Netw Model Anal Health Inform Bioinforma (2013) 2:185–189
DOI 10.1007/s13721-013-0029-7

ORIGINAL ARTICLE

Computational analysis of xanthine dehydrogenase enzyme from different source organisms

Vivek Dhar Dwivedi · Shikha Arora · Akash Kumar · Sarad Kumar Mishra

Received: 2 January 2013 / Revised: 25 January 2013 / Accepted: 11 February 2013 / Published online: 27 February 2013

Abstract In this communication, 30 full length protein sequences of xanthine dehydrogenase enzyme from different species of fungi, bacteria and animals were retrieved from GenPept database and subjected to computational analysis including multiple sequence alignment (MSA), phylogenetic analysis, motif identification, domain identification and discovering individual amino acid composition. MSA revealed that thirteen glycine, three aspartic acid, three glutamic acid, three glutamine, three alanine, two arginine, one valine, one phenylalanine, one tyrosine, one proline, one serine and one asparagine residues which were identically conserved in all analyzed species. Two major sequence clusters were obtained by phylogenetic analysis. One cluster was of animal and fungal origin whereas the other one was of bacterial origin. The amino acid composition results revealed that average frequency of alanine was 9.24 %, which was higher in comparison to other amino acids average frequency. Average frequency of tryptophan residue was 0.84 %, which was very low in all analyzed species. Six domains were identified but only two domains were conserved in all the sequences of bacteria, fungi and animals. These domains were found to be responsible for the functional activity of xanthine dehydrogenase enzyme. In addition, some motifs which were unique for their groups were also identified.

Keywords Xanthine dehydrogenase · MSA · Phylogenetic analysis · Conserved regions · Motifs · Domains

V. D. Dwivedi (✉) · S. Arora
Forest Pathology Division,
Forest Research Institute, Dehradun, India
e-mail: vivek_bioinformatics@yahoo.com

A. Kumar
Department of Bioinformatics,
Uttaranchal College of Science and Technology,
Dehradun, India

S. K. Mishra
Department of Biotechnology,
DDU Gorakhpur University, Gorakhpur, India

1 Introduction

Xanthine dehydrogenase (XDH) is widely distributed enzyme in bacteria to higher organisms (Bray 1975). This enzyme belongs to the group of molybdenum-containing hydroxylases involved in oxidative metabolism of purines. XDH is a homodimer, can be converted to xanthine oxidase by reversible sulfhydryl oxidation or by irreversible proteolytic modification (Della and Stirpe 1972; Amaya et al. 1990). XDH uses NAD^+ as its physiological electron acceptor whereas xanthine oxidase (XO) transfers the reducing equivalents directly to molecular oxygen, generating superoxide radicals and hydrogen peroxide as by-products (Hille and Massey 1981). Superoxide production is also catalyzed by XDH due to its intrinsic NADH-oxidase activity (Sanders et al. 1996). XDH catalyzes the initial oxidation to uric acid. Although uric acid is excreted with urine as the final catabolite in certain animals including humans and other primates, it is further oxidized in higher organisms to urea and ureides such as allantoin and allantoate, and ultimately broken down into CO_2, ammonia and glyoxylate (Stasolla et al. 2003; Zrenner et al. 2006). These end products of purine catabolism are then presumably reutilized or reassimilated as carbon and nitrogen sources. Considering the above facts, a study of amino acid sequences of XDH enzyme from different source organisms is quite challenging. In the present study,

we performed the computational analysis of retrieved amino acid sequences of XDH from bacteria, fungi, and animals to find out the evolutionary relationships, hidden biological information and correlated them on the basis of common features using bioinformatics tools.

2 Materials and methods

The 30 full length amino acid sequences of xanthine dehydrogenase from bacteria, fungi and animals were searched and retrieved from GenPept database available at NCBI (www.ncbi.nlm.nih.gov/protein). The sequences were arranged in bacterial, fungal and animal profiles, respectively. The multiple sequence alignment of the individual profiles was performed using MUSCLE at the European Bioinformatics Institute (Edgar 2004) (http://www.ebi.ac.uk/Tools/msa/muscle/). Motifs were identified in profiles using the expectation maximization approach implemented in Multiple EM for Motif Elicitation server (Bailey and Elkan 1995) (http://meme.nbcr.net/meme/). Domains were identified by batch searching in Pfam (database Punta et al. 2012) (http://pfam.sanger.ac.uk/). The UPGMA approach implemented in the Mega program was employed for constructing phylogenetic relationships among sequences (Kumar et al. 2008). The statistical reliability of the phylogenetic tree was tested by bootstrap analyses with 500 replications. Mega program is also used for discovering individual amino acid composition.

3 Results and discussion

The accession number of retrieved sequences along with the species name and origin is listed in Table 1. MSA showed the presence of some conserved regions in all the sequences from different sources, while others were restricted only to their groups. All the analyzed species of bacteria, fungi and animals possessed thirteen glycine, three aspartic acid, three glutamic acid, three glutamine, three alanine, two arginine, one valine, one phenylalanine, one tyrosine, one proline, one serine and one asparagine residues which were identically conserved in all analyzed species.

Phylogenetic analysis of bacterial sequences profile showed the two major clusters (Fig. 1). Cluster I consisted of six species which was further divided into two subclusters. Subcluster I contains *Halomonas* sp., *Pseudomonas syringae,* and *Ochrobactrum anthropi*. Subcluster II contains *Niastella koreensis*, and *Chitinophaga pinensis. Sulfobacillus acidophilus* was distantly related and not included in any subclusters. Cluster II consists of four species namely *Escherichia coli*, *Oribacterium* sp., *Clostridium hathewayi*, and *Dethiobacter alkaliphilus*.

Table 1 Retrieved sequences and their accession number from GenPept database

Serial number	Source	Name of organisms	Accession number
1.	Bacteria	*Escherichia coli*	ZP_07124277.1
2.	Bacteria	*Niastella koreensis*	YP_005006254.1
3.	Bacteria	*Halomonas sp.*	ZP_08637530.1
4.	Bacteria	*Pseudomonas syringae*	EFW79581.1
5.	Bacteria	*Dethiobacter alkaliphilus*	ZP_03729912.1
6.	Bacteria	*Clostridium hathewayi*	ZP_06113592.1
7.	Bacteria	*Ochrobactrum anthropi*	ZP_10965261.1
8.	Bacteria	*Oribacterium sp.*	ZP_06599899.1
9.	Bacteria	*Chitinophaga pinensis*	YP_003126368.1
10.	Bacteria	*Sulfobacillus acidophilus*	YP_005257444.1
11.	Fungi	*Metarhizium acridum*	EFY88162.1
12.	Fungi	*Neurospora tetrasperma*	EGZ70183.1
13.	Fungi	*Verticillium dahliae*	EGY16571.1
14.	Fungi	*Emericella nidulans*	CAA58034.1
15.	Fungi	*Aspergillus kawachii*	GAA86933.1
16.	Fungi	*Gaeumannomyces graminis*	EJT71940.1
17.	Fungi	*Beauveria bassiana*	EJP61546.1
18.	Fungi	*Exophiala dermatitidis*	EHY58092.1
19.	Fungi	*Cordyceps militaris*	EGX88933.1
20.	Fungi	*Ajellomyces dermatitidis*	EGE82939.1
21.	Animal	*Rattus norvegicus*	AAA42349.1
22.	Animal	*Felis catus*	AAF97949.1
23.	Animal	*Homo sapiens*	AAA75287.1
24.	Animal	*Gallus gallus*	BAA02502.1
25.	Animal	*Mus musculus*	CAA52997.1
26.	Animal	*Mustela putorius furo*	AES09841.1
27.	Animal	*Taeniopygia guttata*	XP_002194980.1
28.	Animal	*Bos taurus*	CAA58497.1
29.	Animal	*Poecilia reticulata*	AAK59699.1
30.	Animal	*Sus scrofa*	AEW10559.1

Phylogenetic analysis of fungal sequences profile showed the two major clusters (Fig. 2). Cluster I consisted of five species namely *Beauveria bassiana*, *Cordyceps militaris*, *Metarhizium acridum*, *Verticillium dahliae* and *Neurospora tetrasperma.* Cluster II also consists of five species which were further divided into two subclusters. Subcluster I contains two species (*Emericella nidulans* and *Exophiala dermatitidis*). Subcluster II also contains two species (*Aspergillus kawachii* and *Ajellomyces dermatitidis*). *Gaeumannomyces graminis* was distantly related and not included in any subclusters.

Phylogenetic analysis of animal sequences profile showed the two major clusters (Fig. 3). Cluster I consisted of five species which were further divided into two subclusters. Subcluster I contains three species (*Gallus gallus*,

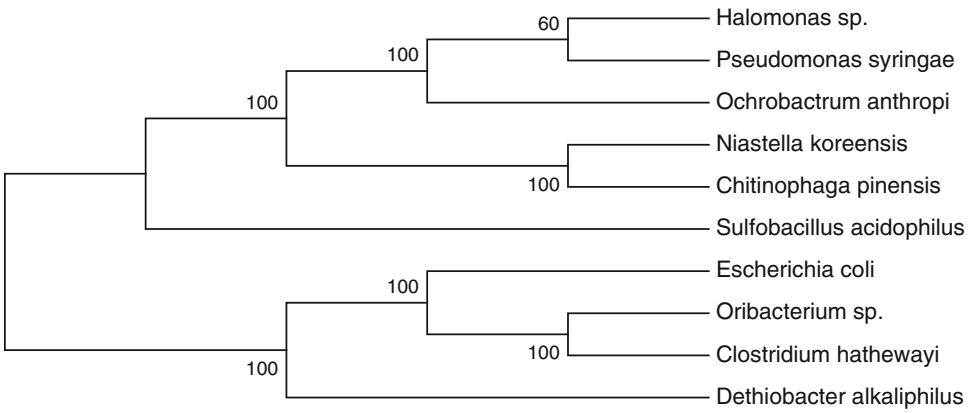

Fig. 1 Phylogenetic analysis of bacterial sequences using UPGMA method

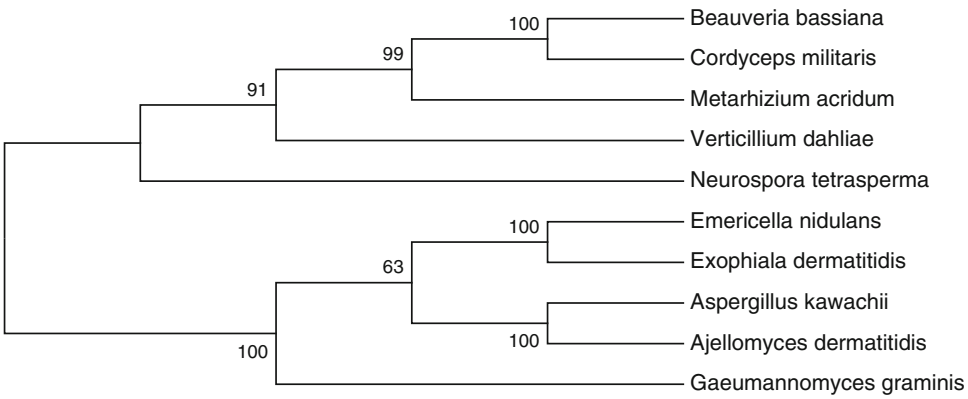

Fig. 2 Phylogenetic analysis of fungal sequences using UPGMA method

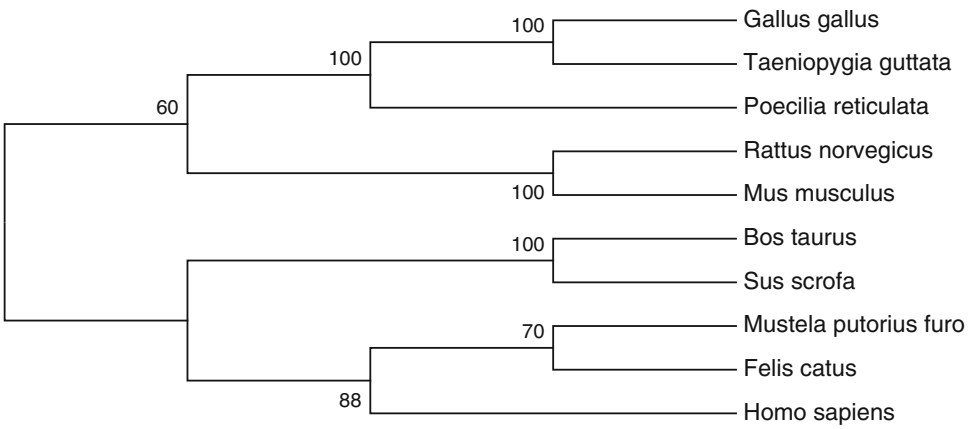

Fig. 3 Phylogenetic analysis of animal sequences using UPGMA method

Taeniopygia guttata, and *Poecilia reticulate*). Subcluster II contains two species (*Rattus norvegicus*, and *Mus musculu*). Cluster II also consists of five species which was further divided into two subclusters. Subcluster I contains two species (*Bos taurus* and *Sus scrofa*). Subcluster II contains three species (*Mustela putorius furo*, *Felis catus*, and *Homo sapiens*).

When complete retrieved sequences of bacteria, fungi and animals were taken for phylogenetic analysis, two major sequence clusters were obtained (Fig. 4). One cluster was of animal and fungal origin, whereas the other one was of bacterial origin. Six domains were identified but only two domains were conserved in all the sequences of bacteria, fungi and animals (Table 2). These domains were found to be responsible for the functional activity of xanthine dehydrogenase enzyme. The amino acid composition results revealed that average frequency of alanine was 9.24 % that was higher in comparison to other amino acids average frequency and average frequency of tryptophan residue was 0.84 % that was very low in all analyzed species. In addition, some motifs which were unique for their groups were also identified (Table 3).

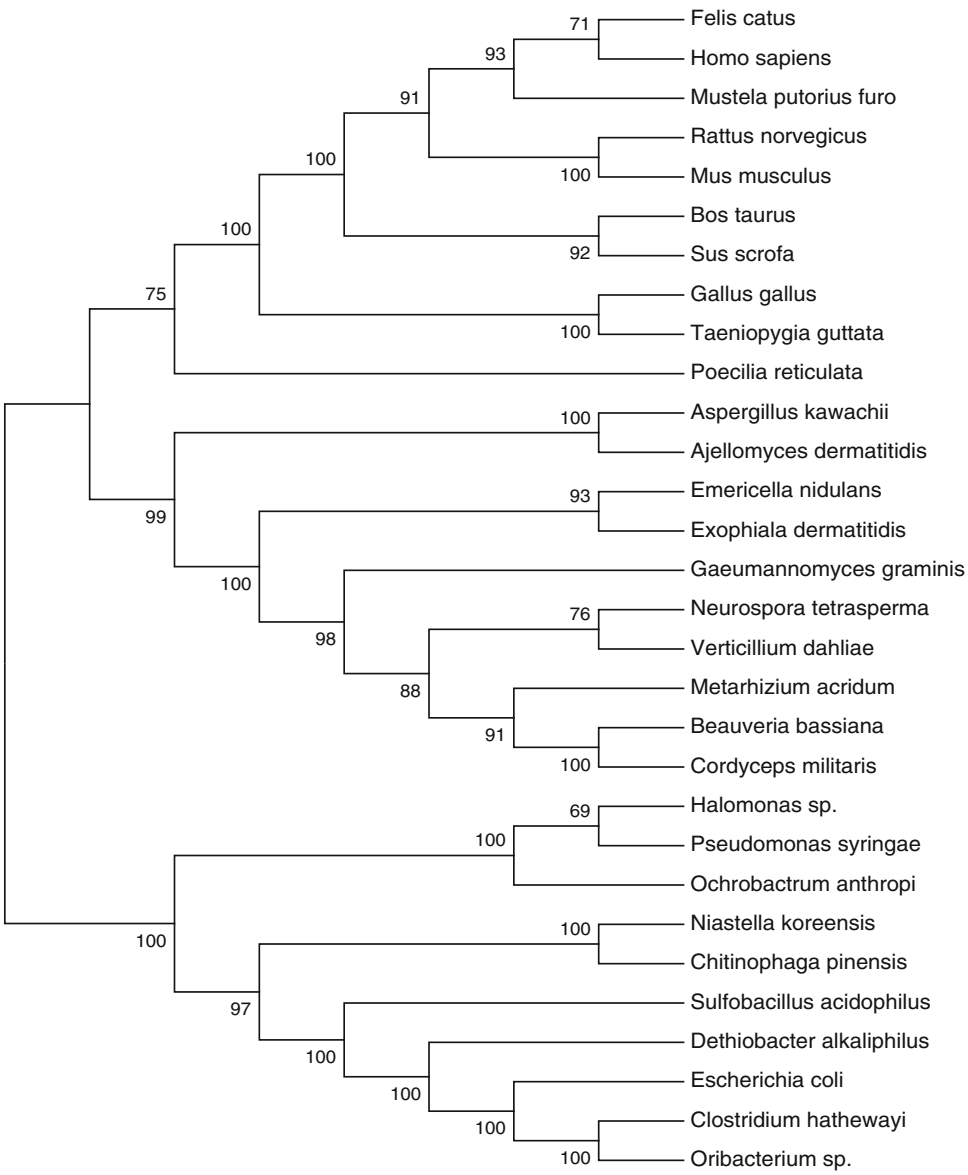

Fig. 4 Phylogenetic analysis of joint profile of bacterial, fungal, and animal sequences using UPGMA method

Table 2 Domains identified along with their family and source by batch searching in Pfam database

S. no.	Domain name	Family	Source
1.	2Fe-2S iron-sulfur cluster binding domain	Fer2	Fungi and animals
2.	[2Fe-2S] binding domain	Fer2_2	Fungi and animals
3.	FAD binding domain in molybdopterin dehydrogenase	FAD_binding_5	Fungi and animals
4.	CO dehydrogenase flavoprotein C-terminal domain	CO_deh_flav_C	Fungi and animals
5.	Aldehyde oxidase and xanthine dehydrogenase, a/b hammerhead domain	Ald_Xan_dh_C	Bacteria, fungi and animals
6.	Molybdopterin-binding domain of aldehyde dehydrogenase	Ald_Xan_dh_C2	Bacteria, fungi and animals

4 Conclusions

Computational analysis of xanthine dehydrogenase sequences from different sources showed sequence-based similarities depending on their source organism. Thirteen glycine, three aspartic acid, three glutamic acid, three glutamine, three alanine, two arginine, one valine, one phenylalanine, one tyrosine, one proline, one serine and one asparagine residues were identically conserved in all analyzed species. This suggests that these conserved amino

Table 3 Motifs identified in XDH protein sequences of bacteria, fungi and animals

S.no.	Motif	Motif width	Motif present in number of sequences	Family	Source
1.	SSTQIPTEVQRLVARALGIPMHKVRVIKPRIGGGFGGKEDQ	41	10	Ald_Xan_dh_C2	Bacteria
2.	VHDSGRSINPAIDEGQVEGGIVQGIGWALSEELVWDKKGRL	41	10	Ald_Xan_dh_C2	Bacteria
3.	CYTHLPSNGAFRGFGGPQGMFAIEAAIDDAAAALGMDPLEIRKKNLYREG	50	10	Ald_Xan_dh_C2	Bacteria
4.	NQAGALVHVYQDGSVGLNHGATEMGQGADTKFTQVAAETFG	41	10	Ald_Xan_dh_C2	Bacteria
5.	DTLTGDWTCLRADIKMDVGRSINPAIDYGQIQGAFVQGMGLFTMEESLWL	50	10	Ald_Xan_dh_C2	Fungi
6.	VDGKHVITIEGIGNTKKPHPAQERVAKGNGSQCGFCTPGIVMSLYALLRN	50	10	Fer2_2	Fungi
7.	TALFLNQAGALVHIYHDGSVLVAHGGTEMGQGLHTKMTMIAAQ	43	10	Ald_Xan_dh_C2	Fungi
8.	KSHGSQCGFCTPGIVMSMYTLLRNQPEPTMEEIEDAFQGNLCRCTGYRPI	50	10	Fer2_2	Animal
9.	YFSYGVACSEVEIDCLTGDHKNLRTDIVMDVGSSLNPAIDIGQVEGAFVQ	50	10	Ald_Xan_dh_C2	Animal
10.	HVYTDGSVLLTHGGTEMGQGLHTKMVQVASRALKIPTSKIYISETSTNTV	50	10	Ald_Xan_dh_C2	Animal
11.	CGEGGCGACTVMISKYDRLQNKIVHFSANACLAPICSLHHVAVTTVEGIG	50	10	Fer2	Animal

acid residues have an important function in XDH sequences and in its evolution from lower organisms (bacteria) to higher organisms (animals). Six domains were identified but only two domains belonging to Ald_Xan_dh_C and Ald_Xan_dh_C2 family were found in all analyzed sequences of bacteria, fungi and animals. These domains were found to be responsible for the functional activity of XDH enzymes in different source organisms. This suggests that these domains were conserved during the evolution of lower organisms (bacteria) to higher organisms (animals) and their existence is important for the functional activity of this enzyme. Any single mutation in these domains can alter the functions of this enzyme. Some motifs which were unique for their group were also identified. In all species of bacteria, fungi, and animals, an average frequency of amino acid alanine was 9.24 % that was higher in comparison to other amino acids average frequency. This suggests that the amino acid glycine plays a very important role in the composition XDHs. Two major sequence clusters were obtained by phylogenetic analysis. One cluster was of animal and fungal origin, whereas the other one was of bacterial origin. This suggests that the fungi and animals are more closely related when compared to bacteria. This classification significantly contributes in the understanding of the evolutionary relationships between the species at molecular level.

Acknowledgments We are thankful to the Head of Forest Pathology Division department, Forest Research Institute, Dehradun, India for providing laboratory facilities and encouragement and to director of U.C.S.T., Dehradun for their cooperation during the study. We are grateful to Mr. Suresh Kumar Dwivedi S.G.S.I.C. Hata, Kushinagar for his kind support and necessary suggestions whenever we needed.

References

Amaya Y, Yamazaki K, Sato M, Noda K, Nishino T, Nishino T (1990) Proteolytic conversion of xanthine dehydrogenase from the NAD-dependent type to the O_2-dependent type. Amino acid sequence of rat liver xanthine dehydrogenase and identification of the cleavage sites of the enzyme protein during irreversible conversion by trypsin. J Biol Chem 265:14170–14175

Bailey TL, Elkan C (1995) Unsupervised learning of multiple motifs in biopolymers using expectation maximization. Mach Learn 21(51–80):33

Bray RC (1975) Molybdenum iron-sulfur flavin hydroxylases and related enzymes. In: The Enzymes, vol 12, 3rd edn, Academic Press, New York, pp 299–419

Della CE, Stirpe F (1972) The regulation of rat liver xanthine oxidase, involvement of thiol groups in the conversion of the enzyme activity from dehydrogenase (type D) into oxidase (type O) and purification of the enzyme. Biochem J 126:739–745

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 19: 32(5): 1792–1797

Hille R, Massey V (1981) Tight binding inhibitors of xanthine oxidase. Pharm Ther 14:249–263

Kumar S, Dudley J, Nei M, Tamura K (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform 9:299–306

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J et al (2012) The Pfam protein families database. Nucleic Acids Res 40:290–301

Sanders SA, Harrison R, Eisenthal R (1996) Oxidation of NADH catalysed by human xanthine oxidase: generation of superoxide anion. Biochem Soc Trans 24:13S

Stasolla C, Katahira R, Thorpe TA, Ashihara H (2003) Purine and pyrimidine nucleotide metabolism in higher plants. J Plant Physiol 160:1271–1295

Zrenner R, Stitt M, Sonnewald U, Boldt R (2006) Pyrimidine and purine biosynthesis and degradation in plant. Annu Rev Plant Biol 57:805–836