

Molecular evolution of β Galactosidase in Thermophiles, Psychrophiles, Mesophiles, Plants and Mammals by *in silico* approach

Rani V. and Dev K.*

Dept. of Biotechnology, Shoolini University of Biotechnology and Management Sciences, Solan, Himachal Pradesh, India
kdb_nih@yahoo.com

Available online at: www.isca.in, www.isca.me

Received 14th September 2014, revised 22st September 2015, accepted 19th November 2015

Abstract

To understand adaptation and the evolution at molecular level, β -galactosidase was studied among thermophiles, psychrophiles, mesophiles, plants and mammals. Conserved domain analysis revealed that β -galactosidase belongs to glycosyl hydrolase family. However, phylogenetic analysis showed higher degree of divergence among bacteria while highly conserved in mammals and plants except *Arabidopsis thaliana*. 3D modeled structures were studied for interaction with lactose, ONPG, PNPG, glucose, galactose, and ONP. Lactose showed tight binding to all the β -galactosidase except in *A. psychrolactophilus*, where maximum interaction was observed with ONPG. Galactose, glucose and ONP exhibited competitive inhibition for lactose, ONPG and PNPG in *H. sapiens*, *A. psychrolactophilus*, and *T. african*, while un-competitive inhibition for *A. thaliana* and *E. coli*.

Keywords Extremophilic microorganisms, Conserved domain, Glycosyl hydrolase family, Ramachandran plot, Modeled proteins, Instability index.

Introduction

Extremophilic microorganisms live under forbidding conditions and their discovery points out the unique adaptability of primitive life-forms. Grouping of these microorganisms are according to their optimal growth conditions in which they exist such as acidophiles (acidic pH), alkaliphiles (alkaline pH), piezophiles (high pressures), halophiles (high salt concentrations), psychrophiles (temperature below 20°C), thermophiles (temperature between 45-80°C), and hyperthermophiles (temperature above 80°C)¹. These microorganisms produce biocatalysts that are functional and adapted under extreme conditions. Study of these biocatalysts has given depth knowledge of stability factors in these microorganisms². Thermophilic proteins showed increased number of ion pairs, strong hydrophobic interior³ and decreased number of cavities (active sites) as compared to their mesophilic counterparts which contributes for stabilizing thermophilic proteins⁴. Archeal organisms acquired thermostability by substituting amino acids on the surface of the proteins like uncharged polar amino acids (glycine) with glutamic acid and lysine and non polar amino acids with isoleucine⁵. Bioinformatics has open new avenues into protein sequence and structural features and provide deeper insights into molecular evolution, protein engineering and designing the enzyme inhibitors and activators. Instead of the characterization of a protein's binding abilities, *In Silico*, protein engineering has applications in bioanalysis and biotechnology. One of the ubiquitous and very well studied enzyme is β -galactosidase (EC 3.2.1.23), a glycoside hydrolase that hydrolyses β -glycosidic bond between two or more carbohydrates or a carbohydrate and

another moiety. Glycoside hydrolases are remarkably a diverse group of enzymes which degrades a huge variety of naturally occurring carbohydrates and glycoconjugates. Within this larger group, the β -galactosidases are members of four families of glycosyl hydrolases namely 1, 2, 35, and 42⁶. Bacterial β -galactosidases are the most extensively studied enzymes among microbial β -galactosidases. β -galactosidase from bacteria, fungi and plants showed homology in the protein sequences and evolutionary relatedness among species at molecular level⁷.

β -galactosidase is widely used in the food technology, mainly in the dairy industry and utilized in the development of new products with hydrolysed lactose which are suitable for lactose intolerant people, for improving non technological properties of non fermented milk products and for removing lactose from whey⁸⁻¹⁰. Industrially this enzyme is used in the production of galacto-oligosaccharides which can be used in variety of foods because they act as growth promoting substrates for intestinal microflora^{10,11}.

There is a long and distinguished history of *E. coli* β -galactosidase (LacZ) in Molecular Biology and Biotechnology. Amino acid sequence of β -galactosidase in *Escherichia coli* was determined¹² and substitution of aspartic acid to glycine 794 caused a dramatic increase in the activity of β -galactosidase when lactose was used as substrate¹³.

The objective of this study was to analyze the divergence of β -galactosidase among different microorganisms (thermophiles, mesophiles and psychrophiles) plants and mammals by *in silico* approach. We conducted an extensive analysis of

comparison of conserved domain analysis, structural modeling and docking, prediction of substrate/product binding and their affinity and amino acid composition.

Materials and Methods

Multiple sequence alignment (amino acid) and conservation: Protein sequence of β -galactosidase for *Escherichia coli* was retrieved from NCBI database¹⁴ and homologous protein sequences were selected by performing PSI-BLAST¹⁵. *Escherichia coli* was taken as a template because it is an very well studied organism among the bacteria. Based on amino acid sequence homology, we selected thermophiles (nine), psychrophiles (six), mesophiles (ten), plants (seventeen) and mammals (fifteen) and amino acid sequence conservation among these were studied by using PRALINE program¹⁶. Highly conserved residues were predicted by multiple sequence alignment which might contribute to the catalytic efficiency of β -galactosidase in these different organisms.

Identification of conserved domains and residues of β -galactosidase from thermophiles, mesophiles, psychrophiles, plants and mammals and their evolutionary relationship: Domains analyses were performed by Pfam database¹⁷ and alignment as well as representation of domains in an order was performed by using DOG 2.0 software. Phylogenetic tree was constructed by Phylip-3.68 software¹⁸. Protein maximum likelihood program of phylip was used to construct the tree which uses the Dayhoff probability model of change between amino acids.

Secondary structure determination, molecular modelling and docking of β -galactosidase and its substrates (lactose and ONPG), structural analogue (PNPG) and products (Glucose, ONP and Galactose): Secondary structure was determined by Phyre2 online tool¹⁹ and percentage of helix, and beta sheets were predicted in the selected five organisms (*Thermosipho africans*, *Arthrobacter psychrolactophilus*, *Escherichia coli*, *Homo sapiens* and *Arabidopsis thaliana*). Three dimensional (3D) structures for *Escherichia coli* and *Homo sapiens* were retrieved from protein data bank²⁰ (PDB) database while others (*Thermosipho africans*, *Arthrobacter psychrolactophilus* and *Arabidopsis thaliana*) were modeled by Swissmodel tool²¹. Structures were then verified using Ramachandran plot by Rampage online tool²² and docking was performed to study the interactions between β -galactosidase and substrates like lactose, its structural analogues (ONPG, PNPG) and reaction products (ONP, Glucose and Galactose) by Hex software²³. Molecular visualization and structural interactions among β -galactosidase and different ligands were studied by Pymol software²⁴.

Results and Discussion

Comparison of amino acid sequence of various β -galactosidases and conserved domains analysis: β

galactosidase protein sequences were retrieved from NCBI database¹⁴ and homology was predicted by Praline tool¹⁶. Y344 was conserved in mesophiles, E570 and G609 were conserved in thermophiles, G70, G155, G926, F179, W189, W240, W245, G209, Y353, T433, V456, A487, N519, M527 were conserved in psychrophiles. In general, plants and mammals were found to be more conserved as compared to the bacteria. Mammals were found to be the most conserved. These conserved residues might contribute for the catalytic efficiency of β -galactosidase in these organisms and need further exploration by site directed mutagenesis and catalytic activity. E141 and E312 were predicted as putative catalytic residues in *Thermus thermophilus* A4²⁵ which are not found conserved in the multiple sequence alignment of thermophiles. Nucleophilic displacement of catalytic residues, E537 and proton abstraction by E461 resulted in the formation of allolactose in *Escherichia coli*²⁶. E200 and E299 were reported to be involved in the catalysis of β galactosidase in *Penicillium* sp.²⁷. Three aromatic residues, W240, W243 and Y455 of β galactosidase showed substrate specificity in *Streptococcus pneumoniae*²⁸.

Conserved domains were predicted by Pfam database¹⁷ and this analysis placed all the organisms under glycosyl hydrolase family as shown in Figure-1. Three different conservation patterns were observed among all the bacteria analysed: Glyco_hydro1 (sky blue) domain was conserved throughout evolution in *Pyrococcus furiosus*, *Thermoplasma acidophilum*, *Salfobolus sulfatricus*, *Thermoplasma volcanium* (thermophiles), *Edwardsiella tarda* and *Nakamurella multipartita* (mesophiles). Among thermophiles, this has been split at the N-terminus to accommodate large insertions of 9-20 unconcerned residues, while no splitting was observed among the mesophiles. Few residues (19-53 residues) were not conserved at the N terminus end in *Thermoplasma acidophilum*, *Salfobolus sulfatricus* and *Thermoplasma volcanium*.

Glyco_hydro42 (pink) domain was conserved in *Pyrococcus abyssii*, *Thermoanaerobacterium thermosaccharolyticum* (thermophiles), *Arthrobacter* sp.FB24, *Planococcus* sp.L4 (psychrophiles), *Clostridium lentocellum*, *Klebsiella pneumoniae*, *Clostridium cellulovorance* and *Beutenbergia cavernae* (mesophiles). This domain is followed by Glyco_hydro42M domain (yellow), and Glyco_hydro42C domain (black), except *Pyrococcus abyssii* which is lacking Glyco_hydro42C domain. Few residues (31) were unconserved at the C- terminus region in *Beutenbergia cavernae*.

Glyco_hydro2N domain (green) was conserved at N- terminus region in *Thermosipho africans*, *Thermoanaerobacter ethanolicus*, *Roseflexus castenholzii* (thermophiles), *Pseudoalteromonas haploplanktitis*, *Psychromonas marina*, *Maribacter* sp., *Arthrobacter psychrolactophilus* (psychrophiles), *Citrobacter youngae*, *Vibrio orientilis*, *Vibrio splendidus* and *Escherichia coli* (mesophiles). This domain was followed by Glyco_hydro2 domain (red), Glyco_hydro2C domain (Blue) and Bgal_smallN domain (purple).

Thermoanaerobacter ethanolicus was lacking Bgal_smallN domain. A gap of unconserved residues (268-344) was found in between Glyco_hydro2C (blue) and Bgal_smallN domain (purple). Bgal_smallN domain (purple) was splitted by non-conserved residues (13) in *Maribacter sp.* (psychrophile) and this domain was followed by small region (2-94) of unconserved residues. Small gap of unconserved residues (1-50) was found at the N terminus region in all these organisms except *Thermoanaerobacter ethanolicus*.

Overall, there is high degree of similarity amongst plants and mammals. Glyco_hydro35 domain was conserved in all mammals and plants towards the N- terminus while in plants, Gal_lectin was present at the C- terminus in *Brassica oleracea*, *Mangifera indica*, *Oryza sativa*, *Persea americana*, *Petunia X hybrida*, *Prunus persica*, *Prunus salicina*, *Pyrus pyrifolia*, *Ricinus communis* and *Solanum lycopersicum*, except *A. thaliana*, which showed similarity with bacteria, having glyco_hydro2N, glyco_hydro2, glyco_hydro2C and Bgal_smallN subdomains. *Arabidopsis thaliana* showed conserved domains pattern very much similar to *Roseflexus castenholzii* (thermophile), *Pseudomonas haploplanktis* (psychrophile) and *Vibrio orientalis* (mesophile). Glycoside hydrolases are organized into glycoside hydrolase families (GHFs) and β galactosidases are members of four families in glycoside hydrolase namely: 1, 2, 35, and 42. Most genes encoding GHF 42 enzymes are from prokaryotes (Shipkowski and Brenchley, 2006). Four conserved motifs were found in the β galactosidase of *Arabidopsis thaliana* (GenBank Accession number: NP_001154292), RLGPFQAEWNHGGLPYWL, LFASQGGPILGQIENEYNA, WAANLVESMNLGIPWV MCKQ and DAPGNLINACNGRHC¹⁷. *Thermus thermophilus* A4. β -galactosidase (A4-beta-Gal), was found thermostable and belongs to the glycoside hydrolase family 42 (GH-42). Its crystal structure was determined as free and galactose-bound form showing 1.6 Å and 2.2 Å resolution, respectively²⁴. Crystal structure of β galactosidases was reported in *Penicillium sp.* and its primary structure revealed that β galactosidases belongs to glycoside hydrolase family 35²⁶.

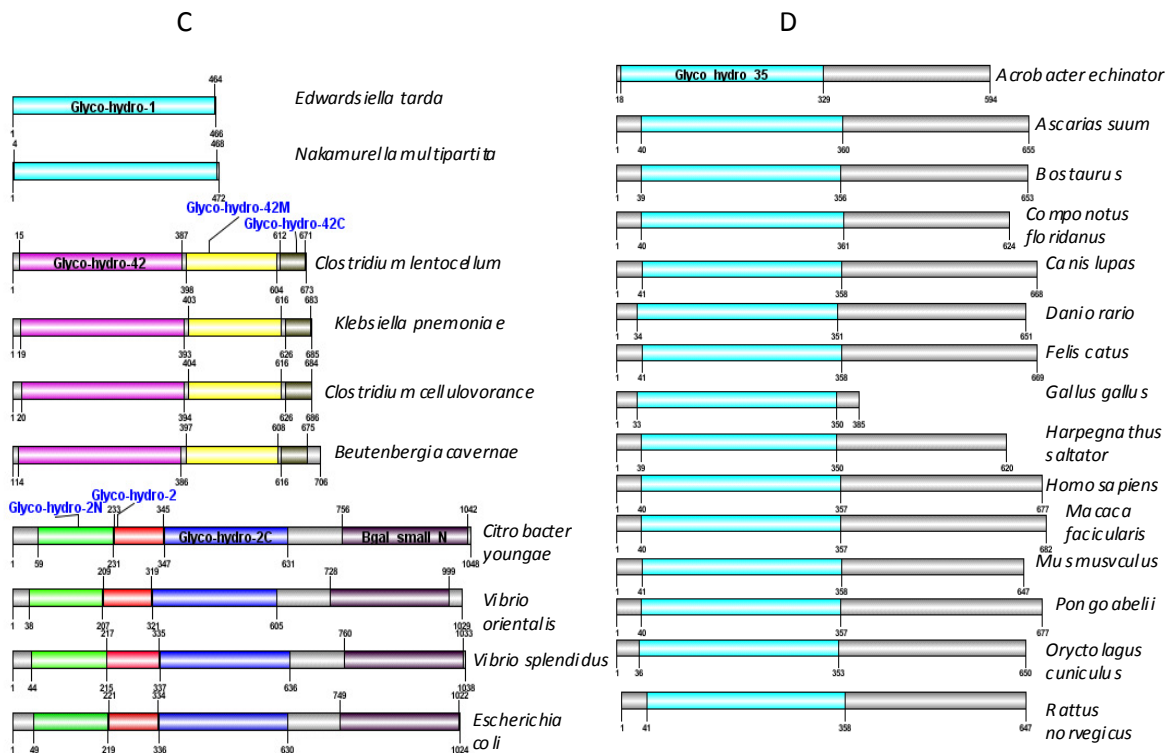
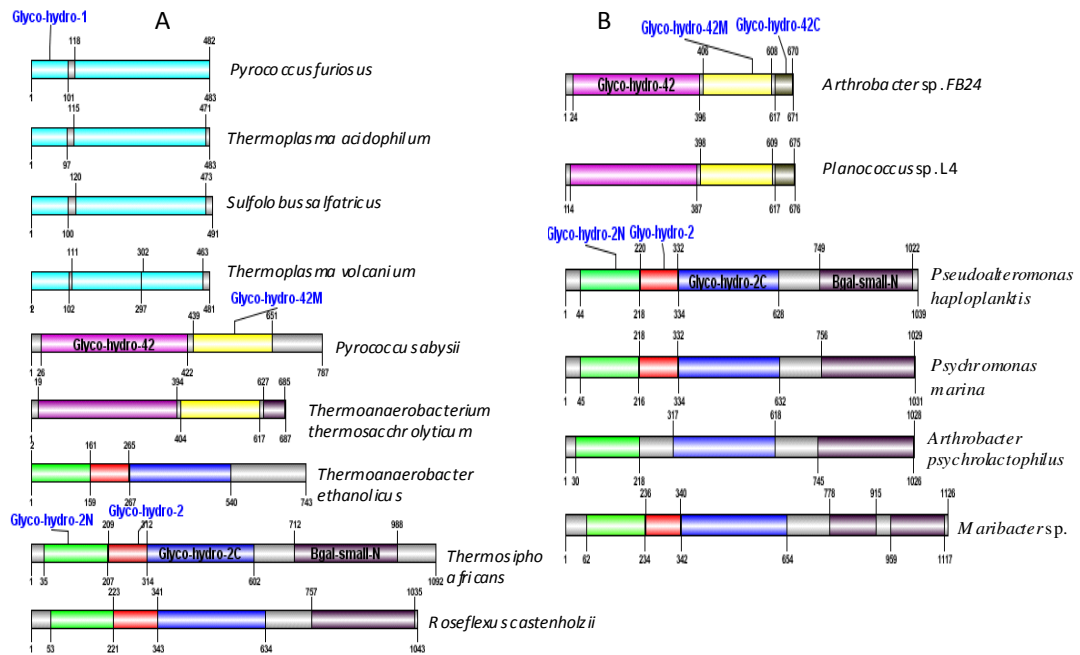
Secondary structure determination, homology modeling and validation: Secondary structure determination revealed that highest percentage of α helix was found in *Homo sapiens* (22%) followed by *Escherichia coli* (14%), *Arabidopsis thaliana* (13%) *Arthrobacter psychrolactophilus* and *Thermosipho africans* (12%) and beta sheets were found to be 43% in *Thermosipho africans*, 41% in *Escherichia coli*, 41% in *Arthrobacter psychrolactophilus*, 40% in *Arabidopsis thaliana* and 24% in *Homo sapiens*. 3D structures of β -galactosidase of *Escherichia coli* (mesophile) and *Homo sapiens* (mammal) were retrieved from PDB database having PDB i.d, 1BGL (*Escherichia coli*) and 3THC (*Homo sapiens*). *Thermosipho africans* (thermophile), *Arthrobacter psychrolactophilus* (psychrophile) and *Arabidopsis thaliana* (plant) were modeled by using Swissmodel tool²⁹.

For modeling their structures, the template used were 1yq2 (β galactosidase from *Arthrobacter sp.* C2-2) for *Arthrobacter psychrolactophilus* having 83% sequence identity, 3lpf (β glucuronidase from *E. coli*) for *Arabidopsis thaliana* having 19% sequence identity and 3bga (β galactosidase from *Bacteroides thetaiotaomicron* VPI-5482) for *Thermosipho africans* having 36% sequence identity. Validation of the modeled structures were performed by Ramachandran plot, using Rampage online tool²² and more than 90% residues were found in the allowed regions. The modeled structures were submitted in Protein model database²⁹ and assigned the accession no. PM0078235 for *Thermosipho africans*, PM0078237 for *Arthrobacter psychrolactophilus* and PM0078243 for *Arabidopsis thaliana*.

In silico docking studies of active site: *In silico* studies were designed to predict binding mode of substrates and reaction products. Structure of ligands like Lactose, ONPG (ortho-Nitrophenyl- β -galactoside), PNPG (para-Nitrophenyl- β -galactoside), ONP (ortho nitro phenol), glucose and galactose were retrieved from Chemspider database in .mol format and .mol files were then converted into .pdb files by Openbabel software. Potential binding sites in β -galactosidase were predicted by Q site finder tool which works by binding hydrophobic (CH₃) probes to the protein³¹ and docking was performed by Hex software²².

Docking analysis predicted that natural substrate (lactose) showed maximum interaction (Etotal = -217.50) with β -galactosidase in *Arabidopsis thaliana* followed by *Escherichia coli* (Etotal = -196.35), *Homo sapiens* (Etotal = -165.32) and *Thermosipho africans* (Etotal = -141.49) except *Arthrobacter psychrolactophilus* which showed maximum interaction (Etotal = -93.91) with ONPG.

Most of the interacting residues are present on the surface of the β -galactosidase protein in *Thermosipho africans* and *Homo sapiens* while all the interacting residues are not present on the surface (they are buried inside the protein surface) in case of *Arthrobacter psychrolactophilus*, *Escherichia coli* and *Arabidopsis thaliana*. Ligands bound in close vicinity with β -galactosidase in *Thermosipho africans*, *Escherichia coli* and *Homo sapiens* while they are scattered on the surface of the β -galactosidase in *Arthrobacter psychrolactophilus* and *Arabidopsis thaliana*. Further, docking studies revealed that galactose and ONP were competitive inhibitors for ONPG, PNPG and lactose in *Thermosipho africans*, glucose for ONPG in *Arthrobacter psychrolactophilus*, ONP for ONPG, PNPG and lactose in *Homo sapiens*. Glucose, galactose and ONP were predicted as uncompetitive inhibitors for *Escherichia coli* and *Arabidopsis thaliana*. Binding affinity of β galactosidase enzyme towards the three different substrates (lactose, ONPG and PNPG) was more for psychrophilic enzyme when compared with its mesophilic and thermophilic counterparts³².



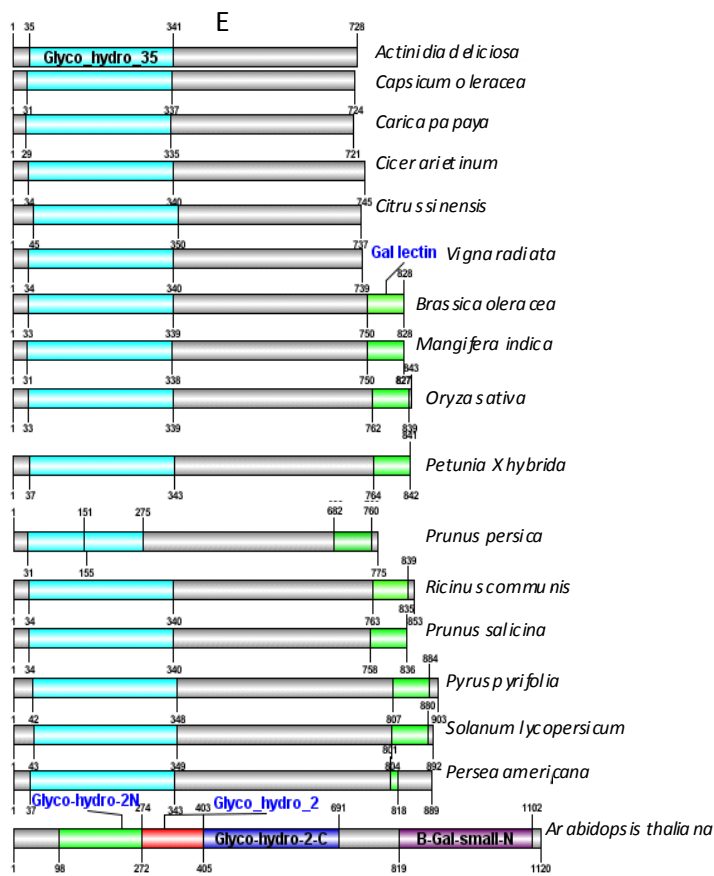


Figure-1

Conserved Domain Analysis: Amino acid sequences of β -galactosidase in thermophiles, psychrophiles, mesophiles, mammals and plants were retrieved from NCBI database. Conserved domains in these organisms (figure1A (Thermophiles), B (Psychrophiles), C (Mesophiles), D (Mammals) and E (Plants)) were analyzed by using Pfam database

Table-1

Total interaction energy (Etotal) of β -galactosidase with substrates and reaction products

	Substrates and Products	<i>Thermosiphon africans</i>	<i>Arthrobacter psychrolactophilus</i>	<i>Escherichia coli</i>	<i>Homo sapiens</i>	<i>Arabidopsis thaliana</i>
Substrates	Lactose	-141.49	-86.69	-196.35	-165.32	-217.5
	ONPG	-134.46	-93.91	-179.64	-159.88	-188.24
	PNPG	-110.25	-50.95	-179.31	-144.8	-165.94
Products	Galactose	-85.18	-68.36	-130.34	-166.52	-146.31
	ONP	-78.33	-87.37	-137	-115.37	-129.97
	Glucose	-81.97	-86.65	-137	-105.48	-145.73

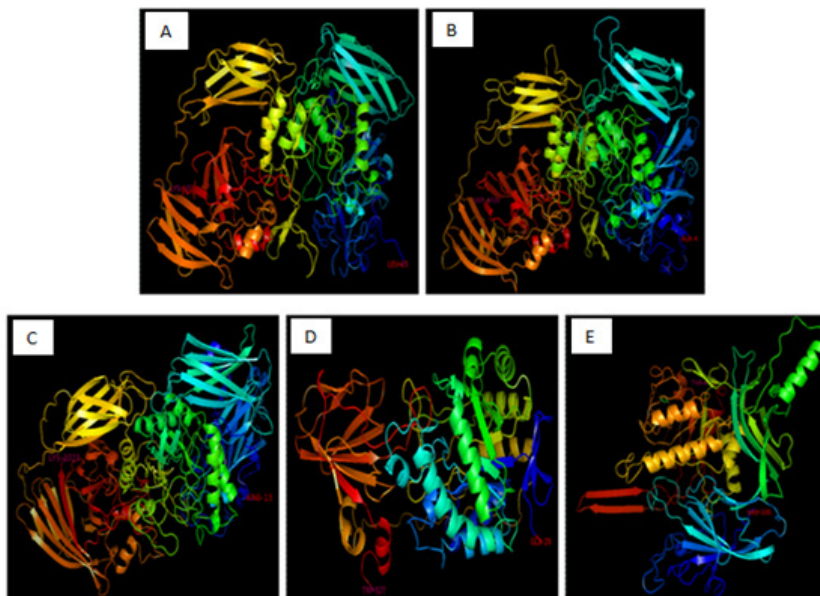


Figure-2

3D Modelled structures. Modelled structures of β -galactosidase of *Thermosipho africans* (A), *Arthrobacter psychrolactophilus* (B), *Escherichia coli* (C), *Homo sapiens* (D) and *Arabidopsis thaliana* (E) by Swissmodel tool (<http://swissmodel.expasy.org/>). Structures of β -galactosidase in *Escherichia coli* and *Homo sapiens* were retrieved from pdb database having pdb id 1BGL and 3THC while the modeled structures for *Thermosipho africans*, *Arthrobacter psychrolactophilus* and *Arabidopsis thaliana* were submitted in protein model database

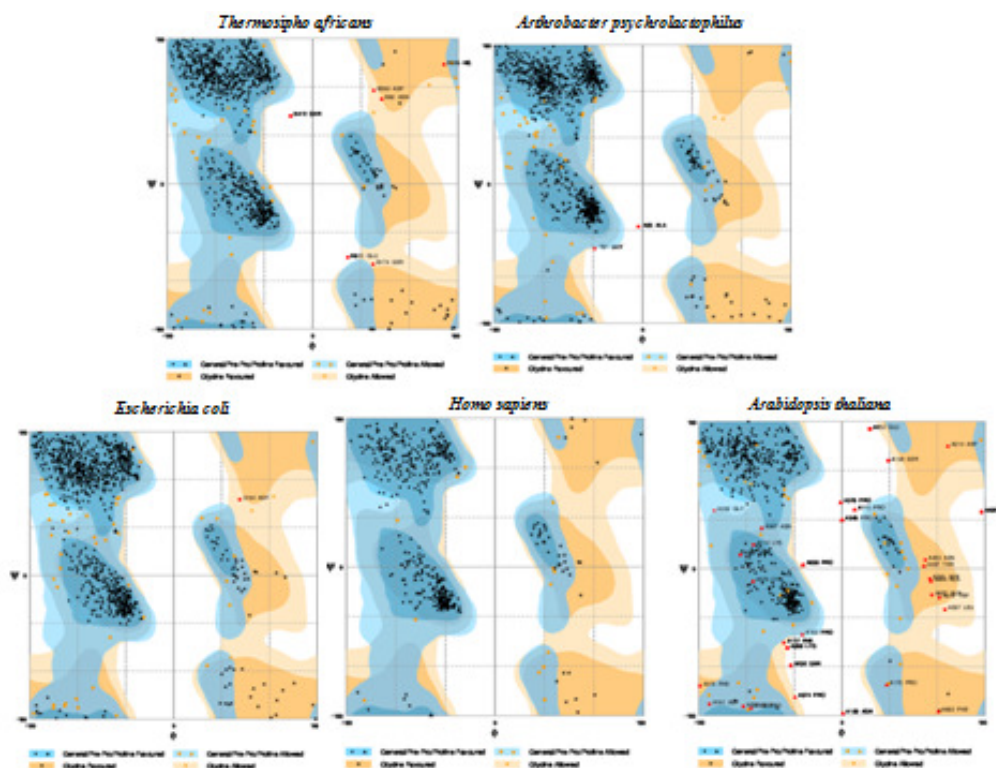


Figure-3

Ramachandran Plots. Ramachandran plots in *Thermosipho africans* (A), *Arthrobacter psychrolactophilus* (B), *Escherichia coli* (C), *Homo sapiens* (D) and *Arabidopsis thaliana* (E) plotted by Rampage online tool

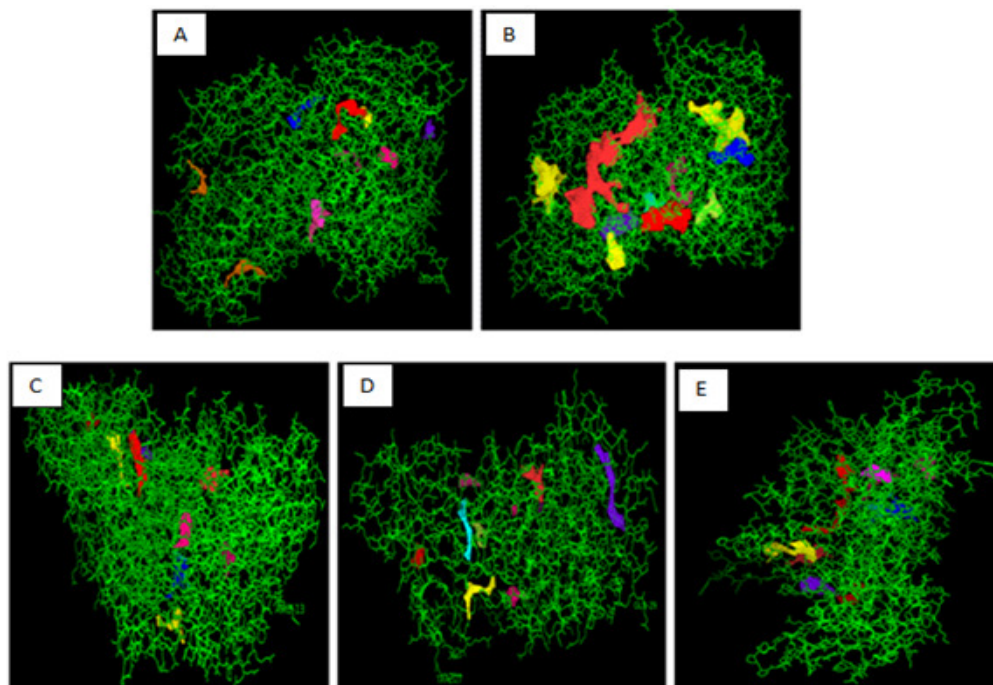


Figure-4

Predicted substrates/products binding sites. Potential active binding sites (shown in different colors) of β -galactosidase in *Thermosipho africans* (A), *Arthrobacter psychrolactophilus* (B), *Escherichia coli* (C), *Homo sapiens* (D) and *Arabidopsis thaliana* (E), predicted by Q-Site Finder tool

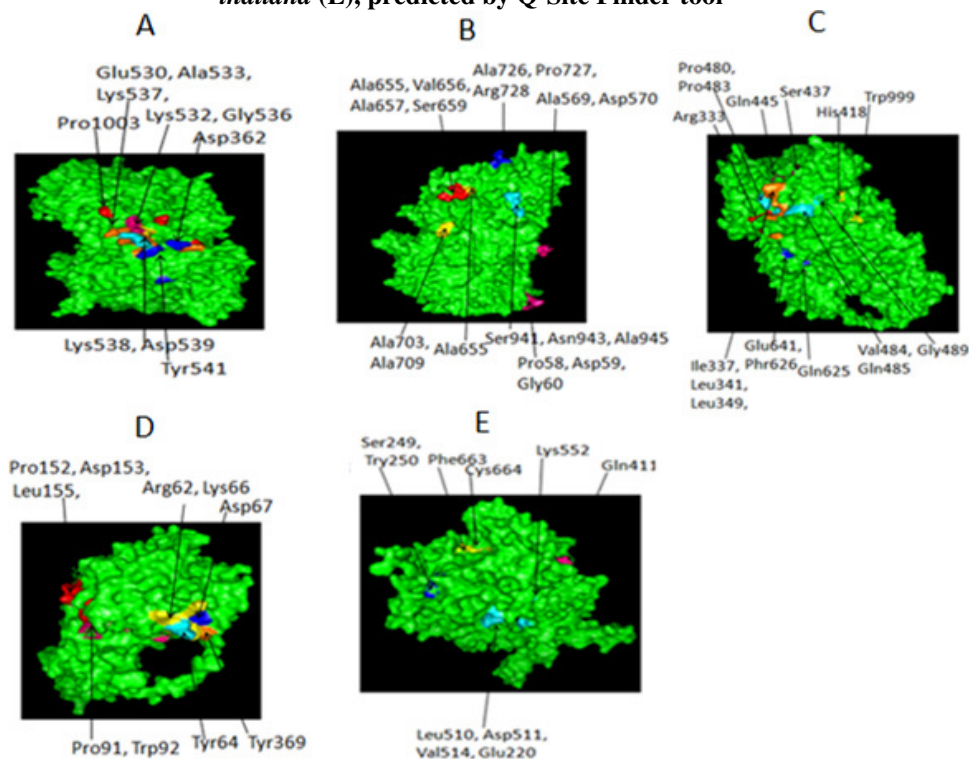


Figure-5

Predicted docking sites for substrates and products of different β -galactosidase. Interacting residues of β -galactosidase with lactose (sky blue), ONPG (yellow), PNPG (royal blue), ONP (orange), galactose (hot pink) and glucose (red) after docking studies in *Thermosipho africans* (A), *Arthrobacter psychrolactophilus* (B), *Escherichia coli* (C), *Homo sapiens* (D) and *Arabidopsis thaliana* (E)

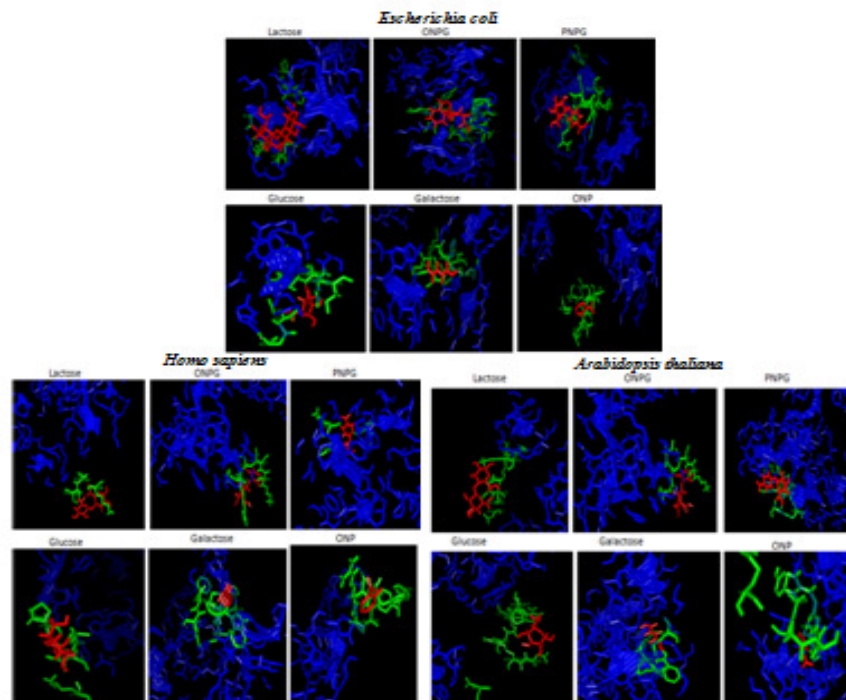


Figure-6

Figure showing interacting residues (green), ligand (red) and active site (blue) in *Thermosipho africans*, *Arthrobacter psychrolactophilus*, *Escherichia coli*, *Homo sapiens* and *Arabidopsis thaliana*

Maximum interacting residues belongs to glyco_hydro_2C domain in *Thermosipho africans* (glutamic acid at 530, alanine at 533, lysine at 537, lysine at 532, glycine at 536, asparatic acid at 362, lysine at 538, aspartic acid at 539 and tyrosine at position 541), *Arthrobacter psychrolactophilus* (alanine at 569 and aspartic acid at position 570), *Escherichia coli* (isoleucine at 337, leucine at 341 and 349, histidine at 418, serine at 437, proline at 480 and 483, valine at 484, glutamic acid at 485 and phenyl alanine at position 626) and *Arabidopsis thaliana* (glutamic acid at 411, leucine at 510, aspartic acid at 511, valine et 514, phenyl alanine at 663 and cysteine at position 664) while glyco_hydro_35 domain in *Homo sapiens* (arginine at 62, tyrosine at 64, lysine at 66, aspartic acid at 67, proline at 91 and 152, tryptophan at 92, aspartic acid at 153 and leucine at position 155). Three residues (proline at 58, aspartic acid at 59 and glycine at position 60) belongs to glyco_hydro_2N domain in *Arthrobacter psychrolactophilus*. One interacting residue belongs to B_gal_small_N domain in *Arthrobacter psychrolactophilus* (proline at position 1003) and *Escherichia coli* (tryptophan at position 999). Three residues (glutamic acid at 220, tyrosine at 250 and serine at position 249) belongs to glyco_hydro_2N domain in *Arabidopsis thaliana* and few residues lies in the unconserved regions in *Arthrobacter psychrolactophilus* (alanine at 655, 657, 703, 709 and 726, valine at 656, serine at 659, proline at 727 and arginine at position 728), *Thermosipho africans* (proline at position 1003) and *Escherichia coli* (glutamic acid at position 641). Glyco_hydro_2C domain was found interacting in *Thermosipho africans*, *Arthrobacter psychrolactophilus*, *Escherichia coli* and

Arabidopsis thaliana except *Homo sapiens* which was having only glyco_hydro_35 domain. Glyco_hydro_2N domain was also found interacting in case of *Arthrobacter psychrolactophilus* and *Arabidopsis thaliana*. B_gal_small_N domain was also found interacting in *Arthrobacter psychrolactophilus* and *Escherichia coli*. This can be concluded that glyco_hydro_2C domain might play important role in the catalytic activity of β -galactosidase in *Thermosipho africans*, *Arthrobacter psychrolactophilus*, *Escherichia coli* and *Arabidopsis thaliana* while glyco_hydro_35 domain in *Homo sapiens*, along with glyco_hydro_2N domain in *Arthrobacter psychrolactophilus* and *Arabidopsis thaliana* and B_gal_small_N domain in *Arthrobacter psychrolactophilus* and *Escherichia coli*.

Phylogenetic analysis: A β -galactosidase hypothetical evolutionary cladogram was constructed using thermophiles, mesophiles, psychrophiles, plants and mammals. Infile (input) in Phylip format was constructed by Clustalw program³³ and unrooted tree was constructed by proml program of Phylip software. Phylogenetic analysis showed higher degree of sequence divergence amongst bacterial β -galactosidase, while plants and mammals revealed high degree of sequence conservation except *Arabidopsis thaliana*, which evolved along with bacteria. β -galactosidase of mammals evolved along with thermophilic fungi (*Talaromyces emersonii*, *Aspergillus nidulans*, *Talaromyces stipitatus* and *Aspergillus fumigatus*), thermophilic bacteria (*Pyrococcus abyssi*) and some model organisms (*Neurospora crassa*, *Drosophila melanogaster* and

Xenopus laevis). In plants β -galactosidase is divergent in *Actinidia deliciosa* and *Carica papaya* as compared to other selected plants while one exception is also there for *Arabidopsis thaliana* which is evolving together with bacteria. Five model organisms were also included to study the molecular evolution of β -galactosidase. *Caenorhabditis elegans* and *Saccharomyces cerevisiae* evolved together with bacteria while *Neurospora crassa* evolved with thermophilic fungi and the remaining two, *Drosophila melanogaster* and *Xenopus laevis* evolved with mammals. It is very interesting to note that thermophilic bacteria are the most primitive and rapidly evolving along with psychrophiles and mesophiles.

Multiple sequence analysis and domain analysis were further supported by this phylogenetic analysis. It is reported that *A. thaliana* β galactosidase has a close relationship with some plants' β galactosidases from different families such as Malvaceae, Solanaceae, and Poaceae³⁴.

In relation to the domains, thermophilic organisms having gene i.d. 13541516 (*Thermoplasma volcanium*), 81118 (*Salfolobus sulfatricus*), 1608231 (*Thermoplasma acidophilum*), 18976728 (*Pyrococcus furiosus*) and mesophiles having gene i.d. 2585536 (*Nakamurella multipartita*) and 2910917 (*Edwardsilla tarda*) have glyco_hydro_1 domain. These organisms were followed by gene i.d. 3874645 (*Caenorhabditis elegans*) having glyco_hydro_35 domain followed by beta dom_4 5. Organisms having gene i.d. 2838363 (*Citrobacter youngae* - mesophile), 2170776 (*Thermosiphon africanus* - thermophile), 3056654 (*Maribacter* sp. - psychrophile), 8438722 (*Vibrio splendidus* - mesophile), 4589389 (*Psychromonas marina* - psychrophile), 114939 (*Escherichia coli*), 4079639 (*Pseudoalteromonas haploplanktis* - psychrophile), 3341859 (*Arabidopsis thaliana* - plant), 1562337 (*Roseflexus castenholzii* - thermophile), 261252 (*Vibrio orientalis* - mesophile), 2567506 (*Thermoanaerobacter ethanolicus* - thermophile), 119407378 (*Neosartora fisheri* - thermophilic fungus) and 261252 (*Vibrio orientalis* - mesophile) belongs to domain glycol_hydro_2N followed by glyco_hydro2, glyco_hydro_2C and β gal_smallN. *Thermoanaerobacter ethanolicus* was lacking β gal_smallN domain. *Saccharomyces cerevisiae* (173115) was also included in this group having transposon protein domain. Plants form a cluster which are evolving together having glycol_hydro_35 domain. Mammals were evolving with *Drosophila melanogaster*, *Xenopus laevis* and *Pyrococcus abyssii* in the order, 3071885 (*Camponatus floridans*), 3320300 (*Acromyrmex echinator*), 3071979 (*Harpegnathus saltator*), 2294572 (*Drosophila melanogaster*), 8341508 (*Canis lupus*), 5761908 (*Felis catus*), 179419 (*Homo sapiens*), 2070292 (*Pongo abelii*), 5278222 (*Macaca fascicularis*), 192187 (*Mus musculus*), 157829 (*Rattus norvegicus*), 7189650 (*Gallus gallus*), 6295506 (*Danio rario*), 7804254 (*Bos taurus*), 4925628 (*Xenopus laevis*), 32465076 (*Ascaris suum*) and 1452173 (*Pyrococcus abyssii*) having glyco_hydro_35 domain. *Xenopus laevis* has one additional domain named beta_gal dom_4 5 while *Pyrococcus abyssii* belongs to glyco_hydro_42 domain followed by

glyco_hydro_42M. Thermophilic fungi evolved together with *Neurospora crassa* in an order like 2891951 (*Neurospora crassa*), 2594887 (*Aspergillus nidulans*), 248113 (*Talaromyces stipitatus*), 7099606 (*Aspergillus fumigatus*) and 3244879 (*Talaromyces emersonii*) having glyco_hydro35 domain followed by betagal_dom_2, betagal_dom_3 and betagal_dom_4 5 while *Neurospora crassa* have glyco_hydro35 domain and betagal_dom_4 5. This was followed by one cluster having 2551109 (*Thermoanaerobacter ethanolicus*), 2422630 (*Clostridium cellulovorans*), 2065798 (*Klebsiella pneumoniae*), 2962510 (*Clostridium lentocellum*), 2295677 (*Beutenbergia cavernae*), 114359 (*Planococcus* sp. L4) and 1166119 (*Arthrobacter* sp. FB 24) which belongs to domain glyco_hydro_42, glyco_hydro_42M and glyco_hydro_42C except *Thermoanaerobacter ethanolicus* which belongs to domain glyco_hydro_2N followed by glyco_hydro_2, glyco_hydro_2C and β gal_smallN.

Conclusion

Based on *In silico* studies of β galactosidase, it is observed that bacteria are more divergent and evolving at a much higher rate as compared to plants and mammals.

Acknowledgment

We thank Department of Science and Technology, Govt. of India for providing financial assistance to carry out the project work. We would also like to thank Shoolini University of Biotechnology and Management Sciences, Solan, Himachal Pradesh, India for providing infrastructure and support to carry out this work.

Funding: This work was supported by Department of Science and Technology, Government of India (INSPIRE Fellow code IF110061) to Varsha Rani.

References

1. <http://www.britannica.com/EBchecked/topic/1515406/extremophile> (2015).
2. Niehaus F., Bertoldo C., Kaehler M. and Antranikian G. (1999). Extremophiles as a source of novel enzymes for industrial application, *Applied Microbiology and Biotechnology*, 51, 711-729.
3. Taylor T.J. and Vaisman I.I. (2009). Discrimination of thermophilic and mesophilic proteins, *BMC Structural Biology*, 10, S5.
4. Szilagyi A. and Zavodszky P. (2000). Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure*, 8, 493-504.
5. Mizuguchi K., Sele M. and Cubellis M.V. (2006). Environment specific substitution tables for thermophilic proteins, *BMC Bioinformatics*, 8, S15.

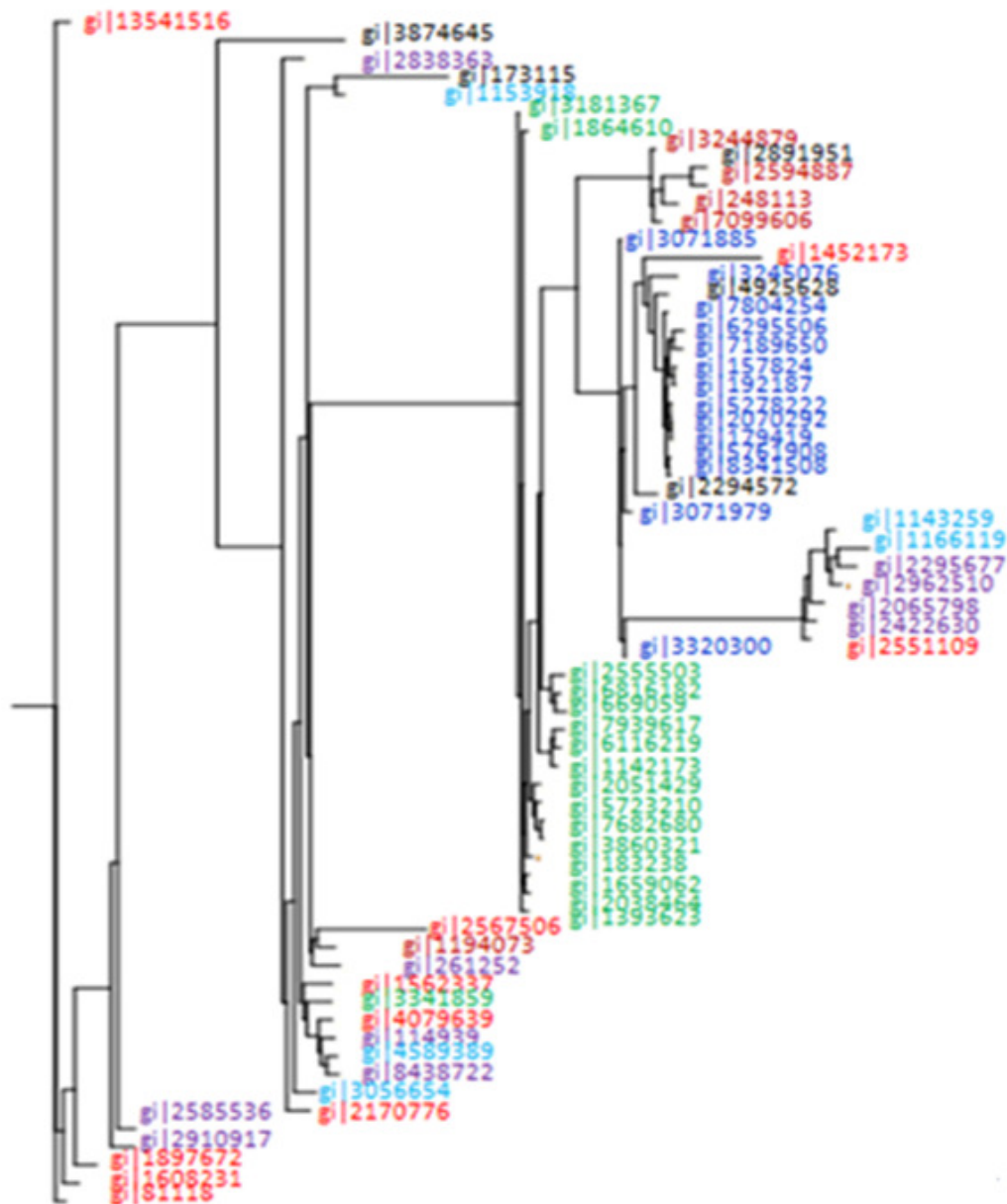


Figure-7

Phylogenetic tree. Phylogenetic tree was constructed by using Phylip software. Thermophiles are shown in red shades, psychrophiles are shown in blue shade, mesophiles are shown in purple shades, mammals are shown in royal blue shades, plants are shown in green shades and model organisms are shown in black shades (gi 3874645 (*Caenorhabditis elegance*), gi 22945722 (*Drosophila melanogaster*), gi 49256283 (*Xenopus laevis*), gi 173115 (*Saccharomyces cerevisiae*), and gi 2891512 (*Neurospora crassa*))

6. Shipkowski S. and Brenchley J.E. (2006). Bioinformatic, genetic, and biochemical evidence that some glycoside hydrolase family 42 β galactosidases are arabinogalactan type I Oligomer hydrolases, *Applied and Environmental Microbiology*, 72, 7730-7738.
7. Bose R., Arora S., Dwivedi V.D. and Pandey A. (2013). Amino acid based *in silico* analysis of β galactosidases, *International journal on Bioinformatics and Biosciences*, 3, 37- 44.
8. Kern F.J. and Struthers J.E.J. (1996). Intestinal lactose deficiency and lactose intolerance in adults, *The Journal of American Medical Association*, 195, 143-147.
9. Tumerman L., Fraw H. and Corneley K.W. (1954). The effect of lactose crystallization of protein stability in frozen concentrated milk, *Journal of Dairy Science*, 37, 830-838.

10. Mlichova Z. and Rosenberg M. (2006). Current trends of β galactosidases application in food technology, *Journal of food and nutrition research*, 45, 47–54.
11. Park H.Y., Kim H.J. and Lee J.K. (2008). Galactooligosaccharide production by a thermostable β galactosidase from *Sulfolobus solfataricus*, *World Journal of Microbiology and Biotechnology*, 24, 1553-1558.
12. Fowler A.V. and Zabin I. (1977). The amino acid sequence of β galactosidase of *Escherichia coli*, *Biochemistry*, 4, 1507-1510.
13. Bilbao M.M., Holdsworth R.E., Edwards L.A. and Huber R.E. (1991). Highly reactive β galactosidase (*Escherichia coli*) resulting from a substitution of an Aspartic acid for Gly 794, *The journal of Biological Chemistry*, 266, 4979–4986.
14. <http://www.ncbi.nlm.nih.gov>. (2014)
15. <http://www.ebi.ac.uk/Tools/sss/psiblast>. (2014)
16. <http://www.ibi.vu.nl/programs/pralinewww>. (2014)
17. <http://pfam.sanger.ac.uk>. (2014)
18. <http://evolution.genetics.washington.edu/phylip.html>. (2014)
19. <http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>. (2014)
20. <http://www.rcsb.org/pdb/home/home.do>. (2014)
21. <http://swissmodel.expasy.org>. (2014)
22. <http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>. (2014)
23. <http://hex.loria.fr>. (2014)
24. <https://www.pymol.org>. (2014)
25. Hidaka M., Fushinobu S. and Ohtsu N. *et al.* (2002). Trimeric crystal structure of the glycoside hydrolase family 42 beta-galactosidase from *Thermus thermophilus* A4 and the structure of its complex with galactose, *Journal of Molecular Biology*, 322(1), 79-91.
26. Jures D.H., Mathews B.W. and Huber R.E. and Lac Z (2012). β galactosidase: Structure and function of an enzyme of historical and molecular biological importance, *Protein Science*, 21, 1792–1807.
27. Rojas A.L., Nagem R.A.P., Neustroev K.N., Arand M., Adamska M., Eneyskaya E.V., Kulminskaya A.A., Garratt R.C., Golubev A.M. and Polikarpov I. (2004). Crystal structure of β galactosidase from *Penicillium* sp. and its complex with galactose, *Journal of Molecular Biology*, 343, 1281–1292.
28. Cheng W., Wang L., Jiyang Y.L., Bai X.H., Chu J., Li Q., Yu G., Liang Q.L., Zhou C.Z. and Chen Y. (2012). Structural insights into the substrate specificity of *Streptococcus pneumoniae* β 1.3 galactosidase BgaC, *The Journal of Biological Chemistry*, 287, 22910-22918.
29. <http://bioinformatics.cineca.it/PMDB>. (2014)
30. <http://www.modelling.leeds.ac.uk/qsitefinder>. (2014)
31. Kumar P.S., Pulicherla K.K., Ghosh M., Kumar A. and Rao K.R.S.S. (2011). Structural prediction and comparative docking studies of psychrophilic β galactosidase with lactose, ONPG and PNPG against its counterparts of mesophilic and thermophilic enzymes, *Bioinformation*, 6, 311-314.
32. <http://www.ebi.ac.uk/Tools/msa/clustalw2>. (2014)
33. Seddigh S. and Darabi M. (2014). Comprehensive analysis of beta-galactosidase protein in plants based on *Arabidopsis thaliana*, *Turkish Journal of Biology*, 38, 140-150.
34. <http://www.rcsb.org/pdb/home/home.do>. (2014).
35. <http://www.chemspider.com>. (2014).