

# ASRDb: A comprehensive resource for archaeal stress response genes

Rajendra Kumar Labala<sup>1</sup>, Santasabuj Das<sup>1</sup> & Surajit Basak<sup>2\*</sup>

<sup>1</sup>Biomedical Informatics Center, National Institute of Cholera and Enteric Diseases, P-33, C.I.T Road, Scheme-XM, Beliaghata, Kolkata-700010, India; <sup>2</sup>Department of Molecular Biology & Bioinformatics, Tripura University, Suryamaninagar- 799022, Tripura, India; Surajit Basak – E-mail: basaksurajit@gmail.com; Phone: 91-9862924152; \*Corresponding author

Received March 30, 2013; Accepted April 04, 2013; Published July 12, 2013

## Abstract:

An organism's survival strategy under the constantly changing environment depends on its ability to sense and respond to changes in its environment. Archaea, being capable to grow under various extreme environmental conditions, provide valuable model for exploring how single-celled organisms respond to environmental stresses. However, no such approach has ever been made to make an integrated classification of various archaeal stress responses.

Archaeal Stress Response Database (ASRDb) is a web accessible (<http://121.241.218.70/ASRDb>) database that represents the first online available resource providing a comprehensive overview of stress response genes of 66 archaeal genomes. This database currently contains almost 6000 stress specific genes of 66 archaeal genomes. All the stress specific genes are grouped into 17 different stress categories. A user-friendly interface has been designed to examine data using query tools. This database provides an efficient search engine for random and advanced database search operations. We have incorporated BLAST search options to the resulting sequences retrieved from database search operations. A site map page representing the schematic diagram will enable user to understand the logic behind the construction of the database. We have also provided a very rich and informative help page to make user familiar with the database. We sincerely believe that ASRDb will be of particular interest to the life science community and facilitates the biologists to unravel the role of stress specific genes in the adaptation of microorganisms under various extreme environmental conditions.

**Key words:** Archaea, Stress, Genome, Database.

## Background:

Microscopic organisms live in a constantly changing environment and are subject to a variety of environmental stresses. They have evolved adaptive networks to face the challenges of changing environments and to survive under conditions of stress [1]. It is difficult to provide a universal definition for stress. Under adverse environmental conditions, for example depletion of nutrients, changes in pH, temperature etc., bacteria launch "stress responses" which significantly improve their chances of survival under, or successful adaptation to, the challenges posed by such unfavorable environments [2]. These responses have significant implications for many of our interactions with bacteria.

Most of the archaea are known to be extremophiles and can survive in extreme environments. It suggests that archaea should be well equipped with several stress response machineries that help them to survive in extreme conditions. The ability of archaea to sense and respond (correctly) to spontaneous alterations in the environment is crucial to their survival [3]. Here we have chosen archaea for the construction of this database because this group shows most unique and diverse stress responses.

## Objectives

From the above discussion it has become clear that different stress responses play their respective roles to provide important

survival strategies to individual archaeal species. Many stress response proteins are involved and regulated carefully throughout these processes. Though there are quite a few studies on stress responses but still this field has got huge dimensions to offer. However, some troubles are often encountered:

- 1) Different stress responses had been identified at different times but no such approach has ever been made to classify all the stress responses.
- 2) Lack of an effective system to identify,
  - a. All the stress responses present in a specific archaea.
  - b. In how many archaea a particular stress response is present?
  - c. Whether some archaea have more stress responses than others.
  - d. Whether stress responsive proteins have some other functions.
  - e. Frequency of occurrences of different stress responses in different archaea.

A stress response database for archaea can provide answers to all the questions. With the availability of a large number of completely sequenced archaeal genome in public domain and rapid advancement of Bioinformatics, it is possible to create a archaeal stress response database which will not only help to provide more detail insight into different archaeal stress responses, but will make researchers jobs comparatively easy. Keeping these facts in mind we have developed an archaeal stress response database as no such database is publicly available till date.

## Methodology:

A thorough examination of literature and existing online resources including other large databases like GenBank was performed and relevant data were retrieved (**Figure 1**). 66 archaeal genomes have been used for the construction of the database. With adequate literature searching, an integrated classification of 17 different types of archaeal stress responses had been made. They are: Heat shock, Cold shock, Envelope stress, General stress, Osmotic stress, pH & acid tolerance, Universal stress, Sodium stress, Metal resistance, Nitrosative stress, Oxidative stress, Oxygen deprivation, SOS response, Organic solvent tolerance, Starvation & Stringent induced shock, Phage shock, Stationary phase response (please check the web database for complete reference of each stress category).

In the next step, we have generated "key search terms" for every stress category through extensive study of literatures **Table 1 (see supplementary material)**. These "key search terms" were then used to retrieve sequences from publicly available sequence databases. While retrieving the sequences corresponding to any particular stress category using "key search term" we took special attention to remove all false positive results through manual curation. Information on accession numbers, nucleotide and protein sequences were extracted and included in ASRDb.

## Database Architecture:

ASRDb (Archaeal Stress Response Database) is built on Internet Information Services (IIS) 7.0 Server in Microsoft Windows Vista Business SP1 Platform. Easy and most accessible front-end

was designed using Microsoft Visual Studio 2005 version 8.0 with Microsoft .NET Framework version 2.0 SP1. Most lucrative frames are made using Visual basics embedded in Active Server Pages (ASP) and to make bright and keen web pages JavaScript are used along with HTML tags. An overall database support is maintained by Microsoft Access 2007 which is server supportive and also having flexible environment for securing the unique data stored in the database.

The architecture of ASRDb is shown in **Figure 1**. We have provided several useful tools for database searching, online BLAST, genome download and online reference. Sequence information can be retrieved in various forms, the whole genome sequences of all archaeal genomes as well as individual sequences (protein and nucleotide) for a particular stress. A short sequence description with its NCBI Gene ID can also be retrieved.

## Database Interfaces:

ASRDb is designed to be a publicly accessible online resource with user-friendly interface. Various useful features have been built into ASRDb. Users can easily navigate and use various tools integrated in the database. The database interface includes: Home, Organism list, Stress categories, Database search engine, Database statistics, Help, and Site map.

There are 66 archeal genomes and 17 different archaeal stress categories included in the database. The search features in ASRDb are designed to accommodate all possible queries of the users. There are two types of search options available, one is random search and another is advanced search. In random search, users can search the entire database with any keywords of his/her choice. The advanced search feature enables specific searches by using the advanced query form. Users can limit their search to a particular organism with a specific stress category. Database statistics page gives the total number of genomes, stresses and sequences included in the ASRDb database. In addition we have also provided some important database characteristics. Users can find the percentage of different stress categories present in a given archaeal genome in the form of a bar diagram. Detailed statistics of the percentage of all genes distributed in the 17 stress categories among 66 archaeal genomes has been displayed through pie chart and bar chart.

We have also provided a very rich and informative help page detailing every aspect of the database. A set of frequently asked questions (FAQ) has also been provided to make user acquainted with the database. A site map page representing the schematic diagram will enable user to understand the logic behind the construction of the database.

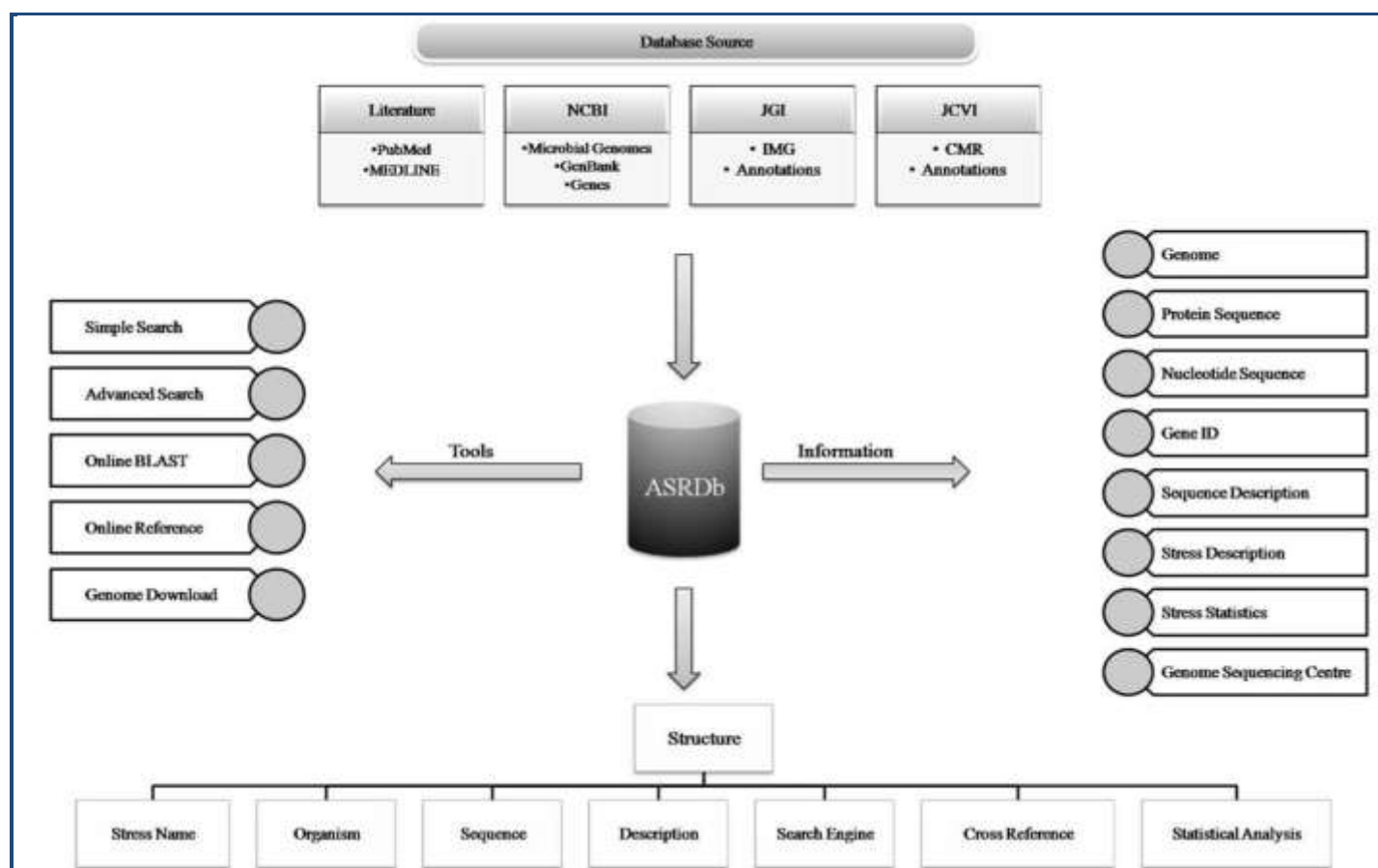
## Utility:

Microorganisms have evolved adaptive networks to face the challenges of changing environments and to survive under conditions of stress. Their response to environmental stress was first elucidated about 25 years ago, through pioneering proteomic studies of the response to temperature shift-up - now known as the heat shock response [4, 5]. Since then many important studies have been published on the stress responses. However, after the availability of a large number of completely sequenced archaeal genomes, it is the need of the hour to

develop an exhaustive database on stress response genes to provide constructive support to research community.

ASRDB is a repository for 17 different stress specific genes of 66 archaeal genomes. It is the first of its kind, and should prove of value to a variety of researchers. Altogether, 6295 gene sequences have been classified into 17 different stress categories. From the database statistics page users can get some more additional information. Allowing users to compare between different stresses categories is another helpful element of ASRDB. Comparative statistics of all the genes among 17

different stress categories has been provided. The average percentage of all the different stress response genes for a given genome has also been provided. Through systematic data mining, ASRDB offers researchers new means for inspecting and analysis using comparative genomics approach. We believe that ASRDB represents a new and important model for stress response gene database. We are hopeful that the rich content in ASRDB will allow researchers to uncover answers to many common questions about different categories of stresses in archaeal genomes.



**Figure 1:** Schematic diagram for overall design and data transfer of ASRDB.

### Future Directions:

Over the coming years we also plan to bring in additional data and to add new features to this database, with the intent of making ASRDB a more comprehensive database for stress response genes. The ASRDB will continue solicitation of feedback from users in order to continue improving all aspects of the database. We are eager for scientists to visit the database webpage and let us know what you think. We will surely try to accommodate constructive suggestions and upgrade the database accordingly. We sincerely hope that this thorough and comprehensive database will be extended to effective completeness, and then maintained and its content expanded, with constantly enhanced search and analysis features added on a rolling basis.

### Acknowledgment:

This work is funded through an extramural project (IRIS ID: 2001-0496D) of ICMR, Government of India.

### References:

- [1] Watson K, *Adv Microb Physiol.* 1990 **31**: 183 [PMID: 2264523]
- [2] Abee T & Wouters JA, *Int J Food Microbiol.* 1999 **50**: 65 [PMID: 10488845].
- [3] Pedone E *et al. Front Biosci.* 2004 **9**: 2909 [PMID: 15353325].
- [4] Neidhardt FC *et al. Annu Rev Genet.* 1984 **18**: 295 [PMID: 6442118]
- [5] Rosen R & Ron EZ, *Mass Spectrom Rev.* 2002 **21**: 244 [PMID: 12533799]

Edited by P Kanguane

Citation: Labala *et al.* Bioinformatics 9(12): 650-655 (2013)

**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

## Supplementary material:

**Table 1:** Key search terms used for stress categories

Sl. No.	Stress Category	Key Search Terms	Reference(s) in web database
1	Heat Shock Response	GroES GroEL Peptidyl-prolyl Heat shock protein Dna Clp Hsc Htp Grp Chaperon Msb Omp Tor Hsp chaperonin	1-7, 11, 23
2	Acid & pH Tolerance	acid tolerance RpoS Crp HNS CysB AdiY IHF PhoP OmpR GadC CagA WbcJ	8-12, 30, 59
3	Cold Shock Response	cold shock protein csp Nus Ribosomal protein L Ribosomal protein S6 Peptidyl-prolyl IF-2 Csd Trigger factor Dihydroliipoamide H-NS	1-3, 13-19
4	General stress	general stress	7, 25
5	Envelope stress	gsp envelope stress envelope resistance thiol disulphide oxidoreductase peptide-propyl cis-tran isomerase sigma E Cpx Bae Rse Deg(P) skp Dsb Lia Ppi YvcPQ YxdJK	19-24
6	Metal & Metal Ion Resistance	copper zinc	38-44

		cobalt	
		cadmium	
		Molybdenum	
		Fe	
		Iron	
		Arsenic	
		CZC	
		mercury	
		Aluminium	
		Tellurite	
		Nickle	
7	Nitrosative Stress	AhpC	45-49
		Hmp	
		Vgb	
		Nor	
		Cgb	
		HbN	
		NrtA	
		ytfE	
		SrrA	
		YjeB	
		Nsr	
		Fnr	
8	Organic Solvent Tolerance	organic solvent tolerance	65-68
		Ost	
		toluene	
		phenol	
9	Osmotic Stress	osmo	26-30
		EnvZ	
		OmpR	
		MicF	
		OmpC	
		OmpF	
		ProP	
		ProU	
		PutU	
		Trk	
		Kdp	
		Kup	
10	Oxidative Stress Response	Icd	41, 48-52
		Suc	
		Fdn	
		GlpA	
		GlpB	
		GlpC	
		GlpD	
		Hya	
		Cyd	
		Cyo	
		Dms	
		NarG	
		NarH	
		NarI	
		NapF	
		Nif	
		Nir	
		Foc	
		Acn	
		Fum	
11	Oxygen Deprivation	Icd	41, 48-51, 53-56
		Suc	
		Fdn	

		GlpA	
		GlpB	
		GlpC	
		GlpD	
		Hya	
		Cyd	
		Cyo	
		Dms	
		NarG	
		NarH	
		NarI	
		NapF	
		Nif	
		Nir	
		Foc	
12	Phage Shock	phage shock protein	19, 69-71
		Psp	
13	Sodium Stress	F ATPase	11-12, 32-37
		V ATPase	
		ABC type transporter	
		Methylmalonyl-CoA decarboxylase	
		Malonate decarboxylase	
		Oxaloacetate decarboxylase	
		N5-Methyltetrahydromethanopterin:CoM methyltransferase	
		NADH:ubiquinone oxidoreductase	
		Cytochrome bo-type ubiquinol oxidase	
		Nha	
		TetA(L)	
		mrpA	
		NapA	
14	SOS Response to DNA	Dna	12, 57-64
		Rec	
		Din	
		Lex	
		Ruv	
		Uvr	
		Suh	
15	Universal Stress Response	universal stress protein	31, 32
		Usp	
16	Starvation & Stringent Response	stringent	72-76
		CstA	
		CstB	
		PhoH	
		RelA	
		SpoT	
17	Stationary Phase Survival	Stationary phase survival	77-79
		SurE	

---