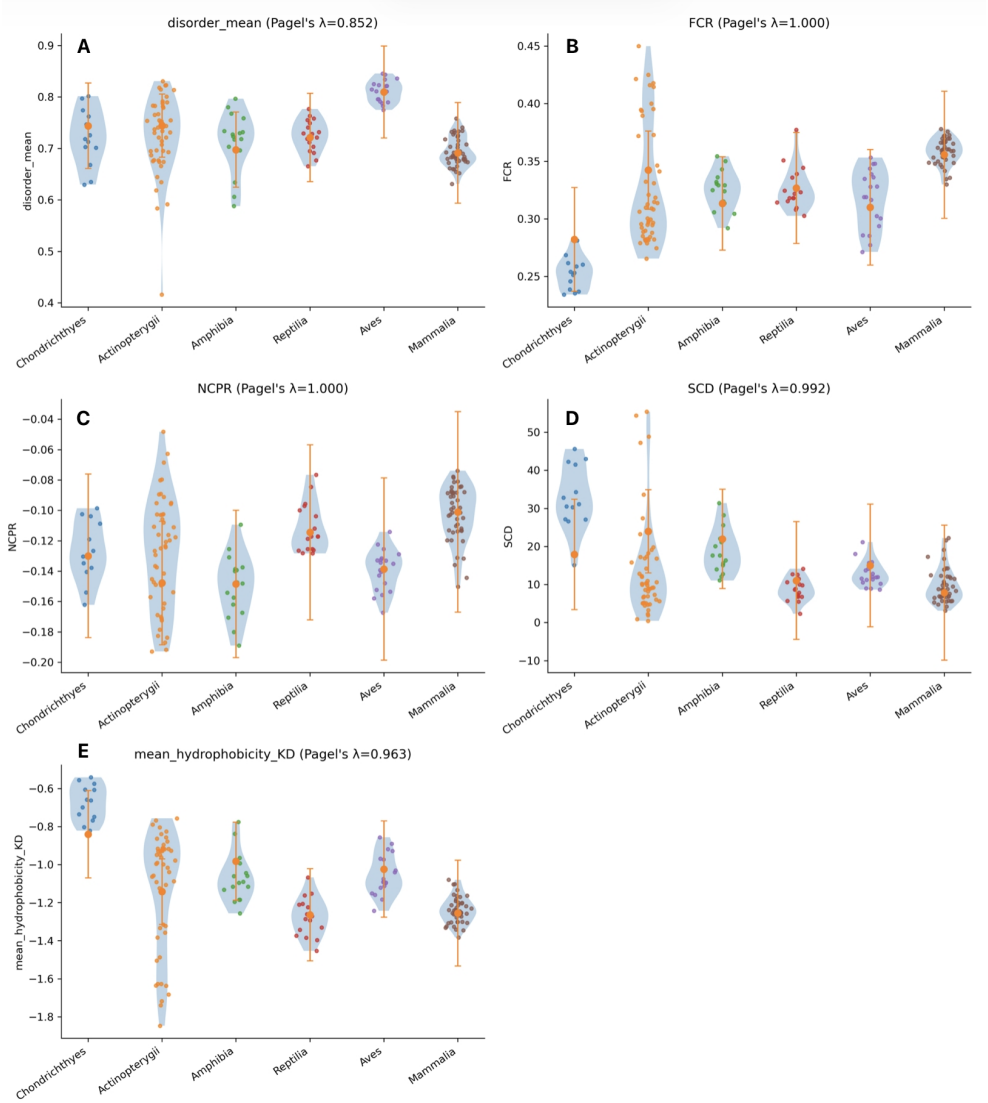

Figure 6. Evolutionary changes in intrinsic disorder, mean hydrophobicity and electrostatic properties of vertebrate Ermin. Violin plots show the distribution of sequence-level properties across major vertebrate clades (*Chondrichthyes*, *Actinopterygii*, *Amphibia*, *Reptilia*, *Aves*, *Mammalia*). Points represent individual species; violins indicate kernel density estimates. Orange circles and error bars denote phylogenetically corrected clade means ± 95% confidence intervals estimated using PGLS. (A) Mean disorder score, (B) fraction of charged residues (FCR), (C) net charge per residue (NCPR), (D) sequence charge decoration (SCD), (E) mean hydrophobicity. Pagel's λ values are shown for each trait, indicating strong phylogenetic signal, particularly for electrostatic properties.

Taken together, these analyses suggest that variation in ermin sequence properties is structured along the phylogeny in a gradual, lineage-dependent manner rather than partitioned into sharply distinct clade-specific regimes. The data therefore support a model of distributed evolutionary tuning across vertebrate lineages rather than discrete categorical shifts among major clades.

Root-to-tip analyses reveal limited directional trends

Root-to-tip PGLS regressions were performed to assess directional evolutionary trends across the vertebrate Ermin phylogeny (Table 2). Mean Kyte–Doolittle hydrophobicity exhibited a highly significant positive slope ($p = 0.0008$), indicating a progressive decrease in hydrophobicity (shift toward more negative KD values) with increasing phylogenetic distance from the root. Fraction of charged residues (FCR) also showed a statistically significant but modest negative slope ($p = 0.0152$), suggesting a weak directional reduction in overall charge density. In contrast, net charge per residue (NCPR), sequence charge decoration (SCD), and mean disorder propensity did not display significant monotonic trends. Pagel's λ values were high for all traits, indicating substantial phylogenetic signal.

Collectively, these findings indicate that while hydrophobicity — and to a lesser extent overall charge density — exhibits directional evolutionary change, variation in disorder propensity and charge patterning is better explained by lineage-specific restructuring rather than by uniform trends across the vertebrate phylogeny.

Multivariate determinants of Ermin sequence evolution

To assess the independent contributions of multiple sequence features, a multivariate PGLS regression was fitted including sequence length, mean hydrophobicity, FCR, NCPR, and SCD (Fig. 7). Standardized regression coefficients revealed that mean hydrophobicity had a strong negative association with evolutionary variation in Ermin sequence properties, whose confidence intervals lay entirely below zero, indicating that reduced hydrophobicity independently characterizes Ermin evolution after accounting for correlated traits.

Fraction of charged residues (FCR) also showed a significant negative independent contribution, consistent with a modest reduction in overall charge density across vertebrate evolution. Sequence charge decoration (SCD) exhibited a significant positive independent effect, whereas sequence length and NCPR showed confidence intervals overlapping zero, indicating no independent contribution when considered jointly with other predictors.

Table 2. Root-to-tip PGLS trends in Ermin sequence evolution. Results of phylogenetic generalized least squares (PGLS) regressions assessing directional evolutionary trends in Ermin sequence properties across vertebrates. For each trait, the regression slope with respect to root-to-tip phylogenetic distance (Slope) and its associated p-value are reported, together with the maximum-likelihood estimate of Pagel's λ and the model log-likelihood (logLik). The log-likelihood denotes the maximum log-likelihood of the fitted PGLS model under the estimated phylogenetic covariance structure. Positive slope values indicate increasing trait values with increasing phylogenetic divergence from the root, whereas negative slope values indicate decreasing trait values. Because Kyte–Doolittle hydrophobicity values are expressed on a negative scale, positive slope values for hydrophobicity correspond to decreasing hydrophobicity over evolutionary divergence.

Trait	**N**	Pagel's λ	**logLik**	**Slope (root-to-tip)**	**p (slope)**
FCR	159	0.9999	428.8252	-0.0633	0.0152
NCPR	159	0.9999	409.018	0.0448	0.1293
SCD	159	0.9846	-508.504	17.4608	0.0598
Mean disorder	159	0.8732	287.877	-0.0825	0.1444
Mean hydrophobicity	159	0.9875	146.0139	0.504	0.0008

These results demonstrate that Ermin sequence evolution is primarily driven by compositional and patterning changes rather than by gross changes in protein length or net charge alone.

Taken together, the phylogeny-aligned trait mapping, clade-wise PGLS contrasts, root-to-tip analyses, and multivariate regression converge on a consistent evolutionary picture. Ermin sequences are intrinsically disordered across vertebrates, but disorder propensity is modulated in a lineage-specific, non-monotonic manner. Hydrophobicity shows a clearer directional decrease over evolutionary time, whereas electrostatic properties — especially charge patterning — undergo pronounced clade-specific restructuring. These results indicate that Ermin evolution is dominated by lineage-dependent tuning of sequence composition and electrostatic organization rather than by a simple, uniform increase in intrinsic disorder.

Discussion

Long IDRs as a conserved structural core of Ermin

The predominance of long intrinsically disordered regions across all vertebrate Ermin orthologs demonstrates that extensive disorder is a fundamental and conserved structural feature of the protein. Long IDRs are commonly associated with scaffolding functions, multivalent interactions, and regulatory flexibility (Wright and Dyson, 2015; Uversky, 2019), consistent with Ermin's role in cytoskeletal organization and oligodendrocyte biology (Zhang *et al.*, 2005; Meng *et al.*, 2010; Wang *et al.*, 2011). The near-universal presence of a C-terminal long IDR further suggests strong functional constraint, likely reflecting conserved interaction interfaces encompassing the characterized actin-binding segment (Zhang *et al.,* 2005; Ruskamo *et al.*, 2012) and potentially additional regulatory modules.

Fragmentation as an independent evolutionary dimension of disorder

By explicitly quantifying IDR fragmentation, we demonstrate that disorder organization evolves independently of overall disorder abundance. Birds represent a low-fragmentation regime in which long disordered regions remain largely continuous, whereas mammals exhibit a high-fragmentation architecture characterized by multiple long IDRs separated by ordered segments. This architectural distinction would not be detected by disorder fraction alone and highlights fragmentation as a distinct evolutionary parameter. These findings align with emerging views that IDP function depends critically on spatial organization and modular segmentation rather than solely on disorder content (van der Lee *et al.*, 2014; Darling and Uversky, 2018).

IDR amino acid composition: conserved constraints and lineage-specific diversification

Across vertebrates, IDR composition is shaped by conserved physicochemical constraints, reflected in consistent depletion of aromatic and bulky hydrophobic residues. The stability of these depletion patterns — particularly for tryptophan, phenylalanine, isoleucine, and leucine — supports the view that avoidance of structure-stabilizing chemistry represents a fundamental requirement for intrinsic disorder (Williams *et al.*, 2001; Lieutaud *et al.*, 2016; Oldfield *et al.*, 2019).

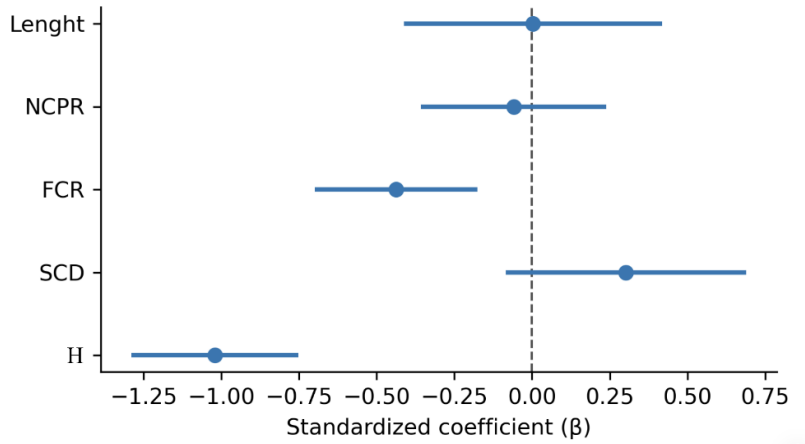

Figure 7. Multivariate PGLS regression coefficients for Ermin sequence properties. Standardized regression coefficients (β ± 95% confidence intervals) from a phylogenetic multiple regression including sequence length, fraction of charged residues (FCR), net charge per residue (NCPR), sequence charge decoration (SCD), and mean hydrophobicity (H) as predictors. The dashed vertical line indicates β = 0. Positive and negative coefficients represent independent contributions to evolutionary variation in Ermin sequence properties after accounting for phylogenetic relatedness. Pagel's λ = 0.93 corresponds to the maximum-likelihood estimate of phylogenetic signal for the fitted multivariate PGLS model.

In contrast, enrichment patterns vary markedly among clades, indicating that intrinsic disorder can be achieved through multiple compositional strategies. Early-diverging vertebrates rely predominantly on polar uncharged residues, whereas actinopterygians exhibit pronounced acidic enrichment. Amphibians and reptiles display increased compositional heterogeneity, while birds and mammals show the strongest low-complexity signatures. When analysed jointly, lineage-specific enrichments partially average out, revealing a conserved vertebrate IDR signature dominated by proline, glutamate, and polar residues, consistent with proteome-wide analyses of disordered regions (Djulbegovic and Uversky, 2022). Thus, vertebrate IDRs appear to be governed by universal biophysical constraints combined with clade-dependent compositional diversification.

Decoupling disorder from compositional evolution

A central outcome of this study is the decoupling of intrinsic disorder from compositional evolutionary trends. Root-to-tip analyses reveal a strong directional decrease in mean hydrophobicity and a weaker but statistically significant reduction in overall charge density (FCR) across vertebrate evolution, whereas disorder propensity and sequence charge decoration (SCD) do not

exhibit consistent monotonic trends. Multivariate analyses further demonstrate that hydrophobicity, global charge density, and charge patterning independently contribute to evolutionary divergence.

Because Kyte–Doolittle hydrophobicity values for disordered proteins are typically negative, shifts toward more negative values reflect progressive selection against hydrophobic residues rather than increased structural order. The concurrent reduction in global charge density suggests gradual compositional retuning, while the independent effect of SCD indicates lineage-specific restructuring of charge distribution along the sequence. Together, these results show that Ermin maintains a conserved disordered scaffold while undergoing directional compositional shifts and clade-specific electrostatic remodelling.

Phylogenetic structuring of sequence evolution

High Pagel's λ values across traits indicate substantial phylogenetic signal, confirming that variation in Ermin sequence properties is strongly structured by shared ancestry. Importantly, strong phylogenetic signal does not preclude adaptation; rather, it indicates that evolutionary changes accumulate in a lineage-dependent manner such that closely related species resemble one another more than expected under independence (Pagel, 1999; Blomberg et al., 2003). Although clade-wise distributions and phylogeny-aligned heatmaps visually suggest lineage-specific differences, formal PGLS likelihood-ratio tests indicate that inclusion of categorical clade identity does not significantly improve model fit once phylogenetic covariance is accounted for. Similarly, pairwise PGLS contrasts do not remain significant after conservative multiple-testing correction. These findings suggest that variation in Ermin sequence architecture is not partitioned into statistically discrete clade-level regimes but instead reflects gradual, phylogenetically structured divergence along evolutionary lineages. Thus, the observed clustering in heatmaps and trait distributions is best interpreted as the outcome of shared evolutionary history and distributed lineage-dependent tuning rather than abrupt evolutionary shifts confined to particular vertebrate clades. Ermin sequence evolution therefore appears to proceed through incremental modifications along the phylogeny rather than through categorical regime changes.

Functional context: Ermin among myelin-associated IDPs

Placing Ermin within the broader context of myelin-associated intrinsically disordered proteins underscores its functional distinctiveness. Classical myelin proteins such as myelin basic protein (MBP) and the P0 cytoplasmic tail undergo partial folding upon membrane association and act as molecular

adhesives within compact myelin (Boggs, 2006; Harauz and Musse, 2007). In contrast, Ermin functions primarily as an actin-associated cytoskeletal regulator, favoring sustained disorder, low hydrophobicity, and finely tuned electrostatic organization. The evolutionary patterns observed here — directional reduction in hydrophobicity, modest decrease in global charge density, and lineage-specific modulation of charge patterning — are consistent with a highly dynamic interaction scaffold rather than a membrane-anchored structural protein.

Evolutionary implications

Collectively, phylogenetic mapping, root-to-tip regressions, and multivariate PGLS analyses converge on a model in which Ermin evolution is dominated by gradual, lineage-structured compositional tuning of an already intrinsically disordered scaffold. Intrinsic disorder itself is deeply conserved across vertebrates and does not exhibit a uniform directional increase. Instead, early-diverging lineages already display high disorder levels, indicating that extensive intrinsic disorder represents an ancestral structural state rather than a derived innovation.

In contrast, mean hydrophobicity shows a consistent directional shift along the phylogeny, reflecting progressive selection against hydrophobic residues over vertebrate divergence. Electrostatic properties, particularly charge patterning, vary in a phylogenetically structured but non-discrete manner, suggesting distributed modulation of ensemble behavior rather than categorical regime shifts among major clades. The absence of statistically separable clade effects indicates that evolutionary change in Ermin is best understood as incremental divergence accumulating along branches rather than as abrupt transitions at major taxonomic boundaries.

These findings reinforce the view that intrinsic disorder is an emergent property shaped by multiple interacting sequence features, including hydrophobicity, charge density, and charge patterning, rather than by a single evolutionary axis. Ermin evolution therefore exemplifies how intrinsically disordered proteins diversify through subtle compositional redistribution and electrostatic retuning while maintaining a conserved disordered structural framework. By integrating phylogenetic comparative methods with polymer- and disorder-based metrics, this study provides a quantitative framework for understanding how IDPs evolve across deep evolutionary timescales through gradual lineage-dependent modification rather than discrete structural innovation.

Conclusion

This study provides a comprehensive phylogenetic and sequence-level analysis of 159 vertebrate Ermin orthologs and establishes Ermin as a deeply conserved intrinsically disordered protein. Across all examined clades, Ermin consistently meets criteria for high intrinsic disorder, indicating that extensive disorder represents an ancestral and fundamental structural feature rather than a lineage-specific innovation.

Despite this conserved disordered framework, Ermin evolution is characterized by lineage-dependent remodeling of disorder architecture and composition. Mammals exhibit increased fragmentation of long intrinsically disordered regions (IDRs), whereas birds maintain largely continuous long-IDR architectures, demonstrating that IDR partitioning evolves independently of overall disorder abundance. Amino-acid composition analyses reveal conserved depletion of aromatic and bulky hydrophobic residues, combined with clade-structured variation in charged and disorder-promoting residues.

Phylogenetic comparative analyses further indicate a directional decrease in mean hydrophobicity across vertebrate divergence, consistent with progressive selection against hydrophobic residues within an already disordered scaffold. In contrast, disorder propensity and charge patterning do not exhibit uniform monotonic trends, and formal clade-wise tests do not support discrete categorical regime shifts after accounting for phylogenetic covariance. Instead, evolutionary variation in Ermin sequence properties is best explained by gradual, lineage-structured divergence along the phylogeny.

Together, these findings support a model in which Ermin evolution proceeds through incremental architectural restructuring and electrostatic retuning of a conserved intrinsically disordered framework. This highlights IDR fragmentation, compositional redistribution, and charge organization as key evolutionary dimensions shaping diversification of intrinsically disordered proteins across vertebrate lineages.

References

Aggarwal, S., Yurlova, L. & Simons, M., 2011. Central nervous system myelin: structure, synthesis and assembly. *Trends Cell Biol.* 21, 585–593. https://doi.org/10.1016/j.tcb.2011.06.004

Bah, A. & Forman-Kay, J.D. 2016. Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications. *J. Biol. Chem.* 291, 6696–6705. https://doi.org/10.1074/jbc.R115.695056

Baumann, N. & Pham-Dinh, D., 2001. Biology of Oligodendrocyte and Myelin in the Mammalian Central Nervous System. *Physiol. Rev.* 81, 871–927. https://doi.org/10.1152/physrev.2001.81.2.871

Bellay, J., Han, S., Michaut, M., Kim, T., Costanzo, M., Andrews, B.J., Boone, C., Bader, G.D., Myers, C.L. & Kim, P.M., 2011. Bringing order to protein disorder through comparative genomics and genetic interactions. *Genome Biol.* 12, R14. https://doi.org/10.1186/gb-2011-12-2-r14

Blomberg, S.P., Garland, T. & Ives, A.R., 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57, 717–745. https://doi.org/10.1111/j.0014-3820.2003.tb00285.x

Boggs, J.M., 2006. Myelin basic protein: a multifunctional protein. *Cell. Mol. Life Sci.* 63, 1945–1961. https://doi.org/10.1007/s00018-006-6094-7

Brockschnieder, D., Sabanay, H., Riethmacher, D. & Peles, E., 2006. Ermin, A Myelinating Oligodendrocyte-Specific Protein That Regulates Cell Morphology. *J. Neurosci.* 26, 757–762. https://doi.org/10.1523/JNEUROSCI.4317-05.2006

Burnham, K.P. & Anderson, D.R., 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd ed. Springer, New York. http://doi.org/10.1007/b97636

Cooper, N., Thomas, G.H., Venditti, C., Meade, A. & Freckleton, R.P., 2016. A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. *Biol. J. Linn. Soc.* 118, 64–77. https://doi.org/10.1111/bij.12701

Darling, A.L. & Uversky, V.N., 2018. Intrinsic Disorder and Posttranslational Modifications: The Darker Side of the Biological Dark Matter. *Front. Genet.* 9, 158. https://doi.org/10.3389/fgene.2018.00158

Das, R.K. & Pappu, R.V., 2013. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U.S.A.* 110, 13392–13397. https://doi.org/10.1073/pnas.1304749110

Das, R.K., Ruff, K.M. & Pappu, R.V., 2015. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 32, 102–112. https://doi.org/10.1016/j.sbi.2015.03.008

Dayhoff, G.W. & Uversky, V.N., 2022. Rapid prediction and analysis of protein intrinsic disorder. *Protein Sci.* 31, e4496. https://doi.org/10.1002/pro.4496

Djulbegovic, M. & Uversky, V.N., 2022. The aqueous humor proteome is intrinsically disordered. Biochemistry and *Biophys. Rep.* 29, 101202. https://doi.org/10.1016/j.bbrep.2022.101202

Dosztányi, Z., Csizmók, V., Tompa, P. & Simon, I., 2005. The Pairwise Energy Content Estimated from Amino Acid Composition Discriminates between Folded and Intrinsically Unstructured Proteins. *J. Mol. Biol.* 347, 827–839. https://doi.org/10.1016/j.jmb.2005.01.071

Erdős, G. & Dosztányi, Z., 2020. Analyzing Protein Disorder with IUPred2A. *Curr. Prot. Bioinform.* 70, e99. https://doi.org/10.1002/cpbi.99

Farris, J.S., 1972. Estimating Phylogenetic Trees from Distance Matrices. *Am. Nat.* 106, 645–668. https://doi.org/10.1086/282802

Freckleton, R.P., Harvey, P.H. & Pagel, M., 2002. Phylogenetic Analysis and Comparative Data: A Test and Review of Evidence. *Am. Nat.* 160, 712–726. https://doi.org/10.1086/343873

Ghosh, K., Huihui, J., Phillips, M. & Haider, A., 2022. Rules of Physical Mathematics Govern Intrinsically Disordered Proteins. *Annu. Rev. Biophys.* 51, 355–376. https://doi.org/10.1146/annurev-biophys-120221-095357

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

Gsponer, J., Futschik, M.E., Teichmann, S.A. & Babu, M.M., 2008. Tight Regulation of Unstructured Proteins: From Transcript Synthesis to Protein Degradation. *Science 322, 1365–1368*. https://doi.org/10.1126/science.1163581

Harauz, G. & Boggs, J.M., 2013. Myelin management by the 18.5-kDa and 21.5-kDa classic myelin basic protein isoforms. *J. Neurochem.* 125, 334–361. https://doi.org/10.1111/jnc.12195

Harauz, G. & Musse, A.A., 2007. A Tale of Two Citrullines—Structural and Functional Aspects of Myelin Basic Protein Deimination in Health and Disease. *Neurochem. Res.* 32, 137–158. https://doi.org/10.1007/s11064-006-9108-9

Hess, P.N. & De Moraes Russo, C.A., 2007. An empirical test of the midpoint rooting method: a test for the midpoint rooting method. *Biol. J. Linn. Soc.* 92, 669–674. https://doi.org/10.1111/j.1095-8312.2007.00864.x

Holm, S., 1979. A Simple Sequentially Rejective Multiple Test Procedure. *Scand. J. Stat.* 6, 65–70. https://doi.org/10.2307/4615733

Holehouse, A.S., & Kragelund, B.B., 2024. The molecular basis for cellular function of intrinsically disordered protein regions. *Nat. Rev. Mol. Cell. Biol.* 25, 187–211. https://doi.org/10.1038/s41580-023-00673-0

Homs, A., Codina-Solà, M., Rodríguez-Santiago, B., Villanueva, C.M., Monk, D., Cuscó, I. & Pérez-Jurado, L.A., 2016. Genetic and epigenetic methylation defects and implication of the ERMN gene in autism spectrum disorders. *Transl. Psychiatry* 6, e855–e855. https://doi.org/10.1038/tp.2016.120

Hunt, G., 2006. Fitting and comparing models of phyletic evolution: random walks and beyond. *Paleobiology* 32, 578–601. https://doi.org/10.1666/05070.1

Jahn, O., Tenzer, S. & Werner, H.B., 2009. Myelin Proteomics: Molecular Anatomy of an Insulating Sheath. *Mol. Neurobiol.* 40, 55–72. https://doi.org/10.1007/s12035-009-8071-2

Koszorus, I. & Kósa, F., 2025. Evolutionary patterns of structural disorder and post-translational modifications in the 18.5 kDa myelin basic protein. Studia UBB Biologia 70, 115–150. https://doi.org/10.24193/subbbiol.2025.2.08

Kragelund, B.B., Jensen, M.K. & Skriver, K., 2012. Order by disorder in plant signaling. *Trends. Plant. Sci.* 17, 625–632. https://doi.org/ 10.1016/j.tplants.2012.06.010

Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K., 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* 35, 1547–1549. https://doi.org/10.1093/molbev/msy096

Kyte, J. & Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105–132. https://doi.org/10.1016/0022-2836(82)90515-0

Liang F, Hwang JH, Tang NW & Hunziker W., 2018. Juxtanodin in retinal pigment epithelial cells: Expression and biological activities in regulating cell morphology and actin cytoskeleton organization. *J. Comp. Neurol.* 526(2):205-215. https://doi.org/10.1002/cne.24301

Li X, Romero P, Rani M, Dunker AK & Obradovic Z. Predicting Protein Disorder for N-, C-, and Internal Regions. *Genome Inform. Ser. Workshop Genome Inform.* 1999;10:30-40. PMID: 11072340.

Lieutaud, P., Ferron, F., Uversky, A.V., Kurgan, L., Uversky, V.N. & Longhi, S., 2016. How disordered is my protein and what is its disorder for? A guide through the "dark side" of the protein universe. *Intrinsically Disord. Proteins* 4, e1259708. https://doi.org/10.1080/21690707.2016.1259708

Martins, E.P. & Hansen, T.F., 1997. Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am. Nat.* 149, 646–667. https://doi.org/10.1086/286013

Meng J, Xia W, Tang J, Tang BL & Liang F., 2010. Dephosphorylation-dependent inhibitory activity of juxtanodin on filamentous actin disassembly. *J. Biol. Chem.* 10;285(37):28838-49 https://doi.org/10.1074/jbc.M110.117887

Monzon, A.M., Necci, M., Quaglia, F., Walsh, I., Zanotti, G., Piovesan, D. & Tosatto, S.C.E., 2020. Experimentally Determined Long Intrinsically Disordered Protein Regions Are Now Abundant in the Protein Data Bank. *IJMS.* 21, 4496. https://doi.org/10.3390/ijms21124496

Nave, K.-A. & Werner, H.B., 2014. Myelination of the Nervous System: Mechanisms and Functions. *Annu. Rev. Cell Dev. Biol.* 30, 503–533. https://doi.org/10.1146/annurev-cellbio-100913-013101

Nawaz, S., Sánchez, P., Schmitt, S., Snaidero, N., Mitkovski, M., Velte, C., Brückner, B.R., Alexopoulos, I., Czopka, T., Jung, S.Y., Rhee, J.S., Janshoff, A., Witke, W., Schaap, I.A.T., Lyons, D.A. & Simons, M., 2015. Actin Filament Turnover Drives Leading Edge Growth during Myelin Sheath Formation in the Central Nervous System. *Dev. Cell* 34, 139–151. https://doi.org/10.1016/j.devcel.2015.05.013

Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C.J. & Dunker, A.K., 2003. Predicting intrinsic disorder from amino acid sequence. *Proteins* 53, 566–572. https://doi.org/10.1002/prot.10532

Oldfield, C.J., Uversky, V.N., Dunker, A.K. & Kurgan, L., 2019. Introduction to intrinsically disordered proteins and regions, in: Intrinsically Disordered Proteins. *Elsevier*, pp. 1–34. https://doi.org/10.1016/B978-0-12-816348-1.00001-6

Pagel, M., 1999. Inferring the historical patterns of biological evolution. *Nature* 401, 877–884. https://doi.org/10.1038/44766

Peng, K., Radivojac, P., Vucetic, S., Dunker, A.K. & Obradovic, Z., 2006. Length-dependent prediction of protein intrinsic disorder. *BMC. Bioinformatics* 7, 208. https://doi.org/10.1186/1471-2105-7-208

Peng, Z., Yan, J., Fan, X., Mizianty, M.J., Xue, B., Wang, K., Hu, G., Uversky, V.N. & Kurgan, L., 2015. Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell. Mol. Life Sci.* 72, 137–151. https://doi.org/10.1007/s00018-014-1661-9

Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H. & Ferrin, T.E., 2021. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. 30, 70–82. https://doi.org/10.1002/pro.3943

Raasakka, A. & Kursula, P., 2020. Flexible Players within the Sheaths: The Intrinsically Disordered Proteins of Myelin in Health and Disease. *Cells* 9, 470. https://doi.org/10.3390/cells9020470

Radivojac, P., Obradovic, Z., Brown, C.J. & Dunker A.K., 2003. Prediction of boundaries between intrinsically ordered and disordered protein regions. *Pac. Symp. Biocomput*. 216–227

Rajagopalan, K., Mooney, S.M., Parekh, N., Getzenberg, R.H. & Kulkarni, P., 2011. A majority of the cancer/testis antigens are intrinsically disordered proteins. *J. Cell. Biochem. 112, 3256–3267.* https://doi.org/10.1002/jcb.23252

Ruskamo, S., Chukhlieb, M., Vahokoski, J., Bhargav, S.P., Liang, F., Kursula, I. & Kursula, P., 2012. Juxtanodin is an intrinsically disordered F-actin-binding protein. *Sci. Rep.* 2, 899. https://doi.org/10.1038/srep00899

Salek Esfahani, B., Gharesouran, J., Ghafouri-Fard, S., Talebian, S., Arsang-Jang, S., Omrani, M.D., Taheri, M. & Rezazadeh, M., 2019. Down-regulation of ERMN expression in relapsing remitting multiple sclerosis. *Metab. Brain Dis.* 34, 1261–1266. https://doi.org/10.1007/s11011-019-00429-w

Sawle, L. & Ghosh, K., 2015. A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins. *J. Chem. Phys.* 143, 085101. https://doi.org/10.1063/1.4929391

Shiva, S., Gharesouran, J., Sabaie, H., Asadi, M.R., Arsang-Jang, S., Taheri, M. & Rezazadeh, M., 2021. Expression Analysis of Ermin and Listerin E3 Ubiquitin Protein Ligase 1 Genes in Autistic Patients. *Front. Mol. Neurosci.* 14, 701977. https://doi.org/10.3389/fnmol.2021.701977

Saitou N. & Nei M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406-425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

Uversky, V.N., 2019. Intrinsically Disordered Proteins and Their “Mysterious” (Meta)Physics. *Front. Phys.* 7, 10. https://doi.org/10.3389/fphy.2019.00010

Uversky, V.N., Gillespie, J.R. & Fink, A.L., 2000. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins* 41, 415–427. https://doi.org/10.1002/1097-0134(20001115)41:3<415::AID-PROT130>3.0.CO;2-7

Vacic, V. & Iakoucheva, L.M., 2012. Disease mutations in disordered regions—exception to the rule? *Mol. BioSyst.* 8, 27–32. https://doi.org/10.1039/C1MB05251A

Vacic, V., Markwick, P.R.L., Oldfield, C.J., Zhao, X., Haynes, C., Uversky, V.N. & Iakoucheva, L.M., 2012. Disease-Associated Mutations Disrupt Functionally Important Regions of Intrinsic Protein Disorder. *PLoS. Comput. Biol.* 8, e1002709. https://doi.org/10.1371/journal.pcbi.1002709

Vacic, V., Uversky, V.N., Dunker, A.K. & Lonardi, S., 2007. Composition Profiler: a tool for discovery and visualization of amino acid composition differences. *BMC. Bioinformatics* 8, 211. https://doi.org/10.1186/1471-2105-8-211

Van Der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., Kim, P.M., Kriwacki, R.W., Oldfield, C.J., Pappu, R.V., Tompa, P., Uversky, V.N., Wright, P.E. & Babu, M.M., 2014. Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* 114, 6589–6631. https://doi.org/10.1021/cr400525m

Wang, S., Wang, T., Liu, T., Xie, R., Zhao, X., Wang, L., Yang, Q., Jia, L. & Han, J., 2020. Ermin is a $p116^{RIP}$ -interacting protein promoting oligodendroglial differentiation and myelin maintenance. *Glia* 68, 2264–2276. https://doi.org/10.1002/glia.23838

Wang, T., Jia, L., Lv, B., Liu, B., Wang, W., Wang, F., Yang, G., Bu, X., Yao, L. & Zhang, B., 2011. Human Ermin (hErmin), a new oligodendrocyte-specific cytoskeletal protein related to epileptic seizure. *Brain Res.* 1367, 77–84. https://doi.org/10.1016/j.brainres.2010.10.003

Williams, R.M., Obradovic, Z., Mathura, V., Braun, W., Garner, E.C., Young, J., Takayama, S., Brown, C.J. & Dunker, A.K., 2001. The protein non-folding problem: Amino acid determinants of intrinsic order and disorder. In: *Biocomputing 2001*. Proceedings of the Pacific Symposium on Biocomputing. World Scientific, Mauna Lani, Hawaii, pp. 89–100. https://doi.org/10.1142/9789814447362_0010

Wright, P.E. & Dyson, H.J., 2015. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* 16, 18–29. https://doi.org/10.1038/nrm3920

Wright, P.E. & Dyson, H.J., 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* 293, 321–331. https://doi.org/10.1006/jmbi.1999.3110

Xue, B., Dunbrack, R.L., Williams, R.W., Dunker, A.K. & Uversky, V.N., 2010. PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta Proteins Proteom.* 1804, 996–1010. https://doi.org/10.1016/j.bbapap.2010.01.011

Zhang, B., Cao, Q., Guo, A., Chu, H., Chan, Y.G., Buschdorf, J.P., Low, B.C., Ling, E.A. & Liang, F., 2005. Juxtanodin: An oligodendroglial protein that promotes cellular arborization and 2′,3′-cyclic nucleotide-3′-phosphodiesterase trafficking. *Proc. Natl. Acad. Sci. U.S.A.* 102, 11527–11532. https://doi.org/10.1073/pnas.0500952102

Ziaei, A., Garcia-Miralles, M., Radulescu, C.I., Sidik, H., Silvin, A., Bae, H., Bonnard, C., Yusof, N.A.B.M., Ferrari Bardile, C., Tan, L.J., Ng, A.Y.J., Tohari, S., Dehghani, L., Henry, L., Yeo, X.Y., Lee, S., Venkatesh, B., Langley, S.R., Shaygannejad, V., Reversade, B., Jung, S., Ginhoux, F. & Pouladi, M.A., 2022. Ermin deficiency leads to compromised myelin, inflammatory milieu, and susceptibility to demyelinating insult. *Brain Path.* 32(5):e13064. https://doi.org/10.1111/bpa.13064

Supplementary materials

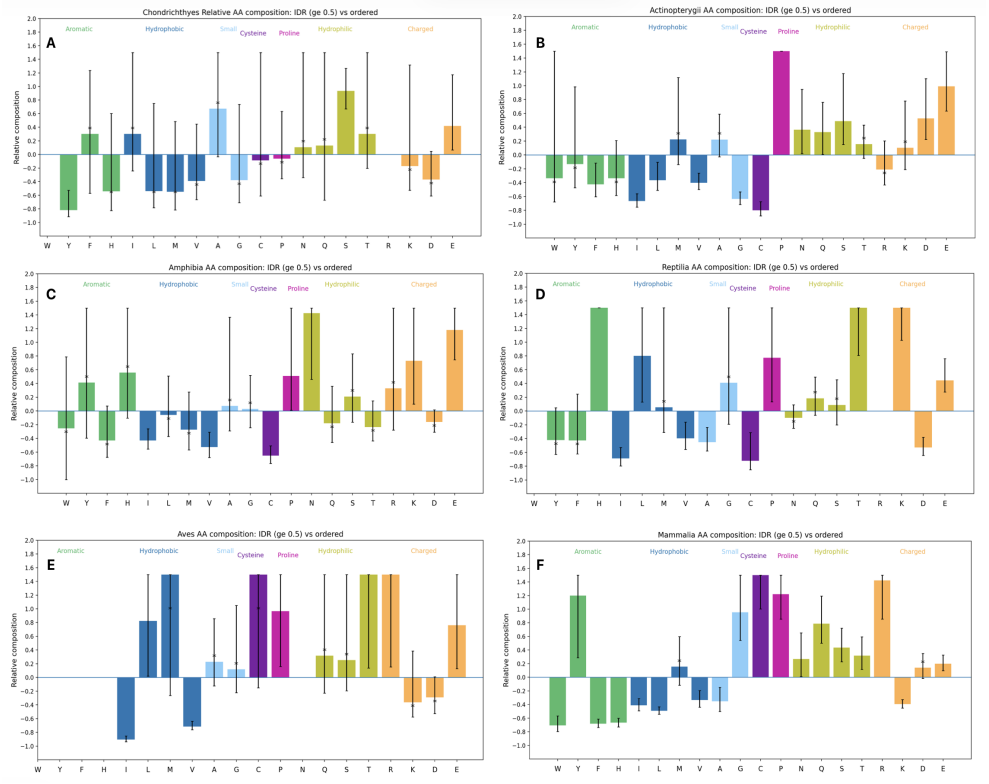

Figure S1. Relative amino-acid composition of intrinsically disordered regions (IDRs) versus ordered regions in Ermin proteins. A. *Chondrichthyes*, B. *Actinopterygii*, C. *Amphibia*, D. *Reptilia*, E. *Aves*, F. *Mammalia.* Relative composition was calculated as $(f_{IDR} - f_{ordered}) / f_{ordered}$; positive values indicate enrichment in IDRs and negative values indicate depletion. Amino acids are grouped by physicochemical properties, with cysteine and proline shown separately. Error bars represent 95% bootstrap confidence intervals (10,000 sequence resamplings). Bootstrap resampling was performed at the sequence level. Asterisks denote residues whose confidence intervals overlap zero (non-significant enrichment or depletion).

Table S1. Pairwise PGLS clade contrasts for Ermin sequence properties.
Differences represent estimated contrasts between clade means while accounting for phylogenetic covariance. Column definitions: trait — trait analyzed (FCR, NCPR, SCD, disorder_mean, mean_hydrophobicity_KD); Clade A — first clade in the comparison; Clade B — second clade in the comparison; Mean difference (A − B) — estimated difference between clade means from the fitted PGLS model (effect size in the original trait units); p (raw) — unadjusted p-value for the clade contrast; p (Holm-adjusted) — Holm–Bonferroni–adjusted p-value controlling the family-wise error rate across pairwise contrasts. Note: Positive values of Mean difference (A − B) indicate higher trait values in Clade A relative to Clade B; negative values indicate the opposite. No pairwise comparisons remained significant after Holm-Bonferroni correction.

Trait	Clade A	Clade B	Mean difference (A − B)	p (raw)	p (Holm-adjusted)
FCR	*Actinopterygii*	*Chondrichthyes*	0.0601	0.0414	0.6215
FCR	*Chondrichthyes*	*Mammalia*	-0.0735	0.0424	0.6215
FCR	*Aves*	*Mammalia*	-0.0456	0.1152	1.0000
FCR	*Amphibia*	*Mammalia*	-0.0423	0.1750	1.0000
FCR	*Chondrichthyes*	*Reptilia*	-0.0446	0.1873	1.0000
FCR	*Actinopterygii*	*Amphibia*	0.0289	0.2748	1.0000
FCR	*Actinopterygii*	*Aves*	0.0322	0.2903	1.0000
FCR	*Amphibia*	*Chondrichthyes*	0.0312	0.3168	1.0000
FCR	*Mammalia*	*Reptilia*	0.0290	0.3446	1.0000
FCR	*Aves*	*Chondrichthyes*	0.0279	0.4197	1.0000
FCR	*Aves*	*Reptilia*	-0.0166	0.5560	1.0000
FCR	*Actinopterygii*	*Reptilia*	0.0156	0.5994	1.0000
FCR	*Amphibia*	*Reptilia*	-0.0134	0.6329	1.0000
FCR	*Actinopterygii*	*Mammalia*	-0.0134	0.6830	1.0000
FCR	*Amphibia*	*Aves*	0.0033	0.9093	1.0000
NCPR	*Amphibia*	*Mammalia*	-0.0474	0.2044	1.0000
NCPR	*Actinopterygii*	*Mammalia*	-0.0468	0.2333	1.0000
NCPR	*Aves*	*Mammalia*	-0.0375	0.2787	1.0000
NCPR	*Amphibia*	*Reptilia*	-0.0340	0.3093	1.0000
NCPR	*Actinopterygii*	*Reptilia*	-0.0334	0.3458	1.0000
NCPR	*Aves*	*Reptilia*	-0.0241	0.4751	1.0000
NCPR	*Chondrichthyes*	*Mammalia*	-0.0289	0.5046	1.0000
NCPR	*Actinopterygii*	*Chondrichthyes*	-0.0179	0.6124	1.0000
NCPR	*Amphibia*	*Chondrichthyes*	-0.0185	0.6205	1.0000
NCPR	*Chondrichthyes*	*Reptilia*	-0.0155	0.7004	1.0000
NCPR	*Mammalia*	*Reptilia*	0.0134	0.7155	1.0000
NCPR	*Amphibia*	*Aves*	-0.0099	0.7752	1.0000
NCPR	*Actinopterygii*	*Aves*	-0.0093	0.7993	1.0000
NCPR	*Aves*	*Chondrichthyes*	-0.0086	0.8353	1.0000
NCPR	*Actinopterygii*	*Amphibia*	0.0006	0.9851	1.0000
SCD	*Actinopterygii*	*Mammalia*	16.1180	0.1260	1.0000
SCD	*Amphibia*	*Mammalia*	14.1134	0.1590	1.0000

Trait	Clade A	Clade B	Mean difference (A − B)	p (raw)	p (Holm-adjusted)
SCD	*Actinopterygii*	*Reptilia*	12.9310	0.1742	1.0000
SCD	*Amphibia*	*Reptilia*	10.9264	0.2237	1.0000
SCD	*Actinopterygii*	*Aves*	8.9645	0.3594	1.0000
SCD	*Chondrichthyes*	*Mammalia*	10.0325	0.3886	1.0000
SCD	*Aves*	*Mammalia*	7.1535	0.4416	1.0000
SCD	*Amphibia*	*Aves*	6.9599	0.4526	1.0000
SCD	*Actinopterygii*	*Chondrichthyes*	6.0855	0.5214	1.0000
SCD	*Chondrichthyes*	*Reptilia*	6.8455	0.5285	1.0000
SCD	*Aves*	*Reptilia*	3.9665	0.6621	1.0000
SCD	*Amphibia*	*Chondrichthyes*	4.0809	0.6839	1.0000
SCD	*Mammalia*	*Reptilia*	-3.1870	0.7460	1.0000
SCD	*Aves*	*Chondrichthyes*	-2.8790	0.7957	1.0000
SCD	*Actinopterygii*	*Amphibia*	2.0046	0.8137	1.0000
mean disorder	*Aves*	*Mammalia*	0.1183	0.0222	0.3326
mean disorder	*Amphibia*	*Aves*	-0.1119	0.0306	0.4282
mean disorder	*Aves*	*Reptilia*	0.0880	0.0819	1.0000
mean disorder	*Actinopterygii*	*Aves*	-0.0651	0.2316	1.0000
mean disorder	*Aves*	*Chondrichthyes*	0.0656	0.2957	1.0000
mean disorder	*Actinopterygii*	*Amphibia*	0.0468	0.3254	1.0000
mean disorder	*Actinopterygii*	*Mammalia*	0.0532	0.3629	1.0000
mean disorder	*Amphibia*	*Chondrichthyes*	-0.0462	0.4179	1.0000
mean disorder	*Chondrichthyes*	*Mammalia*	0.0527	0.4190	1.0000
mean disorder	*Mammalia*	*Reptilia*	-0.0303	0.5793	1.0000
mean disorder	*Amphibia*	*Reptilia*	-0.0239	0.6342	1.0000
mean disorder	*Actinopterygii*	*Reptilia*	0.0229	0.6650	1.0000
mean disorder	*Chondrichthyes*	*Reptilia*	0.0224	0.7154	1.0000
mean disorder	*Amphibia*	*Mammalia*	0.0064	0.9083	1.0000
mean disorder	*Actinopterygii*	*Chondrichthyes*	0.0006	0.9917	1.0000
mean hydrophobicity	*Chondrichthyes*	*Reptilia*	0.4230	0.0135	0.2028
mean hydrophobicity	*Chondrichthyes*	*Mammalia*	0.4140	0.0238	0.3328
mean hydrophobicity	*Actinopterygii*	*Chondrichthyes*	-0.3015	0.0448	0.5830
mean hydrophobicity	*Amphibia*	*Reptilia*	0.2816	0.0461	0.5830
mean hydrophobicity	*Amphibia*	*Mammalia*	0.2726	0.0834	0.9176
mean hydrophobicity	*Aves*	*Reptilia*	0.2403	0.0921	0.9214
mean hydrophobicity	*Aves*	*Mammalia*	0.2313	0.1134	1.0000
mean hydrophobicity	*Actinopterygii*	*Amphibia*	-0.1600	0.2317	1.0000
mean hydrophobicity	*Aves*	*Chondrichthyes*	-0.1827	0.2972	1.0000
mean hydrophobicity	*Amphibia*	*Chondrichthyes*	-0.1415	0.3717	1.0000
mean hydrophobicity	*Actinopterygii*	*Reptilia*	0.1216	0.4162	1.0000
mean hydrophobicity	*Actinopterygii*	*Aves*	-0.1187	0.4397	1.0000
mean hydrophobicity	*Actinopterygii*	*Mammalia*	0.1126	0.4964	1.0000
mean hydrophobicity	*Amphibia*	*Aves*	0.0413	0.7768	1.0000
mean hydrophobicity	*Mammalia*	*Reptilia*	0.0090	0.9536	1.0000

Table S2. Likelihood-ratio tests of clade effects in univariate PGLS models. Summary of univariate PGLS models testing whether vertebrate clade identity improves model fit for each Ermin sequence trait. For each trait, a full model including clade identity was compared to a reduced (intercept-only) model using a likelihood-ratio test (LRT). Pagel's λ was estimated by maximum likelihood for both full and reduced models. Row definitions: Trait — analyzed sequence property (mean disorder, mean hydrophobicity, FCR, NCPR, or SCD); n — number of species included in the model; Pagel's λ (full) — maximum-likelihood estimate of phylogenetic signal in the full model; logLik (full) — log-likelihood of the model including clade identity; logLik (reduced) — log-likelihood of the intercept-only model; LR statistic — likelihood-ratio statistic calculated as 2 × (logLik_full − logLik_reduced); df — degrees of freedom corresponding to the number of clade parameters; p (LRT, clade effect) — p-value of the likelihood-ratio test evaluating the significance of clade identity. Significant LRT p-values indicate that clade membership explains trait variation beyond phylogenetic covariance alone.

Trait	FCR	NCPR	SCD	Mean disorder	Mean hydrophobicity
n	159	159	159	159	159
λ (full model)	1.0000	1.0000	0.9919	0.8516	0.9633
logLik (full model)	437.6059	409.2020	-484.9054	290.5840	162.0218
logLik (reduced model)	434.0968	407.8885	-486.6154	287.1705	156.9697
LR statistic	7.0182	2.6270	3.4201	6.8270	10.1044
df	5	5	5	5	5
p (LRT, clade effect)	0.2193	0.7573	0.6355	0.2338	0.0723