Bangladesh J Microbiol, Volume 40, Number 1, June 2023, pp 15-24

Original Article

DOI: https://doi.org/10.3329/bjm.v40i1.71706

In silico Evolutionary Insights into Prokaryotic and Eukaryotic Malate Dehydrogenases (MDH) Support the Archaebacterial Origin of Eukaryotes

Israt Dilruba Mishu[1], Nowshin Kabir[1], Zimam Mahmud[2]*

[1]Department of Microbiology, University of Dhaka, Dhaka-1000, Bangladesh, [2]Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka-1000, Bangladesh

Malate Dehydrogenase (MDH) stands as a pivotal enzyme crucial for cellular energy metabolism, orchestrating the conversion of malate to oxaloacetate in both prokaryotic and eukaryotic organisms. This study delves into the evolutionary trajectories of MDH1 (cytoplasmic) and MDH2 (mitochondrial), offering substantial evidence supporting the archaebacterial origin of eukaryotes. The dataset spans nine groups, encompassing human MDH1 and MDH2, mammalian MDH1 and MDH2, amphibian MDH2, arthropod MDH2, amoeba MDH1, archaea, and bacteria. Protein BLAST analysis revealed significant sequence homology in mammalian MDH1, particularly among primates, underlining a close evolutionary connection. Conversely, lower eukaryotes, including amoeba, arthropods, and amphibians, exhibited marked divergence from human MDH1 and MDH2. Phylogenetic analysis unveils distinct clusters for MDH1 and MDH2, accentuating significant genetic diversity between mitochondrial and cytoplasmic MDH enzymes. Prokaryotic MDH sequences cluster with human mitochondrial MDH2, while MDH1 forms a separate cluster. Staphylococcus MDH aligns with archaeal MDH, emphasizing the diversity within bacterial MDH evolution. Protein variability analysis indicates noteworthy divergence of human MDH1 from prokaryotic MDH, while MDH2 displays comparatively lower divergence. Pairwise evolutionary divergence analysis sheds light on complex relationships among MDH protein sequences. Human MDH1 shows close evolutionary ties to mammalian MDH1, whereas MDH2 exhibits a unique pattern, aligning closely with mammalian MDH2, arthropods, and amphibians. Furthermore, MDH2 demonstrates closer proximity to bacterial MDH, supporting a bacterial origin of mitochondrial MDH. In contrast, MDH1 displays less divergence to archaeal MDH than its bacterial counterpart, endorsing an archaeal origin for cytoplasmic MDH. In conclusion, this study provides compelling support for the archaebacterial origin of eukaryotes, suggesting a bacterial endosymbiont within an archaeal host that evolved into mitochondria. It contributes valuable insights into MDH evolution, unraveling the intricate relationships and unique adaptations shaping the evolutionary history of eukaryotic cells.

Keywords: In Silico, Malate Dehyderogenase, Endosymbiont, Archaea

Introduction

Malate dehydrogenase (MDH) is a pivotal enzyme found in all eukaryotic cells [1], orchestrating the conversion of L-malate to oxaloacetate with nicotinamide adenine dinucleotide (NAD) as a coenzyme [2]. In eukaryotes, MDH manifests as two distinct isozymes: cytoplasmic (MDH1) and mitochondrial (MDH2) [3]. With a dual presence in mitochondria and the cytoplasm, MDH influences key metabolic pathways. This enzyme plays a vital role in cellular processes, contributing to the citric acid cycle and the malate-aspartate shuttle. The structural and functional aspects of MDH have been subject to intensive investigation, revealing that MDH exists as a homodimeric molecule with two subunits, each weighing approximately 30 to 35 kDa [4]. The subunits exhibit distinct domains responsible for NAD^+ binding and substrate interaction. MDH's significance extends to its involvement in gluconeogenesis and the final step of the tricarboxylic acid (TCA) cycle [5].

Multiple studies suggest that MDH is a highly conserved enzyme [5-7] present in organisms ranging from prokaryotes to eukaryotes. Its ubiquity and evolutionary conservation make it an ideal candidate for studying relationships and divergence among species. Playing a vital role in the tricarboxylic acid (TCA) cycle that connects various metabolic pathways, MDH offers insights into the metabolic adaptations and energy needs of different organisms. Furthermore, the availability of cytoplasmic (MDH1) and mitochondrial (MDH2) variants makes it a reliable candidate for studying the distinct adaptations and evolutionary paths of these two variants. Considering all these phenomenon, MDH was chosen in the current study to gain valuable insights into the archaebacterial origin of eukaryotes.

The concept of the archaebacterial origin of eukaryotes [8] is a fascinating idea in the study of evolutionary biology. According to this theory, eukaryotic cells, which make up complex organisms like plants, animals, and fungi, have ancient origins connected to

Corresponding author:
Dr. Zimam Mahmud, Associate Professor, Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka-1000, Bangladesh. Email: zimam@du.ac.bd

a group of microorganisms known as archaea. Specifically, the proposal suggests that a particular member of the archaea, part of the Asgard group, formed a symbiotic relationship with a bacterial cell. This bacterium, belonging to the Alphaproteobacteria group, possessed the unique ability to use oxygen for respiration [9]. This partnership was advantageous because it allowed the combined entity to survive and flourish in the presence of oxygen, unlike other organisms adapted to reducing conditions. The endosymbiotic bacteria eventually evolved into the mitochondria, a vital component within eukaryotic cells responsible for energy production [10]. The narrative of archaebacterial origin sheds light on the intricate relationships between different microorganisms that eventually led to the development of complex life forms.

The current study focuses on exploring the divergence between mitochondrial and cytoplasmic isozymes of MDH from prokaryotic and eukaryotic origins, as well as elucidating the evolutionary trajectory of MDH. The aim of this study is to verify the archaebacterial origin of eukaryotes by scrutinizing the MDH sequences from various organisms, including bacteria, archaea and eukaryotes. Through multiple sequence alignment, phylogenetic tree construction, and in-depth analyses, we seek to contribute insights into the evolutionary dynamics of MDH and its potential connection to the endosymbiotic theory.

Methods and materials

Sequence retrieval and dataset generation

27 MDH (Malate Dehydrogenase) sequences were retrieved from NCBI (National Center for Biotechnology Information) database. The sequences represent a wide range of species, encompassing both prokaryotic and eukaryotic origins. Within this dataset, 7 sequences are attributed to the prokaryotic domain, with 4 from bacteria and 3 from archaea, while the remaining others represent eukaryotes. Among the eukaryotic MDH sequences, 11 are of mitochondrial (MDH2) and 9 of cytoplasmic (MDH1) origin. The complete dataset is categorized into 9 groups, namely *H. sapiens*_MDH1, *H. sapiens*_MDH2, Mammalia_MDH1, Mammalia_MDH2, Amphibia, Arthropoda, Amoeba, Archaea, and Bacteria. The dataset encompasses sequences reported between 2000 and 2023. Table 1 provides an overview of the dataset.

BLAST Analysis

Basic Local Alignment Search Tool (BLAST) was used to analyze the sequence similarity between the sequences of the selected database where *H. sapiens* MDH1 (NP_005908.1) and MDH2 (NP_005909.2) sequences were used as the reference sequences. Protein-protein blast (Blastp) was performed to analyze the percent identity, query coverage and accession length of the selected sequences.

Phylogenetic analysis

The evolutionary relationships between MDH from different organisms were visualized by phylogenetic analysis of the translated protein sequence of nine previously mentioned groups. Multiple sequence alignment of the full dataset was performed and subsequent phylogenetic analysis was carried out in MEGA 11.0 using ClustalW algorithm[11]. The phylogenetic analysis was inferred by using the Maximum Likelihood method and JTT matrix-based model[12].

Calculation of Protein Variability Index

For calculating the protein variability index of MDH sequences, Protein Variability Server (PVS) was used [13]. Three analysis were performed, i) Variability analysis of all the 27 MDH protein sequences of nine different groups (reference: consensus sequence) ii) Variability analysis between human MDH1 (reference: NP_005908.1) with prokaryotic MDH iii) Variability analysis between human MDH2 (reference: NP_005909.2) with prokaryotic MDH. The Shannon variability coefficient, with a variability threshold of 1.0, was calculated to find out the maximum variable positions in the MDH proteins.

Pairwise genetic distance matrix calculation between groups and within groups

The study conducted pairwise genetic distance matrix calculations for the comparison of 27 MDH sequences across nine distinct groups. Amino acid substitutions per site were determined by averaging overall sequence pairs between groups. Similarly, distance matrices were computed for sequences within the same group, and the number of amino acid substitutions per site was determined by averaging overall sequence pairs within each group. The analysis focused on 24 amino acid sequences representing six groups, including Human MDH1 and MDH2, Mammalia MDH1 and MDH2, and Bacteria and Archaea. Ambiguous positions were excluded for each sequence pair using the pairwise deletion option. The final dataset comprised a total of 365 positions. The evolutionary analyses were conducted using the Poisson correction model [14] in MEGA11.

Result

Sequence homology of human MDH1 with mammalian MDH1

The Protein Blast analysis revealed notable sequence similarity among mammalian Malate Dehydrogenase 1 (MDH1) sequences. The accession length of all the mammalian MDH1 sequences were 334 amino acids, ensuring 100% query coverage with *H. sapiens* MDH1 protein sequence. *Pan paniscus* (pygmy chimpanzee) exhibited the highest percent identity (99.70%) with *H. sapiens* MDH1, closely followed by *Symphalangus syndactylus* and *Pongo pygmaeus*, all belonging to the Primate order. The rodent model *Rattus norvegicus* displayed a sequence identity of 97.01% with human MDH1, while similarity with pig, zebra, and black rhinoceros ranged between 95.51% to 96.71%. A significant contrast in sequence identity was observed between the human MDH1 sequence and lower eukaryotes, exemplified by the notably lower sequence identity of 59.82% with amoeba. The detailed results are presented in Table 2.

Table 1: *Overview of the Dataset used in the current study*

SL.	Scientific name	Common name	Group	Domain	MDH type	Accession ID	Location	Year
1.	*Staphylococcus*	Bacteria	Bacteria	Prokaryota		WP_221162212.1	Unknown	2022
2.	*Proteus vulgaris*	Bacteria	Bacteria	Prokaryota		WP_285715176.1	Unknown	2023
3.	*Proteus faecis*	Bacteria	Bacteria	Prokaryota		WP_235378802.1	Unknown	2022
4.	*Escherichia coli*	Bacteria	Bacteria	Prokaryota		MCF1957383.1	South Africa	2018
5.	*Methanosarcina mazei*	Archaea	Archaea	Prokaryota		WP_048045487.1	Unknown	2019
6.	*Haloferax volcanii*	Archaea	Archaea	Prokaryota		AAF43044.1	Unknown	2000
7.	*Haloarcula*	Archaea	Archaea	Prokaryota		WP_004959949.1	Unknown	2022
8.	*Homo sapiens*	Human	*H. sapiens*_ MDH2	Eukaryota	MDH2	NP_ 005909.2	Unknown Chromosome 7	2022
9.	*Symphalangus syndactylus*	Primate (Siamang)	Mammalia_ MDH2	Eukaryota	MDH2	XP_ 055148793.1	Jambi Chromosome 9	2023
10.	*Pan troglodytes*	Primate (Chimpanzee)	Mammalia_ MDH2	Eukaryota	MDH2	XP_ 001156205.1	Unknown Chromosome 7	2023
11.	*Pongo pygmaeus*	Primate (Bornean orangutan)	Mammalia_ MDH2	Eukaryota	MDH2	XP_ 054352535.1	Unknown Chromosome 7	2023
12.	*Sus scrofa*	Pig	Mammalia_ MDH2	Eukaryota	MDH2	NP_ 001231082.1	Unknown Chromosome 3	2022
13.	*Rattus norvegicus*	Rodent	Mammalia_ MDH2	Eukaryota	MDH2	NP_ 112413.2	Unknown Chromosome 12	2022
14.	*Pteropus vampyrus*	Large flying fox	Mammalia_ MDH2	Eukaryota	MDH2	XP_ 011369846.1	Lubee Bat Conservancy	2018
15.	*Pteropus alecto*	Black flying fox	Mammalia_ MDH2	Eukaryota	MDH2	XP_ 006918628.1	Austrailia	2008
16.	*Balaenoptera musculus*	Blue Whale	Mammalia_ MDH2	Eukaryota	MDH2	XP_ 036681156.1	Pacific Ocean: Santa Barbara	2016
17.	*Coptotermes formosanus*	Formosan subterranean termite	Arthropoda	Eukaryota	MDH2	AGM32513.1	Unknown	2013
18.	*Xenopus laevis*	African Clawed Frog	Amphibia	Eukaryota	MDH2	NP_001085326.1	Unknown Chromosome 25	2020
19.	*Homo sapiens*	Human	*H. sapiens*_ MDH1	Eukaryota	MDH1	NP_005908.1	Unknown Chromosome 2	2022
20.	*Symphalangus syndactylus*	Primate(Siamang)	Mammalia_ MDH1	Eukaryota	MDH1	XP_055097553.1	Jambi Chromosome 14	2023
21.	*Pongo pygmaeus*	Primate (Bornean orangutan)	Mammalia_ MDH1	Eukaryota	MDH1	XP_054331114.1	Unknown Chromosome 2A	2023
22.	*Pan paniscus*	Primates (pygmy chimpanzee)	Mammalia_ MDH1	Eukaryota	MDH1	XP_003830935.1	Unknown Chromosome 2A	2023
23.	*Sus scrofa*	Pig	Mammalia_ MDH1	Eukaryota	MDH1	NP_999039.1	Unknown Chromosome 3	2022
24.	*Rattus norvegicus*	Rodent	Mammalia_ MDH1	Eukaryota	MDH1	NP_150238.1	Unknown Chromosome 14	2022
25.	*Equus quagga*	Zebra	Mammalia_ MDH1	Eukaryota	MDH1	XP_046517947.1	Namibia: Etosha	2008
26.	*Diceros bicornis minor*	Black rhinoceros	Mammalia_ MDH1	Eukaryota	MDH1	XP_058407102.1	USA Chromosome 12	2023
27.	*Dictyostelium discoideum*	Slime mold	Amoeba	Eukaryota	MDH1	XP_641333.1	Unknown Chromosome 3	2010

MDH1 (cytoplasmic), MDH2 (mitochondrial)

Table 2: *Protein BLAST analysis result of eukaryotic MDH1 sequences with reference to Human MDH1 (NP_005908.1)*

Organism	Group	Query coverage	Percent identity	Accession length
Symphalangus syndactylus	Mammalia_MDH1	100%	99.10%	334
Pongo pygmaeus	Mammalia_MDH1	100%	99.10%	334
Pan paniscus	Mammalia_MDH1	100%	99.70%	334
Sus scrofa	Mammalia_MDH1	100%	95.51%	334
Rattus norvegicus	Mammalia_MDH1	100%	97.01%	334
Equus quagga	Mammalia_MDH1	100%	96.71%	334
Diceros bicornis minor	Mammalia_MDH1	100%	95.81%	334
Dictyostelium discoideum	Amoeba	99%	59.82%	333

Sequence homology of human MDH2 with prokaryotic MDH and mammalian MDH2

The analysis of Malate Dehydrogenase 2 (MDH2) sequences revealed distinct patterns of sequence identity across various organisms. Mammalian MDH2 exhibited the highest sequence identity with human MDH2, with *Symphalangus syndactylus* MDH2 displaying complete identity and 100% query coverage (accession length 338 aa). Other mammalian MDH2 sequences showed a range of similarity, spanning from 94.08% to 99.70%. *Xenopus laevis* (Frog) displayed 100% query coverage but with a lower percent identity of 84.02%. In contrast, *Coptotermes formosanus*, an insect from the arthropod group, exhibited 71.79% sequence similarity with 79% query coverage.

Bacterial MDH sequences demonstrated approximately 58% sequence similarity with human MDH2, except for *Staphylococcus*, which showed only 30.10% sequence similarity. Bacterial MDH sequences also had shorter lengths than human MDH2, with query coverage ranging from 81% to 92%. Archaeal MDH sequences exhibited even shorter lengths compared to bacterial counterparts, with query coverage ranging from 62% to 82%. Archaeal MDH was the most divergent from human MDH2, with percent identity ranging between 26.83% and 29.35%. These results, presented in Table 3, underscore the varying degrees of sequence conservation and divergence among MDH sequences with reference to human MDH2.

Evolutionary relationship among the MDH of nine different groups

The analysis of evolutionary relationships among MDH sequences from nine diverse groups unveiled distinct phylogenetic clusters, signifying significant genetic diversity between mitochondrial and cytoplasmic MDH enzymes. Prokaryotic MDH sequences exhibited clustering with human mitochondrial MDH2 rather than the cytoplasmic variant. MDH1 sequences, in turn, formed an isolated cluster, highlighting their pronounced genetic divergence from MDH2 and prokaryotic MDH. Notably, archaeal MDH displayed a closer evolutionary proximity to MDH1 groups compared to bacterial MDH and MDH2 groups. This observation provides valuable insights into prokaryotic evolution within higher eukaryotes, particularly concerning MDH2, and underscores the evolutionary

Table 3: *Protein BLAST analysis result of eukaryotic MDH2 and prokaryotic MDH sequences with reference to Human MDH2 (NP_005909.2)*

Organism	Group	Query coverage	Percent identity	Accession length
Symphalangus syndactylus	Mammalia_MDH2	100%	100%	338
Pan troglodytes	Mammalia_MDH2	100%	99.70%	338
Pongo pygmaeus	Mammalia_MDH2	100%	99.41%	338
Sus scrofa	Mammalia_MDH2	100%	94.08%	338
Rattus norvegicus	Mammalia_MDH2	100%	94.38%	338
Pteropus vampyrus	Mammalia_MDH2	100%	94.38%	338
Pteropus alecto	Mammalia_MDH2	100%	94.97%	338
Balaenoptera musculus	Mammalia_MDH2	100%	95.27%	338
Xenopus laevis	Amphibia	100%	84.02%	338
Coptotermes formosanus	Arthropoda	79%	71.79%	273
Staphylococcus	Bacteria	81%	30.10%	313
Proteus vulgaris	Bacteria	91%	58.84%	312
Proteus faecis	Bacteria	92%	58.47%	312
Escherichia coli	Bacteria	91%	58.84%	312
Methanosarcina mazei	Archaea	82%	29.35%	307
Haloferax volcanii	Archaea	62%	28.25%	303
Haloarcula	Archaea	79%	26.83%	304

development of a cytoplasmic MDH1 variant relating to archaeal origins. Intriguingly, *Staphylococcus* MDH clustered with Archaeal MDH, sharing a common cluster with *Methanosarcina mazei*, while other bacterial MDH sequences formed a distinct and separate cluster, distant from *Staphylococcus.*

Protein Variability Index

The protein variability analysis encompassing 27 selected MDH sequences across nine distinct groups revealed that 89.59% (327 out of 365) amino acids displayed variability. The protein variability coefficient for human MDH1 and prokaryotic MDH exhibited a divergence of 82.2% (n=365) amino acid residues. Notably, the variability coefficient for human MDH2 with prokaryotic MDH was comparatively lower, standing at 77.8% divergence, in contrast to the other two analyses. Intriguingly, the variability analysis across all nine groups, as well as for human MDH1 and prokaryotic MDH individually, did not reveal any conserved fragment of ≥6 amino acids in length. However, a singular conserved fragment of 9 amino acids (AGIPRKPGM), spanning from position 104 to 112, was identified in both human MDH2 and prokaryotic MDH, with a variability coefficient ≤1. The Shannon variability plot depicted in Figure 2 visually represented the protein variability coefficients.

Pair wise distance matrix between groups

The analysis of pairwise evolutionary divergence among Malate Dehydrogenase (MDH) protein sequences across nine groups reveals nuanced relationships. For human MDH1, its divergence pattern suggests a notably close evolutionary connection with Mammalia_MDH1, followed by Amoeba_MDH1. Similarly, human MDH2 shows a distinct pattern, with closer ties to Mammalia_MDH2 (0.0357), followed by Amphibia_MDH2, and Arthropoda_MDH2.

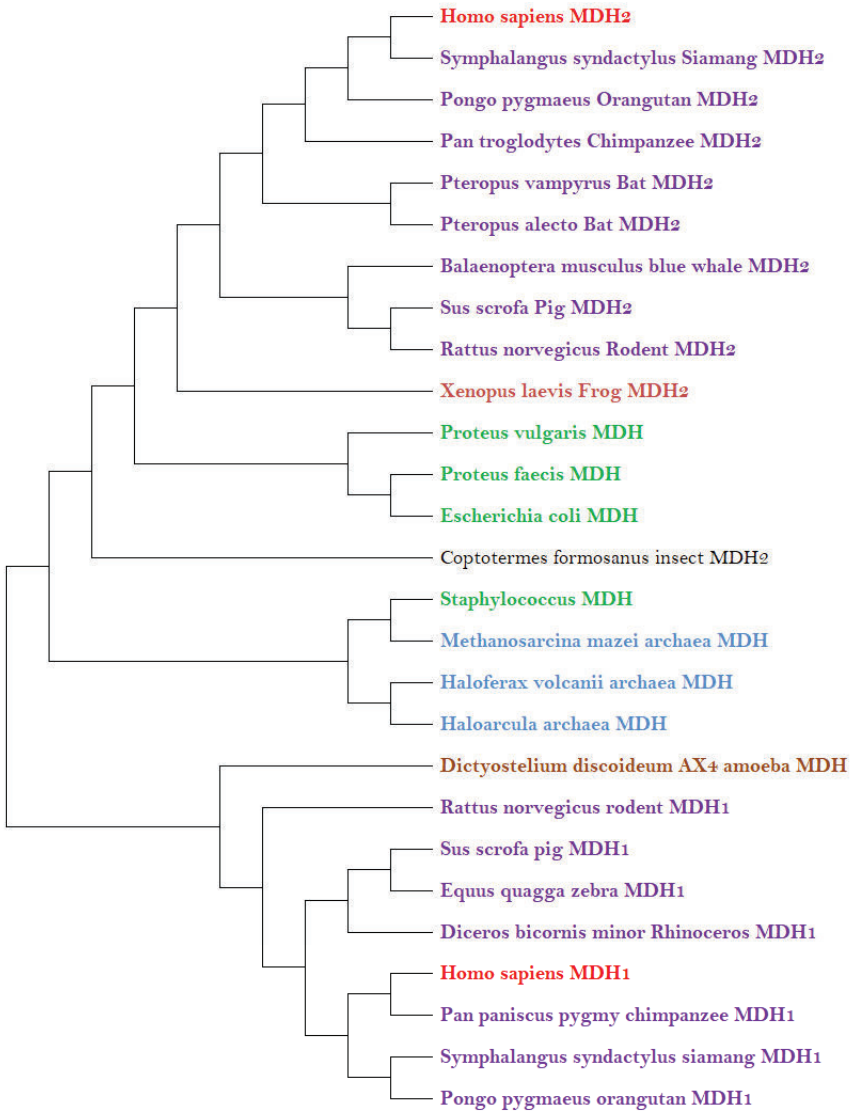

Figure 1: *Evolutionary analysis of 27 MDH sequences representing nine different groups by phylogenetic study. The evolutionary history was inferred by using the Maximum Likelihood method and JTT matrix-based model. There were a total of 365 positions in the final dataset. Evolutionary analyses were conducted in MEGA11. MDH1 and MDH2 denotes cytoplasmic and mitochondrial MDH, respectively.*

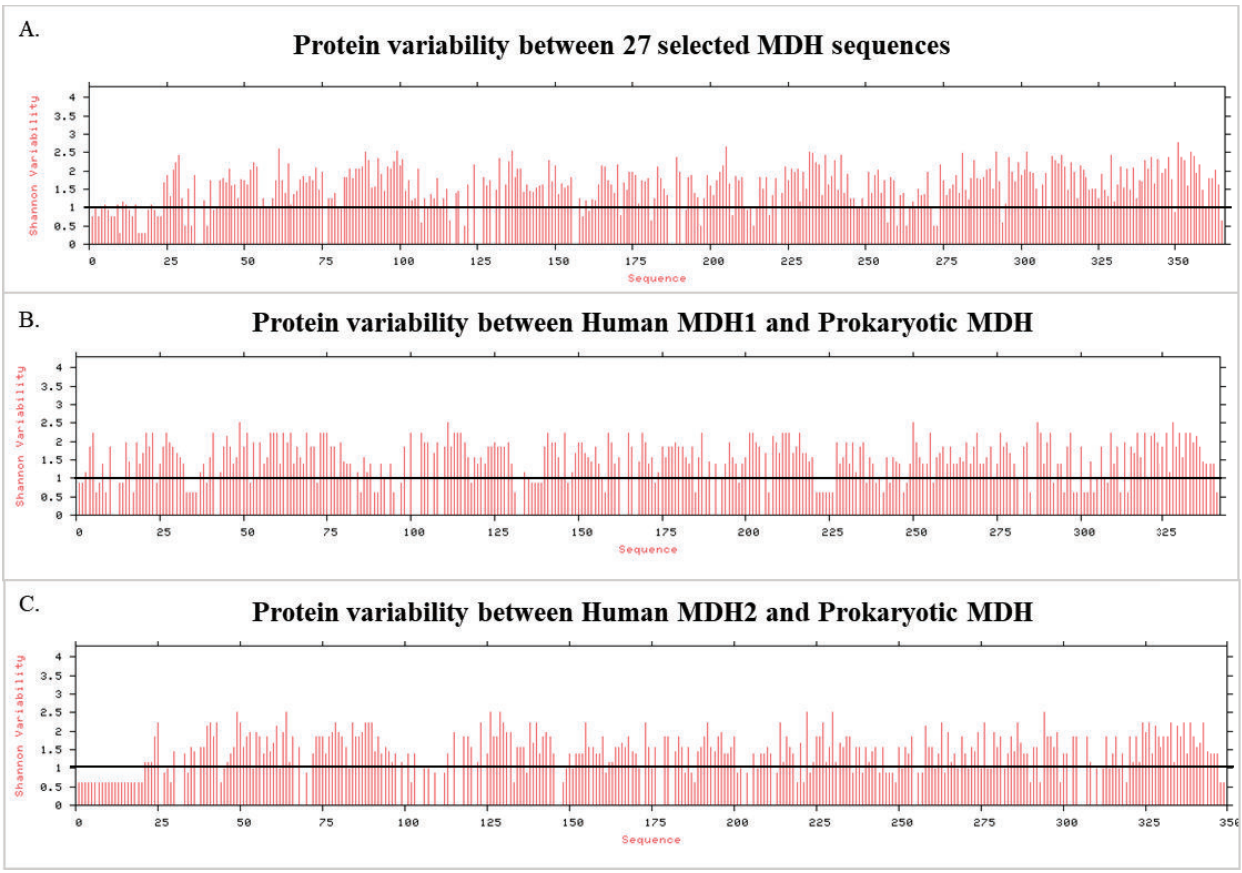

Fig. 2: *Shannon variability plot for MDH protein sequences using Protein Variability Server (PVS). The plots were generated using (A) 27 MDH sequences of nine different groups, (B) Human MDH1 and Prokaryotic MDH (C) Human MDH2 and Prokaryotic MDH. MDH1 and MDH2 denotes cytoplasmic and mitochondrial MDH, respectively.*

Pairwise distance between *H. sapiens* MDH2 and bacterial MDH was 0.7511 which was much lower than archaeal MDH (1.3867). Within the mammalian group, MDH1 of *H. sapiens* appears more closely related to Amoeba, while MDH2 of *H. sapiens* demonstrates a closer evolutionary connection with amphibia, arthropoda and bacteria. Notably, both MDH1 and MDH2 within the mammalian group show closer relationships with each other than with other groups. Amphibia_MDH2 displays a closer evolutionary affinity with Mammalia_MDH2, *H. sapiens*_MDH2, and Arthropoda_MDH2. Divergence of Amphibia_MDH2 with Bacteria_MDH (0.7337) was much less than Archaea_MDH and MDH1 of other eukaryotic organisms. Arthropoda_MDH2, on the other hand, showcases a distinctive divergence pattern, with a closer evolutionary relationship to *H. sapiens*_MDH2 and Mammalia_MDH2 and greater divergence with bacterial and archaeal MDH, as well as eukaryotic MDH1. Amoeba_MDH1 appears more closely related to Mammalia_MDH1 and *H. sapiens*_MDH1, while exhibiting greater evolutionary divergence with all the other groups. Archaeal MDH shows the most evolutionary divergence from all other groups, with 1.2729 divergence from the closest relative which is bacteria. MDH2 from all origins exhibited a closer evolutionary proximity to archaeal MDH than MDH1. Lastly, Bacterial MDH exhibits a divergence pattern with closer ties to Arthropoda_MDH2, Amphibia_MDH2, *H. sapiens*_MDH2, and Mammalia_MDH2. Evolutionary divergence of bacterial MDH with archaeal MDH and MDH1 was significantly higher.

MDH1 variants derived from various origins (including Human, Mammalian, and Amoeba) manifest less evolutionary divergence to Archaeal MDH than to their bacterial counterparts. Conversely, MDH2 sequences from diverse sources (such as Human, Mammalian, Amphibia, and Arthropoda) demonstrate a more conspicuous evolutionary proximity to Bacterial MDH than to Archaeal MDH. This insight, derived from the pairwise distance matrix analysis among distinct groups, underscores the unique evolutionary trajectories of MDH1 and MDH2, hinting at an archaeal origin for MDH1 and a bacterial origin for MDH2. Figure 3A depicts the results of the complex evolutionary relationships and divergence patterns among MDH protein sequences across diverse biological groups. The order of pairwise evolutionary divergence between MDH protein sequences of nine groups are represented here.

1. The divergence order of Human MDH1 was Mammalia_MDH1< Amoeba_MDH1< Arthropoda_MDH2< Archaea< Amphibia_MDH2< Bacteria< Mammalia_MDH2< *H. sapiens*_MDH2.
2. The divergence order of Human MDH2 was Mammalia_MDH2 < Amphibia_MDH2 < Arthropoda_MDH2 < Bacteria_MDH < Archaea_MDH < Mammalia_MDH1 < *H. sapiens*_MDH1 < Amoeba_MDH1
3. The divergence order of Mammalian MDH1 was *H. sapiens*_MDH1 < Amoeba_MDH1 < Arthropoda_MDH2 <

Amphibia_MDH2 < Archaea_MDH <Mammalia_MDH2 < *H. sapiens*_MDH2 < Bacteria_MDH

4. The divergence order of Mammalian MDH2 was *H. sapiens*_MDH2 < Amphibia_MDH2 < Arthropoda_MDH2 < Bacteria_MDH < Archaea_MDH < Mammalia_MDH1 < *H. sapiens*_MDH1 < Amoeba_MDH1
5. The divergence order of Amphibia_MDH2 was Mammalia_MDH2 < *H. sapiens*_MDH2 < Arthropoda_MDH2 < Bacteria_MDH < Archaea_MDH < Mammalia_MDH1 < *H. sapiens*_MDH1 < Amoeba_MDH1
6. The divergence order of Arthropoda_MDH2 was *H. sapiens*_MDH2 < Mammalia_MDH2 < Amphibia_MDH2 < Bacteria_MDH < Archaea_MDH < Amoeba_MDH1 < *H. sapiens*_MDH1 < Mammalia_MDH1
7. The divergence order of Amoeba_MDH1 was Mammalia_MDH1 < *H. sapiens*_MDH1 < Arthropoda_MDH2 < Archaea_MDH < Amphibia_MDH2 < Bacteria_MDH < *H. sapiens*_MDH2 < Mammalia_MDH2
8. The divergence order of Archaea_MDH was Bacteria_MDH < Arthropoda_MDH2 < Amphibia_MDH2 < *H. sapiens*_MDH2 < Mammalia_MDH2 < Amoeba_MDH1 < *H. sapiens*_MDH1 < Mammalia_MDH1
9. The divergence order of Bacteria_MDH was Arthropoda_MDH2 < Amphibia_MDH2 < *H. sapiens*_MDH2 < Mammalia_MDH2 < Archaea_MDH < *H. sapiens*_MDH1 < Mammalia_MDH1 < Amoeba_MDH1

Pair wise distance matrix within groups

The pairwise distance matrix within groups provides insights into the evolutionary relationships among Malate Dehydrogenase (MDH) sequences within distinct taxonomic groups. Analyzing the data, it is evident that the MDH sequences within bacterial species exhibit smaller pairwise distances compared to those within archaeal species. Moving up the taxonomic hierarchy, the pairwise distances within mammals (MDH1 and MDH2) are greater than those observed within bacterial and archaeal species but smaller than the distances within human (MDH1 and MDH2). Finally, the MDH sequences within human (MDH1 and MDH2) show the largest pairwise distances within this context. The order of increasing pairwise distances is as follows: Bacterial species < Archaeal species < Mammals (MDH1 and MDH2) < Human (MDH1 and MDH2). This gradient suggests a pattern of increasing evolutionary divergence, with the MDH sequences within bacterial species being the most conserved, followed by archaeal species, mammals, and finally, *H. sapiens*. Figure 3B represents the data of pair wise distance within groups.

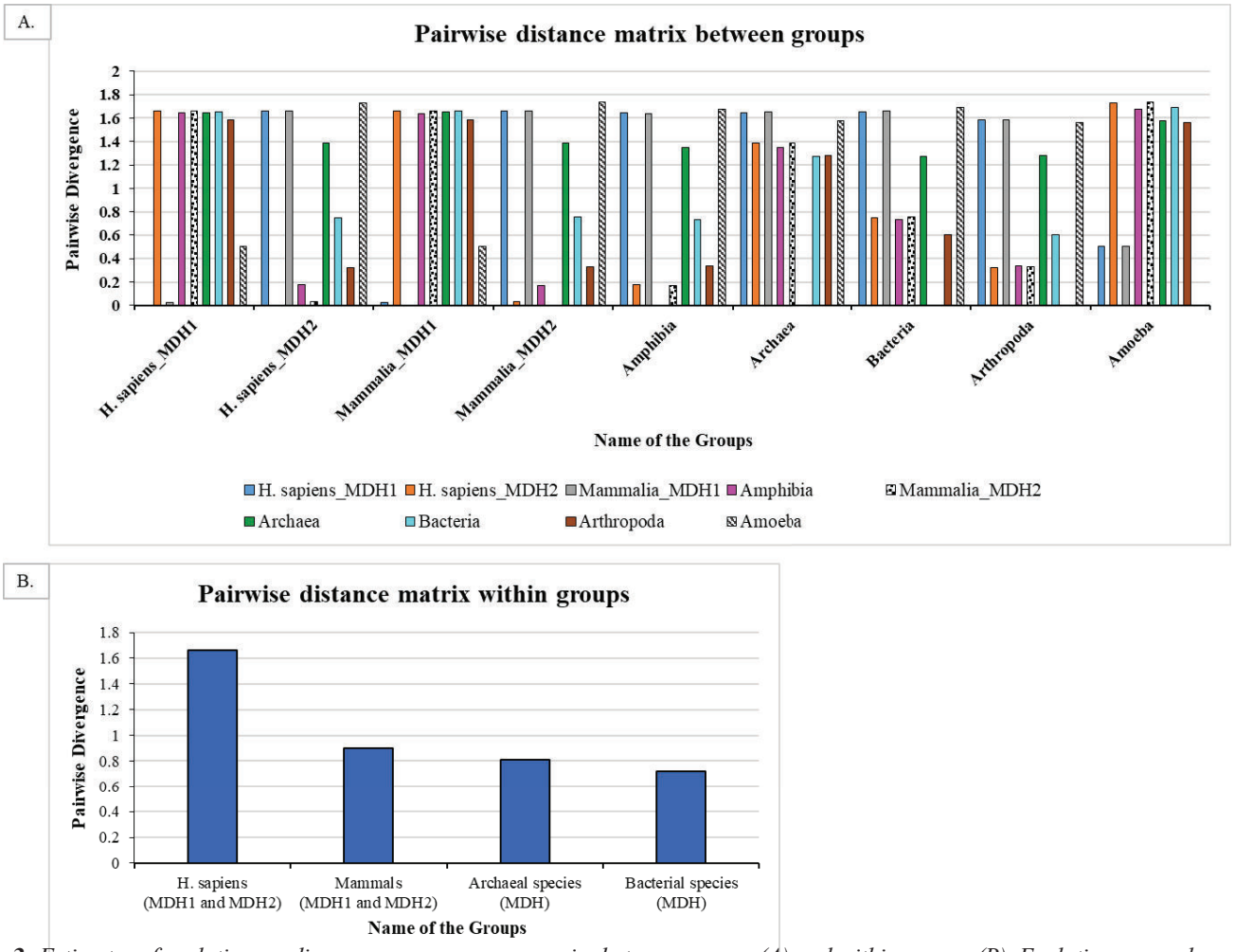

Fig. 3: *Estimates of evolutionary divergence over sequence pairs between groups (A) and within groups (B). Evolutionary analyses were conducted in MEGA11 using the Poisson correction model. MDH1 and MDH2 denotes cytoplasmic and mitochondrial MDH, respectively.*

Discussion

The diversity and evolutionary conservation of MDH across various organisms provide valuable insights into the intricate pathways of molecular evolution. This study delves into the comparative analysis of MDH sequences, focusing on the divergence patterns between cytoplasmic (MDH1) and mitochondrial (MDH2) malate dehydrogenase variants.

The Protein Blast analysis of cytoplasmic (MDH1) and mitochondrial (MDH2) malate dehydrogenase with reference to *H. sapiens* demonstrated a high sequence similarity among mammals, with primate species showing close conservation. Comparisons with non-primate mammals also indicated significant conservation. Woo et al. also noticed similar evolutionary conservation in mammals for histone modifications [15]. Similar finding was noticed for breast cancer type 1 (BRCA1) gene which exhibits identical homologue, functional similarity and high conservancy in mammalian species [16]. Our study results along with previous reports suggest mammals, precisely primates to have close evolutionary link. However, lower eukaryotes, such as amoeba, exhibited marked divergence from human MDH1. In alignment, MDH2 sequences from Amphibia and Arthropoda demonstrated a diminished percent identity to human, particularly when compared to mammals. This observation emphasizes the evolutionary divergence between lower eukaryotes and human, reflecting distinct functional requirements for MDH in these groups. Bacterial MDH sequences demonstrated an intermediate level of sequence similarity (approximately 58%) with human MDH2, with notable variations for *Staphylococcus*. The shorter lengths of bacterial MDH sequences, coupled with varying degrees of conservation, suggest adaptive changes in MDH2 across bacterial species, possibly driven by environmental and functional considerations. Archaeal MDH sequences, on the other hand, exhibited even shorter lengths compared to bacterial counterparts, with query coverage ranging from 62% to 82%. Archaeal MDH was identified as the most divergent from human MDH2, with percent identity ranging between 26.83% and 29.35%. This indicates the archaeal MDH to have distinct evolutionary entity than mitochondrial MDH, supporting the notion of unique adaptations in the metabolic pathways of archaea [17].

The phylogenetic analysis unveiled distinct clusters for MDH1 and MDH2, indicative of substantial genetic diversity between mitochondrial and cytoplasmic MDH enzymes (Figure 1). Corroborating to our finding, Goward et al. also noticed distinct phylogenetic clusters of MDH1 and MDH2 (Goward and Nicholls 1994). One significant observation is the clustering of prokaryotic MDH sequences with human mitochondrial MDH2, rather than cytoplasmic MDH1. This finding extends previous understanding reporting that the mitochondrial enzyme is more closely related to its prokaryotic counterpart than to the cytoplasmic MDH enzyme [18]. This finding suggests a closer evolutionary connection between prokaryotic MDH and mitochondrial MDH2, shedding light on a more recent divergence from a common ancestral gene. The formation of a distinct cluster by all MDH1 sequences underscores their unique genetic divergence from MDH2 and prokaryotic MDH. This supports the idea of a cytoplasmic version of MDH1, indicating distinct evolutionary trajectories for these variants. This insight contributes to our understanding of the evolutionary dynamics and functional diversification of MDH enzymes in different cellular compartments. The specific clustering of *Staphylococcus* MDH with archaeal MDH and separation of other bacterial MDH sequences into a distinct cluster, apart from *Staphylococcus*, highlights the diversity within bacterial MDH evolution. This divergence likely reflects bacterial adaptations to diverse ecological niches and metabolic pathways through horizontal genetic transfer [19]. The observed clustering patterns suggests distinct evolutionary paths for mitochondrial and cytoplasmic MDH, and highlight the intricate evolutionary history of MDH across diverse organisms.

The protein variability analysis of 27 MDH sequences across nine groups reveals a high variability of 89.59% in amino acids, emphasizing MDH's adaptability. Human MDH1 shows substantial divergence (82.2%) from prokaryotic MDH, while MDH2 exhibits lower divergence (77.8%). This discrepancy suggests a differential evolutionary pressure on MDH2, potentially reflecting a more conserved role or functional constraints compared to MDH1. In alignment with this, structural homology studies by William et al. indicated that MDH enzymes from *Escherichia coli*, plants, and mammals share high sequence homology with human MDH2, ranging from 55–95%. Furthermore, the study reports greater structural divergence between human MDH1 and MDH enzymes from *Escherichia coli, plants, and mammals*, with only 25–30% structural homology [6].

The pairwise evolutionary divergence analysis of MDH protein sequences across nine groups unravels intricate relationships, providing insights into the evolutionary dynamics of MDH in diverse biological contexts. For human MDH1, the divergence order indicates a close evolutionary association with Mammalia_MDH1, followed by Amoeba_MDH1. Conversely, Human MDH2 exhibits a distinctive pattern, displaying closer ties to Mammalia_MDH2, Amphibia_MDH2, and Arthropoda_MDH2. Arthropoda_MDH2 showcases a distinct divergence pattern, with a closer relationship to *H. sapiens*_MDH2 and Mammalia_MDH2, while exhibiting greater divergence with bacterial and archaeal MDH, as well as eukaryotic MDH1. Amoeba_MDH1 appears more closely related to Mammalia_MDH1 and H. *sapiens*_MDH1, exhibiting greater evolutionary divergence from all other groups. Archaeal MDH shows the most evolutionary divergence from all other groups, particularly with bacteria as its closest relative. Bacterial MDH exhibits a divergence pattern with closer ties to Arthropoda_MDH2, Amphibia_MDH2, *H. sapiens*_MDH2, and Mammalia_MDH2.

Remarkably, the pairwise distance between *H. sapiens* MDH2 and bacterial MDH exhibited a significantly lower value compared to the distance with archaeal MDH. This observation implies a closer evolutionary relationship between mitochondrial MDH and bacterial MDH. In contrast, the pairwise divergence order for Human MDH1 revealed a closer association with archaeal MDH when compared to its bacterial counterpart, providing support for an evolutionary link between cytoplasmic and archaeal MDH. This discovery sheds light on the endosymbiotic theory, suggesting a common ancestral organelle for all mitochondria, which might be an endosymbiotic alphaproteobacterium, while the host cell was related to Asgard Archaea [20 21]. This notion finds further support in the cases of all other MDH2 (Mammalia, Amphibia, Arthropoda), where bacterial MDH consistently exhibited closer proximity to MDH2 in the divergence order. Additionally, MDH1 of the lower eukaryote amoeba displayed less divergence to archaea compared to bacteria.

The MDH sequences within bacterial species exhibited smaller pairwise distances. This suggests a higher degree of conservation within bacterial MDH, indicating a relatively stable evolutionary history. Moving up the taxonomic hierarchy, the pairwise distances within archaeal species were greater than those in bacterial species but smaller than those in mammals and humans. Archaeal MDH, while more diverse than bacteria, still maintains a certain level of conservation, possibly reflecting shared functional constraints. The pairwise distances within mammals, encompassing both MDH1 and MDH2, were greater than those in bacterial and archaeal species. This increased divergence indicates a more dynamic evolutionary landscape, likely influenced by the complex physiological and ecological roles of MDH in multicellular organisms. The MDH sequences within human (MDH1 and MDH2) showed the largest pairwise distances. This heightened level of divergence observed in MDH1 and MDH2 of mammalian species and human suggests distinct evolutionary pathway of mitochondrial and cytoplasmic MDH, aligning with and extending previous studies[22].

Conclusively, the comprehensive analysis of Malate Dehydrogenase (MDH) sequences across diverse biological groups has revealed intriguing patterns of diversity and evolutionary relationships. The divergence patterns of human MDH1 and MDH2 indicate greater similarity within the mammalian group, particularly with primate MDH1 and MDH2, suggesting a more conserved evolutionary history in these lineages. Conversely, human MDH1 and MDH2 both exhibits high diversity when compared to lower eukaryotes, such as amoeba, amphibian and arthropods. The greater pairwise evolutionary divergence between MDH1 and MDH2, coupled with their distinct placement in separate phylogenetic clusters, provides insight into the distinct evolutionary trajectories of mitochondrial and cytoplasmic MDH. The pairwise divergence order points towards an archaeal origin for cytoplasmic MDH and a bacterial origin for mitochondrial MDH. This discovery aligns with the concept of the archaebacterial origin of eukaryotes, proposing an archaeal origin for cytoplasmic MDH (MDH1) and a bacterial origin for mitochondrial MDH (MDH2). Beyond a mere recapitulation of existing theories, the research integrates prior knowledge with novel findings, employing a multidimensional dataset for comprehensive comparisons across multiple groups. This approach enhances our understanding of Malate Dehydrogenase evolution, providing valuable insights into complex relationships and unique adaptations that have shaped the evolutionary history of these enzyme variants.

Despite its valuable contributions, the study acknowledges certain limitations. The inclusion of only three archaeal sequences may limit the comprehensive understanding of evolutionary relationships, particularly regarding the proposed archaeal origin of cytoplasmic MDH (MDH1). A more extensive representation of archaeal sequences would enhance the robustness of the findings. Additionally, the absence of a 3D structural homology analysis among MDH variants, while focusing on sequence-based assays, may limit a holistic understanding of MDH evolution. Future studies incorporating such analyses could offer valuable insights into the functional implications of sequence variations.

References

1. Yao Y-X, Li M, Zhai H, You C-X and Hao Y-J. 2011. Isolation and characterization of an apple cytosolic malate dehydrogenase gene reveal its function in malate synthesis. *J. plant . physio.* **168** (5): 474-480.
2. Dasika SK, Vinnakota KC and Beard DA. 2015. Determination of the catalytic mechanism for mitochondrial malate dehydrogenase. *Biophys J.* **108** (2): 408-419.
3. Minárik P, Tomásková N, Kollárová M and Antalík M. 2002. Malate dehydrogenases— structure and function. *Gen Physiol Biophys.* **21** (3): 257-265.
4. Banaszak LJ and Bradshaw RA. 1975. 6 Malate Dehydrogenases. In. *The Enzymes* eds), edn, pp 369-396. Elsevier.
5. Imran M, Munir MZ, Ialhi S, Abbas F, Younus M, Ahmad S, Naeem MK, Waseem M, Iqbal A, Gul S, Widemann E and Shafiq S. 2022. Identification and Characterization of Malate Dehydrogenases in Tomato (Solanum lycopersicum L.). *Int J Mol Sci.* **23** (17).
6. McCue WM and Finzel BC. 2022. Structural Characterization of the Human Cytosolic Malate Dehydrogenase I. *ACS Omega.* **7** (1): 207-214.
7. Imran M, Tang K and Liu JY. 2016. Comparative Genome-Wide Analysis of the Malate Dehydrogenase Gene Families in Cotton. *PLoS One.* **11** (11): e0166341.
8. Cox CJ, Foster PG, Hirt RP, Harris SR and Embley TM. 2008. The archaebacterialorigin of eukaryotes. *Proc Natl Acad Sci U S A.* **105** (51): 20356-20361.
9. López-García P and Moreira D. 2019. Eukaryogenesis, a syntrophy affair. *Nat Microbiol.* **4** (7): 1068-1070.
10. Latorre A, Durbán A, Moya A and Peretó J. *Role of symbiosis in eukaryotic evolution*. Cambridge University press: Cambridge UK, 2011; p 326-339.
11. Tamura K, Stecher G and Kumar S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol.* **38** (7): 3022-3027.

12. Jones DT, Taylor WR and Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci.* **8** (3): 275-282.

13. Garcia-Boronat M, Diez-Rivero CM, Reinherz EL and Reche PA. 2008. PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. *Nucleic Acids Res.* **36** (Web Server issue): W35-41.

14. Zuckerkandl E and Pauling L. 1965. Evolutionary divergence and convergence in proteins. In. *Evolving genes and proteins* eds), edn, pp 97-166. Elsevier.

15. Woo YH and Li WH. 2012. Evolutionary conservation of histone modifications in mammals. *Mol Biol Evol.* **29** (7): 1757-1767.

16. Ikpeme E, Udensi O, Kooffreh M, Etta H, Ushie B, Echea E and Ozoje M. 2016. *In silico* analysis of BRCA1 gene and its phylogenetic relationship in some selected domestic animal species. *Trends in Bioinformatics.* **10** (1): 1-10.

17. Siebers B and Schönheit P. 2005. Unusual pathways and enzymes of central carbohydrate metabolism in Archaea. *Current opinion in microbiology.* **8** (6): 695-705.

18. McAlister-Henn L. 1988. Evolutionary relationships among the malate dehydrogenases. *Trends Biochem Sci.* **13** (5): 178-181.

19. Wiedenbeck J and Cohan FM. 2011. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol rev.* **35** (5): 957-976.

20. Roger AJ, Muñoz-Gómez SA and Kamikawa R. 2017. The origin and diversification of mitochondria. *Current Biol.* **27** (21): R1177-R1192.

21. Koonin EV. 2015. Origin of eukaryotes from within archaea, archaeal eukaryome and bursts of gene gain: eukaryogenesis just made easier? *PhiloTrans Royal Society B: Biol Sci.* **370** (1678): 20140333.

22. Takahashi-Íñiguez T, Aburto-Rodríguez N, Vilchis-González AL and Flores ME. 2016. Function, kinetic properties, crystallization, and regulation of microbial malate dehydrogenase. *J. Zhejiang Univ - SCIENCE B.* **17** (4): 247-261.