Phylogenetic analysis of uroporphyrinogen III synthase (UROS) gene

Abjal Pasha Shaik[1,$,*], Abbas H Alsaeed[1$] & Asma Sultana[2$]

[1]Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, King Saud University, Riyadh, Kingdom of Saudi Arabia; [2]Janaki Nagar Colony, No: 45, Tolichowki, Hyderabad, India; Abjal Pasha Shaik – Email: afzalshaik@gmail.com; Phone: +966544023766; Fax: +96614684995; *Corresponding author
$ All authors contribute equally

Abstract:
The uroporphyrinogen III synthase (UROS) enzyme (also known as hydroxymethylbilane hydrolyase) catalyzes the cyclization of hydroxymethylbilane to uroporphyrinogen III during heme biosynthesis. A deficiency of this enzyme is associated with the very rare Gunther's disease or congenital erythropoietic porphyria, an autosomal recessive inborn error of metabolism. The current study investigated the possible role of UROS (Homo sapiens [EC: 4.2.1.75; 265 aa; 1371 bp mRNA; Entrez Pubmed ref NP_000366.1, NM_000375.2]) in evolution by studying the phylogenetic relationship and divergence of this gene using computational methods. The UROS protein sequences from various taxa were retrieved from GenBank database and were compared using Clustal-W (multiple sequence alignment) with defaults and a first-pass phylogenetic tree was built using neighbor-joining method as in DELTA BLAST 2.2.27+ version. A total of 163 BLAST hits were found for the uroporphyrinogen III synthase query sequence and these hits showed putative conserved domain, HemD superfamily (as on 14th Nov 2012). We then narrowed down the search by manually deleting the proteins which were not UROS sequences and sequences belonging to phyla other than Chordata were deleted. A repeat phylogenetic analysis of 39 taxa was performed using PhyML and TreeDyn software to confirm that UROS is a highly conserved protein with approximately 85% conserved sequences in almost all chordate taxons emphasizing its importance in heme synthesis.

Key words: Uroporphyrinogen III synthase, Protein sequences, Phylogeny, Sequence alignment

Background:
Congenital erythropoietic porphyria (CEP) or Gunther's disease is a rare inherited autosomal recessive trait and is caused due to the deficiency of uroporphyrinogen III synthase (UROS), the fourth enzyme in heme biosynthesis. Only about 150 cases of CEP have been reported to date **[1-4].** CEP symptoms are heterogeneous, ranging from severe hemolytic anemia in utero to milder, later onset forms, with skin lesions due to cutaneous photosensitivity in adult life **[4].** The deficiency of functional UROS causes buildup of porphyrins to toxic levels in red blood cells. The excess porphyrins can then accumulate in the skin causing oversensitivity to sunlight **[4, 5].** UROS is the central point for the synthesis of heme, a complex organic molecule, is

a cyclic tetrapyrrole that contains a centrally chelated Fe and functions in the transport oxygen. The human UROS gene located on chromosome 10 (10q26.2; Cytogenetic Location: 10q25.2-q26.3; EC 4.2.1.75; 1371 bp mRNA; Entrez Pubmed ref NP_000366.1, NM_000375.2) and expressed as a 265 amino acid UROS enzyme catalyzes the cyclization of the linear tetrapyrrole, hydroxymethylbilane, to the macrocyclic uroporphyrinogen III which is eventually converted to heme. Most organisms have the ability to synthesize their own heme molecules by this largely conserved metabolic pathway **[4, 6, 7].** UROS enzyme is localized in the cytosol and plays a critical part in production of heme which is an essential prosthetic group in many cellular reactions in prokaryotes and eukaryotes.

BIOINFORMATION

The presence of this pathway across Bacteria, Archaea, and Eukarya suggests that heme performs a central function in the evolution of life. The acquisition by eukaryotes of the heme is one of the most important events in cellular evolution and any interference to its synthesis can thus have dire consequences to the survival of these organisms. Because of the inherent role of UROS in the synthesis of heme, and considering its role in many other inherent metabolic pathways in the cell, we aimed to elucidate if variance in this gene exists in various species during evolution by using a phylogenetic analysis of published protein sequences of these genes.
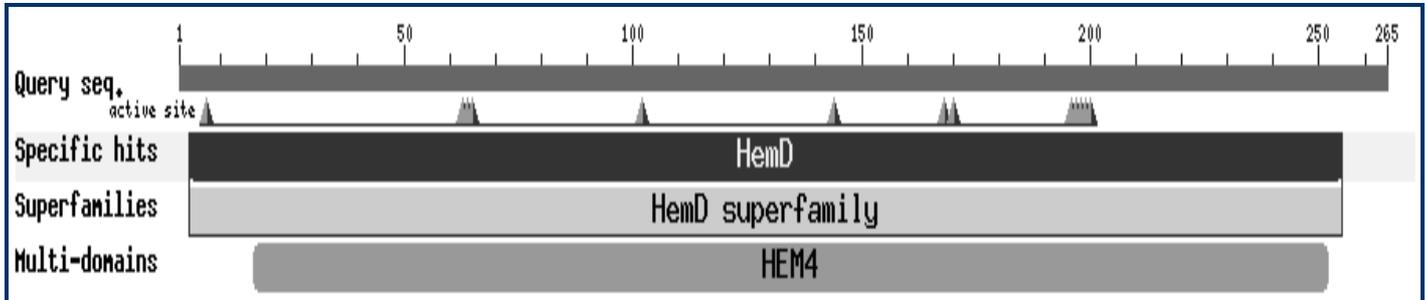

Figure 1: Putative sequence of UROS in super families; UROS (HemD) catalyzes the production of uroporphyrinogen-III, the fourth step in the biosynthesis of heme. This ubiquitous enzyme is present in eukaryotes, bacteria and archaea. Cd 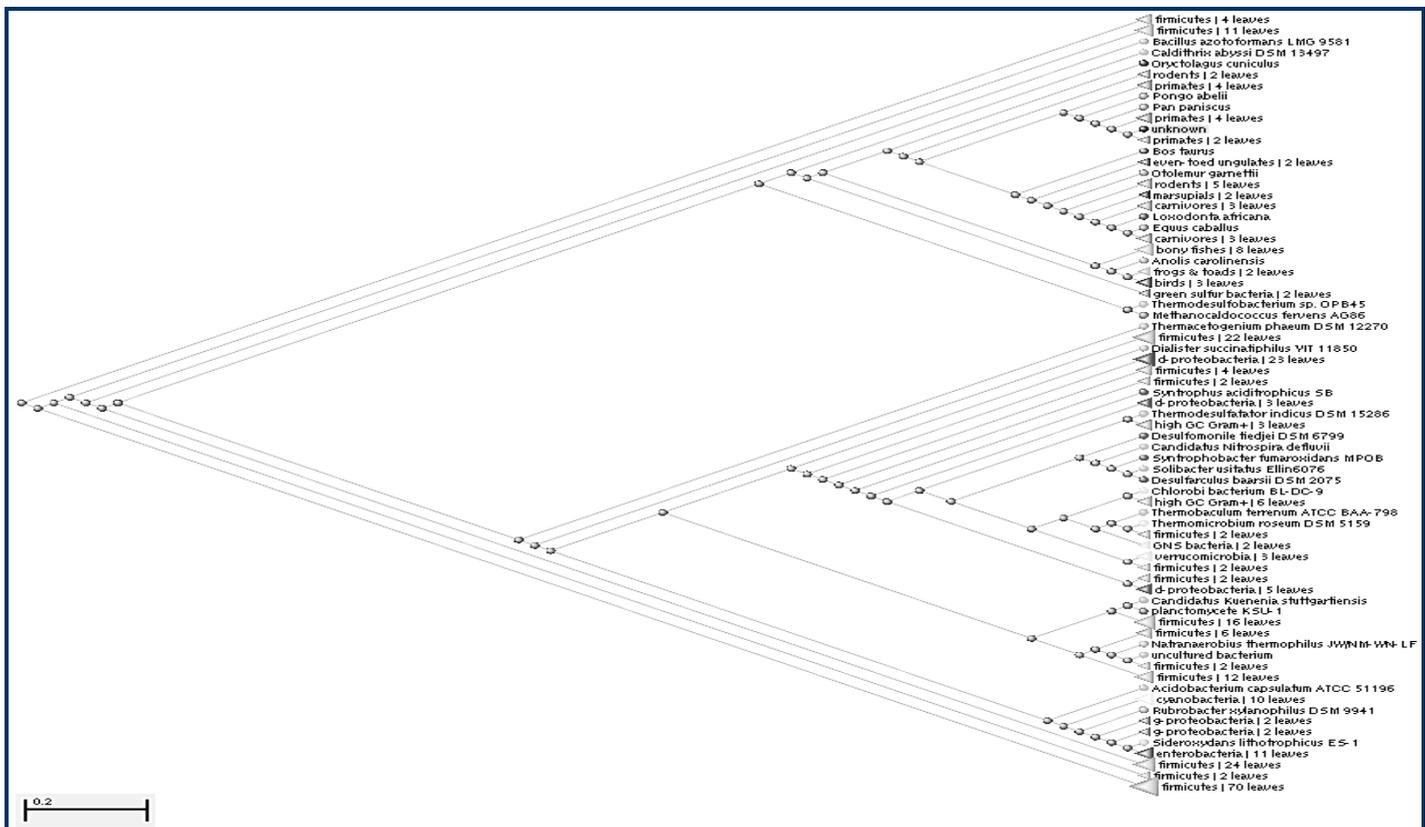Length: 239 Bit Score: 171.72 E-value: 5.15e-52; http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?RID=B6X0K5J901R&mode=all

Figure 2: First pass phylogenetic tree constructed by multiple alignment using BLAST pair wise alignments: Results presented using Taxonomic name [163 hits]

Methodology:
Data Set, Sequence Alignment and Construction of Phylogenetic Tree
The GenBank database **[8]** was queried to retrieve all available protein sequences of the UROS protein. These sequences were retrieved and saved in FASTA sequence format. These sequences were then aligned using Clustal W **[9]** algorithm using default parameters. The initial first-pass phylogenetic tree was constructed using Neighbour Joining **[10]** method (maximum sequence difference of 0.85) using Domain Enhanced Lookup Time Accelerated Basic Local Alignment Search Tool [DELTA BLAST] pairwise alignments between a query and the database sequences searched **[11]**. The evolutionary distance between two retrieved sequences modeled as expected fraction of amino acid substitutions per site given the fraction of mismatched amino acids in the aligned region was taken by the software using Grishin computation **[12]**. Using the results from DELTA BLAST, we created a first-pass phylogenetic tree after which we used a purpose-built computational phylogenetic method using Phylogeny.fr

software. Sequences were aligned with MUSCLE (v3.7) [13, 14] configured for highest accuracy (MUSCLE with default settings). After alignment, ambiguous regions (i.e. containing gaps and/or poorly aligned) were removed with Gblocks (v0.91b) using the following parameters: minimum length of a block after gap cleaning: 10; no gap positions were allowed in the final alignment; all segments with contiguous nonconserved positions bigger than 4 were rejected and 85% minimum number of sequences for a flank position [15-17]. The phylogenetic tree was reconstructed using the maximum likelihood method implemented in the PhyML program (v3.0 aLRT) [13, 18]. The Jones-Taylor-Thornton (JTT) substitution model was selected assuming an estimated proportion of invariant sites (of 0.197) and 4 gamma-distributed rate categories to account for rate heterogeneity across sites. The gamma shape parameter was estimated directly from the data (gamma=1.199). Reliability for internal branch was assessed using the aLRT test (SH-Like). The graphical representation and edition of the phylogenetic tree were performed with TreeDyn (v198.3) [19].

Figure 3: MUSCLE (3.7) multiple sequence alignment [CLUSTAL format] of short-listed sequences [39 taxons] Similar residues are colored as the most conserved one (according to BLOSUM62).

Results:
From the NCBI GenBank database, 163 sequences of UROS covering the putative conserved HemD domain were used to construct a first-pass phylogenetic tree. The tree however had some of the repetitive and unrelated sequences which were deleted. The short-listed sequences belonged to Porifera, Arthropoda, Echinodermata, Tunicates, Cephalochordates, Bacteria, Cnidaria, Fishes, Amphibia, Reptiles, Aves and

BIOINFORMATION

Mammalia. The accession information for these sequences is available in **Table 1 (see supplementary material)**. Analysis of the sequences revealed that there is a high degree of sequence similarity of UROS enzyme in many of the selected sequences used for the phylogeny reconstruction. Putative conserved domains were observed in many taxa at the HemD region **(Figure 1) [20]**. The actual alignment was detected with superfamily member pfam02602 (E-value: 3.02e-41). BLAST produced 163 hits **(Figure 2)**; these sequences were screened manually and only those related to the sequence in question (HPSE-1) from different taxa were retained for further analyses. This produced a total of 39 sequences from 31 taxa. Multiple sequence alignment results of these short-listed sequences is presented in **(Figure 3)**. Using the PhyML program a tree was constructed for these sequences, the results of which are presented in **(Figure 4)**.
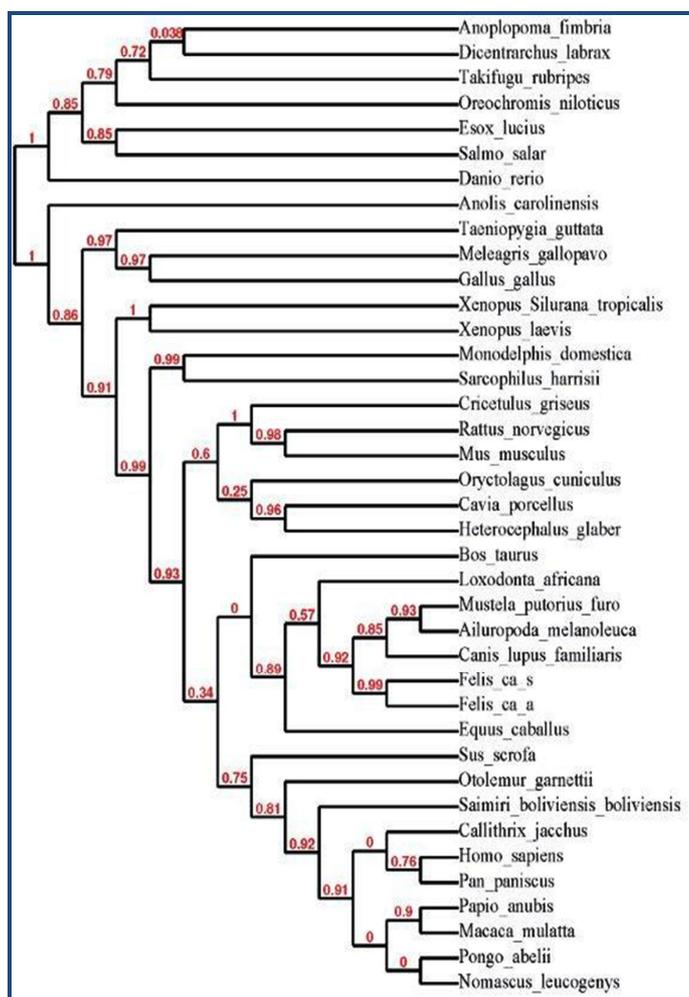

Figure 4: Phylogenetic tree using short-listed sequences [cladogram] built using the PhyML software [39 taxons].

Discussion:

The evolutionary relationship of UROS enzyme in various species, taxa and phyla was evaluated using computational phylogenetics to identify similar genes in the short-listed organisms. A mix of algorithms and programs were used to construct a phylogenetic tree **[21, 22]**; the neighbor-joining method was used to calculate genetic distance and ClustalW to create trees based on multiple sequences. The JTT matrix method was used to generate mutation data matrices from

protein sequences and the set sequences were clustered at the 85% identity level. In addition, the sequences were aligned, and the observed exchanges amino acids were tallied **[23]**. Following this, the final phylogenetic tree was constructed **[18]**. We have used MUSCLE (multiple sequence comparison by log-expectation) method to achieve the highest scores. Close to four benchmarks showed that MUSCLE achieved the highest ranking of any method available at the time of publication. Using Gblocks, poorly aligned positions and divergent regions of sequences could be eliminated – these positions may not be homologous or may have been saturated by multiple substitutions **[24, 25]**. In addition, this program helped to reduce the necessity of manually editing multiple alignments with very fast processing. The PhyML software was used since it was shown to be at least as accurate and slightly faster as other existing phylogeny programs using simulated data. The DELTA BLAST algorithm which uses a heuristic method to identify homologous sequences helped to produce high scoring sequence alignment between the query and database sequences. We have used BLAST as a first pass sequence alignment tool to narrow down the target and most relevant sequences.

The phylogenetic tree indicated that UROS protein is conserved and plays an important role in organismal evolution **(Figures 3 & 4)**. It is interesting to note that the conserved regions as shown in Homo sapiens are similar to those found in some other organisms that have this conserved gene. Presence of UROS in the major organisms indicates that it is crucial for the development of the physiology of cells. Its high conservation at certain domains indicates that its function is preserved. In conclusion, the evolutionary relationship of UROS gene was established based on the sequence alignment, conserved sequences and phylogenetic trees. The results of the published data on protein sequences of the above genes showed that the sequences are highly conserved especially at certain domains. Human sequences consistently clustered with their mammal orthologs within these genes clearly indicate the importance of these genes in evolution **[26]**. The phylogenetic reconstruction of the metabolic pathways of many organisms is one of the major goals of genomics. Reconstructions made through comparative genomics, and results from experiments on model systems help in understanding the biochemical diversity of life. Thus, analysis of phyletic patterns re-emphasizes the importance of certain metabolic enzymes in evolution.

References:
[1] Gross U, *J Inherit Metab Dis.* 2000 **23**: 641 [PMID: 11117426]
[2] Anderson KE *et al. The Molecular and Metabolic Bases of Inherited Disease.* CS Scriver, AL
[3] Beaudet, WS & Slyand D, *Br JHaematol.* 2002 **117**: 779 [PMID: 12060112]
[4] Bickers DR & Frank J, *Dermatology in general medicine.* 2003 pp 1435.
[5] Kauppinen R, *Porphyrias Lancet.* 2005 **365**: 241 [PMID: 15652607]
[6] http://ghr.nlm.nih.gov/gene/UROS

BIOINFORMATION

[7] van Dooren GG & Kennedy AT, *Antioxidants Redox signal.* DOI: 10.1089/ars.2012.4539 [PMID: 22320355]

[8] Benson DA *et al. Nucleic Acids Res.* 2011 **39:** D32 [PMID: 21071399]

[9] Thomson JD *et al. Nucleic Acids Res.* 1994 **22**: 4673 [PMID: 7984417]

[10] Saitou N & Nei M, *Mol Biol Evol.* 1987 **4**: 406 [PMID: 3447015]

[11] Boratyn GM *et al. Biol Direct.* 2012 **7:** 12 [PMID: 22510480]

[12] Grishin NV, *J Mol Evol.* 1995 **41**: 67 [PMID: 18345592]

[13] Dereeper A *et al. Nucleic Acids Res.* 2008 **36**: W465 [PMID: 18424797]

[14] Edgar RC, *Nucleic Acids Res.* 2004 **32**: 1792 [PMID: 15034147]

[15] Castresana J, *Mol Biol Evol.* 2000 **17**: 540 [PMID: 10742046]

[16] Guindon S & Gascuel O, *Syst Biol.* 2003 **52**: 696 [PMID: 14530136]

[17] Anisimova M & Gascuel O, *Syst Biol.* 2006 **55**: 539 [PMID: 16785212]

[18] Dereeper A *et al. BMC Evol Biol.* 2010 **10**: 8 [PMID: 20067610]

[19] Chevenet F *et al. BMC Bioinformatics.* 2006 **7**: 439 [PMID: 17032440]

[20] Marchler-Bauer A *et al. Nucleic Acids Res.* 2011 **39**: 225 [PMID: 21109532]

[21] Wu CH *et al. Nucleic Acids Res.* 2006 **34**: D187 [PMID: 16381842]

[22] Strait DS & Grine FE, *J Hum Evol.* 2004 **47**: 399 [PMID: 15566946]

[23] Jones DT *et al. Comput Appl Biosci.* 1992 **8:** 275 [PMID: 1633570]

[24] Talavera, G & Castresana J, *Syst Biol.* 2007 **56**: 564 [PMID: 17654362]

[25] Castresana J, *Mol Biol Evol.* 2000 **17**: 540 [PMID: 10742046]

[26] http://www.ncbi.nlm.nih.gov/books/NBK20266/ [PMID: 21089240]

BIOINFORMATION

Supplementary material:

Table 1: Accession details of short-listed uroporphyrinogen-III synthase (UROS) sequences

Description	Accession no:
Pongo abelii – predicted	XP_002821304.1
Macaca mulatta	NP_001253711.1
Papio anubis– predicted	XP_003904442.1
Callithrix jacchus– predicted	XP_002756746.1
Homo sapiens	NP_000366.1
Nomascus leucogenys– predicted	XP_003277827.1
Saimiri boliviensis boliviensis– predicted	XP_003922155.1
Pan paniscus– predicted	XP_003816334.1
Loxodonta africana– predicted	XP_003419757.1
Bos taurus	NP_001178298.1
Rattus norvegicus	NP_001012068.1
Meleagris gallopavo– predicted	XP_003208282.1
Otolemur garnettii– predicted	XP_003794158.1
Oryctolagus cuniculus– predicted	XP_002721286.1
Cavia porcellus– predicted	XP_003479664.1
Sus scrofa	NP_001231296.1
Cricetulus griseus– predicted	XP_003513922.1
Mus musculus	NP_033505.1
Gallus gallus– predicted	XP_423886.2
Canis lupus familiaris– predicted	XP_854192.1
Felis catus– predicted	XP_003994563.1
Ailuropoda melanoleuca– predicted	XP_002916164.1
Heterocephalus glaber	EHB18080.1
Equus caballus– predicted	XP_001490208.1
Taeniopygia guttata– predicted	XP_002193805.1
Xenopus (Silurana) tropicalis	NP_001107324.1
Xenopus laevis	NP_001087355.1
Monodelphis domestica– predicted	XP_001362178.1
Esox lucius	ACO13405.1
Mustela putorius furo	AES09336.1
Salmo salar	ACI66186.1
Anoplopoma fimbria	ACQ58390.1
Anolis carolinensis– predicted	XP_003218642.1
Oreochromis niloticus– predicted	XP_003441824.1
Sarcophilus harrisii– predicted	XP_003755055.1
Takifugu rubripes– predicted	XP_003961197.1
Danio rerio	AAO49476.1
Dicentrarchus labrax	CBN81562.1