Koc. J. Sci. Eng., 4(2): (2021) 146-159 https://doi.org/10.34088/kojose.904914

Kocaeli University

Kocaeli Journal of Science and Engineering

http://dergipark.org.tr/kojose

Searching for the Roots of Bloom Syndrome Protein and Its Homologs Using Phylogenetic Analysis

Tuğcan KORAK [1,*], Murat KASAP [2]

[1] *Department of Medical Biology, Faculty of Medicine, Kocaeli University, Kocaeli, 41001, Turkey,* ***ORCID:*** *0000-0003-4902-4022*
[2] *Department of Medical Biology, Faculty of Medicine, Kocaeli University, Kocaeli, 41001, Turkey,* ***ORCID:*** *0000-0001-8727-2096*

Article Info

Research paper

Received : *March 03, 2021*
Accepted : *August 11, 2021*

Keywords

BLM RecQ Like Helicase (BLM)
Maximum Parsimony
Phylogeny
Phylogenetic Tree

Abstract

Phylogenetic analysis (PA) is used for elucidation of relationships among different species and provides information about their evolution. BLM protein (BLM RecQ like helicase) is responsible for the repair of stalled replication fork during double-strand break repair by homologous recombination. In the current study, phylogenetic analysis was performed using BLM protein sequences, sequences of its homologs and its putative homologs from 34 species including covering the genera of Bacteria, Archaea and Eukaryotes. This study was carried out for the purpose of (1) illustrating and comparing relationships among eukaryotic BLM proteins, their homologs (ATP-dependent DNA helicase RecQs in Bacteria) and their potential putative homologs (ATP-dependent DNA helicase Hel308s in Archaea), (2) evaluating how BLM protein evolution took place, what it brought to the organisms by acquiring functional changes and how future potential changes would occur and (3) gaining the general perspective in the evolution of BLM protein. All analyzed species in Bacteria, Archaea and Eukaryota formed a clear inter-species cluster, except for *P. sinensis* (Reptilia). The results supported that Hjm helicase may be one of the candidate potential ancestors of the BLM proteins and their homologs. Moreover, especially two domains which are Helicase ATP-binding and Helicase C-terminal domain were encountered in the all analyzed species and seem to be strictly conserved in the future. Repair related-highly sophisticated interaction network of BLM indicated that its functional evolution reaches a certain level and it appears to have taken an important place in maintaining genomic stability. However, it should be taken into account that BLM may acquire additional functions or become a cornerstone in different pathways in the future depending on its participation in various metabolic roads.

1. Introduction

The construction of phylogenetic trees is an essential process to uncover various chemical and morphological mechanisms as well as the evolution of life [1,2]. In a similar manner, molecular phylogenetic trees which are performed for the gene or genome data are used to predict phylogeny of focused species [1]. Analysis of phylogeny has become a routine for the understanding novel gene sequences and is also used for RNA and protein sequences [3]. Moreover, the construction of universal phylogenetic trees with the three domains of life, Bacteria, Archaea, Eukaryota, leads to an understanding of all extents of life and stages of evolutionary processes through the roots and branches up to the modern cell types. There are two fundamental perspectives of universal phylogenetic trees to facilitate deduction of evolutionary information; (1) the nature of the species represented by its root and (2) how the species bring about the primary organismal lineages [4].

BLM (BLM RecQ like helicase) helicase, which is encoded by BLM gene located on 15q26.1 [5] chromosomal region, is responsible for the repair of halted replication fork during double-strand break (DSB) repair by homologous recombination (HR). It shows DNA-stimulated ATPase activity and ATP-dependent DNA helicase activity. Also, it regulates branch migration of Holliday junctions by interacting with some other proteins

* Corresponding Author: tugcankorak@gmail.com

for the purpose of HR completion [6,7,8]. Moreover, it contributes to the maintenance of genome integrity by resolving anaphase bridges necessary for accurate chromosome segregation [9]. In addition, BLM was reported to be associated with some tumor suppressor proteins and is thought to play a role in DNA-damage response in the BLM-TOP3A-RM1 complex [10, 11]. On the other hand, studies have shown that mutations in the *BLM* gene may cause Bloom syndrome which is a genetic disease characterized by stature, fertility problems, growth inhibition, light sensitivity and susceptibility to cancer. Moreover, it has been shown that mice models heterozygous for a number of *BLM* exons in which some of them are non-functional are more susceptible to cancer as compared to mice carrying inactivated tumour suppressor genes such as APC. Thus, haploinsufficiency in the BLM gene may stimulate cancer development [6,7,11].

Structural studies on BLM protein which belongs to the RecQ helicase family have revealed that it has a zinc-binding subdomain unique to the RecQ helicase family at the C-terminus of ATPase domain [12-15]. Besides, the N-terminal region has been suggested to involve in oligomerization [16]. In another study, the crystal structure of the human RecQ C-terminal (RQC) domain of BLM protein was elucidated and the results demonstrated that the BLM RQC domain has distinguished it from some other helicases in several aspects. Firstly, the C-terminal region of BLM RQC spreads throughout the domain surface. Secondly, there is no aromatic residue at the end of the β-wing region which is used in DNA-strand separation. Eventually, BLM-specific insertion regions located between N-terminal helices have been shown as important for an accurate angle of β-wing and Holliday junction resolution [17].

Considering the important tasks that BLM protein carries out in the cells, we were interested in knowing how this ubiquitous protein evolved over time. To the best of our knowledge, phylogenetic analysis of BLM protein has not been carried out previously. Thus, phylogenetic analysis was performed based on BLM proteins, their homologs and putative homologs in 34 species that are representative of three genera; Archaea, Bacteria, and Eukaryota. The constructed phylogenetic tree provided insights into BLM evolution.

2. Materials and Methods

2.1. Data Retrieval

For Bacteria (Clade 1): BLASTP software (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE= Proteins) was used to obtain homologous sequences of BLM belonging to bacterial species. The database searched was UniProtKB/SwissProt and the query sequence used was BLM protein sequence of Gorilla gorilla gorilla (Western lowland gorilla) which is more likely to have similar sequences with primitive organisms than the human BLM sequence. All bacterial species which contained protein sequences that have functional similarities with BLM were included in this study.

For Archaea (Clade 2). Since BLM homologs of Archea species could not be obtained in the BLASTP search performed by selecting the UniProtKB/SwissProt database, functional homology search was carried out for these species in the literature. Several studies supported that archaeal Hjm helicase has recq-like functions and may play a role in the repair of stalled replication fork [18-24]. By taking this information into consideration, Hjm helicase sequences were retrieved from the UniprotKB database (http://www.uniprot.org/). All archaeal species which contained protein sequences that have functional similarities with BLM were included in the study.

For Eukaryotes (Clade 3): Phylum Chordata including fishes, amphibians, reptiles, birds, mammals in Animalia kingdom of Eukaryota was used in this study. BLM protein sequences were retrieved from the UniprotKB database. The data with the longest BLM sequence length for each species and with compatible to helicase function of BLM were included. Since the higher mammalian data for BLM is readily available in the databases, more of the mammalian species were taken into consideration.

2.2. Phylogenetic Analysis

For phylogenetic analysis, amino acid sequences were used in order to exclude compositional bias that may be encountered in DNA sequences [25] BLM amino acid sequences from 34 different organisms were retrieved from the UniprotKB database (http://www.uniprot.org/) and listed in Table 1.

Unrooted trees are suitable for determining the degree of inter-species similarity, whereas rooted trees, where a common ancestor is assigned, provide a clearer understanding of the evolutionary path. Fossil data are needed to construct rooted trees [26]. Since there are no fossil data for the phylogenetic analysis carried out in this study, rooted trees were constructed in order to interpret the BLM protein similarity among species better. The rooted tree was constructed using the midpoint rooting approach from the unrooted tree.

The sequences were downloaded in FASTA format and aligned with Clustal Omega (https://www.ebi.ac.uk/Tools/ msa/clustalo/) in which

output format was chosen as PHYLIP and other parameters were remained as default.

Table 1. Uniprot accession numbers of the eukaryotic BLM proteins, their homologs (ATP-dependent DNA helicase RecQs in Bacteria) and their potential putative homologs (ATP-dependent DNA helicase Hel308s in Archaea), and the organisms to which the sequences belong.

	UniprotKB No.	**Species**	
Bacteria	P15043	*Escherichia Coli* (*E. Coli*)	strain K12
	O34748	*Bacillus subtilis* (*B. subtilis*)	strain 168
	P40724	*Salmonella typhimurium* (*S. typhimurium*)	strain LT2 / SGSC1412 / ATCC 700720
	Q9CL21	*Pasteurella multocida* (*P. multocida*)	strain Pm70
	P71359	*Haemophilus influenza* (*H. influenzae*)	strain ATCC 51907 / DSM 11121 / KW20 / Rd
Archaea	O73946	*Pyrococcus furiosus* (*P. furiosus*)	strain ATCC 43587 / DSM 3638 / JCM 8422 / Vc1
	Q974S1	*Sulfurisphaera tokodaii* (*S. tokodaii*)	strain DSM 16993 / JCM 10545 / NBRC 100140 / 7
	Q5JGV6	*Thermococcus kodakarensis* (*T. kodakarensis*)	strain ATCC BAA-918 / JCM 12380 / KOD1
	U3TB36	*Aeropyrum camini* (*A. camini*)	SY1 = JCM 12091
	D1Z2D5	*Methanocella paludicola* (*M. paludicola*)	strain DSM 17711 / JCM 13418 / NBRC 101707 / SANAE
Eukaryota		**Fishes**	
	E7EZY7	*Danio rerio* (*D. rerio*)	Zebrafish
	A0A3B3H755	*Oryzias latipes* (*O. latipes*)	Japanese rice fish / Japanese killifish
	A0A1S3KNZ2	*Salmo salar* (*S. salar*)	Atlantic salmon
		Amphibians	
	Q9DEY9	*Xenopus laevis* (*X. laevis*)	African clawed frog
	A0A6I8PXE0	*Xenopus tropicalis* (*X. tropicalis*)	Western clawed frog
		Reptilia	
	G1KPR4	*Anolis carolinensis* (*A. carolinensis*)	Green anole
	A0A452GZE4	*Gopherus agassizii* (*G. agassizii*)	Agassiz's desert tortoise
	K7FS17	*Pelodiscus sinensis* (*P. sinensis*)	Chinese softshell turtle
		Aves	
	U3IGZ2	*Anas platyrhynchos* (*A. platyrhynchos*)	Mallard
	U3K8X3	*Ficedula albicollis* (*F. albicollis*)	Collared flycatcher
	Q9I920	*Gallus gallus* (*G. gallus*)	Chicken
		Mammalians	
	G1M5K3	*Ailuropoda melanoleuca* (*A. melanoleuca*)	Giant panda
	E1BQ04	*Bos taurus* (*B. taurus*)	Domestic cow
	J9PB86	*Canis lupus familiaris* (*C. l. familiaris*)	Dog
	F6ZL78	*Equus caballus* (*E. caballus*)	Horse
	P54132	*Homo sapiens* (*H. sapiens*)	Human
	O88700	*Mus musculus* (*M. musculus*)	House Mouse
	G1PQL6	*Myotis lucifugus* (*M. lucifugus*)	Little brown bat
	W5PTL9	*Ovis aries* (*O. aries*)	Sheep
	D3ZQW1	*Rattus norvegicus* (*R. norvegicus*)	Brown rat
	G3W1C2	*Sarcophilus harrisii* (*S. harrisii*)	Tasmanian devil
	H2QA33	*Pan troglodytes* (*P. troglodytes*)	Chimpanzee
	G3QXK2	*Gorilla gorilla gorilla* (*G. g. gorilla*)	Western lowland gorilla
	F1RMJ2	*Sus scrofa* (*S. scrofa*)	Pig

The sequence data were bootstrapped using SEQBOOT for 1000 times by randomly selecting columns from the original alignment to infer the reliability of the trees [27]. PROT-PARS program was used to construct trees with maximum parsimony method (multiple data sets or multiple weights: D, The number of data sets: 1000, Random number seed: 3, Number of times to jumble: 5). Consequently, CONSENSUS program was used to produce majority rule consensus trees and RETREE program was used to assign midpoint to the rooted tree. Finally, DRAWTREE and DRAWGRAM were used to construct unrooted and rooted trees, respectively. The outputs of DRAWTREE or DRAWGRAM were used in Adobe Illustrator 2020 to create the figures for publication.

2.3. The Domain and Biological Function Search in the UniprotKB Database

Since the algorithms used for domain identification have some assumptions on the principles of the evolutionary process [28], we initially scanned domain information of all analyzed proteins in sequence depositories. To contribute to the evolutionary history of BLM, we scanned manually the protein domains of Hjm helicase, BLM homologs and BLM proteins in the UniprotKB database (https://www.uniprot.org/) for all the selected species of Archaea, Bacteria and Eukaryota, respectively. The UniprotKB numbers of the species were inputted and the domains of each related protein were listed under the "Domains and Repeats" section. In addition, to assess the major biological functions of the analyzed proteins, the "GO - Biological process" section was analyzed for each protein sequence. The elucidated information was used to understand the fundamental aspects of BLM protein evolution.

2.4. Identification of Domains and Motifs

PFAM database contains large protein families represented by multiple sequence alignments and hidden Markov models. It is used for the classification of domains and protein families; therefore, it helps understand the function of targeted proteins [29]. To expand and support the data obtained from Uniprot KB domain search, we identified the domains and motifs using the Pfam database (Pfam 33.1, 18259 entries: http://pfam.xfam.org/). The FASTA format of protein sequences belonging to Hjm helicase, BLM homologs and BLM proteins for all analyzed species in Archaea, Bacteria and five selected species in Eukaryotes were used as the input sequences. One species was selected from each class (Fishes: *D. rerio*, Amphibians. *X. laevis*, Reptilia: *A. carolinensis*, Aves: *A. platyrhynchos*, Mammalians: *H. sapiens*) to create five representative species of Eukaryota. Consequently, "Significant Pfam-A Matches" were included in the analysis and Adobe Illustrator 2020 was used to combine and edit the results.

2.5. BioGrid and STRING Analysis

The BioGrid and STRING analyses were carried out to obtain predicted interaction networks of BLM protein with other proteins. The BioGrid (https://thebiogrid.org/) and STRING analysis (https://string-db.org/) of BLM was performed by selecting *H. Sapiens* as the target organism. A minimum evidence level of 1 was set as the stringency level in BioGrid analysis. The results were exported in PNG format.

3. Results and Discussion

Phylogenetic analysis is carried out to exhibit relatedness among species and to obtain important insights into the evolution of molecular sequences. General principles derived from PA also lead to predicting future changes in DNA or protein sequences. In general, phylogenetic analysis are used in affiliating taxonomy to an organism, pathogen identification, forensic medicine and determining the cryptic speciation in a species [2, 3, 30]. In the current study phylogenetic analysis of BLM protein was performed for 34 species which include the representative species from three clades, Bacteria, Archaea and Eukaryota. Rather than discussing each proximal or binary relationship and explaining their basements, our aim is to gain an overview of the BLM protein evolution. Based on the analysis of the phylogenetic tree together with the other analysis carried out, evolutionary history and the potential fate of BLM protein will be inferred.

The molecular phylogenetic analysis generated by ribosomal RNA (rRNA) sequences specified the phylogenetic relationships among three domains of life; Bacteria, Archaea and Eukaryota. Apart from rRNA, orthologous genes have been used to construct a phylogenetic tree of life; however, the tree resolution permitted by a single gene remains at the minimum level [31]. Protein sequences can also be preferred in molecular evolution researches depending on their several advantageous aspects. Even if all information required for proteins is found in DNA sequences, the natural selection process does not often directly occur on DNA. However, proteins are the main building blocks and essential components of life on which natural selection occurs. In addition, while four bases create DNA sequences, functional characteristics of proteins are defined by 20

amino acids which bring about much higher resolution in the phylogeny of evolutionary distantly related organisms [32]. Hence, we constructed the phylogenetic tree using protein sequences to evaluate BLM protein evolution in the tree domain of life by considering its homologs (ATP-dependent DNA helicase RecQs in Bacteria) and its potential homologs (ATP-dependent DNA helicase Hel308s in Archaea) processing stalled replication fork.

3.1. Evaluation of Phylogenetic Trees for Elucidation of Relationships Among BLM Proteins

The proximal and distal relationships that were attributed to the species were based on the amino acid sequences of BLM proteins or their homologs or Hjm helicase. In both unrooted and rooted trees, three clades were created. As expected, each clade was formed by the species belonging to the corresponding genus. The clade formed by the bacterial species was more closely related to the species formed by the archaeal ones. The eukaryotes, on the other hand, formed a more distantly related clade. The archaeal species were more closely related to the eukaryotes than the bacterial clade. However, there appeared to be some species that carried the transition form of the protein and thus found a place on the tree between the archeal and the eukaryotic clades. Those species were *A. carolinensis*, *G. agassizii*, *A. platyrhynchos*, *G. gallus* and *F. albicollis*.

By using midpoint rooting, a rooted tree was created that provides a better illustration of relationships among species. (Figure 1 and 2). Except for the bootstrap value (352.3) for *E. caballus* and *A. melanoleuca*-*C. l. familiaris*, the majority of other bootstrap values represented above the nodes were obtained as 1000 or close to1000.

For Clade 1 (Bacteria): All species revealed distant relationships with other clades. Close binaries were revealed between *H. influenza* and *P. multocida* as well as S. *typhimurium* and *E. Coli*. *B. subtilis* were obtained as less close in terms of BLM homologs (Figure 1 and 2).

For the Clade 2 (Archaea): All species revealed distant relationships with other clades (Figure 1 and 2). Close binaries were revealed between *S. tokodaii* and *A. camini*. *T. kodakarensis* were obtained as less close in terms of Hjm helicases (Figure 2).

For the Clade 3 (Eukaryota): Species belonging to different classes in phylum Chordata were assigned to their own classes, except for *P. sinensis* (reptilia) which is acquired as more closely related to the species of Bacteria. In mammalians, the closest relative of *H. Sapiens was P. troglodytes*, and *S. harrisii* was observed as the most distant relative of all species in the mammalian class. Other close binaries in the mammalian class were as follows: *A. melanoleuca*-*C. l. familiaris*, *O. aries*-*B. taurus* and *M. musculus*-*R. norvegicus*. The species of Amphibians, Fishes and Aves were seen as more close relatives as compared to other classes in terms of BLM protein (Figure 1 and 2).

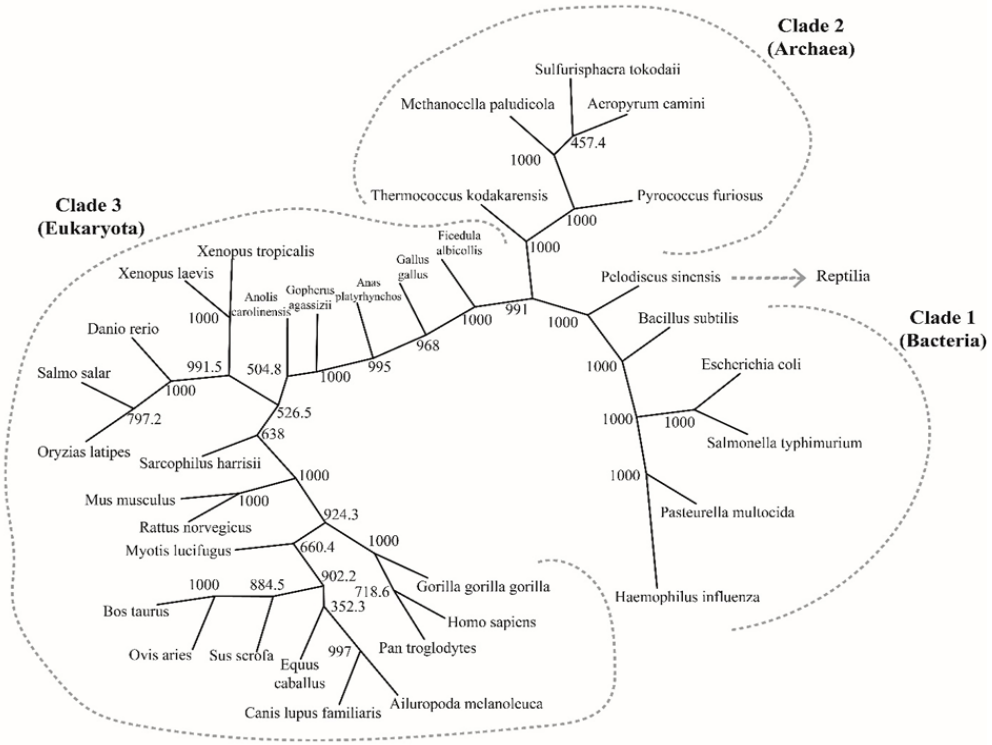

Figure 1. Unrooted maximum parsimony tree constructed for BLM amino acid sequences and its homologs of 34 different species in Bacteria, Archaea, and Eukaryota. Numbers near the nodes are bootstrap values.

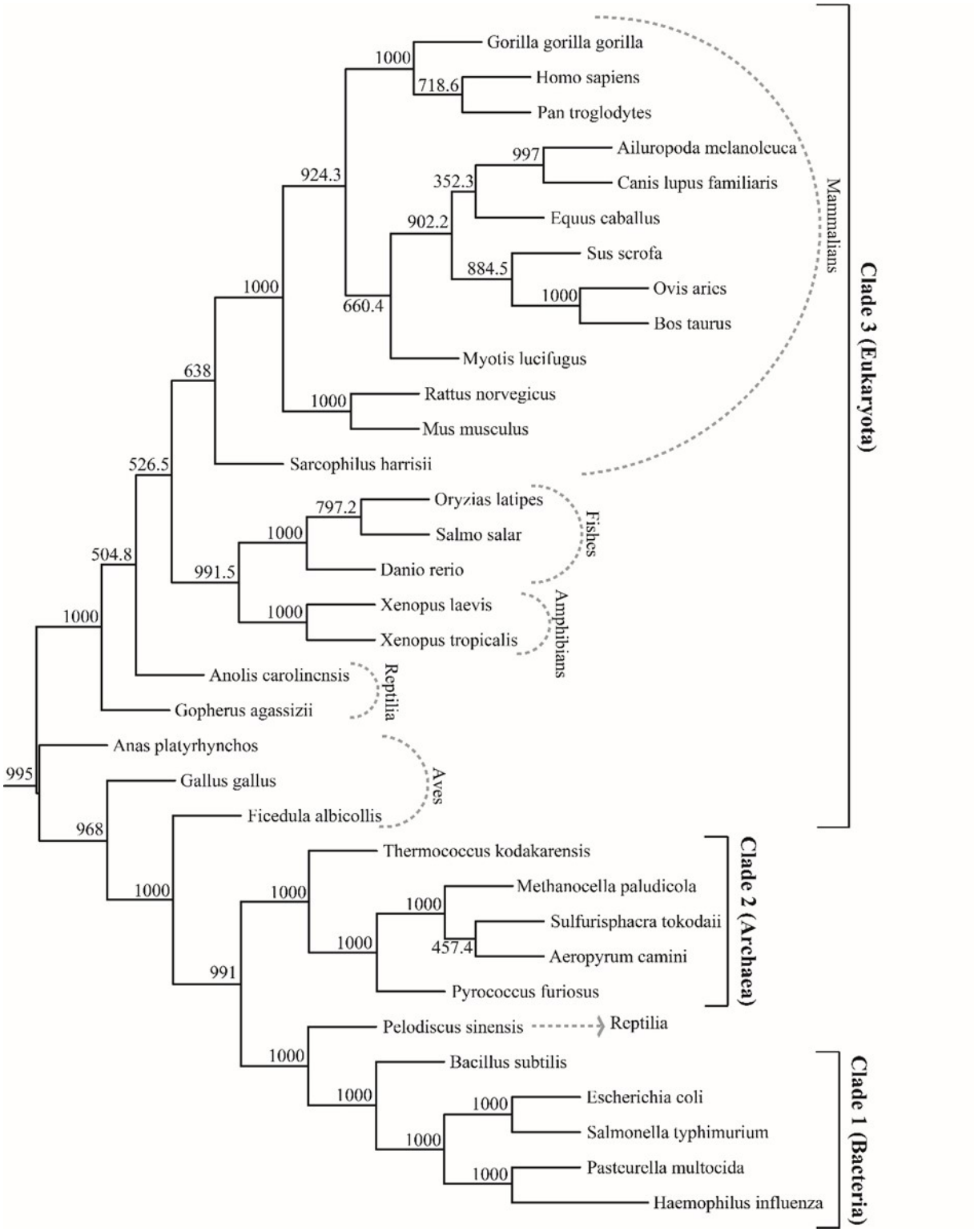

Figure 2. Rooted maximum parsimony tree constructed for BLM amino acid sequences and its homologs of 34 different species in Bacteria, Archaea, and Eukaryota. The different classes including mammalians, amphibians, fishes, reptilians and aves in phylum chordata were indicated. Numbers near the nodes are bootstrap values.

Both rooted and unrooted trees are suitable for determining the degree of inter-species similarity, whereas rooted trees provide a clearer understanding of the evolutionary path by representing ancestral species [26, 33]. In general, the fossil data are necessary to predict divergence dates and also to generate rooted tree [26]. On the other hand, unrooted trees are frequently constructed when there is no sufficient information to assign the root and they still offer insights into evolutionary relationships of organisms without ancestral perspective [33]. Since

there are no fossil data in this phylogenetic analysis, we generated a rooted tree to interpret the notion of evolution of BLM protein better. Also, an unrooted tree was constructed to consider and to reveal the possible alternative root nodes, besides investigating the relatedness of BLM protein among organisms. The generated rooted and unrooted trees commonly demonstrated binary phylogenetic relatedness of organisms in terms of BLM proteins and their homologs (Figure 1 and 2). The bootstrap values represented near the nodes provide to evaluate the reliability of the phylogenetic trees. The bootstrap values of 95% or greater specify statistically significant and mean "support" for a clade while the values less than 5% indicate the rejection of a node. However, the values should not be evaluated as the measurement of the truth or accuracy of the phylogeny. They imply the repeatability rather than accuracy of the phylogenetic trees [27]. In the phylogenetic analysis of BLM protein, the node of *E. caballus* and *A. melanoleuca*-*C. l. familiaris* has the lowest bootstrap value as 352.3, while the majority of others were obtained as 1000 or close to 1000. Although it is not possible to state clearly, the low bootstrap values may be due to the increased taxon sample size [34]. The majority of our branching points were supported by high or moderate bootstrap values (Figures 1 and 2).

Due to the dense species sampling within the mammals in Eukaryota detailed aspects of the BLM protein evolution in mammalian species were obtained (Figures 1 and 2). The closest relative of *H. Sapiens* was found as *P. troglodytes* (chimpanzees). It is possible to associate this unsurprising result with the many phenotypic and genomic similarities that exist between the two species. The whole-genome divergence between two species was reported as approximately 4% resulting from ~1.23% single-nucleotide divergence and ~3% insertion-deletion events. Nearly one of the third proteins is identical and the typical protein alters only by two amino acids. Moreover, orthologous proteins of these two species are still highly similar and less than 1% difference is valid for amino acid sequences [35]. This data explains the closest relationships of BLM protein in these organisms. On the other hand, *G. g. gorilla* (gorilla) was found as branching off *H. Sapiens* and *P. troglodytes.* One of the experimental studies supported these results by demonstrating less divergence for selected DNA segments in *H. Sapiens*-*P. troglodytes* than *H. Sapiens*-*G. g. gorilla*. Also, the divergence of *G. g. gorilla*-*H. Sapiens* was estimated as 1.6-2.2 million years earlier than *H. Sapiens* and *P. troglodytes* divergence [36]. Thus, it seems to be valid for BLM protein evolution among these tree species. Furthermore, the accumulated data obtained from various molecular studies indicate that the transcription factors and the genes related to neural functioning, sexual reproduction, immunity and cell recognition and olfactory receptors evolved faster or evolved under strong positive selection or have different activation states during human speciation and/or after divergence [37-43]. Since the molecular and functional state of BLM protein does not overlap these phenomena, the rate of BLM evolution may proceed at a certain level.

One of the sources of the gene/genome evolution is transposable elements which can lead to mutations, insertions, deletions, rearrangements, copy number variations, etc [44, 45]. The activity of transposable elements (TEs) is strictly regulated by histone modification, DNA methylation and piRNAs. Long interspersed nuclear element-1 (LINE-1) is the member of the class I TE and nearly all LINE-1s are inactive in humans, except for nearly 80-100 ones [46, 47]. However, all LINE-1 elements of *S. harrisii* (Tasmanian devil) were suggested to be nonfunctional. In addition, the Hidden Markov Model strategy supported that the potential sources of functional reverse transcriptase which plays a key role in the mobilization of retrotransposons lack in the *S. harrisii*'s genome [44]. Along a similar line with this background, *S. harrisii* was observed as the most distant relative of all species in mammalian class in terms of BLM protein. Considering less evolutionary changes depending on inactive LINE-1 and lacking functional reverse transcriptase in the genome, BLM protein may firstly branch off right after other classes in Eukaryota.

The evolutionary order of the classes in phylum Chordata from earliest to latest is as follows: Fishes, Amphibians, Reptilia/Aves and Mammalians [48]. The constructed rooted tree did not contain information about the time of evolution, and the order was obtained as Aves/Reptilia, Fishes, Amphibians and Mammalians for BLM protein (Figure 2). Even if the tree was generated with predictions about evolutionary time, the compatible class order might not be obtained. In fact, it should not be considered as an unusual situation depending on the nature of evolution which refers to a process leading to alterations in the genetic material of a population over time. Evolution occurs in different states; one is microevolution that reflects alterations in DNA sequences and allele frequencies within a species over time, and the other one is macroevolution that covers large-scale changes at the species level following the collection of plentiful small alterations on account of microevolution [49]. Given this information, the evolution of the BLM gene/protein and the selected organisms may not proceed directly proportional. It could be one of the reasons why BLM protein of *P. sinensis* (Chinese softshell turtle) was not obtained in its own class, Reptilia, but rather a close relative to bacterial BLM homologs. Although it has been rarely documented and accepted, this result may be a sign

of horizontal gene transfer (HGT) between Bacteria and the Eukaryota. HGT is the movement of genetic material within and/or across species and may occur in all possible directions among three clades [50, 51]. It is an essential force that modulates evolution in the prokaryotes and certain eukaryotes [51]. After HGT events take place from Bacteria to Eukaryota, genes may retain pre-existing functions or provide the eukaryotic recipients with new functions such as protection, nutrition, or adaptation to extreme conditions. Even if HGT between Bacteria and Eukaryota is a controversial event, it should be considered during inferring phylogenetic trees [52]. Thus, closer relatedness of BLM protein in *P. sinensis* to its bacterial homologs might be depending on HGT events from Bacteria to Eukaryota. After that, the evolution to the current BLM gene and thereafter BLM protein may be driven by the evolution mechanisms including mutations, single-nucleotide alterations, gene flow, genetic drift and natural selection in which they enable also the other species to be positioned in the phylogenetic trees [49, 51]. Since each evolutionary approach is valuable to evaluate BLM protein evolution, this rare event should also be taken into account. From another point of view, it is always possible that multiple sequence alignment of a large sequence may result in inaccurate phylogeny depending on misleading signals that are not historically essential [53].

3.2. Domain and Functional Search in UniprotKB

Motifs and domains are functionally essential and evolutionary highly conserved parts of proteins. Motifs (e.g. helix–loop–helix motifs, helix-turn-helix, zinc finger, leucine zipper, helix-hairpin helix) are generally shorter sequences defining the specific functions, whereas domains refer to compact and independently folded structures of proteins and can possess motifs within boundaries. Both are the units of evolution in which they lost, gained, or shuffled as one module [54-56]. According to the knowledge of amino acid sequence, atomic structure and biochemical function of the domain are determined, and this provides an essential perspective to domain evolution. Alterations in the amino acid sequences may result in new protein architectures as well as new folds and functional patterns. They offer valuable aspects to establish evolutionary relationships and to infer phylogenies [28, 57].

Following the protein domain search of Hjm helicase, BLM homologs and BLM proteins in the UniprotKB database for all analyzed species, it was revealed that two domains which are Helicase ATP-binding and Helicase C-terminal were encountered in the all selected 34 species. Three domains- Helicase ATP-binding, Helicase C-terminal and The helicase and RNaseD C-terminal (HRDC)- were detected in all selected species of Bacteria and nearly all species of Eukaryota, except for BLM protein of *G. agassizii* (Agassiz's desert tortoise) and *M. lucifugus* (Little brown bat) in which they contain two following domains; Helicase ATP-binding and Helicase C-terminal (Table 2). These results supported the possibility that Hjm helicase is the ancestor of the BLM protein depending on its common domains (Figure 2). Moreover, these tree domains which are related to helicase functions seem to be very critical for BLM function. It could be deduced that these domains may be preserved in the future and any mutation affecting these domains could alter the BLM function and may affect the natural selection of species. While the DNA repair-focused mechanisms of Hjm helicase have been reported in the UniprotKB database, we have encountered that BLM participates in biological processes on both DNA repair and maintaining genome integrity in a complicated way in Bacteria and especially in Eukaryota. Furthermore, the additional functions such as regulation of cyclin-dependent protein serine/threonine kinase activity, regulation of signal transduction by p53 class mediator, positive regulation of alpha-beta T cell proliferation, negative regulation of thymocyte apoptotic process and protein homooligomerization were declared in this database in some species of Eukaryota. These indicate that BLM will probably continue to conserve genome integrity; furthermore, it may gain different additional functions in the future.

Table 2. The UniprotKB domain search of phylogenetically analyzed proteins (BLM proteins, BLM homologs and Hjm helicase) of the species in Archaea, Bacteria and Eukaryota.

	Helicase ATP-binding	**Helicase C-terminal**	**HRDC**
Archaea	+	+	
Bacteria	+	+	+
Eukaryota	+	+	+*

+ indicates the presence of domains. *There is an exception for BLM protein of *G. agassizii* and *Myotis lucifugus* species. They contain only two domains Helicase ATP-binding and Helicase C-terminal.

On the other hand, when the biological functions of analyzed proteins for each species in, Bacteria, Archaea and Eukaryota were evaluated, the DNA repair and genome integrity maintenance related tasks are the major for all organisms. Moreover, the complexity of the involved biological process revealed an increasing trend

for the species of the following clades in order; Archaea, Bacteria and Eukaryota.

3.3. Classification of Domains and Motifs in PFAM Database

Since domains and motifs provide fundamental evolutionary insights on the identification of the sequences and functions of proteins [54], we expanded our work with the Pfam database in selected 15 species as the representatives of Bacteria, Archaea and Eukaryota. Helix-hairpin-helix motifs which are involved in DNA binding without sequence specificity and some enzymatic activities [58] were acquired only in the species of Archaea (Figure 3). While DEAD (corresponding to the region of Helicase ATP-binding encountered in UniprotKB database) and Helicase C-terminal were obtained in the all analyzed species, four domains- DEAD, Helicase C-terminal, Zn-BD, RQC and HRDC- were found in all species of Bacteria and all analyzed species of Eukaryota (Figure 3). When the position of DEAD domain in all analyzed proteins within Pfam was compared to the position of Helicase ATP-binding domain which was encountered as conserved in UniprotKB database, their positions were detected as overlapped. Thus, the conserved domain results of both databases are compatible with each other's. Hjm helicase was detected to contain two common domains with the other analyzed helicases in Bacteria and Eukaryota and observed to consist a low number of domains as compared to the others. Thus, Pfam results also supported that it has a potential ancestral relationship with BLM protein. Besides the other two, especially DEAD and Helicase C terminal domains seem to be highly conserved during evolution. DEAD box containing proteins that can utilize ATP hydrolysis is vital for nearly all cellular activities in which DNA or RNA is involved [59]. Helicase C terminal, Zn-BD (the domain including Zinc-finger motif), RQC domains and HRDC domain are important for protein localization, DNA binding and protein conformation, dsDNA binding and structural integrity, and the assistance in nucleic acid binding and RecQ substrate specificity, respectively [60-63]. Given those roles, they seem to become the main blocks of BLM proteins and their homologs. Also, it can be deduced that the correct positioning of the enzymes was ensured by these additional domains in Bacteria and Eukaryota.

On the other hand, N-terminal domains have only appeared in Eukaryota. The N-terminal domain of BLM was reported to interact variety of proteins (e.g., BRCA1, Exo1, ATM, TOP3α and DNA2-RPA-MRN complex) and they are generally controlled by phosphorylation mechanisms. Even if it has importance for cellular events, the structure of the N-terminal has not been elucidated in detail yet. One of the studies revealed that the N-terminal domain is loosely structured, lacking conserved three-dimensional structures and it is highly divergent in terms of sequence and structure. In contrast, the BDHCT domain is highly conserved and contributes to BLM oligomerization [62]. Therefore the fewer data available for enzymatic and structural features of the N-terminal domain which is only appeared in analyzed species of Eukaryota suggested that this domain could be the potential region that may involve or explain the complex interaction network of BLM protein (Figure 3 and 4).

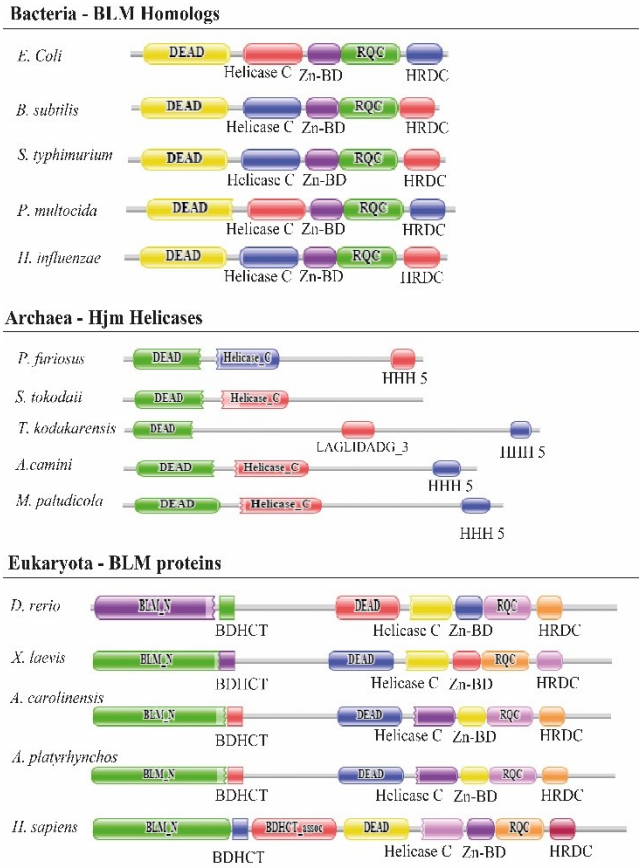

Figure 3. Identification of domains and motifs of BLM homologs, Hjm helicases and BLM proteins in Pfam database for the analyzed species in Bacteria and Archaea as well as five selected species in Eukaryota. (DEAD: DEAD/DEAH box helicase, LAGLIDADG_3: LAGLIDADG-like domain, Helicase C: Helicase conserved C-terminal domain, HHH 5: Helix-hairpin-helix domain, Zn-BD: RecQ zinc-binding domain, RQC: RecQ C-terminal domain, HRDC: The helicase and RNaseD C-terminal, BLM_N: N-terminal region of Bloom syndrome protein, BDHCT: BDHCT (NUC031) domain, BDHCT_assoc: BDHCT-box associated domain).

3.4. BioGrid and STRING Analysis

Besides these various mentioned BLM functions, studies, and database information, we performed BioGrid analysis that also emphasizes the importance of this unique protein by predicting 173 potential interactors and 371 interactions. It has been observed that a significant number of these are proteins responsible for DNA repair and preserving genome integrity (Figure4).

On the other hand, STRING analysis demonstrated the specific and meaningful possible protein associations of BLM protein. The analysis predicted that BLM has 10 different functional partners which are RPA3 (Replication protein A3), MLH1 (MutL homolog 1), RPA1 (Replication Protein A1), FANCM (Fanconi anemia group M protein), APITD1 (apoptosis-inducing, TAF9-like domain 1), RMI1 (RecQ-mediated genome instability protein 1), C14orf70 (Fanconi anemia core complex-associated protein 100), RPA2 (Replication protein A2), FANCG (Fanconi anemia group G protein), FANCA (Fanconi anemia group A protein). These predicted functional partners play a role in DNA repair, cellular response to DNA damage and genome stability (Figure 5).

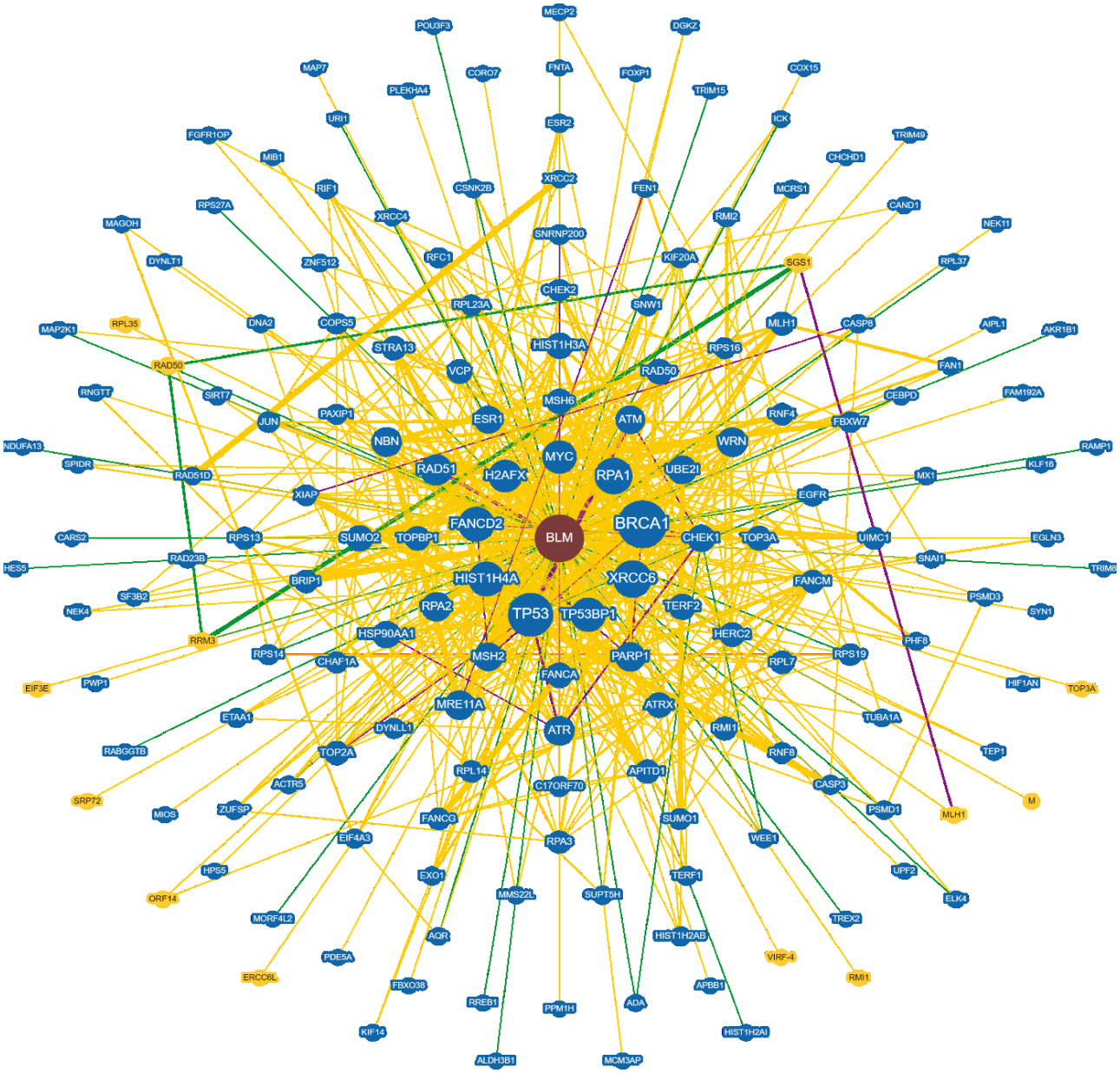

Figure 4. Interactome analysis of BLM protein generated by using BioGrid analysis. The complex interaction network of BLM was demonstrated. The yellow, green and purple lines indicated physical, genetic and physical/genetic interactors, respectively. BioGrid analysis of BLM protein in *H. sapiens* revealed that it totally has 173 potential interactors including 126 physical interactors, 45 genetic interactors and 2 physical/genetic interactors and 371 interactions.

Although BLM has interactors with variable functions, it majorly interacts with genome integrity and/or repair associated proteins such as exonucleases, ligases, histone proteins, checkpoint kinases, topoisomerases, recombinases, dyneins, kinesins, etc. related-highly sophisticated interaction network This repair indicates that functional evolution of BLM reaches to a certain level and it appears to have taken an important place in maintaining genomic stability. However, it should be taken into account that BLM may acquire new functions or become a cornerstone in different pathways in the future due to its various possible metabolic roads to connect distantly related proteins including interferon regulatory factor, diacylglycerol kinases, ubiquitin-conjugating enzymes and ribosomal proteins.

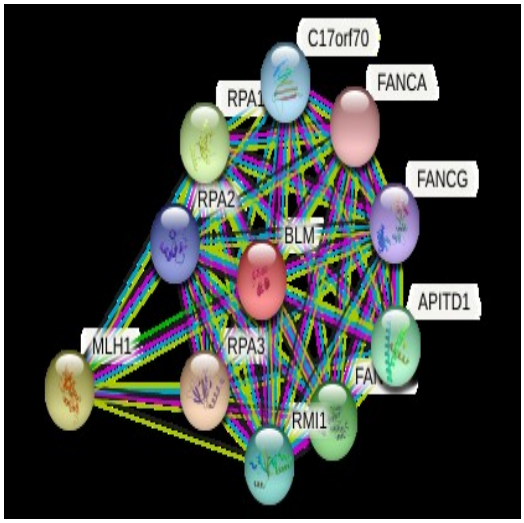

Figure 5. The possible protein interactions network of BLM protein is elucidated by STRING analysis. The light blue and purple lines represent the known interactions in which data were retrieved from curated databases and experiments, respectively. The green lines represent gene neighborhoods from predicted interactions. The yellow and black lines represent the textmining and co-expression data, respectively.

4. Conclusions

All in all, the universal phylogenetic tree construction for a targeted protein opens the doors from the past to the current time. This approach also provides a broad perspective on the notion of evolution as well as the future fate of the focused protein. However, there are some challenges faced by researchers when they attempt to establish science-based and logical associations between the homologous or potentially homologous proteins belonging to quite different organisms living in absolutely different conditions and/or creations. For instance, the BLM protein of *P. sinensis* (reptilia, Chinese softshell turtle) in our study was acquired as more closely related to the Bacteria. Contributing to understanding the evolution of the BLM protein, the current study indicated that Hjm helicase may be one of the candidate potential ancestors of the BLM proteins or their homologs. Moreover, especially two domains- Helicase ATP-binding (or DEAD domain) and Helicase C-terminal BLM proteins- encountered in all species seem to be strictly conserved in the future besides the HRDC domain encountered only in Bacteria and Eukaryota. Therefore, BLM protein is also open to the evolution of a new domain related to helicase function. Considering its highly sophisticated network, BLM protein may acquire additional functions or play a key role in other cellular networks. To clarify these data phylogenetic analysis could be extended with much more species; moreover, it could be supported by DNA data or different algorithms that predict the evolutionary time of BLM proteins or their homologs. Furthermore, comparative experimental functional analysis can be performed to illustrate how molecular evolution affects the biological function of BLM among species.

Declaration of Ethical Standards

The authors of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

Conflict of Interest

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

References

[1] Horiike T., Minai R., Miyata D., Nakamura Y., Tateno Y., 2016. Ortholog-Finder: A Tool for Constructing an Ortholog Data Set. Genome Biology and Evolution, **8**, pp. 446-457.

[2] Soltis D.E., Soltis, P.S., 2003. The role of phylogenetics in comparative genetics. Plant physiology, **132**, pp.1790–1800.

[3] Holder M., Lewis P.O., 2003. Phylogeny estimation: traditional and Bayesian approaches. Nature Reviews. Genetics, **4**, pp. 275-84.

[4] Woese C.R., 2000. Interpreting the universal phylogenetic tree. Proceedings of the National Academy of Sciences of the United States of America, **97**, pp. 8392–8396.

[5] BLM, Bloom syndrome RecQ like helicase. https://ghr.nlm.nih.gov/gene/BLM#location. Accessed October 20, 2020.

[6] Ding S.L., Yu J.C., Chen S. T., Hsu G.C., Kuo S.J., Lin Y.H., Wu P.E., Shen, C.Y., 2009. Genetic variants of BLM interact with RAD51 to increase breast cancer susceptibility. Carcinogenesis, **30**, pp. 43–49.

[7] Shen M., Menashe I., Morton L.M., Zhang Y., Armstrong B., Wang S.S., Lan Q., Hartge P., Purdue M.P., Cerhan J.R., Grulich A., Cozen W., Yeager M., Holford T.R., Vajdic C.M., Davis S., Leaderer B., Kricker A., Severson R.K., Zahm S.H., Chatterjee N., Rothman N, Chanock S.J., Zheng T., 2010. Polymorphisms in DNA repair genes and risk of non-Hodgkin lymphoma in a pooled analysis of three studies. British journal of haematology, **151**, pp. 239–244.

[8] Karow J.K., Constantinou A., Li J.L., West S.C., Hickson I.D., 2000. The Bloom's syndrome gene product promotes branch migration of holliday junctions. Proceedings of the National Academy of Sciences of the United States of America, **97**, pp. 6504–6508.

[9] Frank B., Hoffmeister M., Klopp N., Illig T., Chang-Claude J., Brenner H., 2010. Colorectal cancer and polymorphisms in DNA repair genes WRN, RMI1 and BLM. Carcinogenesis, **31**, pp. 442–445.

[10] Wang Z., Xu Y., Tang J., Ma H., Qin J., Lu C., Wang X., Hu Z., Wang X., Shen H., 2009. A polymorphism in Werner syndrome gene is associated with breast cancer susceptibility in Chinese women. Breast cancer research and treatment, **118**, pp. 169–175.

[11] Broberg K., Huynh E., Schläwicke Engström K., Björk J., Albin M., Ingvar C., Olsson H., Höglund, M., 2009. Association between polymorphisms in RMI1, TOP3A, and BLM and risk of cancer, a case-control study. BMC cancer, **9**, pp. 140.

[12] Vindigni A., Marino F., Gileadi, O., 2010. Probing the structural basis of RecQ helicase function. Biophysical Chemistry, **149**, pp. 67–77.

[13] Pike A.C., Shrestha B., Popuri V., Burgess-Brown N., Muzzolini L., Constantini S., Vindigni A., Gileadi O., 2009. Structure of the human RECQ1 helicase reveals a putative strand-separation pin. Proceedings of the National Academy of Sciences of the United States of America, **27**, pp. 1039-1044.

[14] Bernstein D.A., Zittel M.C., Keck J.L., 2003. High-resolution structure of the E.coli RecQ helicase catalytic core. The EMBO Journal, **22**, pp. 4910–4921.

[15] Hoadley K.A., Keck J.L., 2010. Werner helicase wings DNA binding. Structure, **18**, pp. 149–151.

[16] Beresten S.F., Stan R., van Brabant A.J., Ye, T., Naureckiene, S., Ellis, N. A., 1999. Purification of overexpressed hexahistidine-tagged BLM N431 as oligomeric complexes. Protein Expression and Purification, **17**, pp. 239-248.

[17] Kim S.Y., Hakoshima T., Kitano K., 2013. Structure of the RecQ C-terminal domain of human Bloom syndrome protein. Scientific Reports, **21**, pp. 3294.

[18] Fujikane R., Shinagawa H., Ishino Y., 2006. The archaeal Hjm helicase has recQ-like functions, and may be involved in repair of stalled replication fork. Genes to cells : devoted to molecular & cellular mechanisms, **11**(2), pp. 99–110.

[19] Guy C.P., Bolt E.L., 2005. Archaeal Hel308 helicase targets replication forks in vivo and in vitro and unwinds lagging strands. Nucleic acids research, **33**(11), pp. 3678–3690.

[20] Fujikane R., Komori K., Shinagawa H., Ishino Y., 2005. Identification of a novel helicase activity unwinding branched DNAs from the hyperthermophilic archaeon, *Pyrococcus furiosus*. Journal of Biological Chemistry, **280**(13), pp. 12351–12358.

[21] Li Z., Lu S., Hou G., Ma X., Sheng D., Ni, J., Shen Y., 2008. Hjm/Hel308A DNA helicase from Sulfolobus tokodaii promotes replication fork regression and interacts with Hjc endonuclease in vitro. Journal of bacteriology, **190**(8), pp. 3006–3017.

[22] Hong Y., Chu M., Li Y., Ni J., Sheng D., Hou G., She Q., Shen Y., 2012. Dissection of the functional domains of an archaeal Holliday junction helicase. DNA Repair, 11(2), pp. 102-111.

[23] Liew L.P., Lim Z.Y., Cohen, M., Kong, Z., Marjavaara L., Chabes A., Bell, S.D., 2016. Hydroxyurea-Mediated Cytotoxicity Without Inhibition of Ribonucleotide Reductase. Cell reports, **17**(6), pp. 1657–1670.

[24] Zhai B., DuPrez K., Han X., Yuan Z., Ahmad S., Xu C., Gu L., Ni J., Fan L., Shen Y., 2018. The archaeal ATPase PINA interacts with the helicase Hjm via its carboxyl terminal KH domain remodeling and processing replication fork and Holliday junction. Nucleic acids research, **46**(13), pp. 6627–6641.

[25] Foster P.G., Hickey D.A., 1999. Computational bias may affect both DNA-based and protein based phylogenetic reconstructions. Journal of molecular evolution, **48**(3), 284–290.

[26] Heath T.A., Huelsenbeck J.P., Stadler T., 2014. The fossilized birth-death process for coherent calibration of divergence-time estimates. Proceedings of the National Academy of Sciences of the United States of America, **111**, pp. E2957-66.

[27] Felsenstein J., 1985. Confidence limits on phylogenies: an approach usingthe bootstrap. Evolution, **39**, 783–791.

[28] Bagowski C.P., Bruins W., Te Velthuis A.J., 2010. The nature of protein domain evolution: shaping the interaction network. Current genomics, **11**(5), 368–376.

[29] Mistry J., Chuguransky S., Williams L., Qureshi M., Salazar G.A., Sonnhammer E., Tosatto S., Paladin L., Raj S., Richardson L.J., Finn R.D., Bateman A., 2021. Pfam: The protein families database in 2021. Nucleic acids research, **49**(D1),pp. D412–D419.

[30] Challa S., Neelapu N.R.R., 2019. Phylogenetic Trees: Applications, Construction, and Assessment. In: Hakeem K., Shaik N., Banaganapalli B., Elango R. (eds) Essentials of Bioinformatics, Volume III. Springer, Cham, Switzerland.

[31] Yokono M., Satoh S. Tanaka A., 2018. Comparative analyses of whole-genome protein sequences from multiple organisms. Scientific Reports, **8**, pp. 6800.

[32] Opperdoes, F.R., 2003. Phylogenetic analysis using protein sequences. In The Phylogenetics Handbook: A Practical Approach to DNA and Protein Phylogeny, 1st ed. Salemi, M., Vandamme, A.-M., Eds., Cambridge University Press, Cambridge, London, United Kingdom.

[33] Bogdanowicz D., Giaro K., 2010. Comparing arbitrary unrooted phylogenetic trees using generalized matching split distance. 2nd International Conference on Information Technology, (2010 ICIT), Gdansk, Poland, pp. 259-262.

[34] Soltis P., Soltis, D., 2003. Applying the Bootstrap in Phylogeny Reconstruction. Statistical Science, **18**(2), pp. 256-267.

[35] Varki A., Altheide T.K., 2005. Comparing the human and chimpanzee genomes: searching for needles in a haystack. Genome research, **15**(12), pp. 1746–1758.

[36] Chen F.-C., Li W.-H., 2001. Genomic Divergences between Humans and Other Hominoids and the Effective Population Size of the Common Ancestor of Humans and Chimpanzees. The American Journal of Human Genetics, **68**(2), pp. 444-456.

[37] Suntsova M.V., Buzdin A.A., 2020. Differences between human and chimpanzee genomes and their implications in gene expression, protein functions and biochemical properties of the two species. BMC Genomics, **21**, pp. 535.

[38] Chimpanzee Sequencing and Analysis Consortium, 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature, **437**(7055), pp. 69–87.

[39] Dorus S., Vallender E.J., Evans P.D., Anderson J.R., Gilbert S.L., Mahowald M., Wyckoff G.J., Malcom C.M., Lahn, B.T., 2004. Accelerated evolution of nervous system genes in the origin of Homo sapiens. Cell, **119**(7), pp. 1027–1040.

[40] Evans P.D., Gilbert S.L., Mekel-Bobrov N., Vallender E.J., Anderson J.R., Vaez-Azizi L.M., Tishkoff S.A., Hudson R.R., Lahn B.T., 2005. Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. Science (New York, N.Y.), **309**(5741), pp. 1717–1720.

[41] Zhang J., Webb D.M., Podlaha O., 2002. Accelerated protein evolution and origins of human-specific features: Foxp2 as an example. Genetics, **162**(4), pp. 1825–1835.

[42] Wyckoff G.J., Wang W., Wu C.I., 2000. Rapid evolution of male reproductive genes in the descent of man. Nature, **403**(6767), pp. 304–309.

[43] Go Y., Niimura Y., 2008. Similar numbers but different repertoires of olfactory receptor genes in humans and chimpanzees. Molecular biology and evolution, **25**(9), pp. 1897–1907.

[44] Gallus S., Hallström B.M., Kumar V., Dodt W.G., Janke A., Schumann G.G., Nilsson, M.A., 2015. Evolutionary histories of transposable elements in the genome of the largest living marsupial carnivore, the Tasmanian devil. Molecular biology and evolution, **32**(5), pp. 1268–1283.

[45] Kazazian H.H., Jr, Moran J.V., 2017. Mobile DNA in Health and Disease. The New England journal of medicine, **377**(4), pp. 361–370.

[46] Wang P.J., 2017. Tracking LINE1 retrotransposition in the germline. Proceedings of the National Academy of Sciences of the United States of America, **114**(28), pp. 7194–7196.

[47] Ostertag E.M., Kazazian H.H.,Jr, 2001. Biology of mammalian L1 retrotransposons. Annual review of genetics, **35**, pp. 501–538.

[48] Peat J.R., Ortega-Recalde O., Kardailsky O., Hore, T.A., 2017. The elephant shark methylome reveals conservation of epigenetic regulation across jawed vertebrates. F1000Research, **6**, pp. 526.

[49] Evolution, 2014. Scitable by Nature Education. https://www.nature.com/scitable/definition/evolution-78/. Accessed January 10, 2021.

[50] Mozhayskiy V., Tagkopoulos I., 2012. Horizontal gene transfer dynamics and distribution of fitness effects during microbial *in silico* evolution. BMC Bioinformatics, **13**, pp. S13.

[51] Boto L., 2010. Horizontal gene transfer in evolution: facts and challenges. Proceedings. Biological sciences, **277**(1683), pp. 819–827.

[52] Husnik F., McCutcheon J.P., 2018. Functional horizontal gene transfer from bacteria to eukaryotes. Nature reviews. Microbiology, **16**(2), pp. 67–79.

[53] Naylor G.J.P., Brown W.M., 1998. Amplhioxus mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences. Systematics Biology, **47**, pp. 61–76.

[54] Xiong J., 2006. Protein Motifs and Domain Prediction. In Essential Bioinformatics (pp. 85-94). Cambridge: Cambridge University Press. Cambridge, London, United Kingdom.

[55] Pavlov A.R., Belova G.I., Kozyavkin S.A., Slesarev, A.I., 2002. Helix-hairpin-helix motifs confer salt resistance and processivity on chimeric DNA polymerases. Proceedings of the National Academy of Sciences of the United States of America, **99**(21), pp. 13510–13515.

[56] Alberts B., Johnson A., Lewis J., Raff M., Roberts K., Walter P., 2002. Molecular Biology of the Cell. In DNA-Binding Motifs in Gene Regulatory Proteins. 4th ed. Garland Science, New York, USA.

[57] Newman J.A., Savitsky P., Allerston C.K., Bizard A.H., Özer Ö., Sarlós K., Liu Y., Pardon E., Steyaert J., Hickson I.D., Gileadi O., 2015. Crystal structure of the Bloom's syndrome helicase indicates a role for the HRDC domain in conformational changes. Nucleic acids research, **43**(10), pp. 5221–5235.

[58] Shao X., Grishin N.V., 2000. Common fold in helix-hairpin-helix proteins. Nucleic acids research, **28**(14), pp. 2643–2650.

[59] Umate P., Tuteja N., Tuteja R., 2011. Genome-wide comprehensive analysis of human helicases. Communicative & integrative biology, **4**(1), pp. 118–137.

[60] Yankiwski V., Noonan J.P., Neff N.F., 2001. The C-terminal domain of the Bloom syndrome DNA helicase is essential for genomic stability. BMC cell biology, **2**, pp. 11.

[61] Guo R.B., Rigolet P., Zargarian L., Fermandjian S., Xi X.G., 2005. Structural and functional characterizations reveal the importance of a zinc binding domain in Bloom's syndrome helicase. Nucleic acids research, **33**(10), pp. 3109–3124.

[62] Shi J., Chen W.F., Zhang B., Fan S.H., Ai X., Liu N.N., Rety S., Xi X.G., 2017. A helical bundle in the N-terminal domain of the BLM helicase mediates dimer and potentially hexamer formation. The Journal of biological chemistry, **292**(14), pp. 5909–5920.

[63] Manthei K.A., Keck, J.L., 2013. The BLM dissolvasome in DNA replication and repair. Cellular and molecular life sciences : CMLS, **70**(21), pp. 4067–4084.