Pak. J. Bot., 54(1): 363-369, 2022. DOI: http://dx.doi.org/10.30848/PJB2022-1(35)

COMPUTATIONAL ANALYSIS OF CATALASE FROM DIFFERENT SOURCE ORGANISMS

MUHAMMAD NAVEED SHAHID[1*], MAHNOOR AMJAD[1], UMAIR ASHRAF[1], ADIL JAMAL[2] AND JAVED IQBAL WATTOO[3]

[1]*Department of Botany, Division of Science and Technology, University of Education, Lahore, Pakistan*
[2]*Sciences and Research College of Nursing, Umm Al Qura University, Makkah-715, Kingdom of Saudi Arabia*
[3]*Department of Biotechnology, Faculty of life sciences, University of Central Punjab, Lahore, Pakistan*
**Corresponding author's email: naveed.shahid@ue.edu.pk*

Abstract

The catalase enzyme is present widely in all living organisms. It is the key enzyme in terms of signaling and the metabolism of hydrogen peroxide and belongs to a large superfamily. The main reason to study evolution is the great role of catalase in the degradation of hydrogen peroxide as it is very necessary for all aerobic eukaryotes and prokaryotes. It has great importance due to its linkage with cancer and aging. The sequences of catalase enzymes are present in a wide range of public databases. For the assessment of phylogeny of catalase sequences, 45 full-length amino acid sequences of catalase were searched and collected for further study. The sequences were used for multiple sequence alignment, phylogenetic construction, motif discovery, domain identification, and discovering individual amino acid composition. MSA revealed amino acid glycine and aspartic acid play an important role in the evolution from prokaryotic to eukaryotic organisms with concern for catalase. Two main sequence clusters were obtained by phylogenetic analysis. One cluster comprised of 15 plant species, 8 bacterial species, and 3 fungal species. Cluster II contained 7 bacterial species and 12 fungal species. All the observed organisms showed similar domain structures but some fungal and bacterial sequences possessed the c-terminal domains. The average amino acid frequency was 7.35% that was very much high in contrast to other amino acids. This idea about amino acid frequencies demonstrated that alanine played a vital role in the composition of catalase.

Key words: Catalase, Pfam analysis, Motif elicitation server, Motif, Domain, MEME tool.

Introduction

Catalase is found in all living organisms involves the breakdown of hydrogen peroxide (H_2O_2) to water and oxygen. It is a tetrameric compound with a heme-group in its subunit and has great importance due to its old origin history (Chelikani *et al.,* 2005). It is of great importance due to the presence of events of pathologies linked with their malfunction as involves in enhanced susceptibility to apoptosis, stimulation of tumors, regulated aging and inflammation. It utilizes H_2O_2 for the oxidation of toxic substances such as alcohols and phenols (Ashokan *et al*., 2011). It protects the cell from oxidative damage and also helps in defense mechanisms. Another important characteristic of catalase is its high potent catalysts. It uses H_2O_2 to oxidize toxins comprising of formic acid, phenols, formaldehyde, and alcohols (Muthuraj, 2015). The catalase is the first enzyme to be crystallized and purified and has accomplished great attention in recent years due to its linkage to diabetes, cancer, and aging in animals and humans (Bloch *et al.,* 2007). It has been characterized and purified from a variety of organisms such as plants, like van apple, tobacco, black gram, and parsley, and also from the liver of many mammals such as goat, bacteria, bovine, and dog (Yoruk *et al.,* 2005; Lokman *et al.,* 2007; Kandukuri *et al.,* 2012). Due to its great importance and well-known existence in all mankind the computational study of the amino sequence of catalase was done which indicated the conserved secondary structure in the sequences that playing important role in the evolution (Sheoran *et al.,* 2013). The primary survey about catalases was done to study the properties and important cellular functions. The catalase is present in all aerobic organisms. It was purified and isolated from different organisms (Hoseinian *et al.,* 2006).

The molecular phylogeny of catalases, cytochrome c peroxidases, and ascorbate peroxidases was observed. The characteristics features of catalase have been studied in many eukaryotes and prokaryotes. It is present in several forms in many plant species like tobacco, mustard, pepper, saffron, maize, castor bean, sunflower, cotton, wheat, and Spinach (Garcia *et al.,* 2000; Keyham *et al.,* 2002; Lee & An, 2005). Evidence showed that catalase also has greater importance due to its role in aging, senescence, and plant defense (Mura *et al.,* 2007). Further study is required to elaborate on the sequences of catalase in contrast with other organisms (Boon *et al.*, 2007). It has been also studied to explore the phylogenetic relationship between prokaryotic and eukaryotic catalase. For this purpose, a comparative analysis of fungi, plant, and bacterial sequences is thought to be done by their phylogenetic reconstruction (Zhu *et al.*, 2019). Physico-chemical characteristics and 3D model of protein (Swiss-model server) were analyzed. This suggests which type of amino acids are taking part predominately in the plant sequences as polar, non-polar, acidic, or basic amino acids (Ashokan *et al.*, 2011).

This computational analysis aims to provide an understanding of the contribution of catalase in the evolutionary relationship between the species at the molecular level. The sequences of the amino acid in the individual profiles align to compare the amino acid composition and the role of individual amino acids in the evolution of catalase. The phylogenetic tree construction is to check the evolution of catalase from prokaryotes (bacteria) to eukaryotes (Fungi and Plants).

Materials and Methods

Retrieval of sequences: The sequences of amino acids were searched and retrieved from the NCBI at source https://www.ncbi.nlm.nih.gov/protein (Table 1). The 45 full-length amino acid sequences of the catalase enzyme were collected from different taxonomic groups of plants, fungi, and bacteria. The collected sequences were set into plants, fungal and bacterial profiles. Accession numbers of the sequences were taken for further data information.

Table 1. Retrieval of sequences from NCBI database.

S. No.	Source	Species	Accession No.
1.	Bacteria	*Helicobacter pylori*	AFJ81452
2.	Bacteria	*Cellulophaga algicola*	ADV50285.1
3.	Bacteria	*Capnocytophaga canimorsus*	AEK23441.1
4.	Bacteria	*Riemerella anatipestifer*	ADQ82581.1
5.	Bacteria	*Pseudopedobactersaltans*	ADY51732.1
6.	Bacteria	*Rubinisphaerabrasiliensis*	ADY60883.1
7.	Bacteria	*Deinococcus maricopensis*	ADV66323.1
8.	Bacteria	*Deinococcus proteolyticus*	ADY27184.1
9.	Bacteria	*Deinococcus proteolyticus*	ADY27165.1
10.	Bacteria	*Haliscomenobacterhydrossis*	AEE53489.1
11.	Bacteria	*Desulfobulbus propionicus*	ADW18085.1
12.	Bacteria	*Desulfobacca acetoxidans*	AEB09031.1
13.	Bacteria	*Syntrophobotulus glycolicus*	ADY55709.1
14.	Bacteria	*Weeksella virosa*	ADX68727.1
15.	Bacteria	*Cellulophaga algicola*	ADV50835.1
16.	Fungi	*Rhizoctonia solani*	CEL54483.1
17.	Fungi	*Blumeria graminis*	AAL56982.1
18.	Fungi	*Trichoderma reesei*	EGR44304.1
19.	Fungi	*Wickerhamomyces ciferrii*	XP_011273754.1
20.	Fungi	*Bipolarismaydis*	AAR17472.1
21.	Fungi	*Aspergillusnidulans*	AAC49713.1
22.	Fungi	*Taiwanofungus camphorates*	AAY87889.1
23.	Fungi	*Botrytis cinerea*	CAA91159.1
24.	Fungi	*Schizosaccharomyces japonicus*	XP_002175898.1
25.	Fungi	*Coprinopsiscinerea*	XP_001831809.2
26.	Fungi	*Sclerotiniasclerotiorum*	XP_001589420.1
27.	Fungi	*Aspergillusfischeri*	XP_001265572.1
28.	Fungi	*Xanthophyllomyces dendrorhous*	CED83116.1
29.	Fungi	*Verticillium dahliae*	EGY21639.1
30.	Fungi	*Sclerotinia sclerotiorum*	EDN93275.1
31.	Plant	*Glycine max*	NP_001240021.1
32.	Plant	*Zea mays*	CAA38588.1
33.	Plant	*Prunuspersica*	CAB56850.1
34.	Plant	*Capsicum annuum*	NP_001311603.1
35.	Plant	*Klebsormidium nitens*	GAQ78361.1
36.	Plant	*Saccharum hybrid*	AIU99487.1
37.	Plant	*Triticum aestivum*	ADF83496.1
38.	Plant	*Hevea brasiliensis*	AAG43363.1
39.	Plant	*Arabidopsis thaliana*	CAB80226.1
40.	Plant	*Oryza sativa*	BAA34205.1
41.	Plant	*Populusdeltoides*	CAI43948.1
42.	Plant	*Vigna radiate*	NP_001304079.1
43.	Plant	*Vitisvinifera*	NP_001268098.1
44.	Plant	*Musa acuminate*	SIW58963.1
45.	Plant	*Litchi chinensis*	AEQ30070.1

Multiple sequence alignment: To check the alignment accuracy the alignment of protein profiles was developed by multiple sequence alignment (https://www.ebi.ac.uk/ Tools/msa/muscle/). The PSI-BLAST tool that is based on the BLAST algorithm was developed to check the similar (conserved) regions in the sequences and their profile was constructed by scanning the sequences in the database (Altschul *et al.,* 1997; Park *et al.,* 1998). In the sequence analysis of catalase, the profiles were aligned by using the MUSCLE. The multiple sequence alignment of the sequences was obtained by the MUSCLE program.

Muscle: MUSCLE-fast was designed for the high rate accuracy in sequence alignments. The accuracy was compared with other alignment tools such as CLUSTALW, Progressive POA, and FFTNSI. The MUSCLE-fast tool can align 1,000 sequences in 21 seconds (Edgar, 2004). The individual profiles of plants, fungi, and bacteria were aligned by the MUSCLE tool. The sequences were subjected to the MUSCLE tool by copy-paste method or by inserting a sequence file in the input option available at source https://www.ebi.ac.uk/ Tools/msa/muscle/.

Motifs discovery: The multiple distinct motifs were discovered in a set of sequences. Different methodologies were developed to discover the unique motifs in the sequences of proteins. The expectation-maximization approach in the MEME algorithm was implemented to discover the motifs (Bailey *et al.,* 2009).

MEME tool: MEME tool aimed to discover motifs in a dataset of sequences about which little or nothing is known. The unaligned sequences were implemented to the expectation-maximization algorithm in the MEME program (Bailey *et al.,* 2009). The sequences were subjected to MEME input data and processed to discover motifs (https://meme-suite.org/meme/).

Domain identification: The protein sequences were implemented to identify and demonstrate a novel clustering-based approach to domain identification. It allows clustering of vectors in secondary structural elements (domains) and uses average linkage clustering to bury alpha-carbon positions. The obtained cluster represents the domain in the specific protein sequence (Feldman, 2012). The sequences were subjected for domain identification to analyze the evolutionary events concerning catalase (https://www.ncbi. nlm.nih.gov/Structure/cdd/wrpsb.cgi).

Pfam analysis: The Pfam includes the database of protein families and comprises 14,831 entries that were created manually. It was also described that 1182 families were entered and regulated sequence coverage of UniProtKB (Finn *et al.,* 2014). Pfam analysis of the catalase sequences was performed at source https://pfam. xfam.org/.

Phylogenetic reconstruction and bootstrap analysis: The relationship among protein sequences was demonstrated by the tree (Atkinson *et al.,* 2009). The tree was constructed by the UPGMA approach in the MEGA X and together with its bootstrap analysis with 500 replicates (Dwivedi & Mishra, 2014).

Mega X: MEGA X was previously used to analyze genome sequencing, track the evolutionary origin, and biological divergence. It includes many features for computing data in statistical analysis such as sequence alignments, genetic diversity, and distances, constructing the tree, analyzing amino acid composition that codes for a specific sequence (https://www.megasoftware.net/) (Kumar *et al.,* 2018).

Analysis of amino acids composition: The amino acid composition that coded for catalase was also assessed by MEGA X. The frequency of individual amino acid was obtained by the implementation of sequences in the MEGA X. To determine the amino acid composition "compute amino acid composition" was selected in the MEGA X window and the individual amino acid frequencies of all species were obtained (https://www.megasoftware.net/).

Results

Data collection tools: To analyze the evolution from prokaryotes to eukaryotes due to catalase, 45 full-length amino acid sequences of catalase from fungi, bacteria, and plants were searched and collected from database NCBI (National center for biotechnology information) Accession number of the retrieved sequences were obtained from sourcehttps://www.ncbi.nlm.nih.gov/protein and arranged into plants, bacterial, and fungal profiles.

Assessment of MSA: All the analyzed species of bacteria had 25 residues that are identically conserved in all observed sequences. All the plant species possessed 87 conserved residues. All the bacterial species possessed 67 residues which were identically conserved in all sequences profile. All the fungal species possessed 61 conserved residues. This suggested that all the plants, bacterial and fungal species were closely related to each other while bacteria and fungus were more closely linked about catalase in contrast to plants. In all analyzed species four aspartic acids, four glycines, three arginines, two glutamines, two phenylalanines, one tyrosine, one glutamic acid, one leucine, one isoleucine, one threonine, one lysine, one proline, one serine, and one valine were identically conserved. This revealed that amino acid aspartic acid and glycine played an important role in the evolution of catalase sequences from prokaryotes to eukaryotes (fungi and plants).

Phylogenetic analysis: Phylogenetic study of the sequences of plants revealed two main clusters (Fig. 1). Cluster I contained nine species that were more divided into two sub-clusters. Sub-cluster I contained five species (*Glycine max, Vigna radiate, Arabidopsis thaliana, Populus deltoids,* and *Musa acuminate*). Subcluster II contains four species (*Prunus persica, Hevea brasilieusis, Vitis vinifera,* and *Saccharum hybrid*). Cluster II comprises three species which were further classified into two subclusters. Subcluster I contains one species (*Zea mays*). Subcluster II contains two species (*Triticum aestivum* and *Oryza sativa*). *Litchi chinensis, Capsicum annuum* and *Klebsormidium nitens* were distantly related from all other species, that's why they were not included in any cluster.

Phylogenetic study of the sequences of bacteria demonstrated two major clusters (Fig. 2). Cluster one consists of eight species which were further divided into two subclusters. Subcluster I contains five species (*Capnocytophaga canimorus, Desulfobulbus propionicus, Helicobacter pyroli, Riemerella anatipestifer,* and *Desulfotomaculum acetoxidans*). Subcluster II contains tree species (*Cellulophaga algicola, Deinococcus maricopensis,* and *Deinococcus proteolyticus*). Cluster two contains seven species which were further divided into two subclusters. Subcluster I contains one species (*Deinococcus proteolyticus*). Subcluster II contains six species (*Rubinispaera brasiliensis, Weeksella virosa, Syntrophobotulus glycolicus, Cellulophaga algicola, Pseudopedobacter saltans,* and *Haliscomenobacter hydrossis).* Phylogenetic analysis of the fungal sequences showed two major clusters (Fig. 3). Cluster I contain thirteen species which were further classified into two subclusters. Sub-cluster I contains seven species (*Trichoderma reesei, Verticillium dahliae, Bipolaris maydis, Aspergillus fischeri, Coprinopsis cinerea, Rhizoctonia solani* and *Xanthophyllomyces dendrorhous*). Sub-cluster II contains six species (*Taiwanofungus camhoratus, Schizosaccharomyces japonicus, Wickerhamomyces cinerea, Botrytis cinerea*, and *Sclerotinia sclerotiorum*). Cluster two contains two species one in subcluster I (*Blumeria graminis*) and one in subcluster II (*Aspergillus nidulans*).

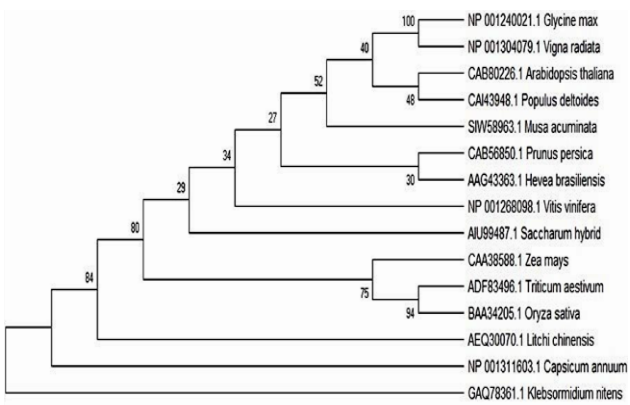

Fig. 1. Phylogenetic tree of the Plant Sequences by UPGMA Method. The bootstrap analysis with were performed with 500 replicates.

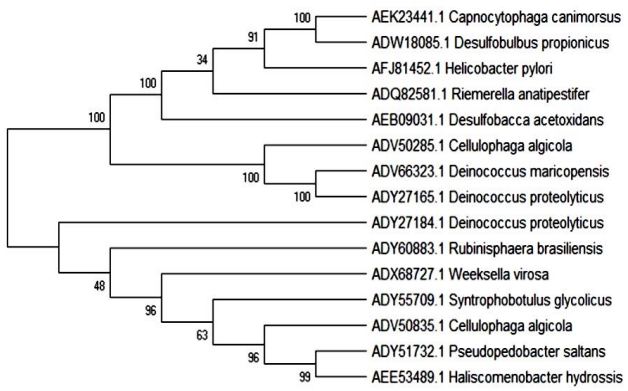

Fig. 2. Phylogenetic tree of the Bacterial Sequences by UPGMA Method. The bootstrap analysis with were performed with 500 replicates.

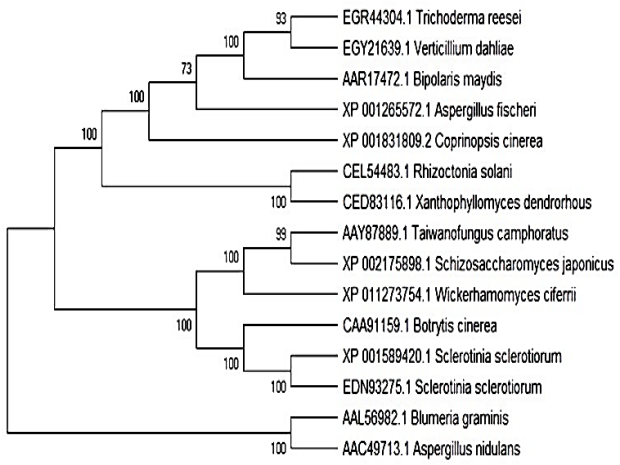

Fig. 3. Phylogenetic tree of the fungal sequences by the UPGMA method. The bootstrap analysis with were performed with 500 replicates.

When the complete sequences were taken for phylogenetic reconstruction two major clusters were obtained (Fig. 4). Cluster 1 contained twenty tree species which were further divided into two sub-clusters. Sub-cluster I contained fourteen plant and three bacterial species. Sub-cluster II contains five bacterial, three fungal, and one plant species. Cluster two contained nineteen species that were further classified into two clusters. Sub-cluster I contained two fungal and 7 bacterial species. Sub-cluster II contained ten fungal species. This phylogenetic construction reveals that plants, fungal and bacterial species were linked to each other but bacteria and fungi were more closely linked to each other concerning catalases, in contrast to plants.

Motif discovery: MEME tool was used to discover unique motifs in the sequences (Bailey *et al.*, 2009). The sequences that we used earlier for alignment and phylogenetic reconstruction were then used to discover motifs from the source (http://meme-suite.org/). Fourteen motifs were discovered that were unique in their taxonomic groups. The motifs width and also the number of sequences possessing motifs were also observed (Table 2).

Pfam analysis: The analysis of their protein family reveals that all of them belonged to the catalase family. A domain is a basic element which plays a fundamental structural role in proteins and gives unique characteristic to its related group (Feldman, 2012). The domains were obtained using Pfam analysis at the Sanger institute from the source (https://pfam.xfam.org/).

Amino acids composition assessment: The amino acid sequences individually were subjected to MEGA X to calculate their composition. The normal amino acid frequency of alanine was 7.35% that was very much high in contrast to other amino acids. This idea about amino acid frequencies demonstrated that alanine amino acid plays an important role in the composition of catalases. The amino acid frequency in all species is given in (Table 3).

Discussion

Multiple sequence alignment of 45 full-length sequences of plants, fungi, and bacteria showed sequence-based similarities. MSA revealed the correlation between plants, bacteria, and fungi. But the fungal and bacterial profiles were more closely linked to each other. During the assessment of sequences four aspartic acid and four glycine residues were identically conserved and thought to take part in evolution. The phylogenetic analysis gives a clearer vision about the pattern, processes of evolution, and comparative genetic research (Soltis & Soltis, 2003). The phylogenetic analysis of the catalase sequences in plants, fungi, and bacteria revealed two major clusters. Cluster 1 contains twenty-six species five bacterial, three fungal, and one plant species. Cluster two contains nineteen species in which twelve fungal, 7 bacterial species were present. The analysis demonstrated the linkage present between plants, fungal and bacterial sequences. In contrast to plants, the fungal and bacterial profiles were more closely related concerning to catalases. The statistical reliability of the constructed tree was checked by bootstrap analysis by using 500 replicates (Dwivedi & Mishra, 2014). The phylogenetic reconstruction reveals a high degree of similarity between the certain observed species. As the bacterial and fungal catalases were more closely linked and thought to share the origin in the phylogeny. The phylogenetic analysis also suggests that catalase is widely spread and evenly present in all the species belonging to the different taxonomic groups (Ashokan *et al.*, 2011). The percentage of replicates in which different organisms were clustered together in the bootstrap analysis is shown next to the branches. In this way, classification can considerably add to the understanding of the evolutionary relations between the species at the molecular level.

The domain identification of the aligned sequences suggested that all these sequences showed the catalase-related immune-responsive domain (Wang *et al.*, 2019). But some catalases possess the domain with c-terminal. Some bacterial and fungal profiles have been shown to possess large catalase enzymes in which the c-terminal domain was observed. The average amino acid frequency of alanine was 7.35% that was very high compared to other amino acids (Dwivedi and Mishra, 2014). The analysis of amino acid frequencies showed that alanine amino acid plays an important role in the composition of catalases. This suggested that the amino acid played a key role in the evolution of enzyme sequences from prokaryotes to eukaryotes (plants and fungi).

Table 3. Amino acids frequencies (%) of catalase of different source organisms.

	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr	Total
Helicobacter pylori	6.1	0.4	7.5	5.9	6.1	6.3	4.0	3.8	9.1	5.0	3.0	4.6	5.9	3.4	5.3	5.1	5.0	6.3	2.2	5.0	505
Cellulophaga algicola	6.1	0.0	6.7	4.9	5.3	7.1	1.6	6.1	8.1	6.7	2.0	5.9	5.3	5.7	3.8	5.9	7.1	5.3	1.6	4.5	506
Capnocytophaga canimorsus	7.7	0.6	6.7	6.3	6.7	7.1	3.9	4.5	6.3	6.1	3.1	6.3	5.7	4.1	5.9	3.5	4.9	5.1	1.6	3.7	508
Riemerella anatipestifer	5.5	0.6	6.7	5.9	6.9	6.3	4.0	4.8	7.1	5.7	3.2	7.9	4.4	4.0	4.6	6.1	5.1	5.9	1.4	4.0	495
Pseudopedobacter saltans	8.5	0.1	6.3	7.2	6.3	5.6	3.0	6.3	6.9	6.2	2.8	5.2	5.6	3.6	5.1	5.5	5.0	7.8	0.8	2.1	727
Rubinisphaera brasiliensis	8.7	0.7	7.4	6.9	5.9	6.9	3.2	4.0	5.0	6.9	2.9	4.3	6.6	4.2	5.7	5.7	5.0	7.0	1.3	1.7	698
Deinococcus maricopensis	6.0	0.4	6.8	7.2	5.4	8.3	4.1	3.3	3.9	6.8	2.1	4.9	5.8	6.0	6.6	3.7	6.2	7.6	1.2	3.7	515
Deinococcus proteolyticus	9.5	0.3	7.3	6.5	4.3	8.0	4.1	4.2	4.5	9.1	2.8	3.2	6.0	5.1	5.5	5.5	4.5	6.2	1.5	1.9	739
Deinococcus proteolyticus	6.3	0.6	6.5	7.4	4.8	8.4	3.6	4.0	4.0	5.7	2.1	5.2	5.9	7.1	6.7	4.6	5.7	5.7	1.5	4.2	524
Haliscomenobacter hydrossis	9.0	0.3	6.2	6.1	6.1	6.8	4.0	5.8	5.7	7.1	2.5	5.0	5.7	4.6	4.4	6.9	4.8	6.8	1.0	1.4	723
Desulfobulbus propionicus	8.6	0.8	6.6	5.7	5.7	7.4	4.1	3.7	5.1	6.8	2.5	5.1	6.2	4.5	6.6	4.5	5.1	5.5	1.8	3.5	487
Desulfobacca acetoxidans	8.7	0.0	6.9	5.1	5.5	7.9	3.9	3.6	4.7	7.3	1.8	5.5	6.9	3.6	6.9	3.9	6.1	5.9	1.4	4.3	507
Syntrophobotulus glycolicus	9.7	0.4	7.1	6.3	6.5	6.3	3.6	4.6	4.8	6.5	2.6	4.6	5.8	4.4	5.0	6.5	5.3	7.1	1.0	2.0	703
Weeksella virosa	6.6	0.6	7.8	6.6	5.8	6.4	3.0	5.7	8.0	7.0	1.9	5.6	4.7	4.4	3.8	6.1	4.6	7.7	1.3	2.6	702
Cellulophaga algicola	8.9	0.6	6.9	7.6	6.1	6.2	3.9	6.9	8.0	6.9	2.5	4.8	4.4	2.7	4.5	5.4	4.8	5.2	1.3	2.4	709
Glycine max	6.3	1.2	6.7	5.7	6.9	5.5	4.7	5.1	4.7	6.7	1.6	6.1	7.3	2.8	7.1	5.7	4.3	6.5	2.0	3.0	492
Zea mays	7.1	1.2	7.7	4.7	5.7	5.3	4.9	4.5	3.3	7.3	2.2	5.1	7.3	2.2	8.4	5.7	6.1	5.7	2.0	3.5	491
Prunus persica	5.9	1.2	7.1	5.7	6.9	4.7	5.4	5.0	5.4	6.9	1.4	5.0	7.3	3.3	6.9	5.9	4.0	6.4	1.9	3.8	423
Capsicum annuum	6.1	1.8	6.7	5.5	6.1	5.3	4.1	5.1	5.1	6.7	1.8	5.3	7.3	2.6	6.9	5.9	5.3	6.9	1.4	4.1	492
Klebsormidium nitens	8.4	0.7	5.9	5.2	4.9	8.2	2.9	2.8	5.3	7.3	2.2	5.2	6.1	4.9	5.6	6.4	5.6	7.3	2.0	2.9	850
Saccharum hybrid	5.5	1.8	6.9	5.1	6.7	5.7	5.5	4.9	4.9	6.9	2.4	5.9	6.9	3.0	6.5	4.9	5.3	6.5	1.8	2.8	492
Triticum aestivum	5.5	0.8	7.3	5.5	6.1	5.5	4.5	4.7	4.7	7.1	2.2	5.5	7.5	2.2	7.1	6.5	5.3	6.1	2.0	3.9	492
Hevea brasiliensis	4.5	2.4	6.3	5.9	6.5	5.7	4.5	5.1	4.3	7.5	1.6	6.5	6.9	3.0	6.9	6.5	4.5	6.5	1.8	3.0	492
Arabidopsis thaliana	5.7	1.2	6.7	5.9	6.7	5.3	4.5	6.1	4.9	6.3	2.0	6.1	7.5	3.0	6.9	5.7	4.5	5.7	1.6	3.7	492
Oryza sativa	5.3	0.8	7.7	4.5	6.3	5.5	4.9	5.1	4.5	7.3	1.8	5.5	7.7	2.4	7.3	6.9	4.9	5.9	2.2	3.5	492
Populus deltoides	6.1	1.6	6.5	5.7	6.5	5.7	4.7	5.1	4.9	6.9	1.6	5.3	7.5	2.8	6.7	6.1	4.3	6.3	2.0	3.7	492
Vigna radiata	6.5	0.6	6.7	5.7	7.3	5.3	4.9	5.3	4.9	6.7	1.6	6.3	7.1	2.6	6.9	5.7	4.1	6.7	2.0	3.0	492
Vitis vinifera	5.7	1.0	6.7	5.9	7.3	5.3	4.1	4.5	5.3	6.3	1.8	5.7	7.3	2.8	6.7	5.5	4.9	7.3	1.4	4.5	492
Musa acuminata	3.5	0.8	7.8	5.4	8.9	6.6	4.7	5.4	4.3	7.4	2.3	5.8	7.4	2.7	5.0	5.0	5.4	7.0	1.6	3.1	258
Litchi chinensis	5.7	1.4	7.4	4.7	6.8	6.4	4.3	5.3	5.3	6.6	1.4	6.0	7.2	2.7	6.6	5.1	4.5	7.0	2.1	3.5	487
Rhizoctonia solani	8.1	0.7	5.7	5.2	6.0	8.1	3.8	4.1	4.8	7.1	1.9	5.0	7.1	5.0	5.5	6.4	4.8	5.2	1.5	4.1	581
Blumeria graminis	7.5	0.1	6.5	4.3	6.1	8.2	2.6	4.3	4.3	7.8	1.5	4.5	5.8	4.6	4.9	8.5	6.8	7.4	1.5	2.5	718
Trichoderma reesei	6.6	1.3	7.3	5.9	7.9	7.1	2.9	3.7	5.7	5.3	2.9	5.9	6.6	4.2	5.1	5.1	5.5	5.7	1.5	4.0	547
Wickerhamomyces ciferrii	6.3	0.4	6.3	6.5	4.5	7.5	3.4	4.8	5.7	6.1	0.9	5.6	6.6	5.9	4.8	5.9	5.7	5.6	2.0	5.4	557
Bipolaris maydis	5.7	0.9	6.6	6.6	7.9	7.3	3.5	4.0	6.4	4.4	2.6	5.9	7.1	5.1	4.8	4.8	4.6	6.2	1.6	4.2	547
Aspergillus nidulans	10.4	0.1	7.2	5.8	6.8	7.6	2.1	3.9	3.3	8.0	1.2	6.0	4.9	4.2	5.4	6.2	6.0	7.5	1.0	2.4	721
Taiwanofungus camphoratus	10.0	0.4	6.1	5.3	6.7	6.1	4.9	4.5	4.9	6.7	1.4	4.7	6.3	4.5	4.7	6.1	5.3	6.9	1.2	3.3	509
Botrytis cinerea	8.4	0.4	6.3	5.0	7.9	6.3	2.5	5.0	4.6	6.7	0.8	6.1	5.8	5.6	4.6	5.2	8.6	6.3	1.0	2.9	479
Schizosaccharomyces japonicus	5.0	1.2	6.5	6.3	6.9	6.0	3.5	4.8	6.5	4.8	2.3	6.3	7.9	4.4	4.8	4.4	6.2	7.3	1.2	3.7	520
Coprinopsis cinerea	7.7	0.9	5.8	5.4	6.5	6.8	3.6	3.2	5.4	5.4	3.2	5.4	7.8	4.3	5.3	6.0	5.3	6.3	1.5	4.3	588
Sclerotinia sclerotiorum	8.2	0.7	5.3	5.1	6.8	6.7	2.6	5.6	5.6	5.6	1.4	7.5	5.8	5.0	5.1	5.6	7.9	5.6	0.9	2.9	585
Aspergillus fischeri	9.1	1.0	5.8	5.2	7.4	7.0	3.8	3.1	6.2	6.2	2.4	4.8	7.9	5.2	4.6	5.7	4.0	5.8	0.7	4.1	582
Xanthophyllomyces dendrorhous	7.8	0.9	6.0	4.4	6.1	7.8	3.1	3.4	4.3	5.4	3.4	5.1	6.6	5.1	5.6	8.0	5.6	5.4	1.4	4.6	588
Verticillium dahliae	9.2	1.2	6.8	5.1	7.0	7.7	4.3	2.9	4.8	5.5	2.4	5.1	7.0	4.9	6.0	5.1	4.1	6.1	1.4	3.6	587
Sclerotinia sclerotiorum	8.2	0.7	5.3	5.1	6.8	6.7	2.6	5.6	5.6	5.6	1.4	7.5	5.8	5.0	5.1	5.6	7.9	5.6	0.9	2.9	585
Avg.	7.4	0.8	6.7	5.8	6.3	6.7	3.7	4.6	5.4	6.6	2.1	5.5	6.4	4.2	5.7	5.7	5.3	6.4	1.5	3.3	559

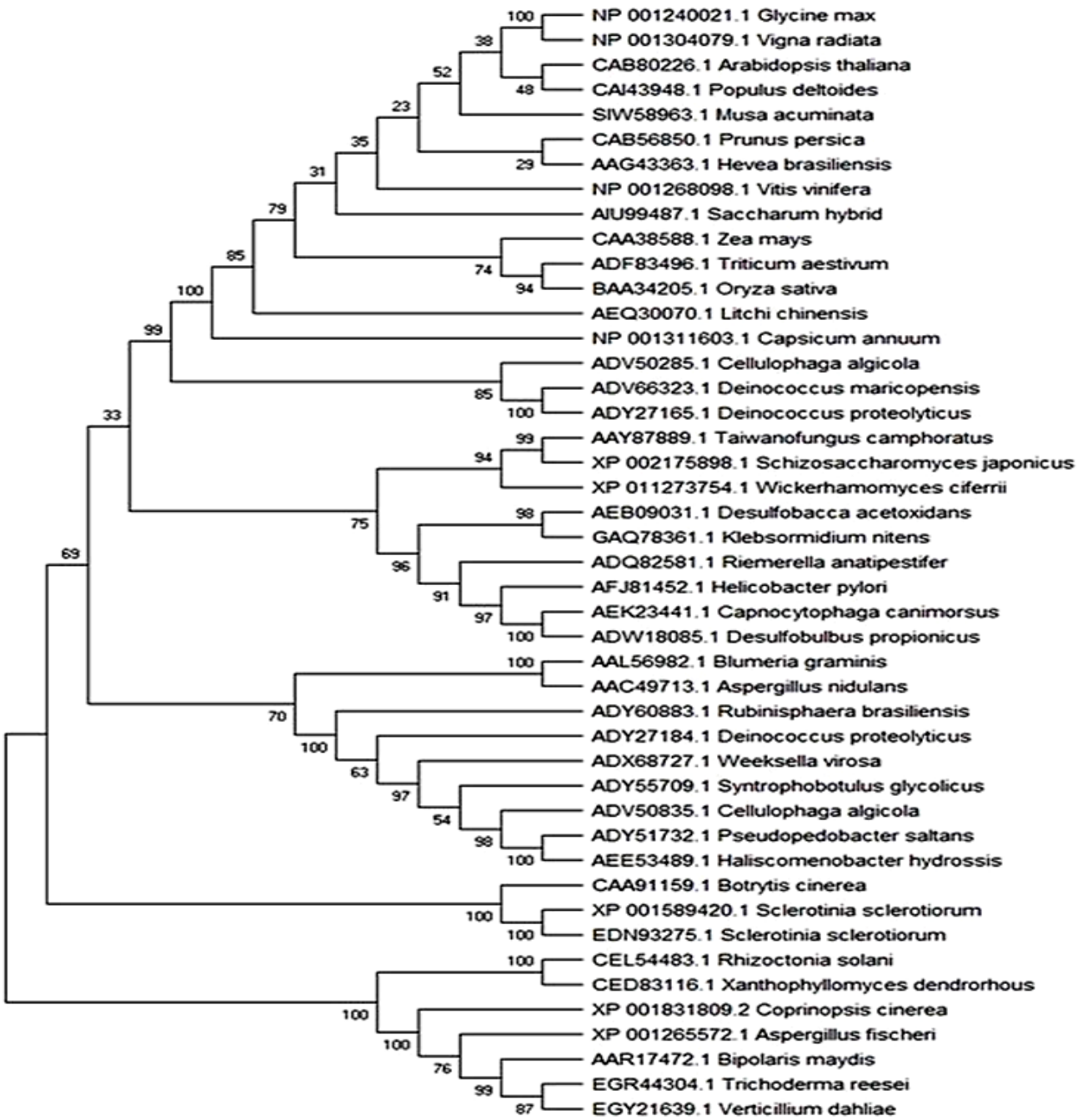

Fig. 4. Phylogenetic tree of complete sequences by the UPGMA method. The bootstrap analyses were performed with 500 replicates.

Table 2. Discovered motifs from MEME.

S. No	Source	Motif width	Motif present in number of sequences	Motif
1.	Bacteria	42	7	TIKSRKIAFLVADGVDEQSVNVVRIALENEGAMVEIVAPKLG
2.	Bacteria	29	7	DAIHFINEAYKHCKAIAADGAGVKLIDAT
3.	Bacteria	21	6	GFTSYAERIDARKVRARSKSF
4.	Bacteria	15	7	FIKAIAQHRFWEREK
5.	Fungi	43	7	VNLHQIPVNCPFMAKSYASLNFDGPMRTDANHAGNKQYAPNSF
6.	Fungi	50	7	EAPYAVSDNIMSRKSHYWHEGKKNDYDQATELWTRVMTSQERQNTCKNTA
7.	Fungi	27	7	EIQLKYLAQLYCIAPEYAKGVYDLLPE
8.	Fungi	41	2	NECKLTVSNGNNIGKAAKLYRNIGKDTKSRVSSPSNGVMRE
9.	Fungi	50	2	TIKQNETGTMTNGMENVKITKQNGIPNGNSHNSETRSRGFAANDCPRPCH
10.	Plants	21	11	MDPYKYRPSSAFNSPFWTTNS
11.	Plants	41	3	MNFMHRDEEVNYFPSRYDPVRHAESFPIPPAVCSGKREKCI
12.	Plants	50	13	KENNFKQPGERYRSWAPDRQDRFNRRWVDALSDPRVTHEIRSVWISYWSQ
13.	Plants	50	7	GPADIIAKVSGVVSGGKREDDSLYFTNNEGIPFPDPAHSKTVGGIPVASD
14.	Plants	8	13	DFLRAPGV

Conclusions

The catalase enzyme was studied in different source organisms. The multiple sequence alignment of the plants, fungal and bacterial profiles were obtained. MSA of catalase sequences revealed that amino acid aspartic acid and glycine play a major role in the evolution from bacteria to plants. Fourteen unique motifs in the plants, bacterial and fungal profiles were discovered that involved specific cellular functions. The domain identification demonstrated that all plants, bacterial and fungal sequences showed catalase-related immune-responsive domain. Some bacterial and fungal catalases exhibited a unique C-terminal domain. The phylogenetic analysis showed that plants, bacterial and fungal profiles were linked with each other but bacterial and fungal catalases were more closely related. The average amino acid frequency of alanine was 7.35% that was very high as compared to other amino acids. The assessment of amino acids revealed that alanine amino acids played an important role in the composition of catalases. It can contribute a significant role in the understanding of the evolutionary relations between the species at the molecular level. Considering the importance of catalases, more contribution is warranted for the detailed investigation of the activity and functional analysis of enzymes.

Conflict of interest

The authors state that the research was steered in the absence of any commercial or financial relationships that could be taken as a potential conflict of interest.

Acknowledgment

The authors are highly grateful to the University of Education, Township, Lahore who has provided us a platform to conduct subject research.

References

Altschul, S.F., T.L.Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W.Miller and D.J.Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res*., 25(17): 3389-402.

Ashokan, K.V., D.S. Mundaganur and Y.D. Mundaganur. 2011. Catalase: phylogenetic characterization to explore protein cluster. *J. Res. Bioinform.*, 1: 001-008.

Atkinson, H.J., J.H. Morris, T.E. Ferrin and P.C. Babbitt. 2009. Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PloSOne*, 4(2): e43-45.

Bailey, T.L., M. Boden, F.A. Buske, M. Frith, C.E., Grant, L. Clementi, J. Ren, W.W. Li and W.S. Noble. 2009. MEME SUITE: tools for motif discovery and searching. *Nucl. Acids Res.*, 37(suppl_2): 202-208.

Bloch, K.E., M.V. Shichman, D. Bloch and P. Vardi. 2007. Catalase expression in pancreatic alpha cells of diabetic and non-diabetic mice. *Histchem.Cell Biol.,* 127(2): 227-232.

Boon, E.M., A. Downs and D. Marcey. 2007. Proposed mechanism of catalase in catalase: H_2O_2: H_2O_2 oxidoreductase. Catalase Structural Tutorial Text http://biology.kenyon.edu/BMB/ Chime/ catalase/ frames/cattx.htm.

Chelikani, P., T. Ramana and T.M. Radhakrishan. 2005. Catalase: A repertoire of unusual features. *Ind. J. Clin.Biochem.*, 20(2): 131-13.

Dwivedi, V.D., and S.K. Mishra. 2014. *Insilico* analysis of L-asparaginase from different source organisms. *Interdiscip. Sci. Comput. Life Sci.,* 6: 93-99.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucl. Acids Res.,* 32(5): 1792-97.

Feldman, H.J. 2012.Identiying structural domains of proteins using clustering.*BMC Bioinform.,* 13: 286.

Finn, R.D., A. Bateman, J. Clements, P. Coggill, R.Y. Eberhardt, S.R. Eddy, A. Heger, K. Hetherington, L. Holm, E.L. Sonnhammer, J. Tate and M. Punta. 2014. Pfam: the protein families database. *Nucl. Acids Res.,* 42(Database issue): D222–D230.

Garcia, R., N. Kaid, C. Vignaud and J. Nicolas. 2000. Purification and some properties of catalase from wheat germ (*Triticum aestivum* L.). *J. Agri. Food Chem.,* 48: 1050-1057.

Hoseinian, G. A., N. Ghaemi and F. Rahimi. 2006. Partial purification and properties of catalase from *Brassiaoleracea capitata*. *Asian J. Plant Sci.*, 5: 827-831.

Kandukuri, S.S., A. Noor, S.S. Ranjini and M. Vijayalakshmi. 2012. Purification and characterization of catalase from sprouted black gram (*Vignamungo*) seeds. *J. Chromatogr. Anal.Technol. Biomed. Life Sci*., 15: 889-890.

Keyham, J., E. Keyhani and J. Kamali. 2002. Thermal stability of catalases active in dormant saffron corms. *Mol. Rep.,* 29(1-2): 125-128.

Kumar, S., G. Stecher, M. Li, C. Knyaz and K. Tamura. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.*, 35(6): 1547-1549.

Lee, S.H. and C.S. An. 2005.Differential expression of three catalase genes in hot pepper. *Mol. Cells*, 20(2): 247-255.

Lokman, O., M. Bulbul, M. Elmastas and M. Ciftci. 2007. Purification and some kinetic properties of catalase from parsley leaves. *Preparat. Biochem. Biotechnol.,* 37: 229-238.

Mura, A., F. Pintus, R. Medda, G. Floris, A.C. Rinaldi and A. Padiglia. 2007. Catalase and antiquitin from *Euphorbia characias*: Two proteins involved in plant defense. *Biochem.*, 72: 501-508.

Muthuraj, M. 2015. Cause of endurance training on catalase of physical education students. *Int. J. Phys. Edu*. Yoga & Health Sci., (IJPEYHS) ISSN 2349- 6312 Vol. 2.

Park, J., K. Karplus, C. Barrett, R. Hughey, D. Haussler, T. Hubbard and C. Chothia. 1998. Sequence comparisons using multiple sequences detect three times as many remote homologues as pairwise methods. *J. Mol. Biol*., 284: 1201-1210.

Sheoran, S., B. Pandey, P. Sharma, S. Narwal, R. Singh, I. Sharma and R. Chatrath. 2013. *Insilico* comparative analysis and expression profile of antioxidant proteins in plants. *Genet. Mol. Res.*, 12(1): 537-51.

Soltis, D.E. and P.S. Soltis. 2003. The role of phylogenetics in comparative genetics. *Plant Physiol.,* 132(4): 1790-1800.

Wang, W., Y. Cheng, D. Chen, D. Liu, M. Hu, J. Dong, X. Zhang, L. Song and F. Shen. 2019. The catalase gene family in cotton: genome-wide characterization and bioinformatics analysis. *Cells*, 8(2): 86.

Yoruk, I.H., H. Demir, K. Ekici and A. Sarvan. 2005. Purification and properties of catalase from van apple (Golden delicious). *Pak. J. Nutr.,* 4: 8-10.

Zhu, T., F. Xin, S. Wei, Y. Liu, Y. Han, J.Xie, Q.Ding and L. Ma. 2019. Genome-wide identification, phylogeny and expression profiling of class III peroxidases gene family in *Brachypodium distachyon*.*Gene*, 700: 149-162.

(Received for publication22 January 2020)