

Research Article

Structural and Evolutionary Adaptations of Nei-Like DNA Glycosylases Proteins Involved in Base Excision Repair of Oxidative DNA Damage in Vertebrates

Hafiz Ishfaq Ahmad ¹, Gulnaz Afzal ², Sehrish Sadia,³ Ghulam Haider,³ Shakeel Ahmed,⁴ Saba Saeed,⁵ and Jinping Chen ⁶

¹Department of Animal Breeding and Genetics, University of Veterinary and Animal Sciences, Lahore, Pakistan

²Department of Zoology, The Islamia University of Bahawalpur, Bahawalpur, Pakistan

³Department of Biological Sciences, University of Veterinary and Animal Sciences, Ravi Campus, Pattoki, Pakistan

⁴Instituto de Farmacia, Facultad de Ciencias, Universidad Austral de Chile, Campus Isla Teja, 5090000 Valdivia, Chile

⁵Institute of Physics, The Islamia University of Bahawalpur, Bahawalpur, Pakistan

⁶Guangdong Key Laboratory of Animal Conservation and Resource Utilization, Guangdong Public Laboratory of Wild Animal Conservation and Utilization, Institute of Zoology, Guangdong Academy of Sciences, Guangzhou, Guangdong, China

Correspondence should be addressed to Hafiz Ishfaq Ahmad; ishfaq.ahmad@uvas.edu.pk and Jinping Chen; chenjp@giz.gd.cn

Received 22 January 2022; Accepted 3 March 2022; Published 4 April 2022

Academic Editor: Ibrahim Hakkı Cigerci

Copyright © 2022 Hafiz Ishfaq Ahmad et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Oxidative stress is a type of stress that damages DNA and can occur from both endogenous and exogenous sources. Damage to DNA caused by oxidative stress can result in base modifications that promote replication errors and the formation of sites of base loss, which pose unique challenges to the preservation of genomic integrity. However, the adaptive evolution of the DNA repair mechanism is poorly understood in vertebrates. This research aimed to explore the evolutionary relationships, physicochemical characteristics, and comparative genomic analysis of the Nei-like glycosylase gene family involved in DNA base repair in the vertebrates. The genomic sequences of NEIL1, NEIL2, and NEIL3 genes were aligned to observe selection constraints in the genes, which were relatively low conserved across vertebrate species. The positive selection signals were identified in these genes across the vertebrate lineages. We identified that only about 2.7% of codons in these genes were subjected to positive selection. We also revealed that positive selection pressure was increased in the Fapy-DNA-glyco and H2TH domain, which are involved in the base excision repair of DNA that has been damaged by oxidative stress. Gene structure, motif, and conserved domain analysis indicated that the Nei-like glycosylase genes in mammals and avians are evolutionarily low conserved compared to other glycosylase genes in other “vertebrates” species. This study revealed that adaptive selection played a critical role in the evolution of Nei-like glycosylase in vertebrate species. Systematic comparative genome analyses will give key insights to elucidate the links between DNA repair and the development of lifespan in various organisms as more diverse vertebrate genome sequences become accessible.

1. Introduction

DNA oxidative damage occurs due to various factors, including external agents, endogenous oxygen species formed during normal cellular respiration, and other essential activities, such as demethylation, that generate intermediates, including basic sites [1]. Incorrectly handled or

repaired sites might cause polymerase to halt replication, eventually leading to mutagenesis and cancer [2]. Mammalian cells utilize a battery of enzymes geared at repairing DNA damage to ensure the genome’s accurate replication and transcript interpretation [3]. The base excision repair (BER) process is initiated by several glycosylases, which safeguard cells against mutagenesis and oxidative DNA damage

by identifying damaged bases and commencing the repair process [4]. This pathway is the primary method for eliminating oxidative DNA damage from the genome, making it an important step in ensuring genome integrity in the first place. The prevention of disorders induced by oxidative DNA damage is therefore dependent on the function of this system [5]. To eliminate oxidative DNA base damage in mammalian cells, at least five distinct DNA glycosylases with overlapping substrate specificities are required, each with a different substrate specificity [6]. The Nei-like DNA glycosylases (NEIL1/2/3), 8-oxoguanine glycosylase 1 (OGG1), and endonuclease three homologue 1 (NTH1) are only a few of the enzymes that exist [7]. One of the functions of NEIL1 is to repair DNA that has been impaired by mutagenic chemicals or oxidation. A DNA glycosylase is an enzyme that detects and eliminates impaired bases from DNA strands. It is particularly favorable to oxidize pyrimidines such as formamidopyrimidine (Fapy), 5-hydroxyuracil, and thymine glycol [8]. Oxidative damage to DNA may be generated in various ways, including endogenously created oxygen species from normal cellular respiration and other critical operations such as demethylation, which form intermediates that include abasic sites [1, 9]. If the polymerase cannot complete its replication cycle due to insufficient site management or repair, this might result in mutagenesis and cancer in the long run [1]. In addition to excising an oxidized base (glycosylase activity), this bifunctional enzyme, according to the researchers, is capable of cutting the DNA backbone (lyase activity), and it can function independently of apurinic endonuclease (APE) [10]. Both double-stranded DNA (dsDNA) and single-stranded DNA (ssDNA) are targets for NEIL2, which has a bias for bubble DNA structures as a substrate and a preference for cytosine oxidation products as an enzyme [11]. Despite their considerable sequence variety, the Fpg/Nei family of proteins exhibits exceptional structural conservation of the two-domain design, notably in the N-terminal domain, largely composed of β -strands and are substantially maintained. C-terminal domain of the protein (which is rather stable across the whole protein family) comprises DNA-binding motifs such as the two-turn helix (H2TH) and the zinc or zincless finger (ZFF) [12].

The study of repair in diverse species can also allow researchers better understand how repair proteins and processes have evolved through time. This is important because the repair is an important biological function and because understanding how repair develops may help direct comparative repair research [13]. Generally, the evolutionary approach to comparative study is useful because it allows researchers to focus on how and why similarities and differences have arisen rather than merely detecting and categorizing them [14]. We feel that understanding the changes in DNA repair between species and the mechanisms and roles of particular DNA repair processes requires an evolutionary viewpoint in DNA repair research [15].

Unfortunately, comparative and evolutionary investigations of DNA repair mechanisms have been limited by a lack of thorough studies across a wide variety of species representing a wide range of ecological and evolutionary heterogeneity [13]. Recently, whole-genome sequences have been

identified to represent a potentially helpful new source of comparative repair data. Entire genome sequencing is predicted to enable the prediction of a strain's or species' phenotype and the collection and analysis of a large amount of data for comparative study [16]. However, extracting useful information from complete genome sequences is quite challenging in practice. Despite the enormous number of studies completed in animals, no comparative genome analysis studies in vertebrates have been conducted to find the DNA repair genes linked to the development of long life [17]. According to current studies, all critical DNA repair pathways are low conserved in mammals [18].

Additionally, many nonmodel vertebrate genomes are becoming available for analysis. There is evidence that stress contributes to major selective factors driving the evolution. Numerous new genomes now enabled evaluations across different lineages to ascertain how certain genes are exposed to positive selection [19]. Pathogens are also the principal selective factor driving evolution. Numerous additional genomes now enable comparisons across lineages to ascertain the extent to which certain genes are exposed to negative selection [20].

The purpose of this study is to examine the evolutionary connection, physiochemical features, and comparative genomics of Nei-like DNA glycosylases gene family in the vertebrate species. We used more than 153 species of vertebrates, including mammalian, avian, and amphibian species, to conduct the systematic comparative analyses of the Nei-like DNA glycosylase genes encoding protein that regulate the DNA repair process in the vertebrate species with various life forms, including long-lived mammalian, avian, and amphibian species. This study examined the genomic sequences of Nei-like DNA glycosylases in the vertebrate species to determine the selection pressure exerted on these genes, which might be important in the adaptive evolution process. In this study, we explore how these genes evolved in various vertebrates and how selection and diversity influenced the evolution of this gene family over the years.

2. Materials and Methods

2.1. Data Collection and Sequence Analysis. Ensembl and NCBI databases were used to retrieve the genomic sequences of Nei-like DNA glycosylase genes (NEIL1, NEIL2, and NEIL3) in the vertebrate species. To extract nonredundant protein sequences, the human Nei-like DNA glycosylase protein sequences (ENSG00000140398, ENSG00000154328, and ENSG00000109674) were utilized in a BLAST search [21]. The accession numbers of these genes were used to search the NCBI and Ensembl databases for coding sequences (CDS) of vertebrate species (Supplementary Tables S1-S3). Furthermore, using tBLASTn and BLASTn searches, orthologs in the vertebrate genome were recognized [22]. The OMA v.1.0.0 tool also analyzed homology patterns across protein-coding genes throughout the sequenced vertebrate species [23]. MAFFT v.7.221 [24] was used to align the sequences. Further analyses were conducted using these aligned sequences of homologous proteins.

2.2. Sequence Alignment and Test for Selection. The PROBCONS version 1.12 was used to create sequence alignments [25]. Gblocks v0.91b [26] was used to identify and screen possibly inaccurate and misaligned sections in the alignments utilized for phylogenetic analysis. MrBayesv3.2.2 (<http://mrbayes.csit.fsu.edu/>) [27] was used to create a phylogenetic tree to find the positive selection in each codon of NEIL1, NEIL2, and NEIL3 genomes. In MrBayes, all analyses were done with a generation of 2000,000. With split frequency values of less than 0.01, convergence was believed to have been identified. This continued with further 2000,000 generations if the split frequency did not decrease below 0.01.

Positive selection signatures were identified by comparing the ratio of sites in positively selected codons across vertebrate species to the nearly neutral evolution model. Sites having greater nonsynonymous-synonymous substitution ratios were recognized as positively selected sites [28]. The ratios of dN/dS per site were examined using maximum likelihood approaches to evaluate positive selection at NEIL genes. Positive selection is shown by a dN/dS ratio greater than one. We employ two machine learning frameworks to increase the rigor of the positive selection: the PAML codeml program [29] and the Datamonkey Web Server's HyPhy package (<http://www.datamonkey.org>) [30]. Two different models (M7 vs. M8) were preferred and compared using log-likelihood values (2lnL) to determine if positive selection was acting on locations within each NEIL gene. The Bayes empirical Bayes (BEB) approach was combined with site models to determine which codons in codeml are flexible [29]. We recognized sites with a *p* value of 0.05 for SLAC and FEL and a Bayes factor 90 for REL to identify candidates for selection. The term "adaptive protein-coding substitutions" refers to those found using two or more machine learning algorithms [31]. Only locations that demonstrated selection signals in at least two of the machine learning approaches were examined to identify robust areas subject to positive selection. We examined codon positions under selection pressure in aligned sequences of targeted genes using Selecton version 2.2 (<http://selecton.tau.ac.il/>) [32]. As determined by the maximum likelihood value calculated by Bayesian inference [33], Selecton allows the ratio to vary between unique codons within a multiple sequence alignment. Furthermore, the outcomes of Selecton tool were graphically represented using color scales that depicted the various types of selections performed [34].

2.3. Structural Analysis and Homology Modeling. We evaluated and compared models to predict protein secondary and tertiary structure using the Phyre2 server [35] and Chimera 1.11.2 [36]. To find protein functional domains, we used the HMMER v3.1 [37], NCBI Conserved Domain Database [38], and Pfam v35.0 tool [39]. We used DOG 1.0 illustrator of protein domain structures [40] to illustrate protein domains. A simple and graphic illustration of protein domain structures with functional motifs assists a vast readership in quickly comprehending the ancient and novel functions of proteins.

2.4. Conservation Analysis. The ConSurf server (<http://consurf.tau.ac.il/>) [41] was used to examine the evolutionary conservation of amino acid residues in the human NEIL1, NEIL2, and NEIL3 proteins. Protein amino acids are more conserved than other amino acids because they are required for protein networks or are positioned in more enzymatic locations. As a result, alterations in conserved amino acids are more detrimental to protein function and structure than polymorphisms in flexible protein areas [41]. The conservation values ranging from 1 to 9 were used to predict conserved amino acids; conservation values between 1 and 4 are considered variable; 5–6 indicate moderate conservation; and 7–9 indicate very high conservation [41, 42].

2.5. Network and Coevolution Analysis. The CoeViz 2, a web-based program [43], was used to conduct a targeted and quick assessment of protein features, such as structural and functional regions, and give a variable analysis and visual representations of paired coevolution of amino acids. We used CoeViz analysis with two covariance metrics [44] to obtain the complete protein sequences of the proteins NEIL1, NEIL2, and NEIL3, which were used for phylogenetic analysis to assess the positions of covarying sites and large coincides with domains of the protein and used circular visuals for the understanding of residue interactions. The coevolving sites were emphasized in three-dimensional structures of protein sequences. Metascape (<https://metascape.org/>) [45] was also used to analyze the data for gene ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) [46]. Drug-target-pathway networks were created using Cytoscape [47].

3. Results

The purpose of this study was to evaluate the gene sequences of Nei-like glycosylase in different "vertebrates" species to estimate the selection vigor in these genes, which may be involved in adaptive evolution. We explored the vertebrate genome for three genes involved in repairing DNA bases initiated by a series of DNA glycosylases such as NEIL1, NEIL2, and NEIL3, damaged by reactive oxygen species. We identified that these genes show the signature of selection during adaptive evolution which involved repairing of DNA bases damaged by reactive oxygen species. Maximum likelihood and phylogenetic analysis of coding sequences from 157 distinct vertebrate species revealed that the NEIL gene family descended from a common vertebrate ancestor. The evolutionary processes, phylogenetic linkages, structural and functional constraints of the NEIL1 and NEIL2 homologs in the vertebrate species, and their structural and functional constraints, were investigated in this study.

3.1. Protein Domains and Selection Analysis. Positive selection sites were found in the H2TH domain (helix-2turn-helix domain), a DNA-binding domain of the NEIL studied. Positive selection was found in the Fapy-DNA-glyco and H2TH domains of human NEIL1 and the H2TH domains of NEIL2 and NEIL3 (Table 1 and Figures 1–3). The H2TH domain (helix-2turn-helix domain) is a DNA-

TABLE 1: Results of positive selection tests for NEIL genes.

Gene	Models	lnL	LRT	PAML	SLAC	FUBAR	MEME
NEIL1	M1a	-11259.31	0	89,90,107,117,143,367,405,471	67, 107, 143, 405, 408, 444	67, 117, 143, 251, 367, 471	90, 117, 143, 405, 471
	M2a	-11259.31					
	M7	-11180.39	6.54				
	M8	-11177.12					
NEIL2	M1a	-7282.58	1.48	65, 67, 84, 94, 103, 105, 116, 191, 256, 301, 329	103, 191, 236, 329	65, 103, 116, 154, 191, 301, 329	103, 105, 191, 256, 329
	M2a	-7281.84					
	M7	-7273.27	9.16				
	M8	-7268.69					
NEIL3	M1a	-17093.12	0	36, 46, 51, 56, 62, 89, 108, 113, 185, 258, 309, 336, 437	46, 62, 108, 113, 185, 309	36, 51, 62, 89, 113, 185, 309, 437	51, 62, 113, 185, 309, 437
	M2a	-17093.12					
	M7	-16962.42	0.48				
	M8	-16962.18					

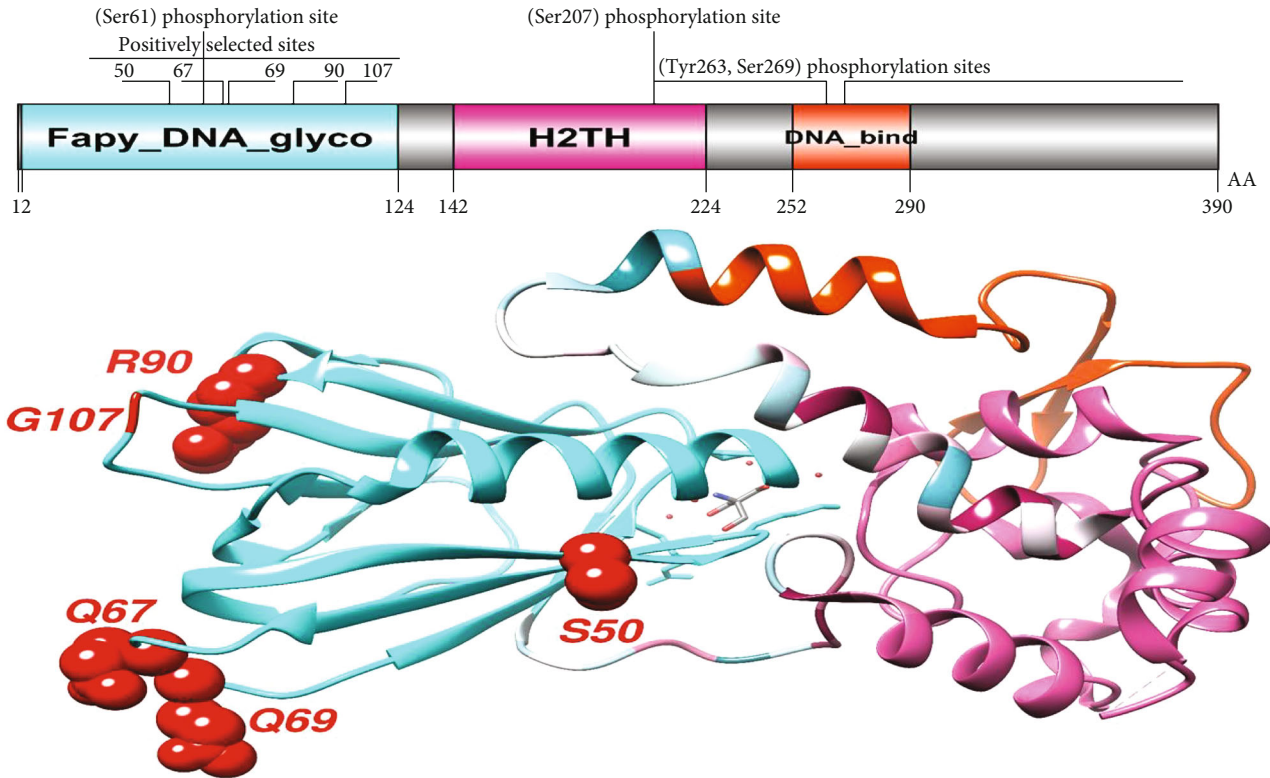
binding domain found in DNA glycosylase/AP lyase enzymes involved in base excision repair of DNA damaged by oxidation or mutagenesis chemicals. It was shown that sequence differences in the H2TH domain of human NEIL genes, particularly around conserved areas or motifs, are predominantly linked with DNA repair during transcription and operate preferentially on cytosine-derived lesions, notably 5-hydroxyuracil and 5-hydroxycytocil. We examined the residues under selective constraints in three genes to evaluate the impact of positive selection on the H2TH domain (Figures 1–3). We found five positively chosen sites (S50, Q67, Q69, R90, and G107) in the H2TH domain of NEIL1, six sites (V190, V251, A254, R256, T269, and P280) in the H2TH domain of NEIL2, and just one site (N187) in the H2TH domain of NEIL3 protein. Positively chosen locations within or near the conserved motifs necessary for lyase activity induce nicks in the DNA strand, cleaving the DNA backbone to form a single-strand break at the site of the deleted base containing both 3' and 5'-phosphates. Motifs involved in H2TH activities are conserved among NEIL proteins, as evidenced by numerous sequence alignments of human sequences (Figures 1–3).

We identified genes that are under positive selection across vertebrate species using a variety of site models. We tested alternative models for the dataset using the phylogenetic tree as input data. We used probability analysis to compare several ratio-based models to find codons subjected to positive selection in linked genes. The parameters linked with gene selection in 51 species were calculated using the codeml program, and the two sets of models (M1a vs. M2a and M7 vs. M8) were used to probe positive selection. The likelihood ratio test (LRT) value of 0 ($p > 0.05$) indicated that the NEIL1 gene test was not significant in M1a-M2a. However, the likelihood ratio test (LRT) value of 6.54 indicated that the NEIL1 gene test was significant in M7-M8. According to the test findings for positive selection model M8, the NEIL1 gene showed signs of positive selection,

which means that M8 was accepted, whereas M7 was rejected ($p > 0.05$). The NEB and BEB analysis revealed that the codon positions in the NEIL2 gene showed positive selection at 95% and 99% probabilities, respectively. However, models M1a vs. M2a and M7 vs. M8 were very significant for the NEIL2 gene, with LRT values of 1.48 and 9.16, respectively (Table 1).

Using SLAC, MEME, and FEL analyses, we analyzed the global values to identify the signals of positive selection during the evolutionary process (Figures 4(a)–4(c)). Our findings showed positive evolutionary selection in vertebrates' NEIL1, NEIL2, and NEIL3 genes. We used the Bayesian technique to identify the locations under selective pressure by computing the posterior probability for each codon. Diversifying selection with higher nonsynonymous/synonymous substitution rates is more likely at sites with higher probability than at locations with lower probabilities (Table 2). Using BEB analysis, we discovered multiple locations in the NEIL1 protein under positive selection with a high posterior probability of 95%. We confirmed positive selection by analyzing the results of PAML using the dataset in the Selecton server, which identifies adaptive selection at specific spots in the protein. The MEC model was used to detect the substitution rates. The results revealed that adaptive selection occurred at several amino acid positions in NEIL1, NEIL2, and NEIL3, among other proteins (Figures 1–3).

3.2. Pathway and Process Enrichment Analysis. Pathway and process enrichment analyses have been performed for each gene list using the following ontology sources: KEGG pathway, GO biological processes, reactome gene sets, PANTHER pathway, and WikiPathways. The enrichment background has been constructed from all of the genes in the genome. GO and KEGG pathway enrichment analyses were carried out on the data to get insight into the functions of the differentially expressed genes. At various periods, GO



The selection scale:

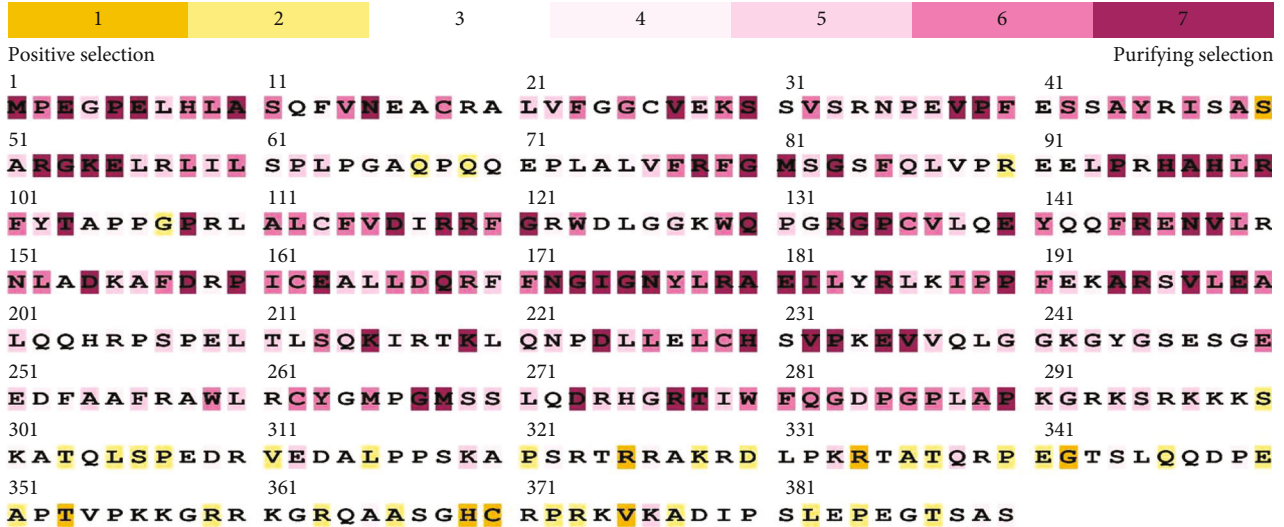
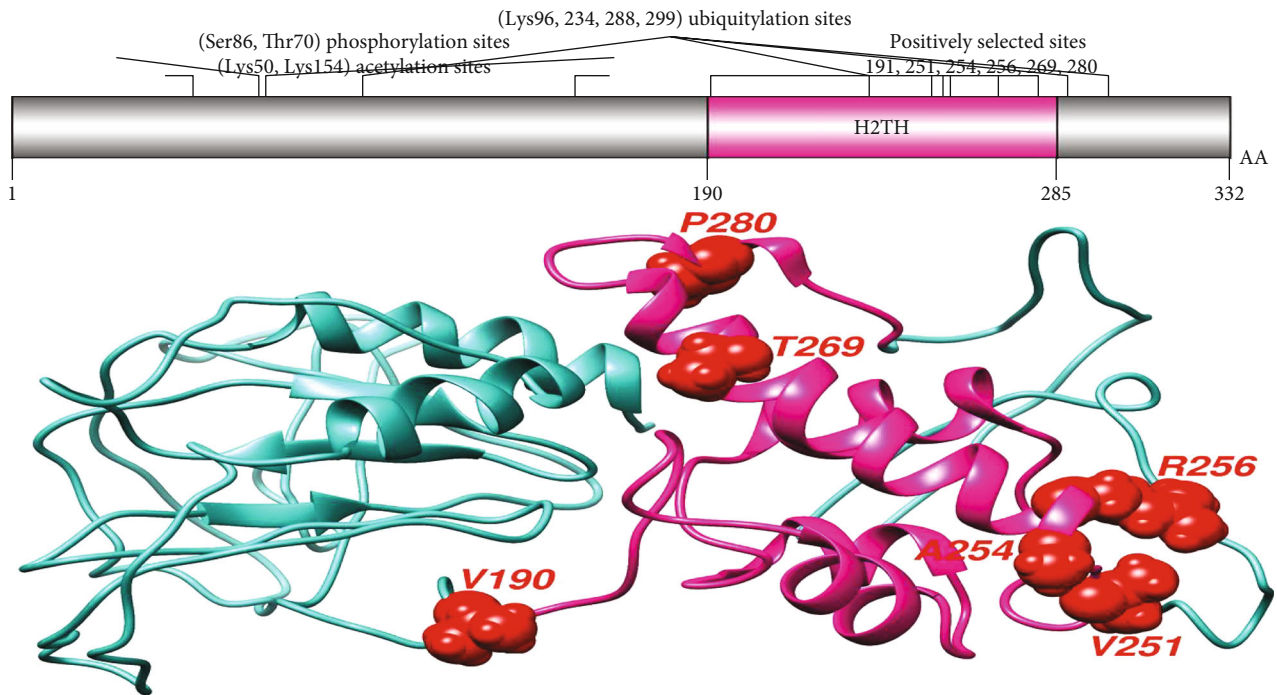


FIGURE 1: Illustration of protein domain structures using DOG 1.0 illustrator. The phosphorylation sites of the NEIL1 protein are shown in the molecular structure and conserved domain analyses. Positively selected amino acid sites were found in conserved domains, especially the H2TH domain. Positively identified locations were drawn into the 3D structure using human NEIL1 as a reference. Selection analyses of human NEIL1 are color-coded and compared to sequences from aligned nucleotide coding sequences. Yellow and brown highlights represent positive selection, the neutral selection is represented by grey and white highlights, and purple highlights on codons represent purifying selection.

annotation indicated that biological processes and molecular activities linked to DNA repair pathways and base excision repair functions were enriched among the genes, indicating that the genes were involved in these processes and functions (Figures 5(a) and 5(b) and Table 3). Genes associated with oxidative stress responses and DNA repair were only significantly enriched in biological processes than their nor-

mal expression (Figures 5(a) and 5(b)). The cluster is represented by the statistically most significant keyword inside the cluster. To further visualize the links between the terms, a subclass of terms was selected and displayed as a linkage plot, in which the terms with a resemblance greater than 0.3 are connected by edges. We choose the GO terms with the best *p* values from each of the twenty clusters, with the



The selection scale:

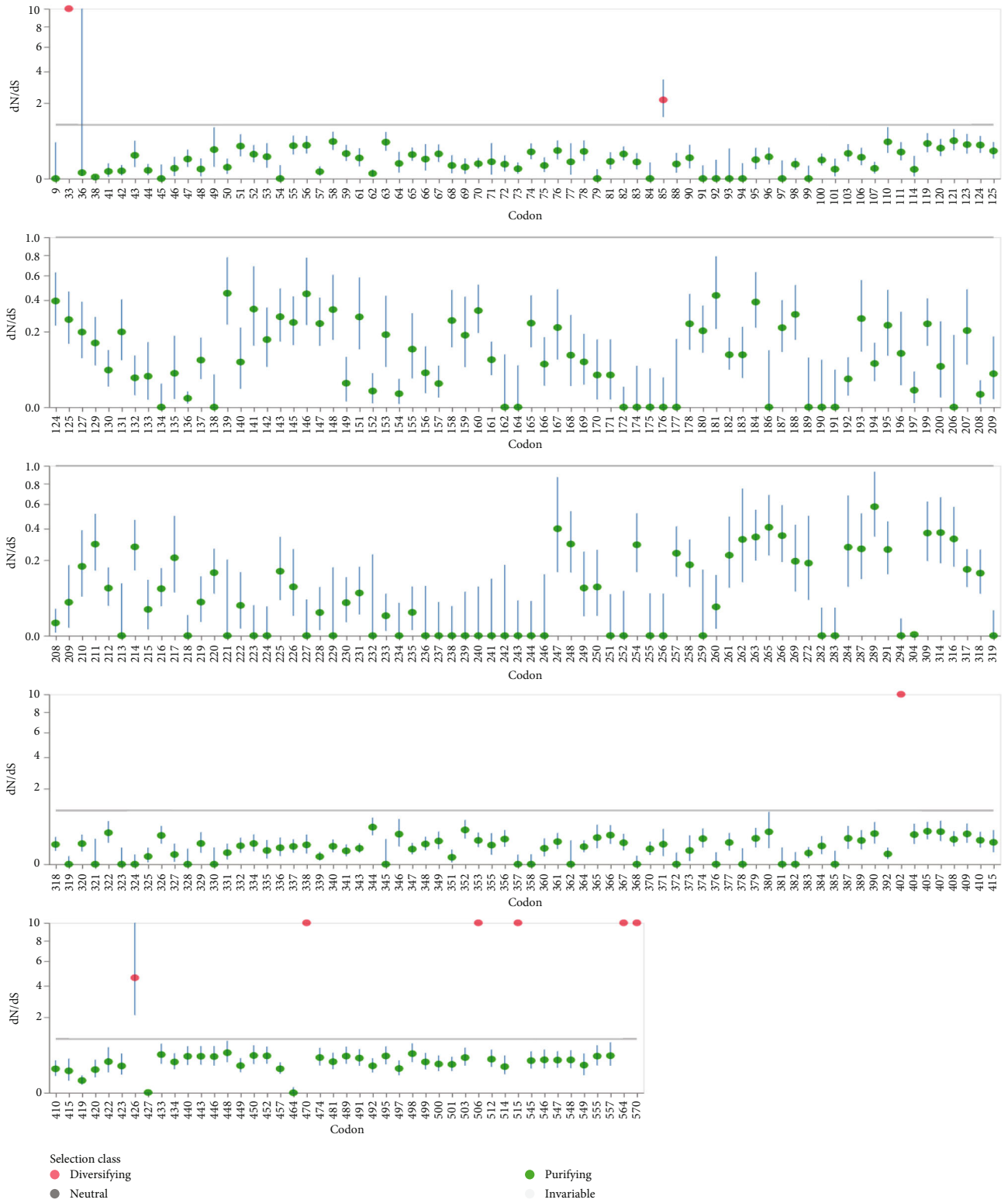
1	2	3	4	5	6	7
Positive selection			Purifying selection			
1	11	21	31	41		
M P E G P L V R K F	H H L V S P F V G Q	Q V V K T G G S S K	K L O P A S L Q S L	W L Q D T Q V H G K		
51	61	71	81	91		
K L F L R F D L D E	E M G P P G S S P T	P E P P Q K E V Q K	E G A A D P K Q V G	E P S G Q K T L D G		
101	111	121	131	141		
S S R S A E L V P Q	G E D D S E Y L E R	D A P A G D A G R W	L R V S E G L F G S	V W V N D F S R A K		
151	161	171	181	191		
K A N K R G D W R D	P S P R L V L H F G	G G G F L A F Y N C	Q L S W S S S P V V	T P T C D I L S E K		
201	211	221	231	241		
F H R G Q A L E A L	G Q A Q P V C Y H L	L D Q R Y F S G L G	N I K N E A L Y R	A G I H E L S L G S		
251	261	271	281	291		
V L S A S R R E V L	V D H V V E E S T A	W L Q G K F Q G R P	Q H T Q V Y Q K E Q	C P A G E Q V M K E		
301	311	321	331			
A F G P E D G L Q R	L T W W C P Q C Q P	Q L S E E P E Q C Q	F S			

FIGURE 2: Illustration of protein domain structures using DOG 1.0 illustrator. The phosphorylation and ubiquitylation sites of the NEIL2 protein are shown in the molecular structure and conserved domain analyses. Positively selected amino acid sites were found in conserved domains, especially in the H2TH domain. Positively identified locations were drawn into the 3D structure using human NEIL2 as a reference. Selection analyses of human NEIL2 are color-coded and compared to sequences from aligned nucleotide coding sequences. Yellow and brown highlights represent positive selection, the neutral selection is represented by grey and white highlights, and purple highlights on codons represent purifying selection.

constraint that there be no more than 15 terms per cluster and a total of 250 terms. The network is shown using Cytoscape, with each node representing an enhanced phrase and colored according to its cluster (Figure 5(b)). Eleven GO keywords were found to be enriched in both genotypes: DNA repair process, DNA repair pathways, full network, DNA repair, base-excision repair, replacement pathway, DNA double-strand break repair, double-strand break repair, hydrolysis, base-excision repair, response via ATR, response to oxidative stress, DNA ligation, and nucleotide-excision repair were detected and enriched (GO:0006281, GO:0006284, GO:0006302, GO:0090305, GO:0006287,

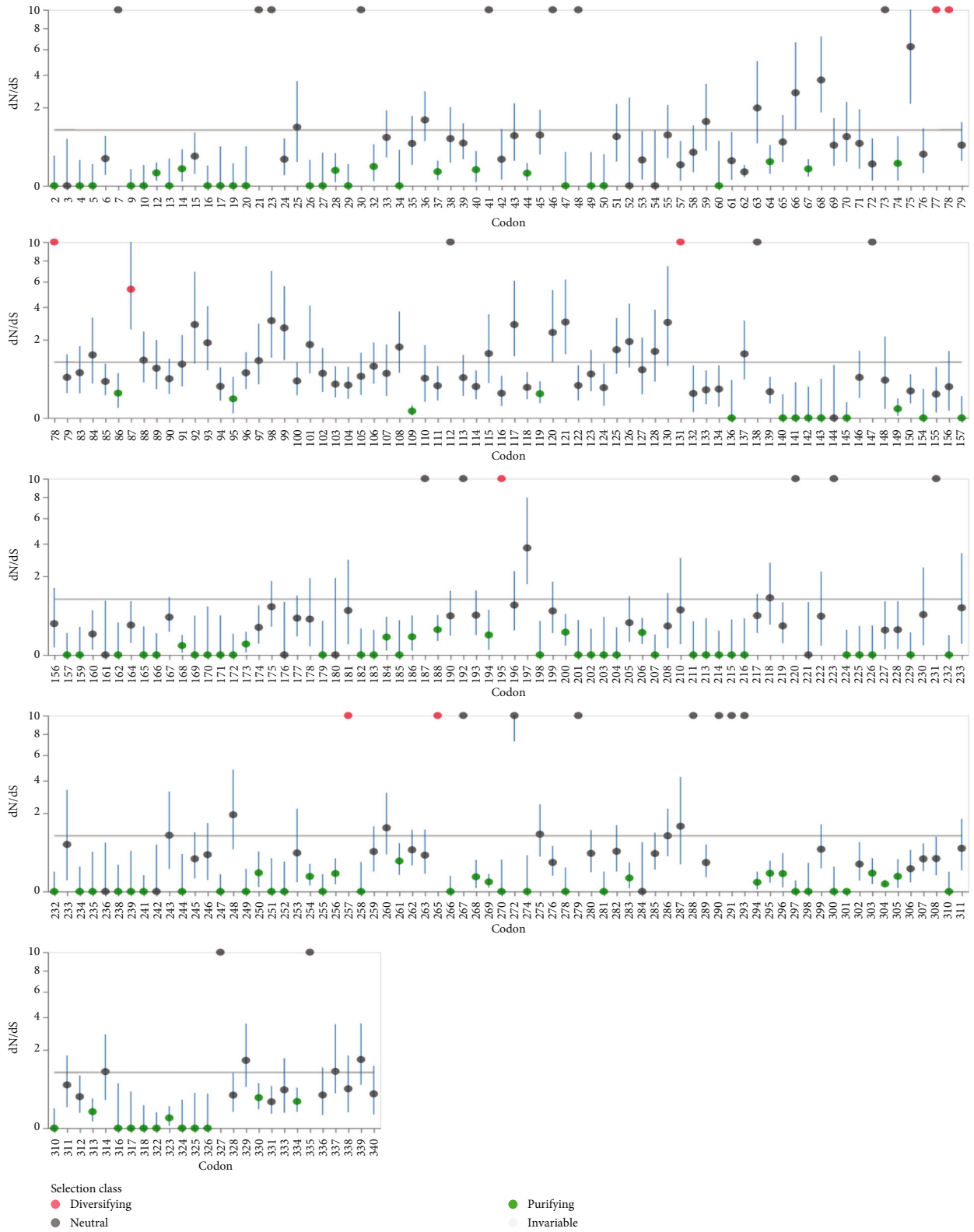
GO:0006979, GO:0006266, GO:0006289, GO:0010332, and GO:0043504). DEGs in each of the three reaction stages were also characterized using GO term enrichment. DEGs in biological processes, molecular function, and cellular components were found in response to cold stress during the early reaction phase (Table 3).

3.3. Conservation and Coevolution Analysis. The functional and structural topographies of positively selected sites were assessed using coevolution analysis, which detects coordinated interactions between residues. In proteins, the coevolution of amino acid sites may be due to structural or



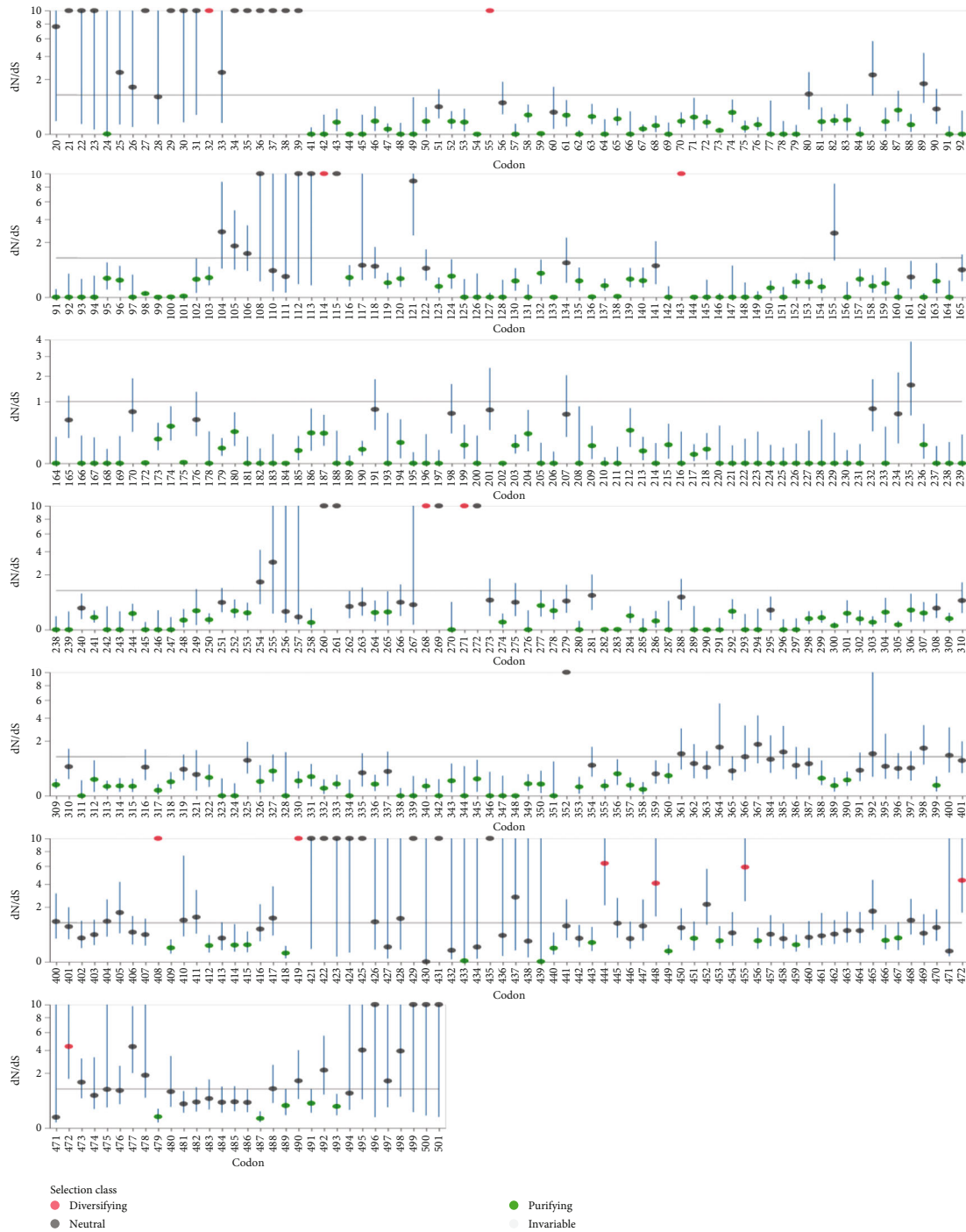
(a)

FIGURE 4: Continued.



(b)

FIGURE 4: Continued.



(c)

FIGURE 4: (a) Maximum likelihood estimations of dN/dS at each site of NEIL1, together with estimated profile confidence intervals. The dN/dS = 1 (neutrality) is depicted as a horizontal gray line. Boundaries between partitions (if present) are shown as vertical dashed lines. Statistical significance is evaluated based on the asymptotic χ^2 distribution. This analysis includes site to site synonymous rate variation. Profile approximate confidence intervals for site-level dN/dS ratios have been computed. (b) Maximum likelihood estimations of dN/dS at each site of NEIL2, together with estimated profile confidence intervals. The dN/dS = 1 (neutrality) is depicted as a horizontal gray line. Boundaries between partitions (if present) are shown as vertical dashed lines. Statistical significance is evaluated based on the asymptotic χ^2 distribution. This analysis includes site to site synonymous rate variation. Profile approximate confidence intervals for site-level dN/dS ratios have been computed. (c) Maximum likelihood estimations of dN/dS at each site of NEIL3, together with estimated profile confidence intervals. The dN/dS = 1 (neutrality) is depicted as a horizontal gray line. Boundaries between partitions (if present) are shown as vertical dashed lines. Statistical significance is evaluated based on the asymptotic χ^2 distribution. This analysis includes site to site synonymous rate variation. Profile approximate confidence intervals for site-level dN/dS ratios have been computed.

TABLE 2: Detailed site-by-site results from the FEL analysis. Statistical significance is evaluated based on the asymptotic χ^2 distribution. This analysis includes site to site synonymous rate variation. Profile approximate confidence intervals for site-level dN/dS ratios have been computed.

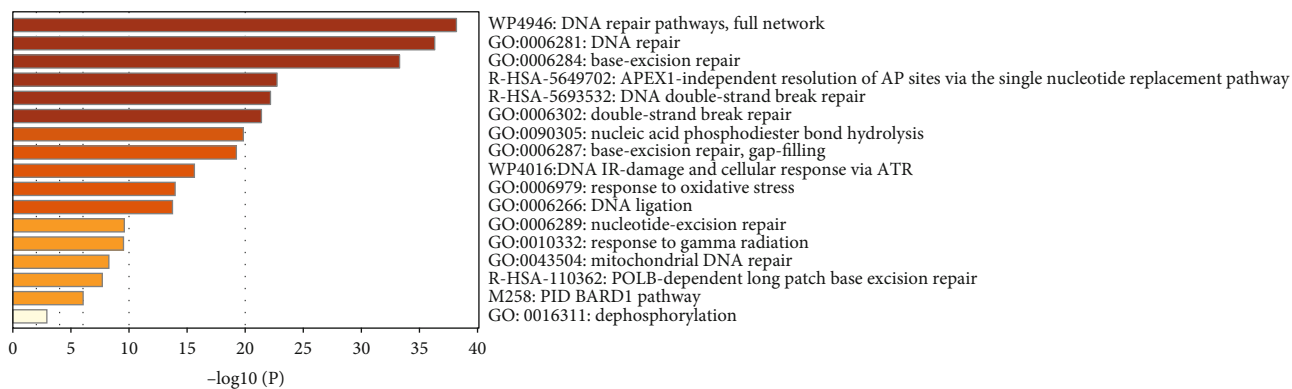
Gene	Codons	α	β	$\alpha = \beta$	LRT	P value	Total branch length	dN/dS LB	dN/dS MLE	dN/dS UB	Selection type
NEIL1	114	0.00	1.519	34.71	2.169	0.1408	296.96	3,827.26	10,00	10,000	Diversifying
	143	0.00	0.538	0.405	3.391	0.0655	3.464	7,865.16	10,00	10,000	Diversifying
	268	0.00	1.002	0.612	1.897	0.1684	5.234	3,827.16	10,00	10,000	Diversifying
	408	0.00	1.123	0.713	4.603	0.0319	6.100	8,626.24	10,00	10,000	Diversifying
	419	0.00	0.495	0.321	1.662	0.1973	2.746	3,826.84	10,00	10,000	Diversifying
	448	0.22	0.885	0.551	1.873	0.1711	4.711	1.347	4.058	13.044	Diversifying
	455	0.24	1.397	0.855	2.04	0.1533	7.314	2.417	5.894	15.896	Diversifying
	77	0.00	0.682	0.404	4.036	0.0445	1.437	6,186.34	10,00	10,000	Diversifying
NEIL2	78	0.00	2.266	1.678	4.706	0.0301	5.968	8,078.11	10,00	10,000	Diversifying
	87	0.70	3.768	2.879	3.02	0.0822	10.24	2.525	5.348	11.869	Diversifying
	131	0.00	1.33	0.851	5.615	0.0178	3.027	7,599.97	10,00	10,000	Diversifying
	195	0.00	0.871	0.605	3.42	0.0644	2.153	6,810.13	10,00	10,000	Diversifying
	257	0.00	0.857	0.558	4.19	0.0407	1.984	6,810.14	10,00	10,000	Diversifying
	265	0.00	0.648	0.415	3.472	0.0624	1.477	6,186.32	10,00	10,000	Diversifying
	32	0.00	1.281	10000	1.093	0.2957	85553.36	3,827.32	10,00	10,000	Diversifying
	55	0.00	0.519	0.406	1.487	0.2227	3.469	7,599.63	10,00	10,000	Diversifying
NEIL3	114	0.00	1.519	55.08	2.169	0.1408	471.262	3,827.26	10,00	10,000	Diversifying
	143	0.00	0.538	0.405	3.391	0.0655	3.464	7,865.16	10,00	10,000	Diversifying
	268	0.00	1.002	0.612	1.897	0.1684	5.234	3,827.16	10,00	10,000	Diversifying
	271	0.00	0.694	0.281	1.615	0.2038	2.407	1,464.80	10,00	10,000	Diversifying
	408	0.00	1.123	0.713	4.603	0.0319	6.100	8,626.24	10,00	10,000	Diversifying
	419	0.00	0.495	0.321	1.662	0.1973	2.746	3,826.84	10,00	10,000	Diversifying
	444	0.74	4.696	3.046	1.594	0.2067	26.059	2.105	6.370	47.341	Diversifying
	448	0.22	0.885	0.551	1.873	0.1711	4.711	1.347	4.058	13.044	Diversifying
455	0.24	1.397	0.855	2.04	0.1533	7.314	2.417	5.894	15.896	Diversifying	
472	0.57	2.486	1.641	1.205	0.2724	14.039	1.59	4.361	13.636	Diversifying	

α : synonymous substitution rate at a site; β : nonsynonymous substitution rate at a site; $\alpha = \beta$: the rate estimate under the neutral model; LRT: likelihood ratio test statistic for beta = alpha, versus beta \neq alpha; P value: asymptotic P value for evidence of selection, i.e., beta \neq alpha; total branch length: the total length of branches contributing to inference at site used to scale dN-dS; dN/dS LB: 95% profile likelihood CI lower bound for dN/dS; dN/dS MLE: point estimate for site dN/dS; dN/dS UB: 95% profile likelihood CI upper bound for dN/dS.

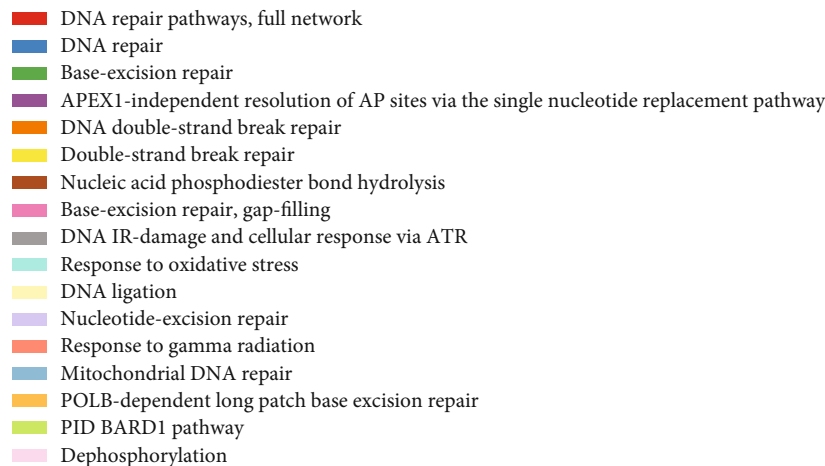
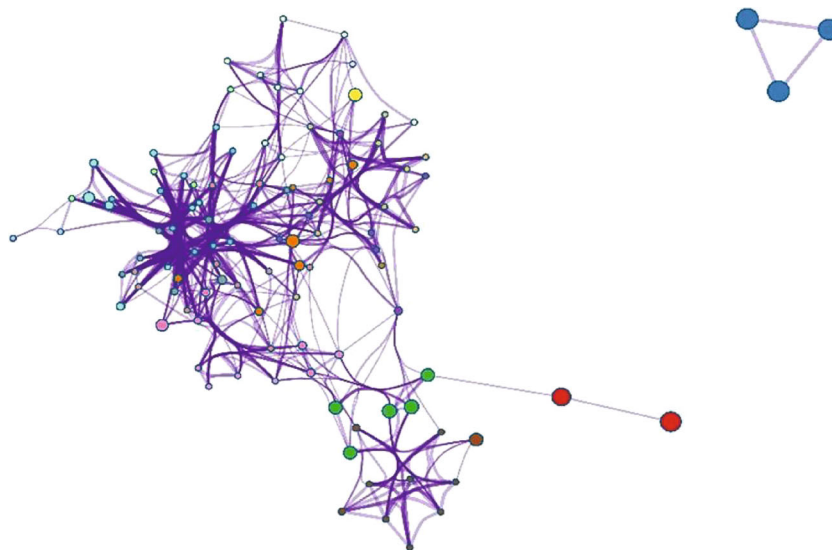
functional links. We employed homologs of NEIL1, NEIL2, and NEIL3 as inputs in a coevolution analysis to identify different coevolving pairs formerly recognized under positive selection. A network diagram displaying the residues' connections was drawn to find a relationship between them (Figure 6). There is a correlation between amino acid residues and the extent of coevolutionary linkages between them. Nodes in a subnetwork containing only amino acids found in the protein's Fapy-DNA-glyco, H2TH, and zf-GRF domains had positively selected residues. NEIL1, NEIL2, and NEIL3 genes were shown to have coevolving probabilities according to Pearson correlation (r). This covariance study uncovered the spans and interactions between positively selected protein domain residues, as well as groups of functionally connected sites (Figure 7). The dynamic 3D panel allows you to zoom and rotate while labelling chosen residues. These findings support the idea that the NEIL protein sections that adhere to these conserved domains are structurally and functionally distinct.

4. Discussion

Animal genomes undergo significant evolutionary changes due to divergence, the adaptation of genetic material from gene duplication, divergent lineages, and epigenesis [48]. We use the findings of this phylogenetic study to infer the evolutionary history of repair pathways and the proteins that make them up and make predictions about the repair phenotypes of species with sequenced genomes [17]. Aside from that, we look at how evolutionary analysis can be used to study full genome sequences, how complete genome sequences may be used in evolutionary research, and the benefits of employing both the phylogenetic technique and evolutionary analysis [13]. Positive selection was discovered in the vertebrate NEIL genes. Positively chosen residues came in a variety of quantities and distributions. In the functional domains of the human NEIL1, NEIL2, and NEIL3 proteins, we discovered two positively chosen codons (Figures 1–3). The Nei-like glycosylases are crucial



(a)



(b)

FIGURE 5: The common target genes of DNA repair pathways. GO annotation and KEGG were used to analyze these target genes.

molecules involved in base excision repair in vertebrate genomes. It activates the range activity of substrates involved in DNA repair and maintenance. These proteins are involved in DNA replication coordination, detecting damage sites for repairing, and modifications in transcriptional regulation for expression of activated genes and recruitment

of multifunctional protein factors for specific functions required for coordination in these activities [49]. The H2TH domain is required for downstream signaling in NEIL proteins since it is involved in DNA damage repair for accurate genome replication and transcript interpretation. Due to functional restrictions, residual with crucial

TABLE 3: Go terms and clusters with enriched keywords (one per cluster). This is the number of genes in the user-supplied lists that belong to the ontology term.

GO	Category	Description	Count	%	Log10(P)	Log10(q)
WP4946	WikiPathways	DNA repair pathways, full network	17	85	-38.21	-33.87
GO:0006281	GO biological processes	DNA repair	20	100	-36.36	-32.31
GO:0006284	GO biological processes	Base-excision repair	13	65	-33.31	-29.57
R-HSA-5649702	Reactome gene sets	Replacement pathway	7	35	-22.77	-19.46
R-HSA-5693532	Reactome gene sets	DNA double-strand break repair	12	60	-22.18	-18.91
GO:0006302	GO biological processes	Double-strand break repair	12	60	-21.4	-18.17
GO:0090305	GO biological processes	Hydrolysis	12	60	-19.86	-16.66
GO:0006287	GO biological processes	Base-excision repair	7	35	-19.23	-16.06
WP4016	WikiPathways	Response via ATR	8	40	-15.64	-12.52
GO:0006979	GO biological processes	Response to oxidative stress	10	50	-14.01	-10.94
GO:0006266	GO biological processes	DNA ligation	5	25	-13.73	-10.68
GO:0006289	GO biological processes	Nucleotide-excision repair	5	25	-9.6	-6.86
GO:0010332	GO biological processes	Response to gamma radiation	5	25	-9.56	-6.83
GO:0043504	GO biological processes	Mitochondrial DNA repair	3	15	-8.3	-5.8
R-HSA-110362	Reactome gene sets	Excision repair	3	15	-7.68	-5.24
M258	Canonical pathways	PID BARD1 PATHWAY	3	15	-6.05	-3.78
GO:0016311	GO biological processes	Dephosphorylation	3	15	-2.97	-0.87

functions may be under modest evolutionary pressure [50]. Core residues that make up the polar surface and hydrophobic core of the Fpg/Nei family of DNA glycosylases are conserved [51]. The Fapy domain of NEIL1 contained a smaller percentage of positively chosen codons and displayed more selection limitations than other NEIL proteins, as predicted [16].

The study of genetics and evolutionary processes is required to understand the regulatory mechanisms of several physiologically relevant genes, such as the NEIL gene family in mammals. It was discovered that positive selection occurred at many locations in three genes across mammals, reptiles, and birds (Table 1). A gene's overall characteristic expression in these lineages was identical to that of genes that were not under selection or only selected in mammals [52]. According to our data, infections appear to represent a constant selection pressure throughout vertebrate clades. As we discovered, NEIL proteins had their amino acids buried and exposed, which did not affect H2TH's electrostatic distribution. Table 1 shows that residues (S50, Q67, R90, G107) of the NEIL2 H2TH domain that interacts with the NEIL3-H2TH domain are under positive selection. These residues are conserved across avians and humans (Figure 8), consistent with its crucial biological functions. In mammals, reptiles, and birds, three genes were identified under positive selection (Table 1). Because of this, we established standards similar to those of lineages where the selection was weak or merely positive in the vertebrates. Based on our findings, infections may exert a selective impact on several vertebrate taxa [53]. The first vertebrate phylogeny shifts in IRF, i.e., the discovery of an adaptive immune system, has been questioned by previous phylogenetic studies [54]. According to some theories, the IRF gene family originally

consisted of two branches, which have been identified in all bilaterians, and cnidarians [55].

Analysis of protein alignments by MAFFT has shown an H2TH domain conserved in all three sequences of the NEIL1, NEIL2, and NEIL3 genes (Figures 2(b), 1(b), and 3(b)). The LRT values for NEIL1, NEIL2, and NEIL3 genes were 0, 1.48, and 0 for the M2 and M1 evolutionary models. Protein alterations in purifying selection zones suffer nonidentical shifts that are harmful to health and so have a low possibility of being fixed throughout evolution [56, 57]. This led to identifying amino acid residues having $\omega > 1$ value in the next phase. Using the evolutionary model, several sites in NEIL1, NEIL2, and NEIL3 genes were identified under positive selection with LRT values 6.54, 9.16, and 0.48, respectively (Table 1). Our findings show that some locations

in other proteins subjected to strong positive selection have developed faster than the mature protein [22, 23]. As a result, dynamic selection induces a modification that aims to increase protein secretion effectiveness, as seen in the case of the NEIL proteins [58]. A branch site test was performed to determine which branches in vertebrate clades were being selectively selected for in the NEIL2 gene, and we found that just a few branches in the mammalian clade were being selectively selected. The NEIL3 gene exhibited signs of positive selection in both mammalian and avian clades. When we studied the NEIL genes, we identified positive selection in the majority of the vertebrate, which was surprising. We used the aBS-REL model to confirm our findings since branch site analysis can lead to ambiguous selection owing to multinucleotide mutations. The aBS-REL and site models both showed comparable patterns of selection. The findings show that the general selection trends we observed are correct [21, 23].

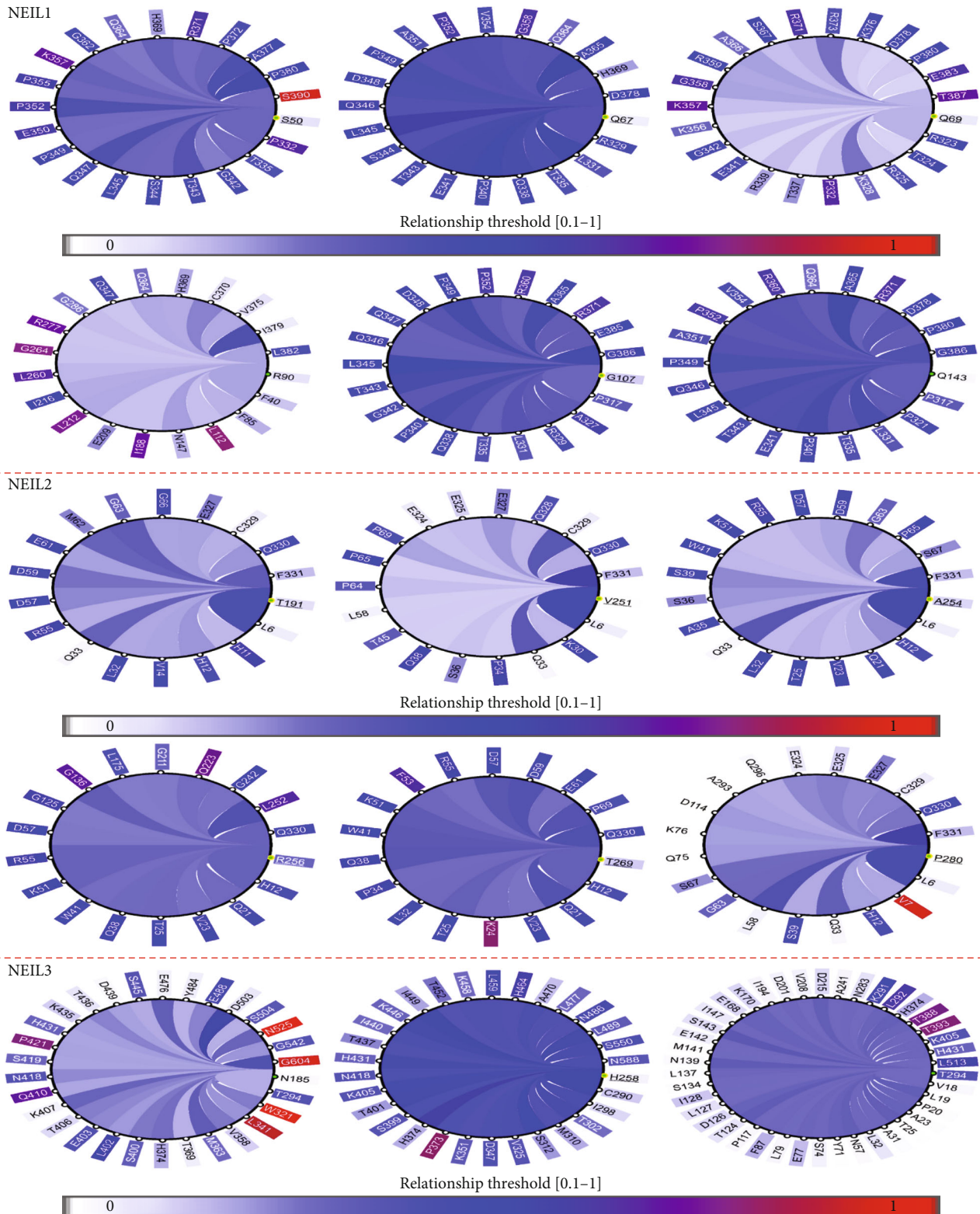


FIGURE 6: Coevolution of conserved domain residues. The circular connection diagram focuses on the residues with cutoffs of the NEIL1, NEIL2, and NEIL3 genes. The graphic shows the positions of amino acids in the protein. The colors of the curves denote covariance scores among spots.

In general, these proteins in vertebrate lineages demonstrated little indication of positive selection, suggesting that genetic diversity is hampered by persistent selection pres-

sure, especially in the avian clade. This might be linked to the idea that the gene's evolutionary history has been free of gene duplication occurrences [59]. Gene duplication is

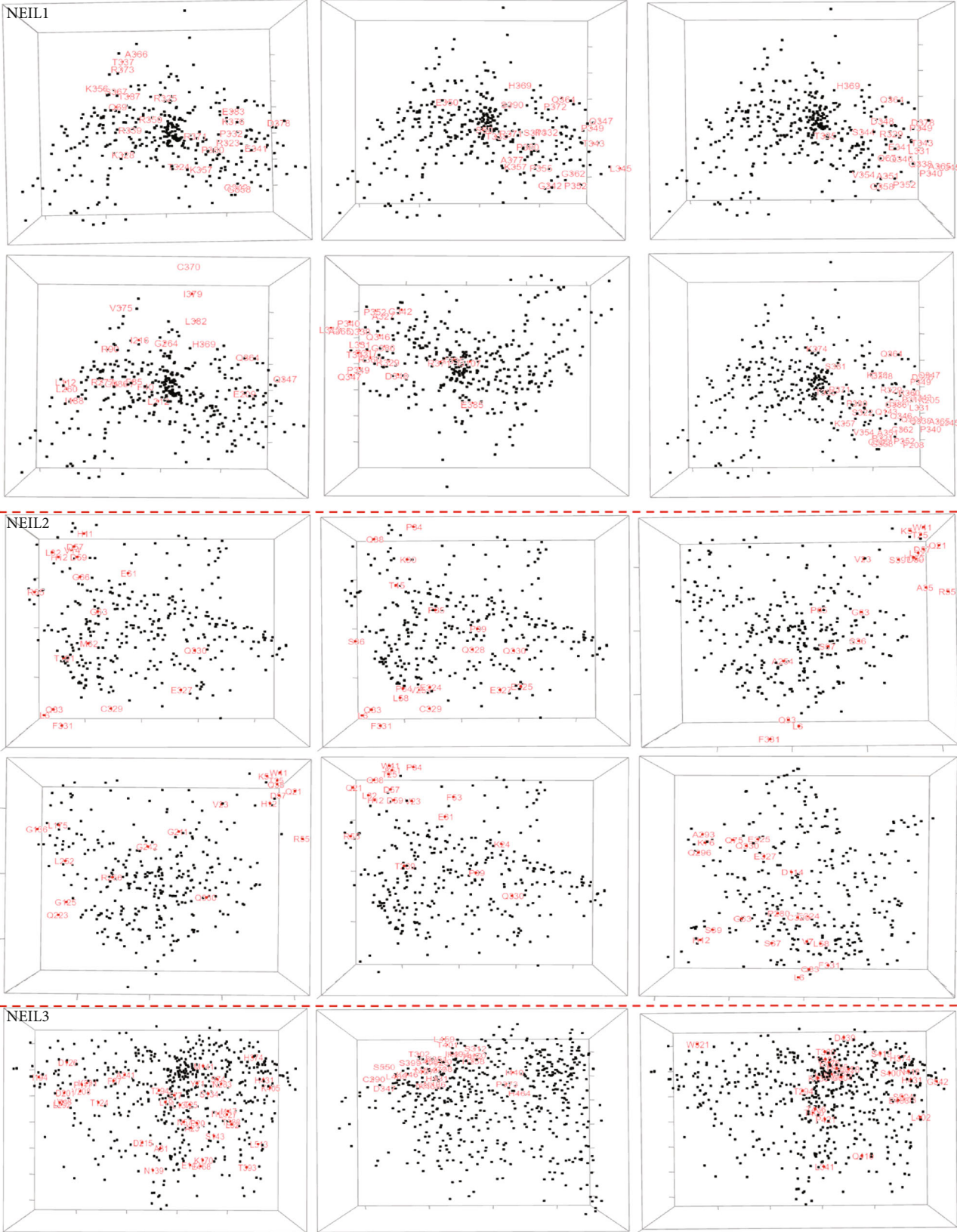


FIGURE 7: 3D multidimensional scaling (MDS) scatterplots of covarying sites in NEIL1, NEIL2, and NEIL3 genes. (A) Highlighted red are the residues corresponding to the positively selected residue. Both black and red spots labeled showing the residues.

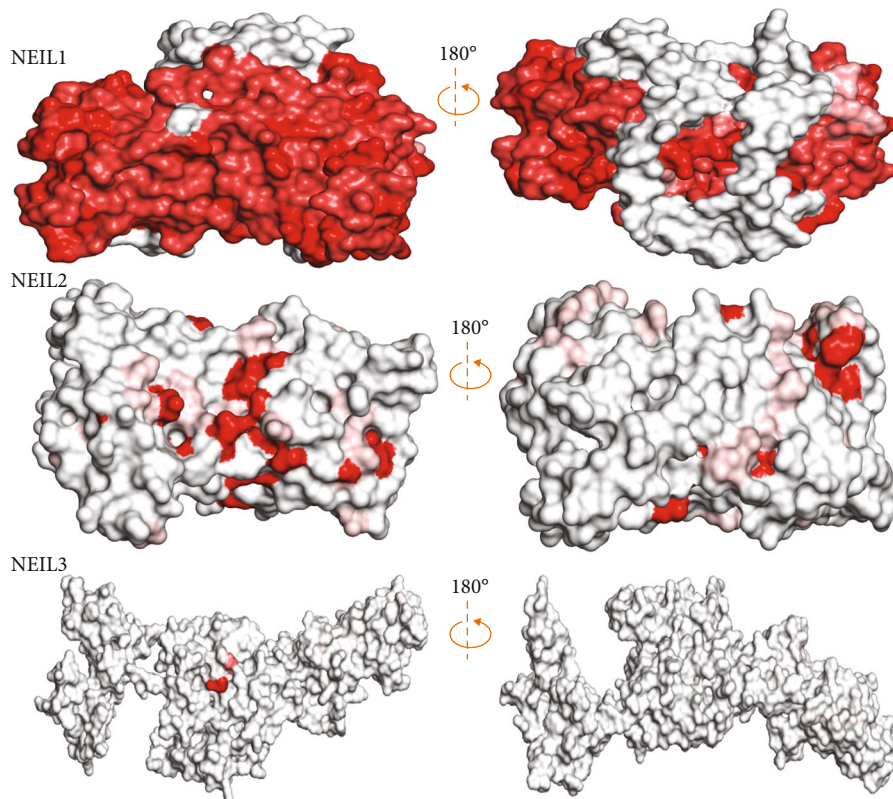


FIGURE 8: Surface representation of the human NEIL1, NEIL2, and NEIL3 proteins colored conferring to sequence conservation. The color ramping from white (low score) to red (high score) allows to quickly identify areas of weak and robust structural conservation of the proteins.

an evolutionary strategy that allows genomes to evolve in various ways. Positive selection occurs following a duplication event for other proteins, implying the selective pressure that supports genetic variety [60]. This relaxation was absent during the development of NEIL genes in avian and different vertebrate lineages, supporting Bayesian phylogenetic techniques. Our data show that purifying selection has traditionally determined the molecular evolution of NEIL genes in vertebrates. Proteins depend on coordinated interactions among their amino acid residues for structure and function [61]. As a result, identifying structural properties of positively chosen amino acid residues might be similar to detecting residues that covary throughout evolution. The structural or functional connections between amino acid residues inside proteins may have resulted in this coevolutionary relationship. Protein coevolution has previously been linked to protein constancy and intermolecular interactions [62, 63]. By measuring the dN/dS of mammalian Neil's sequences, we discovered that S50, Q67, Q69, R90, and G107 in the H2TH domain of NEIL1, six sites (V190, V251, A254, R256, T269, and P280) in the H2TH domain of NEIL2, and just one site (N187) in the H2TH domain of NEIL3 protein may play essential roles in several of the previously reported activities (Figures 1–3). The domains of NEIL1, NEIL2, and NEIL3 are largely conserved among vertebrates, suggesting that their functions are not duplicated and that this selection pressure stems from NEIL specialization for base excision repair pathways [49]. Proteins

that interact directly with other molecules, on the other hand, have a higher probability of adapting to each other's evolutionary changes [64]. In the context of their initial function, NEIL genes may have experienced such coevolutionary features [65].

As a result, we used several sequence alignments obtained for NEIL1, NEIL2, and NEIL3 homologs to undertake a coevolution study, and we discovered positively chosen coevolving residues with significant variability (Figure 6). These findings show that the areas matching to domains in NEIL proteins are structure-function modules. Within H2Th and ZF-GRF domains, there were only a few adaptive substitutions, indicating a stable selection pattern throughout evolution. There has recently been an increase in the number of species with whole-genome sequences, allowing researchers to compare and contrast their DNA repair systems and discover how different repair genes and functional pathways have evolved [13]. It appears that the repair process is very variable among species, as evidenced by a worldwide comparison of DNA repair proteins based on all known genome sequences from microbes, archaea, and eukaryotes [66]. The unique gene insertions are commonly polymorphic (present or absent) at orthologous loci, making them highly useful genetic markers that may be employed in clustering studies, animal evolutionary history, population structure, and demography [67]. As a whole, these elements are known to affect the genome in a wide range of ways, including increasing the size of genome and

instability, epigenetic regulation, and RNA editing [68]. Previously, [69] demonstrated significant structural conservation in the two-domain architecture, despite the considerable sequence divergence in the primarily b-strand-rich N-terminal domain. More study is needed to uniformly apply these potential indicators to different species because of genetic variation and locus distribution [70]. Species can develop their repair system through gene duplication and gene loss that often occurs in their populations [17]. Copy number variations of DNA repair genes will result from the species-specific history of gene duplication and loss, which will have a profound effect on organismal phenotypes such as mutation rates [16], lifespan [71], and adaptation to extreme environments [72, 73].

Furthermore, we reveal interesting data relating genetic changes in transcription activators and repressor elements from an evolutionary approach.

5. Conclusions

We think that the findings presented here may be used to begin experimental research of DNA repair in animals with entire genome sequences and better understand the evolution of proteins involved in DNA repair in the vertebrate genome. In conclusion, these comprehensive comparative assessments of DNA repair genes, particularly NEIL1, NEIL2, and NEIL3 in several species, revealed that these are the strong candidate genes related to tree lifetime; endogenous and external stressors can cause single-strand breaks bulky lesions, which play a critical role in DNA damage repair. As a result, our research may provide a platform on which to build future studies examining the connections between DNA repair and vertebrate species lifespan. Systematic comparative genome investigations will give vital insights to elucidate the links between DNA repair and lifespan development in many creatures when genome sequences of increasingly varied animal species become accessible. The present findings revealed that the conservation of dynamics as a component of a protein fold might have implications that go beyond enzyme catalysis in the future. The limitation of this study is that we used a limited number of species for our analyses may not be sufficient. Further analysis can be performed using genome sequencing technology from a large number of species available in future. Applying our analyses to the larger set of data will uncover the adaptive evolution of gene families involved in cellular longevity.

Data Availability

Data will be available openly to the readers.

Conflicts of Interest

There is no conflict of interest in the conduction of this study.

Authors' Contributions

HIA, GA, SS, and GH inputed the study's data curation. HIA, GA, and SS contributed in the formal analysis. CJ contributed in the funding acquisition. CJ carried out the investigation. HIA, SA, and GH were involved in the methodology. HIA and CJ contributed in the resources. HIA, GH, SS, and SA contributed in the software. CJ was involved in the supervision. HIA and CJ were involved in the validation. GH, MJS, SS, and SA were involved in the visualization. HIA, GA, MJS, and SS carried the writing-original draft. GH, SS, SA, MJS, and CJ carried out the writing-review and editing.

Acknowledgments

This work was supported by the GDAS project of Science and Technology Development (2019-GDASYL-0103059).

Supplementary Materials

Supplementary material showing the list of vertebrate species and NCBI Genbank accession numbers for sequences used to build datasets for hypothesis testing of the NEIL genes. Table S1: List of vertebrate species and NCBI Genbank accession numbers for sequences used to build datasets for hypothesis testing of the NEIL1 gene. Table S2: List of vertebrate species and NCBI Genbank accession numbers for sequences used to build datasets for hypothesis testing of the NEIL2 gene. Table S3: List of vertebrate species and NCBI Genbank accession numbers for sequences used to build datasets for hypothesis testing of the NEIL3 gene. (*Supplementary Materials*)

References

- [1] A. Sharma, P. Gupta, and P. K. Prabhakar, "Endogenous repair system of oxidative damage of DNA," *Current Chemical Biology*, vol. 13, no. 2, pp. 110–119, 2019.
- [2] J. E. Klaunig, Z. Wang, X. Pu, and S. Zhou, "Oxidative stress and oxidative damage in chemical carcinogenesis," *Toxicology and Applied Pharmacology*, vol. 254, no. 2, pp. 86–99, 2011.
- [3] Q. Zhu, Y. Niu, M. Gundry, and C. Zong, "Single-cell damage genome profiling unveils vulnerable genes and functional pathways in human genome toward DNA damage," *Science Advances*, vol. 7, no. 27, 2021.
- [4] M. Liu, V. Bandaru, J. P. Bond et al., "The mouse ortholog of NEIL3 is a functional DNA glycosylase in vitro and in vivo," *Proceedings of the National Academy of Sciences*, vol. 107, no. 11, pp. 4925–4930, 2010.
- [5] O. A. Kladova, I. R. Grin, O. S. Fedorova, N. A. Kuznetsov, and D. O. Zharkov, "Conformational dynamics of damage processing by human DNA glycosylase NEIL1," *Journal of Molecular Biology*, vol. 431, no. 6, pp. 1098–1112, 2019.
- [6] T. K. Hazra, A. Das, S. Das, S. Choudhury, Y. W. Kow, and R. Roy, "Oxidative DNA damage repair in mammalian cells: a new perspective," *DNA Repair*, vol. 6, no. 4, pp. 470–480, 2007.

- [7] V. Rolseth, L. Luna, A. K. Olsen et al., “No cancer predisposition or increased spontaneous mutation frequencies in NEIL DNA glycosylases-deficient mice,” *Scientific Reports*, vol. 7, no. 1, p. 4384, 2017.
- [8] K. Imamura, A. Averill, S. S. Wallace, and S. Doublé, “Structural characterization of viral ortholog of human DNA glycosylase NEIL1 bound to thymine glycol or 5-hydroxyuracil-containing DNA,” *Journal of Biological Chemistry*, vol. 287, no. 6, pp. 4288–4298, 2012.
- [9] I. G. Minko, P. P. Christov, L. Li, M. P. Stone, A. K. McCullough, and R. S. Lloyd, “Processing of $_N^5$ -substituted formamidopyrimidine DNA adducts by DNA glycosylases NEIL1 and NEIL3,” *DNA Repair*, vol. 73, pp. 49–54, 2019.
- [10] A. Das, L. Wiederhold, J. B. Leppard et al., “NEIL2-initiated, APE-independent repair of oxidized bases in DNA: evidence for a repair complex in human cells,” *DNA Repair*, vol. 5, no. 12, pp. 1439–1448, 2006.
- [11] N. Sharma, S. Chakravarthy, M. J. Longley, W. C. Copeland, and A. Prakash, “The C-terminal tail of the NEIL1 DNA glycosylase interacts with the human mitochondrial single-stranded DNA binding protein,” *DNA Repair*, vol. 65, pp. 11–19, 2018.
- [12] S. Doublé, V. Bandaru, J. P. Bond, and S. S. Wallace, “The crystal structure of human endonuclease VIII-like 1 (NEIL1) reveals a zincless finger motif required for glycosylase activity,” *Proceedings of the National Academy of Sciences*, vol. 101, no. 28, pp. 10284–10289, 2004.
- [13] B. E. Eckenroth, V. B. Cao, A. M. Averill, J. A. Dragon, and S. Doublé, “Unique structural features of mammalian NEIL2 DNA glycosylase prime its activity for diverse DNA substrates and environments,” *Structure*, vol. 29, no. 1, pp. 29–42.e4, 2021.
- [14] R. B. Huey, T. Garland Jr., and M. Turelli, “Revisiting a key innovation in evolutionary biology: Felsenstein’s ‘phylogenies and the comparative method,’” *The American Naturalist*, vol. 193, no. 6, pp. 755–772, 2019.
- [15] L. Aravind, D. R. Walker, and E. V. Koonin, “Conserved domains in DNA repair proteins and evolution of repair systems,” *Nucleic Acids Research*, vol. 27, no. 5, pp. 1223–1242, 1999.
- [16] Y. A. Blue, J. Kusumi, and A. Satake, “Copy number analyses of DNA repair genes reveal the role of poly (ADP-ribose) polymerase (PARP) in tree longevity,” *Iscience*, vol. 24, no. 7, article 102779, 2021.
- [17] J. A. Eisen and P. C. Hanawalt, “A phylogenomic study of DNA repair genes, proteins, and processes,” *Mutation Research/DNA Repair*, vol. 435, no. 3, pp. 171–213, 1999.
- [18] A. Robertson, A. Klungland, T. Rognes, and I. Leiros, “DNA repair in mammalian cells: base excision repair: the long and short of it,” *Cellular and Molecular Life Sciences*, vol. 66, no. 6, pp. 981–993, 2009.
- [19] H. I. Ahmad, M. J. Ahmad, M. M. Adeel, A. R. Asif, and X. Du, “Positive selection drives the evolution of endocrine regulatory bone morphogenetic protein system in mammals,” *Oncotarget*, vol. 9, no. 26, pp. 18435–18445, 2018.
- [20] H. I. Ahmad, A. R. Asif, M. J. Ahmad et al., “Adaptive evolution of peptidoglycan recognition protein family regulates the innate signaling against microbial pathogens in vertebrates,” *Microbial Pathogenesis*, vol. 147, article 104361, 2020.
- [21] H. I. Ahmad, G. Afzal, A. Jamal et al., “In silico structural, functional, and phylogenetic analysis of cytochrome (CYPD) protein family,” *BioMed Research International*, vol. 2021, Article ID 5574789, 13 pages, 2021.
- [22] H. I. Ahmad, G. Liu, X. Jiang et al., “Adaptive selection at agouti gene inferred breed specific selection signature within the indigenous goat populations,” *Asian-Australasian journal of animal sciences*, 2017.
- [23] H. I. Ahmad, J. Zhou, M. J. Ahmad et al., “Adaptive selection in the evolution of programmed cell death-1 and its ligands in vertebrates,” *Aging (Albany NY)*, vol. 12, no. 4, pp. 3516–3557, 2020.
- [24] H. McWilliam, W. Li, M. Uludag et al., “Analysis tool web services from the EMBL-EBI,” *Nucleic Acids Research*, vol. 41, no. W1, pp. W597–W600, 2013.
- [25] C. B. Do, M. S. Mahabhashyam, M. Brudno, and S. Batzoglou, “ProbCons: probabilistic consistency-based multiple sequence alignment,” *Genome Research*, vol. 15, no. 2, pp. 330–340, 2005.
- [26] J. Castresana, “Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis,” *Molecular Biology and Evolution*, vol. 17, no. 4, pp. 540–552, 2000.
- [27] X. Ma, B. Liu, Z. Gong, X. Yu, and J. Cai, “Structural and evolutionary adaptation of NOD-like receptors in birds,” *BioMed research international*, vol. 2021, Article ID 5546170, 11 pages, 2021.
- [28] H. I. Ahmad, G. Liu, X. Jiang et al., “Maximum-likelihood approaches reveal signatures of positive selection in BMP15 and GDF9 genes modulating ovarian function in mammalian female fertility,” *Ecology and Evolution*, vol. 7, no. 21, pp. 8895–8902, 2017.
- [29] Z. Yang, “PAML 4: phylogenetic analysis by maximum likelihood,” *Molecular Biology and Evolution*, vol. 24, no. 8, pp. 1586–1591, 2007.
- [30] S. Weaver, S. D. Shank, S. J. Spielman, M. Li, S. V. Muse, and S. L. Kosakovsky Pond, “Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes,” *Molecular Biology and Evolution*, vol. 35, no. 3, pp. 773–777, 2018.
- [31] M. Garber, M. Guttman, M. Clamp, M. C. Zody, N. Friedman, and X. Xie, “Identifying novel constrained elements by exploiting biased substitution patterns,” *Bioinformatics*, vol. 25, no. 12, pp. i54–i62, 2009.
- [32] A. Stern, A. Doron-Faigenboim, E. Erez, E. Martz, E. Bacharach, and T. Pupko, “Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach,” *Nucleic acids research*, vol. 35, no. Web Server, pp. W506–W511, 2007.
- [33] A. R. Lemmon, J. M. Brown, K. Stanger-Hall, and E. M. Lemmon, “The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference,” *Systematic Biology*, vol. 58, no. 1, pp. 130–145, 2009.
- [34] M. A. Raheem, M. Xue, H. I. Ahmad et al., “Adaptation to host-specific bacterial pathogens drive rapid evolution of novel PhoP/PhoQ regulation pathway modulating the virulence,” *Microbial Pathogenesis*, vol. 141, article 103997, 2020.
- [35] L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. Sternberg, “The Phyre2 web portal for protein modeling, prediction and analysis,” *Nature Protocols*, vol. 10, no. 6, pp. 845–858, 2015.
- [36] E. F. Pettersen, T. D. Goddard, C. C. Huang et al., “UCSF chimera—a visualization system for exploratory research and analysis,” *Journal of Computational Chemistry*, vol. 25, no. 13, pp. 1605–1612, 2004.

- [37] A. Prakash, M. Jeffries, A. Bateman, and R. D. Finn, "The HMMER web server for protein sequence similarity search," *Current protocols in bioinformatics*, vol. 60, no. 1, 2017.
- [38] A. Marchler-Bauer, M. K. Derbyshire, N. R. Gonzales et al., "CDD: NCBI's conserved domain database," *Nucleic Acids Research*, vol. 43, no. D1, pp. D222–D226, 2015.
- [39] J. Mistry, S. Chuguransky, L. Williams et al., "Pfam: the protein families database in 2021," *Nucleic Acids Research*, vol. 49, no. D1, pp. D412–D419, 2021.
- [40] J. Ren, L. Wen, X. Gao, C. Jin, Y. Xue, and X. Yao, "DOG 1.0: illustrator of protein domain structures," *Cell Research*, vol. 19, no. 2, pp. 271–273, 2009.
- [41] H. Ashkenazy, E. Erez, E. Martz, T. Pupko, and N. Ben-Tal, "ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids," *Nucleic Acids Research*, vol. 38, no. Web Server, pp. W529–W533, 2010.
- [42] G. Celniker, G. Nimrod, H. Ashkenazy et al., "ConSurf: using evolutionary data to raise testable hypotheses about protein function," *Israel Journal of Chemistry*, vol. 53, no. 3–4, pp. 199–206, 2013.
- [43] D. Corcoran, N. Maltbie, S. Sudalairaj, F. N. Baker, J. Hirschfeld, and A. Porollo, "CoeViz 2: protein graphs derived from amino acid covariance," *Frontiers in Bioinformatics*, vol. 1, 2021.
- [44] F. N. Baker and A. Porollo, "CoeViz: a web-based tool for co-evolution analysis of protein residues," *BMC Bioinformatics*, vol. 17, no. 1, p. 119, 2016.
- [45] Y. Zhou, B. Zhou, L. Pache et al., "Metascape provides a biologist-oriented resource for the analysis of systems-level datasets," *Nature Communications*, vol. 10, no. 1, 2019.
- [46] M. Kanehisa and S. Goto, "KEGG: Kyoto encyclopedia of genes and genomes," *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, 2000.
- [47] M. Kohl, S. Wiese, and B. Warscheid, "Cytoscape: software for visualization and analysis of biological networks," in *Data mining in proteomics*, pp. 291–303, Springer, 2011.
- [48] R. Flatscher, B. Frajman, P. Schönswetter, and O. Paun, "Environmental heterogeneity and phenotypic divergence: can heritable epigenetic variation aid speciation?," *Genetics Research International*, vol. 2012, 9 pages, 2012.
- [49] E. A. Mullins, A. A. Rodriguez, N. P. Bradley, and B. F. Eichman, "Emerging roles of DNA glycosylases and the base excision repair pathway," *Trends in Biochemical Sciences*, vol. 44, no. 9, pp. 765–781, 2019.
- [50] B. Knudsen and M. M. Miyamoto, "A likelihood ratio test for evolutionary rate shifts and functional divergence among proteins," *Proceedings of the National Academy of Sciences*, vol. 98, no. 25, pp. 14512–14517, 2001.
- [51] A. Prakash, S. Doublé, and S. S. Wallace, "The Fpg/Nei family of DNA glycosylases: substrates, structures, and search for damage," *Progress in Molecular Biology and Translational Science*, vol. 110, pp. 71–91, 2012.
- [52] G. Zhang, C. Li, Q. Li et al., "Comparative genomics reveals insights into avian genome evolution and adaptation," *Science*, vol. 346, no. 6215, pp. 1311–1320, 2014.
- [53] L. Parts, F. A. Cubillos, J. Warringer et al., "Revealing the genetic structure of a trait by sequencing a population under selection," *Genome Research*, vol. 21, no. 7, pp. 1131–1138, 2011.
- [54] M. Angeletti, W.-L. N. Hsu, N. Majo, H. Moriyama, E. N. Moriyama, and L. Zhang, "Adaptations of interferon regulatory factor 3 with transition from terrestrial to aquatic life," *Scientific Reports*, vol. 10, no. 1, p. 4508, 2020.
- [55] J. Nehyba, R. Hrdličková, and H. R. Bose, "Dynamic evolution of immune system regulators: the history of the interferon regulatory factor family," *Molecular Biology and Evolution*, vol. 26, no. 11, pp. 2539–2550, 2009.
- [56] H. I. Ahmad, G. Liu, X. Jiang, C. Liu, Y. Chong, and H. Huarong, "Adaptive molecular evolution of MC1R gene reveals the evidence for positive diversifying selection in indigenous goat populations," *Ecology and Evolution*, vol. 7, no. 14, pp. 5170–5180, 2017.
- [57] M. J. Ahmad, H. I. Ahmad, M. M. Adeel et al., "Evolutionary analysis of makorin ring finger protein 3 reveals positive selection in mammals," *Evolutionary Bioinformatics*, vol. 15, p. 117693431983461, 2019.
- [58] M. S. Albelazi, P. R. Martin, S. Mohammed, L. Mutti, J. L. Parsons, and R. H. Elder, "The biochemical role of the human NEIL1 and NEIL3 DNA glycosylases on model DNA replication forks," *Genes*, vol. 10, no. 4, p. 315, 2019.
- [59] J. S. Taylor and J. Raes, "Duplication and divergence: the evolution of new genes and old ideas," *Annual Review of Genetics*, vol. 38, pp. 615–643, 2004.
- [60] G. B. Gillard, L. Grønvold, L. L. Røsæg et al., "Comparative regulomics supports pervasive selection on gene dosage following whole genome duplication," *Genome Biology*, vol. 22, no. 1, p. 103, 2021.
- [61] R. Shayevitch, D. Askayo, I. Keydar, and G. Ast, "The importance of DNA methylation of exons on alternative splicing," *RNA*, vol. 24, no. 10, pp. 1351–1362, 2018.
- [62] N. Segata, D. Boernigen, T. L. Tickle, X. C. Morgan, W. S. Garrett, and C. Huttenhower, "Computational meta'omics for microbial community studies," *Molecular Systems Biology*, vol. 9, no. 1, p. 666, 2013.
- [63] M. AlQuraishi, "End-to-end differentiable learning of protein structure," *Cell systems*, vol. 8, no. 4, 2019.
- [64] A. Gonzalez and G. Bell, "Evolutionary rescue and adaptation to abrupt environmental change depends upon the history of stress," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 368, no. 1610, 2013.
- [65] C.-S. Goh and F. E. Cohen, "Co-evolutionary analysis reveals insights into protein–protein interactions," *Journal of Molecular Biology*, vol. 324, no. 1, pp. 177–192, 2002.
- [66] M. Liu, S. Doublé, and S. S. Wallace, "Neil3, the final frontier for the DNA glycosylases that recognize oxidative damage," *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, vol. 743, pp. 4–11, 2013.
- [67] M. Santagostino, F. M. Piras, E. Cappelletti et al., "Insertion of telomeric repeats in the human and horse genomes: an evolutionary perspective," *International journal of molecular sciences*, vol. 21, no. 8, p. 2838, 2020.
- [68] P. D'Aquila, A. Montesanto, M. Mandalà et al., "Methylation of the ribosomal RNA gene promoter is associated with aging and age-related decline," *Aging Cell*, vol. 16, no. 5, pp. 966–975, 2017.
- [69] C. Zhu, L. Lu, J. Zhang et al., "Tautomerization-dependent recognition and excision of oxidation damage in base-excision DNA repair," *Proceedings of the National Academy of Sciences*, vol. 113, no. 28, pp. 7792–7797, 2016.
- [70] R. P. Barnes, E. Fouquerel, and P. L. Opresko, "The impact of oxidative DNA damage and stress on telomere homeostasis," *Mechanisms of Ageing and Development*, vol. 177, pp. 37–45, 2019.

- [71] H. I. Ahmad, A. Jabbar, M. Imran et al., "Behind the scene: surprises and snags of pseudogenes," *Critical Reviews in Eukaryotic Gene Expression*, vol. 31, no. 5, 2021.
- [72] T. S. Guzella, S. Dey, I. M. Chelo et al., "Slower environmental change hinders adaptation from standing genetic variation," *PLoS Genetics*, vol. 14, no. 11, article e1007731, 2018.
- [73] H. I. Ahmad, M. J. Ahmad, F. Jabbir et al., "The domestication makeup: evolution, survival, and challenges," *Frontiers in Ecology and Evolution*, vol. 8, 2020.