Lineage-Specific Adaptive Evolution of the Mosquito Fibrinogen-Related Protein FBN30 at a Predicted Parasite-Facing Interface

Krishnendu Sinha[1*]

[1]Department of Zoology, Jhargram Raj College, Jhargram-721507

***Corresponding author**

Dr. Krishnendu Sinha
Assistant Professor in Zoology
Jhargram Raj College
Jhargram-721507
dr.krishnendusinha@gmail.com

Abstract:

Fibrinogen-related proteins (FREPs) contribute to mosquito-parasite interactions, yet the evolutionary processes shaping their functional diversification remain poorly resolved. The mosquito protein FBN30 has been implicated in restricting *Plasmodium* development, but its molecular basis of action is unknown. Here, the study examines the evolutionary history of FBN30 across *Anopheles* mosquitoes to test whether lineage-specific adaptive evolution has modified its functional properties. Codon-based analyses of FBN30 orthologs from 29 *Anopheles* species reveal a single episode of strong episodic diversifying selection confined to the *Anopheles darlingi* lineage. Site-level tests identify a positively selected residue within the conserved fibrinogen-like (FBG) domain. Ancestral sequence reconstruction shows that this site underwent a serine-to-asparagine substitution along the *A. darlingi* lineage, with structural modeling indicating only modest local effects on protein stability. Using protein-protein docking and binding affinity prediction as a proxy for functional engagement, the study finds that the reconstructed ancestral FBN30 exhibits significantly stronger predicted affinity for *Plasmodium falciparum* α-tubulin-1 than the extant *A. darlingi* protein, whereas the derived substitution alone does not account for this difference. These results indicate that evolutionary divergence in FBN30 is associated with reduced predicted engagement at a parasite-facing interface and support a model in which inhibitory mosquito proteins undergo fine-scale adaptive refinement under parasite-mediated selective pressures.

Keywords: *Anopheles darlingi, Plasmodium falciparum,* FBN30, α-tubulin-1*,* ookinete

1. INTRODUCTION

Malaria transmission depends on molecular interactions between *Plasmodium* parasites and their *Anopheles* mosquito vectors, particularly during ookinete traversal of the midgut epithelium [1]. This stage of the life cycle represents a critical evolutionary interface, where parasite success is shaped by compatibility with vector-derived immune and recognition factors. Among these, fibrinogen-related proteins (FREPs) constitute a diverse family of mosquito innate immune proteins implicated in modulating parasite development [2–8].

FREPs share a conserved fibrinogen-like (FBG) domain that mediates ligand recognition, yet individual family members differ markedly in their effects on parasite transmission. The best-characterized example, FREP1, facilitates *Plasmodium* invasion by interacting with α-tubulin-1 exposed on the ookinete surface [5–7]. In contrast, other FREPs appear to inhibit parasite development, suggesting functional diversification within the family. One such protein, FBN30, has been shown through functional genetic studies to restrict *Plasmodium* infection in *Anopheles gambiae*, as silencing of FBN30 results in increased parasite loads [9]. Despite this phenotype, the molecular and evolutionary basis of FBN30's inhibitory role remains unresolved.

FBN30 is a fibrinogen-related protein that adopts the conserved FREP/FBN structural architecture, characterized by a signal peptide for secretion and a C-terminal fibrinogen-like (FBG) domain of approximately 200 amino acids[10]. The FBG domain is predicted to form a β-sheet–rich fold with flexible surface-exposed loops that constitute the primary interaction interface for ligand binding, consistent with other invertebrate fibrinogen-related proteins[11]. In vivo, FBN30 assembles into a higher-order octameric complex, indicating quaternary structural organization beyond the monomeric FBG fold and suggesting cooperative or avidity-enhanced binding properties. While atomic-resolution structural data are not yet available, the strong conservation of the FBG core implies structural constraint, whereas naturally occurring substitutions, particularly within the signal peptide and putative surface regions, are expected to modulate protein abundance or interaction efficiency without disrupting overall fold integrity [10,11].

From an evolutionary perspective, FREPs represent compelling candidates for parasite-mediated selection. The FBG domain is structurally constrained yet features surface-exposed regions that may accommodate adaptive modification, enabling fine-scale tuning of

recognition or binding properties without disrupting overall protein architecture. However, it remains unclear whether FBN30 has evolved primarily under purifying selection or whether specific mosquito lineages exhibit signatures of episodic adaptive evolution indicative of host–parasite antagonism.

Here, the study investigates the evolutionary history of FBN30 across *Anopheles* mosquitoes using comparative genomics, codon-based models of molecular evolution, ancestral sequence reconstruction, and structural modeling. By integrating evolutionary inference with predicted functional consequences, the study tests whether lineage-specific adaptive evolution has shaped FBN30 at a putative parasite-facing interface. This approach provides an evolutionary framework for understanding how inhibitory mosquito proteins may be refined under parasite-mediated selective pressures, generating testable hypotheses about their role in mosquito–*Plasmodium* interactions.

2. MATERIALS AND METHODS

2.1. Retrieval of FBG30 sequence and ortholog identification

The protein sequence of FBG30 from *Anopheles gambiae* PEST (VectorBase gene ID: AGAP006914) was retrieved from VectorBase[12]. These 280 amino-acid sequences served as the query for ortholog identification across *Anopheles* clade and the outgroup, *Culex quinquefasciatus* (VectorBase gene ID: CPIJ000937). Complete proteomes for 29 *Anopheles* and one *Culex* species available in VectorBase (release 68, accessed November 2025) were downloaded. Orthologs were identified using a reciprocal BLAST hit (RBH) workflow implemented through an in-house Python script implementing soft masking and Smith–Waterman alignments[13] with an E-value threshold of 1×10^{-5}, and a minimum alignment coverage of 90%. Twenty-nine candidate sequences returning *A. gambiae* FBG30 as the top reciprocal match were accepted as true orthologs[14]. All sequences were further validated for the presence of the fibrinogen-related (FBG) domain using InterProScan[15], and each ortholog was confirmed to contain a canonical FBG region. Corresponding coding sequences (CDS) for all confirmed orthologs were retrieved from VectorBase using gene-level identifiers.

2.2. Multiple sequence alignment and trimming

Protein sequences were aligned using PRANK v.170427 [16]with the codon-aware settings (default parameters) to preserve evolutionary signal and minimize gap misplacement. The

resulting protein alignment was used to generate a codon-preserving nucleotide alignment via PAL2NAL v.14[17], producing an alignment of 1686 nucleotide positions. Both the protein and codon alignments were refined using ClipKIT v1.3 [18]with the kpic-smart-gap mode. For the codon alignment, ClipKIT produced a trimmed alignment of 747 positions, removing 55.69% of sites while preserving informative residues for evolutionary analysis.

2.3. Phylogenetic inference

A maximum likelihood phylogeny of FBG30 protein orthologs was reconstructed using IQ-TREE3 v3.0.1[19]. ModelFinder [20] identified LG+I+G4 as the best-fitting amino acid substitution model under the Bayesian information criterion. Tree reconstruction included 1000 ultrafast bootstrap replicates and 1000 SH-aLRT tests [21,22]. The final ML tree contained 29 taxa and 274 amino-acid sites, with 241 parsimony-informative sites. The tree was used as the fixed topology for all downstream molecular evolutionary analyses [22].

2.4. Detection of positive selection

Episodic diversifying selection across branches was assessed using the adaptive Branch-Site Random Effects Likelihood (aBSREL) model [23] implemented in HyPhy v2.5[24]. The trimmed codon alignment and the ML tree were provided as input. Branch-specific likelihood ratio tests identified the *A. darlingi* FBG30 lineage (ADAR2_011252) as the only branch with significant evidence of episodic diversification ($p = 0.0$ after correction). To further investigate codon-specific selective pressures, a branch-site test in codeml (PAML v4.10.9) [25,26]was performed using the *A. darlingi* branch as the foreground. Bayes Empirical Bayes (BEB) analysis identified several sites with elevated posterior probability, including codon positions 51 (PP = 0.969) and 226 (PP = 0.842). Site-level episodic selection was tested using MEME [24]holding *A. darlingi* FBG30 as foreground, which detected codon 173 (CDS alignment) as significantly evolving under episodic selection ($LRT = 4.63$; $p = 0.05$). Mapping this site using in-house python script revealed that it corresponded to residue 218 in *A. darlingi* and residue 191 in the ancestral node.

2.5. Ancestral sequence reconstruction

Ancestral sequence reconstruction (ASR) was performed using IQ-TREE v3.0.1 to infer the historical amino-acid states of FBG30 across the Anopheles phylogeny. The analysis used the same maximum-likelihood (ML) protein phylogeny that was previously inferred from the 29-sequence, 274-amino-acid alignment, along with the best-fit substitution model (LG+I+G4)

selected by ModelFinder. IQ-TREE's ASR procedure estimates, for every internal node and every alignment position, the most likely ancestral amino acid and its associated posterior probability, based on the fixed tree topology, branch lengths, and substitution model. The internal node representing the most recent common ancestor of *A. darlingi* and its sister taxon *A. aquasalis* was identified from the labeled ML tree and designated "Node 26" following the software's node indexing. For each node, IQ-TREE provides reconstructed amino-acid sequences that can be exported as standard FASTA files using in house python script.

To determine the ancestral state of the positively selected site, the codon identified by MEME (codon 173 in the CDS alignment) was mapped to its corresponding position in the ungapped protein sequence. After accounting for alignment gaps and restoring original residue numbering, this site corresponded to amino-acid position 218 in the *A. darlingi* FBG30 protein. Examination of the reconstructed Node 26 sequence showed that the corresponding position was occupied by a serine (S). In contrast, the extant *A. darlingi* sequence contains an asparagine (N) at the same position, indicating that the S→N substitution occurred along the *A. darlingi* lineage after divergence from *A. aquasalis*. The full ancestral sequence of Node 26 was used for all subsequent structural modeling, stability estimation, and protein–protein docking analyses, enabling direct comparison between the reconstructed ancestral state and the modern *A. darlingi* FBG30 protein.

2.6. Protein structure modeling

Three-dimensional structures of the extant *A. darlingi* FBG30, the ancestral Node 26 variant, and the engineered N218S back-mutation were generated using the AlphaFold 3 prediction server[27]. For the N218S variant, the amino acid substitution was introduced manually using AliView [28] followed by de novo structure prediction. All predicted structures were used as starting models for docking and stability simulations.

2.7. Rosetta-based stability estimation

Protein stability and the energetic effects of individual substitution were estimated using PyRosetta (Rosetta v2025)[29]. A standardized pipeline was applied to all variants using a Python script that performed 50 independent FastRelax replicates per sequence. Each replicate consisted of structure relaxation using the fa_scorefxn scoring function (full-atom score function) followed by calculation of Rosetta Energy Units (REU). ΔΔG values were computed

as the difference between mutant and wild-type energies. This protocol produced stability profiles for the extant *A. darlingi* FBG30, the N218S mutant, and the Node 26 ancestor.

2.8. Protein–protein docking with α-tubulin-1

Protein–protein docking between FBG30 variants and *Plasmodium falciparum* α-tubulin-1 (UniProt Q6ZLZ9) was performed using HADDOCK 2.4[30]. Active residues within FBG30 were defined as those in the FBG domain (positions 92–302) based on InterProScan annotation.The experimentally mapped α-tubulin-1 linear epitope REDLAALEKD (residues 422–431) [31] as the core active site in HADDOCK docking, and expanded this region to residues 419–434 to allow for flanking contacts (passive residues 412–418 and 435–440 were auto-assigned/added). Passive residues were assigned automatically[30]. Docking was run using default parameters, and the resulting structures were clustered based on interface RMSD. Ten clusters were produced for each variant, each containing four water-refined models. Although the entire ensemble was used for binding energy assessment, the top-scoring HADDOCK cluster per variant was used for reporting docking statistics.

2.9. Binding affinity prediction

Binding free energies (ΔG) for all docked complexes were estimated using PRODIGY v2.1 [32,33]installed locally through the Conda Bioconda distribution. For each variant, all 40 structures (10 clusters × 4 models) were processed independently using default temperature (25°C) through inhouse python script. The resulting distributions of predicted ΔG values were compared using two-tailed t-tests, after confirming normality with the Shapiro–Wilk test, implemented in an in-house R script.

2.10. Anopheles-Plasmodium coevolution through PACo and ParaFit

Phylogenetic trees for 29 Anopheles species and 12 Plasmodium species were obtained in Newick format from published genomic resources and converted into patristic distance matrices using cophenetic() in the ape package (R v4.5.2). A natural host–parasite association matrix was constructed using only documented field infections and confirmed vector–parasite pairings, yielding a sparse 28 × 11 binary matrix in which taxa with no associations were removed to produce a final working matrix containing 20 Anopheles hosts, 5 Plasmodium parasites, and 5 confirmed natural links. Cophylogenetic congruence was evaluated using the Procrustean Approach to Cophylogeny (PACo) implemented in the paco package, with Cailliez

correction applied to ensure Euclidean distance matrices and significance assessed using 10,000 permutations. To complement PACo, we applied ParaFit using the ade4 package, again using Cailliez-corrected host and parasite distance matrices and 9,999 permutations to obtain ParaFitGlobal and ParaFitLink statistics. Both analyses were performed on the trimmed natural matrix to ensure compatibility with the underlying algorithms.

3. Results

3.1. Identification of FBN30 orthologs and phylogenetic reconstruction

Using the *Anopheles gambiae* FBN30 sequence as a reference, 29 one-to-one FBN30 orthologs across *Anopheles* species has been identified, with a single ortholog recovered from *Culex quinquefasciatus* as an outgroup. All sequences contained an intact fibrinogen-like (FBG) domain, confirming orthology and functional conservation. After alignment and trimming, the final dataset comprised 274 amino-acid positions and 747 codon sites. Maximum-likelihood phylogenetic inference recovered a topology broadly congruent with established *Anopheles* relationships, with strong branch support across most nodes (Fig 1). The resulting tree was used as a fixed topology for all subsequent evolutionary analyses.

3.2. Lineage-specific episodic diversifying selection on FBN30

To assess whether FBN30 experienced adaptive evolution, branch- and site-based codon models has been applied. aBSREL detected significant episodic diversifying selection on a single branch corresponding to the *Anopheles darlingi* lineage, with no other branches showing evidence of $\omega > 1$. Along this lineage, approximately 24% of sites were inferred to evolve under strong positive selection, indicating a localized episode of adaptive divergence rather than widespread relaxation of constraint (Table 1). Consistent with this result, branch-site analysis using codeml supported positive selection on the *A. darlingi* branch, yielding a significantly better fit than the null model (Table 1). Together, these analyses identify *A. darlingi* FBN30 as the sole lineage exhibiting detectable episodic adaptive evolution.

3.3. Site-level selection and localization within the FBG domain

Site-level tests using MEME identified a single codon evolving under episodic positive selection specifically along the *A. darlingi* lineage (Table 1). This codon maps to a residue within the conserved FBG domain, a region implicated in ligand recognition across fibrinogen-related proteins (Fig 4). Although codeml and MEME differed in the specific sites receiving

highest posterior support-reflecting their distinct statistical sensitivities, both approaches converged on the conclusion that adaptive evolution in FBN30 is limited in scope and concentrated within a functionally relevant domain.

3.4. Ancestral state reconstruction reveals derived substitution in A. darlingi

Ancestral sequence reconstruction was performed to place the positively selected site in an evolutionary context. The reconstructed ancestor shared by *A. darlingi* and its sister taxon *A. aquasalis* carried a serine at the selected position (Fig 1), whereas the extant *A. darlingi* sequence contains an asparagine (Table 1, 2). This serine-to-asparagine substitution therefore represents a derived change unique to the *A. darlingi* lineage and was used as a focal point for downstream structural analyses.

3.5. Structural stability reflects fine-scale evolutionary modulation

To evaluate whether the derived substitution produced major structural effects, predicted protein stability across extant, ancestral, and back-mutated FBN30 variants was compared (Table 2, Figure 2). The reconstructed ancestral protein exhibited substantially reduced predicted stability relative to the extant *A. darlingi* protein, reflecting cumulative divergence across the lineage rather than the effect of any single substitution. In contrast, introducing the ancestral residue into the extant background produced only modest and variable stability changes, indicating that the positively selected site contributes to localized modulation rather than global fold alteration.

3.6. Evolutionary divergence is associated with altered predicted ligand engagement

To explore potential functional consequences of FBN30 evolution, the study used protein-protein docking and binding affinity prediction as a proxy for ligand engagement (Fig 4). Comparative analyses revealed that the reconstructed ancestral FBN30 consistently exhibited stronger predicted affinity for *Plasmodium falciparum* α-tubulin-1 than the extant *A. darlingi* protein (Fig 3). This difference was robust across docking metrics and binding energy estimates (Table 3 & 4). Importantly, the derived substitution alone did not account for the full magnitude of the observed affinity reduction, suggesting that evolutionary divergence along the *A. darlingi* lineage involved cumulative changes affecting a parasite-facing interface.

3.7. Global but not pairwise host-parasite phylogenetic congruence

Cophylogenetic analyses using PACo and ParaFit revealed significant global congruence between *Anopheles* and *Plasmodium* phylogenies based on documented natural associations. However, sparse host–parasite linkages precluded reliable identification of individual coevolving pairs. These results indicate broad phylogenetic structuring of vector-parasite associations rather than strict pairwise co-speciation, consistent with lineage-specific rather than system-wide antagonistic evolution.

4. Discussion

FREPs are central components of mosquito innate immunity, yet their evolutionary dynamics and functional diversification remain incompletely understood. By integrating codon-based evolutionary analyses, ancestral sequence reconstruction, and structural modeling, this study identifies a discrete episode of episodic diversifying selection acting on FBN30 in the *A. darlingi* lineage (Table 1; Fig 1). This adaptive signal is restricted to a single mosquito lineage and absent from other *Anopheles* species, indicating localized evolutionary refinement rather than widespread diversification across the FREP family. Both branch-level and site-level analyses converge on this conclusion, with the positively selected residue mapping to the conserved FBG domain (Fig 4), a region implicated in ligand recognition in fibrinogen-related proteins. The confinement of adaptive evolution to a single lineage suggests that FBN30 has been shaped by lineage-specific selective pressures, potentially reflecting geographically or ecologically structured host–parasite interactions.

Ancestral sequence reconstruction places the derived substitution in a clear evolutionary context, revealing a serine-to-asparagine change unique to *A. darlingi*. Structural stability analyses indicate that this substitution produces only modest and localized effects, whereas the reconstructed ancestral protein differs substantially in overall predicted stability (Table 2; Fig 2). This pattern suggests that the adaptive substitution represents fine-scale functional tuning occurring on a background of broader evolutionary divergence, rather than a single mutation driving major structural reorganization. Such incremental refinement is consistent with expectations for proteins subject to ongoing functional constraint, where adaptive change is accommodated through subtle modulation rather than wholesale innovation.

To explore potential functional consequences of this evolutionary divergence, protein-protein docking and binding affinity prediction as proxies for ligand engagement has been employed. These analyses consistently indicate that the reconstructed ancestral FBN30 exhibits stronger predicted affinity for *Plasmodium falciparum* α-tubulin-1 than the extant *A. darlingi* protein (Table 3 & 4; Fig 3). Although docking-based predictions cannot establish biochemical interaction, the concordance between evolutionary signal and predicted functional modulation suggests that adaptive evolution in FBN30 may have altered a parasite-facing interface. Notably, the derived substitution alone does not account for the full reduction in predicted affinity, implying that cumulative evolutionary changes along the *A. darlingi* lineage contribute to this effect.

Taken together, these findings are consistent with a model in which inhibitory mosquito proteins undergo lineage-specific adaptive refinement under parasite-mediated selective pressures. Rather than indicating the emergence or loss of function, the observed evolutionary pattern suggests modulation of an existing recognition interface, potentially altering the strength or outcome of parasite engagement. Such a scenario aligns with theoretical expectations of antagonistic host-parasite evolution, where reciprocal pressures favor incremental adjustments rather than dramatic functional shifts.

Cophylogenetic analyses reveal significant global congruence between *Anopheles* and *Plasmodium* phylogenies, supporting the view that vector-parasite associations are shaped by broad evolutionary constraints. However, the absence of detectable pairwise co-speciation underscores that adaptive events such as those observed in *A. darlingi* FBN30 likely represent localized evolutionary responses rather than system-wide reciprocal evolution. This distinction highlights the importance of integrating lineage-specific molecular analyses with broader phylogenetic context when interpreting host–parasite evolution.

Several limitations of this study warrant consideration. All functional inferences are based on in silico predictions and should be interpreted as hypotheses rather than demonstrations of molecular interaction. Moreover, the focus on a single mosquito lineage limits generalization across *Anopheles*. Nevertheless, the convergence of evolutionary, structural, and predictive functional analyses provides a coherent framework for understanding how inhibitory mosquito proteins may be shaped by parasite-mediated selection.

5. Conclusion

This study identifies lineage-specific adaptive evolution in FBN30 and suggests that fine-scale evolutionary refinement of a conserved recognition domain may modulate predicted parasite engagement. By emphasizing evolutionary inference over mechanistic assertion, this work contributes to a growing understanding of how host innate immune factors diversify under antagonistic interactions and generates testable hypotheses for future experimental investigation.

6. Funding details

This research received no funding.

7. Conflicts of interest

The author report that there are no competing interests to declare.

8. Data availability statement

All data associated with this study are available on Figshare (DOI: 10.6084/m9.figshare.30618974). The materials provided include both raw and processed datasets, as well as all scripts required to reproduce the analyses described in the manuscript. Supplementary files contained within the repository and the manuscript ensure full transparency and reproducibility of the work.

9. Acknowledgments

I thank the global open-access community for their commitment to freely accessible science, without which this study would not have been possible. I gratefully acknowledge VectorBase for providing high-quality genomic resources essential to this work. I also recognize the contributions of openly available language models that assisted in refining the manuscript. Finally, I am deeply grateful to my wife, Dr. Nabanita Ghosh, Assistant Professor of Zoology at Maulana Azad College, Kolkata, whose insightful discussions helped shape the central idea of this study.

10. References

1. Daily, J. P. & Parikh, S. Malaria. *New England Journal of Medicine* **392**, 1320–1333 (2025).

2. Shaw, W. R., Marcenac, P. & Catteruccia, F. Plasmodium development in Anopheles: a tale of shared resources. *Trends Parasitol.* **38**, 124 (2021).

3. Zhang, G. *et al.* Anopheles Midgut FREP1 Mediates Plasmodium Invasion. *J. Biol. Chem.* **290**, 16490 (2015).

4. Niu, G. *et al.* The fibrinogen-like domain of FREP1 protein is a broad-spectrum malaria transmission-blocking vaccine antigen. *Journal of Biological Chemistry* **292**, 11960–11969 (2017).

5. Zhang, G. *et al.* Targeting plasmodium α-tubulin-1 to block malaria transmission to mosquitoes. *Front. Cell. Infect. Microbiol.* **13**, 1132647 (2023).

6. Zhang, G., Niu, G., Perez, L., Wang, X. & Li, J. Malaria transmission assisted by interaction between Plasmodium α-tubulin-1 and Anopheles FREP1 protein. *bioRxiv* 2019.12.16.878082 (2019) doi:10.1101/2019.12.16.878082.

7. Hanington, P. C. & Zhang, S. M. The Primary Role of Fibrinogen-Related Proteins in Invertebrates Is Defense, Not Coagulation. *J. Innate Immun.* **3**, 17 (2010).

8. Molina-Cruz, A. *et al.* Some strains of Plasmodium falciparum, a human malaria parasite, evade the complement-like system of Anopheles gambiae mosquitoes. *Proc. Natl. Acad. Sci. U. S. A.* **109**, (2012).

9. Li, J. *et al.* Genome-block expression-assisted association studies discover malaria resistance genes in Anopheles gambiae. *Proceedings of the National Academy of Sciences* **110**, 20675–20680 (2013).

10. Niu, G. *et al.* FBN30 in wild Anophelesgambiae functions as a pathogen recognition molecule against clinically circulating Plasmodiumfalciparum in malaria endemic areas in Kenya. *Sci. Rep.* **7**, 8577 (2017).

11. Janjoter, S. *et al.* Deciphering the molecular targets of Plasmodium and Anopheline interactions for malaria control. *FEBS Journal* https://doi.org/10.1111/FEBS.70250;WEBSITE:WEBSITE:FEBS;PAGE:STRING:ARTICLE/CHAPTER (2025) doi:10.1111/FEBS.70250;WEBSITE:WEBSITE:FEBS;PAGE:STRING:ARTICLE/CHAPTER.

12. Giraldo-Calderón, G. I. *et al.* VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. *Nucleic Acids Res.* **43**, D707 (2014).

13. Moreno-Hagelsieb, G. & Latimer, K. Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics* **24**, 319–324 (2008).

14. Hernández-Salmerón, J. E. & Moreno-Hagelsieb, G. Progress in quickly finding orthologs as reciprocal best hits: comparing blast, last, diamond and MMseqs2. *BMC Genomics 2020 21:1* **21**, 741- (2020).

15. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).

16. Löytynoja, A. Phylogeny-Aware Alignment with PRANK and PAGAN. *Methods Mol. Biol.* **2231**, 17–37 (2021).

17. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609 (2006).

18. Steenwyk, J. L., Buida, T. J., Li, Y., Shen, X. X. & Rokas, A. ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biol.* **18**, e3001007 (2020).

19. Wong, T. K. F. *et al.* IQ-TREE 3: Phylogenomic Inference Software using Complex Evolutionary Models. https://doi.org/10.32942/X2P62N (2025) doi:10.32942/X2P62N.

20. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods 2017 14:6* **14**, 587–589 (2017).

21. Sinha, K. Episodic Positive Selection Signatures in Arabidopsis CONSTANS-like Genes COL3 and COL5 Indicating Adaptive Evolution in Red-Light Signaling Pathways. *bioRxiv* 2025.04.03.646976 (2025) doi:10.1101/2025.04.03.646976.

22. Sinha, K., Jana, S., Pramanik, P. & Bera, B. Selection on synonymous codon usage in soybean (Glycine max) WRKY genes. *Scientific Reports 2024 14:1* **14**, 26530- (2024).

23. Smith, M. D. *et al.* Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection. *Mol. Biol. Evol.* **32**, 1342 (2015).

24. Kosakovsky Pond, S. L. *et al.* HyPhy 2.5—A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. *Mol. Biol. Evol.* **37**, 295 (2019).

25. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

26. Álvarez-Carretero, S., Kapli, P. & Yang, Z. Beginner's Guide on the Use of PAML to Detect Positive Selection. *Mol. Biol. Evol.* **40**, (2023).

27. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature 2021 596:7873* **596**, 583–589 (2021).

28. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278 (2014).

29. Sora, V. *et al.* RosettaDDGPrediction for high-throughput mutational scans: From stability to binding. *Protein Sci.* **32**, e4527 (2023).

30. Honorato, R. V. *et al.* The HADDOCK2.4 web server for integrative modeling of biomolecular complexes. *Nature Protocols 2024 19:11* **19**, 3219–3241 (2024).

31. Zhang, G. *et al.* Targeting plasmodium α-tubulin-1 to block malaria transmission to mosquitoes. *Front. Cell. Infect. Microbiol.* **13**, 1132647 (2023).

32. Xue, L. C., Rodrigues, J. P., Kastritis, P. L., Bonvin, A. M. & Vangone, A. PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics* **32**, 3676–3678 (2016).

33. Vangone, A. & Bonvin, A. M. J. J. Contacts-based prediction of binding affinity in protein–protein complexes. *Elife* **4**, (2015).

11. Tables

Table 1. Summary of molecular evolution analyses identifying episodic and site-specific positive selection in *Anopheles* FBN30 orthologs. This table presents the results of aBSREL, the branch-site codeml model, and MEME analysis. The aBSREL test detected a single branch under episodic diversifying selection, corresponding to *A. darlingi* FBN30 (ADAR2_011252_R18153). The branch-site model similarly supported positive selection on this lineage, with one codon site (*) showing elevated posterior probabilities (BEB ≥ 0.95). MEME analysis identified codon 173 (corresponding to residue 218 in *A. darlingi* and residue 191 in Node 26) as evolving under episodic positive selection with a significant LRT. The combined results support the presence of a lineage-specific adaptive substitution along the *A. darlingi* branch.

Table 2. Rosetta ΔΔG stability results for extant *A. darlingi* FBN30, the N218S mutant, and the reconstructed Node 26 ancestor.

Rosetta FastRelax was performed with ten independent replicates to estimate the energetic consequences of the selected substitution. The extant *A. darlingi* FBN30 exhibited more favorable energies than the Node 26 ancestor in all replicates, whereas the N218S mutation introduced into the extant protein produced moderate and variable destabilization.

Analysis	Key Result	Statistical Support	Notes
aBSREL	*A. darlingi* branch under episodic diversifying selection	p = 0.0; ω_2 = 57.25 at 23.6% sites	Strong evidence for branch-specific selection
Branch-site codeml	Several sites under selection in *A. darlingi* foreground	LRT = 6.72; p = 0.01 BEB ≥ 0.969* at codon 51; BEB ≥ 0.842 at codon 226; moderate support at others	Confirms aBSREL signal
MEME	Codon 173 under episodic selection	LRT = 4.63; p = 0.05	Selected residue maps to FBN30 FBG domain

Variant	Mean ΔΔG (REU) ± SD	Interpretation	Notes
N218S (relative to WT)	3.93 ± 7.04	Mild destabilization	Highly variable between replicates
Node 26 ancestor (relative to WT)	237.62 ± 14.68	Strong destabilization	Indicates globally less stable ancestral fold

Table 3. HADDOCK docking statistics for FBN30 variants interacting with *Plasmodium falciparum* α-tubulin-1.

This table summarizes the best-scoring docked clusters for each FBN30 variant. The Node 26 ancestral sequence exhibited the most favorable docking energetics, lower RMSD, and greater buried surface area compared to the WT and N218S forms. These patterns are consistent with stronger and more stable binding in the ancestral FBN30-tubulin complex.

Variant	Best Cluster	HADDOCK Score	Cluster Size	RMSD (Å)	Buried Surface Area (Å²)	Z-score
WT	Cluster 8	125.7 ± 22.8	5	1.3 ± 1.4	2202.0 ± 98.0	–1.7

Node 26 ancestor	Cluster 4	62.5 ± 10.6	22	3.2 ± 0.1	2834.4 ± 38.4	–2.2
N218S mutant	Cluster 1	127.6 ± 4.0	23	22.3 ± 0.4	2194.8 ± 29.6	–1.9

Table 4. PRODIGY binding free energy predictions (ΔG) for the interaction between FBN30 variants and *P. falciparum* α-tubulin-1.

Binding energies were computed for all clusters (40 structures per variant). Node 26 exhibited significantly stronger binding than WT, whereas N218S did not differ significantly from WT.

Comparison	Mean ΔG (kcal·mol⁻¹)	Difference	Statistical Test	p-value	Interpretation
WT vs Node 26	–12.16 vs –13.87	+1.71 kcal·mol⁻¹	Mann-Whitney U	4.52×10^{-10}	Node 26 binds significantly more strongly
WT vs N218S	–12.16 vs –12.50	+0.34 kcal·mol⁻¹	Mann-Whitney U	ns	No significant effect

12. Figures

Figure 1. Maximum likelihood phylogeny of FBN30 across 29 *Anopheles* species. The phylogeny was inferred using IQ-TREE3 under the LG+I+G4 substitution model, with 1000 ultrafast bootstrap replicates and 1000 SH-aLRT tests. Branch lengths represent amino-acid substitutions per site. The tree is rooted using the *Culex quinquefasciatus* FBN30 ortholog (VectorBase gene ID: CPIJ000937), which served as the designated outgroup. The A. darlingi lineage (ADAR2_011252_R18153) is highlighted as the only branch exhibiting significant evidence of episodic diversifying selection in the aBSREL analysis.

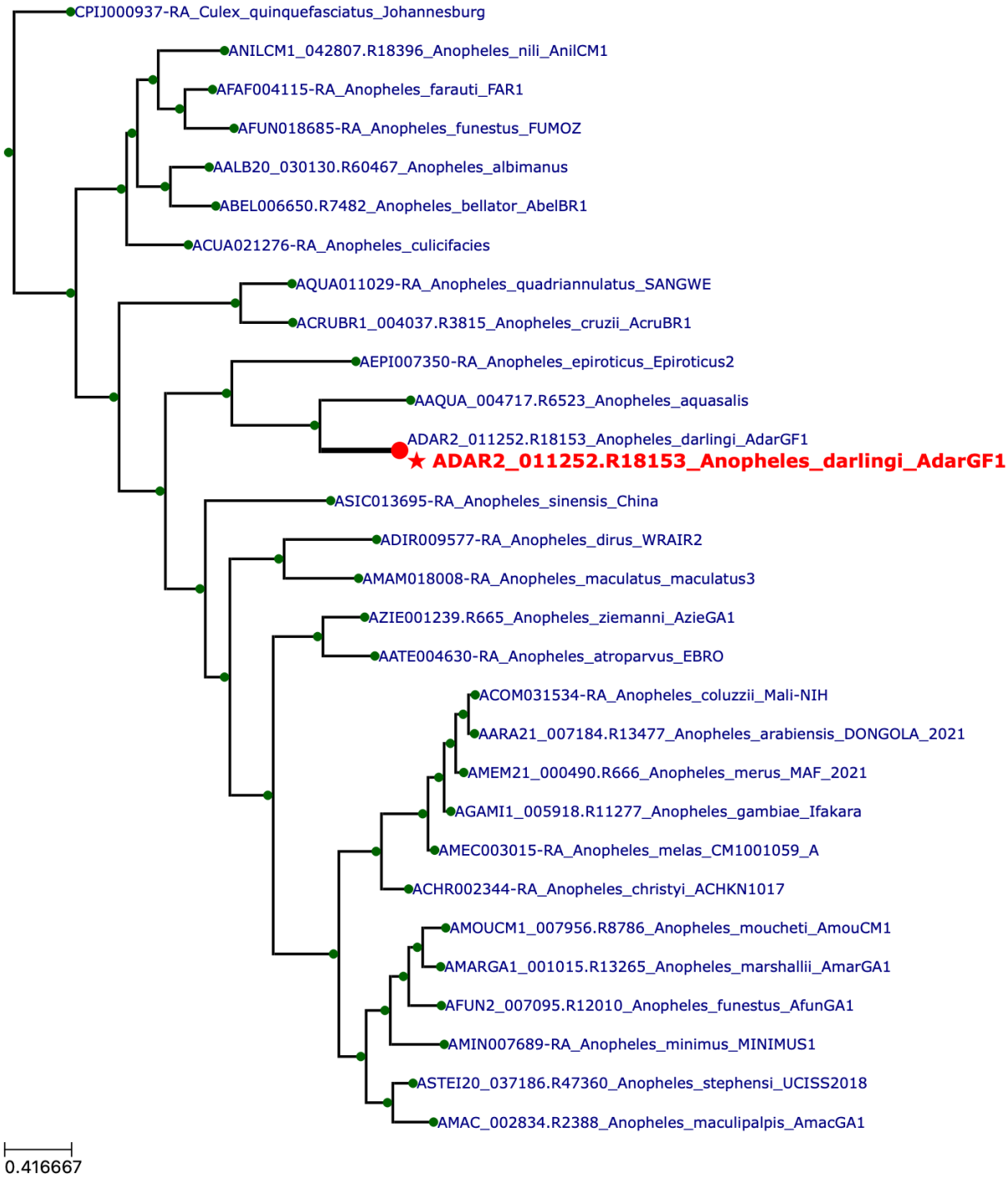


Figure 2. Rosetta stability estimates for WT, N218S, and Node 26 FBN30. Replicate ΔΔG values are shown for each comparison. The Node 26 ancestor exhibits uniformly large positive ΔΔG values relative to WT, consistent with global destabilization (left panel), whereas the N218S mutation shows modest and variable effects (right panel).

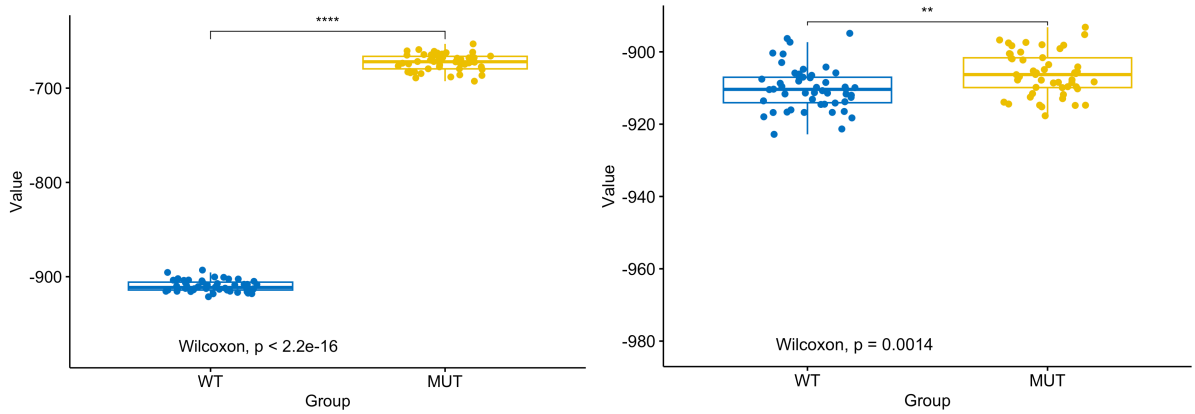


Figure 3. Distribution of PRODIGY-predicted binding energies for WT, N218S, and Node 26 FBN30. Boxplots display ΔG (kcal·mol^{-1}) across 40 docked structures per variant. The Node 26 distribution is significantly shifted toward more negative binding energies, whereas the N218S distribution overlaps the WT (not shown).

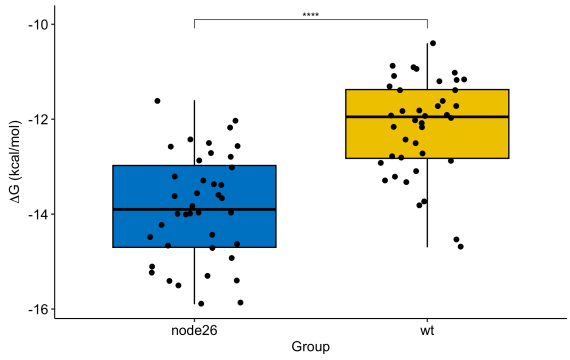
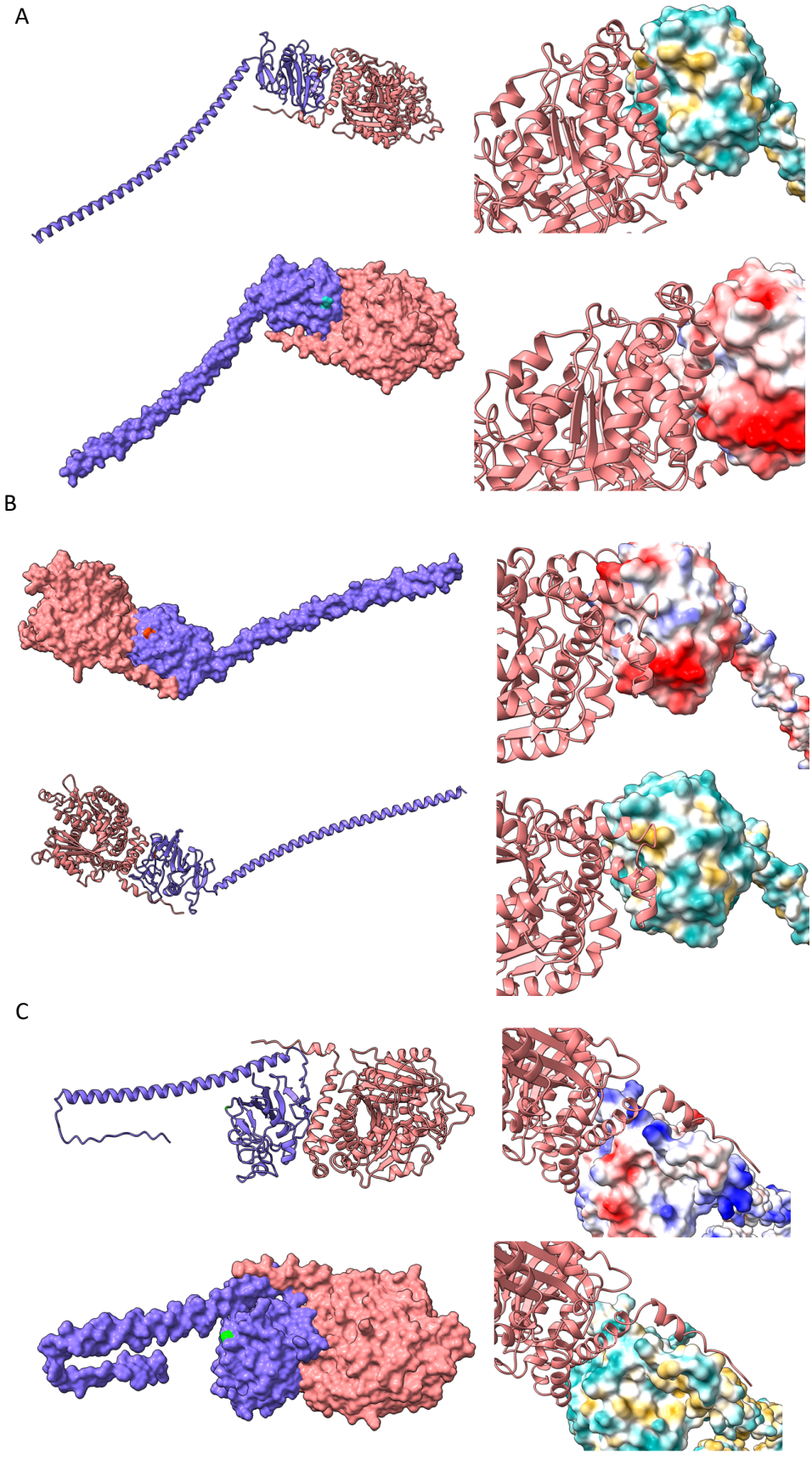


Figure 4. Comparative structural analysis of WT, N218S, and Node 26 FBN30 variants docked to *P. falciparum* α-tubulin-1. HADDOCK best-scoring complexes are shown for the three FREP1 variants: WT (A), N218S mutant (B), and Node 26 ancestral variant (C). For each variant, the overall docked assembly is displayed in ribbon representations, followed by a space-filling dimer view highlighting the variant-specific residue (WT: 218; N218S: Ser218; Node 26: position 191) in distinct color. Corresponding close-up views of the binding interface are shown as electrostatic surfaces and hydrophobicity-mapped surfaces, illustrating differences in charge complementarity, nonpolar contacts, and interface geometry among the three variants. Collectively, these comparisons highlight the expanded buried surface area and altered interaction landscape of the Node 26 complex relative to WT and N218S.


A
B
C