Journal of Computational Intelligence in Bioinformatics.
ISSN: 0973-385X Volume 1 Number 2-3 (2008) pp. 93–108

http://www.ripublication.com/jcib.htm

DLact: An Antimicrobial Resistance Gene Database

Reema Singh
Scientist-I
Biomedical Informatics Center
Indian Council of Medical Research
Ansari Nagar, New Delhi - 110029, India
E-mail:reemas@icmr.org.in

Suchir
Scientist-I
Biomedical Informatics Center
Indian Council of Medical Research
Ansari Nagar, New Delhi - 110029, India
E-mail:suchir@icmr.org.in

Harpreet Singh*
Scientist B
Biomedical Informatics Center
Indian Council of Medical Research
Ansari Nagar, New Delhi - 110029, India
E-mail:hsingh@icmr.org.in

Abstract

Lactamases are the main cause of antimicrobial resistance in the bacterial pathogens. Here, we introduce DLact, a curated collection of lactamase like proteins from bacterial genomes. The database currently contains 2020 lactamase like genes from 814 sequenced bacterial genomes. Of which 1971 (97.57%) were present on chromosome and only 49 (2.42%) were present on plasmids. The database can be searched using text and sequence queries. Diversity at the taxonomic, microbial ecology, domain length and sequence was studied. DLact database may be used for variety of biomedical research such as to develop diagnostic primers and probes and in identifying pathways controlling/affecting the expression of antimicrobial genes, comparative protein modeling and drug designing, as well as those interested in broad subjects such as lateral gene transfer and Codon usage. The database will be updated bi-yearly using in-house

developed perl scripts. Database is available at **http://59.160.102.202/Dlact/**.

Keyword: Beta-lactamase genes, Lactamase database, antimicrobial resistance genes.

Introduction

Resistance to antimicrobial is a serious clinical problem [1, 2]. Antibiotic resistance is the ability of a microorganism to withstand the effects of an antibiotic. Beta-lactam antibiotics are one of the most widely used antibacterial agents in the present chemotherapeutic regime, and beta lactamase, the enzymes that hydrolyze beta lactam antibiotics are the major cause of resistance to these compounds [3, 4]. The growing problem of antimicrobial resistance has become a significant public health concern worldwide and it involves practically all types of pathogens, including viruses, bacteria, mycobacteria, fungi, and parasites [5]. The gene for beta lactamases may be chromosomal, plasmid borne or found on transposable elements. Bacteria acquire exogenous genetic material that leads to antimicrobial resistance. Species such as *pneumococci* and *meningococci* can take up foreign DNA and incorporate it into their chromosomes [6]. Hence many of the genes that mediate resistance are found on transferable plasmids or on transposons that can be disseminated among various bacteria by conjugation [7]. Transposons are mobile pieces of DNA that can insert themselves into various locations on the bacterial chromosome, as well as move into plasmids or bacteriophage DNA. Some transposons or plasmids have genetic elements termed integrons that enable them to capture exogenous genes [8]. A number of genes may therefore be inserted into a given integron, resulting in resistance to multiple antimicrobial drugs or possibly allowing the accumulation of both regulatory and structural genes in the same transposon. A similar mechanism may have been involved in the assembly of the genetic elements that code for vancomycin resistance in *enterococci* [9]. It appears that many of the genes determining resistance have been present in nature and predate the clinical use of antimicrobial drugs [10].

Various classification schemes have been proposed for beta lactamases based on the characteristics of the enzymes and/or their substrate profiles [11, 12]. However a functional classification scheme for beta lactamases proposed by ambler has found common usage [12, 13]. Ambler classifies these enzymes into four classes i.e. A, B, C and D [13, 14, and 15]. Class A, C and D have evolved dependence on an active site serine as their key mechanistic feature. Class B enzymes are zinc dependent and hence different from A, C and D. It is noteworthy that these enzymes do not share any sequence homologies, structural similarities or mechanistic features with serine dependent proteases. A series of recently discovered beta lactamases exhibit wide breadth for their substrate preferences, which often include penicillins, cephalosporins and carbapenems among other substrates [16]. These so called extended-spectrum-beta lactamases (ESBLs) are being identified among all classes of beta lactamases. It is now accepted that beta lactamases evolved from penicillin binding proteins (PBP),

which experience covalent modification by penicillins and other beta lactams [17]. These biosynthetic enzymes assemble the bacterial cell wall and regulate its function. Certain PBPs carry out the cross-linking reaction, which imparts rigidity to the bacterial cell wall [18]. Penicillins bind to PBPs and acylate an active site serine. The resultant acyl-enzyme species is sufficiently stable to provide effective inhibition of the biological function of the PBP, an inhibition event that leads to bacterial cell death. The development of beta lactamase is a relatively recent events, which must have taken place after the evolution of the first biosynthetic pathways for the natural beta-lactam antibiotics [19,20 and 21].The process of evolution of beta lactamases has been accelerated by the extensive use of beta-lactams in clinic over the past few years [22,23] . It is the use of antimicrobial drugs for prophylactic or therapeutic purposes in humans or for veterinary or agricultural purposes that provided the selective pressure favoring the overgrowth of resistant organisms [24].

There are a number of resources of data from surveillance study of antimicrobial resistance available on the World Wide Web [25] e.g. Infectious Disease society of America ,European Society of Clinical Microbiology and Infectious diseases and www.lahey.org that may be useful to scientists with a research interest in the field of antimicrobial resistance. To date over 600 novel beta lactamases have been identified (www.lahey.org) to complicate their therapeutic use. List of beta lactamase genes available (www.lahey.org) were compiled using experimental data obtained from clinical studies. This can be used to study the mutational and substitution rates. Hence the information is present about the lactamase individual gene so this can not be further used for network and regulatory studies. The data present in static table hence not queried using sequence based search.

Resistance is so complex and dynamic at the genetic level that there is a needed to understand the diversity and prevalence of resistance gene families, both in nature and in the animal microflora that are the bridge to human contact and to discern the origin of these genes and how they spread from one organism to other. We also need to understand the regulation of resistance genes in order to develop novel class of resistance inhibitors that work by either activating the resistance gene suppressors or inhibiting the activators.

In this study we report a database of lactamase genes identified from sequenced plasmids and chromosomes of 457 bacterial strains.

Result

Taxonomic distribution

DLact contains 2020 lactamase like genes identified from 814 sequenced bacterial genomes, which much larger than related databases (www.lahey.org). Of these 1971 were present on chromosome and only 49 were present on plasmids. Table 1 shows the taxonomic distribution of lactamase like proteins at the class and family level. As can be seen from the table the lactamase like proteins are approximately uniformly distributed across various taxonomic groups, however, number of genes per chromosome varied across different taxonomic groups. Similarly, distribution at the family and genus level also showed uniform distribution of lactamase like proteins.

The near uniform taxonomic distribution of lactamase like proteins indicates that these genes have evolved from an essential gene which was present very early in the evolution.

Table 1: Taxonomic distribution of lactamase like proteins at the class and family level.

Class	**Family**	**No. of species Studied**	**No. of species containing putative lactamase.**	**No. of Genes**
Acidobacteria (class)	*Acidobacteriaceae*	1	1	21
Solibacteres	*Solibacteraceae*	1	1	29
Actinobacteria (class)	*Acidothermaceae*	1	0	0
	Micrococcaceae	7	4	21
	Bifidobacteriaceae	3	1	1
	Corynebacteriaceae	6	4	16
	Frankiaceae	2	2	13
	Microbacteriaceae	1	1	1
	Mycobacteriaceae	15	12	182
	Nocardiaceae	7	5	35
	Nocardioidaceae	2	1	4
	Propionibacteriaceae	1	1	2
	Rubrobacteraceae	1	1	10
	Streptomycetaceae	5	2	35
	Nocardiopsaceae	1	1	4
	Cellulomonadaceae	2	0	0
Aquificae (class)	*Aquificaceae*	2	1	3
Bacteroidetes	*Bacteroidaceae*	6	3	14
	Porphyromonadaceae	1	1	1
Flavobacteria	*Flavobacteriaceae*	1	1	5
Sphingobacteria	*Flexibacteraceae*	1	1	3
	Crenotrichaceae	2	1	3
Chlamydiae	*Parachlamydiaceae*	1	1	2
	Chlamydiaceae	14	0	0
Chlorobia	*Chlorobiaceae*	4	4	5
Thermoprotei	*Desulfurococcaceae*	1	1	2
	Pyrodictiaceae	1	1	1
	Thermoproteaceae	2	2	2
	Sulfolobaceae	3	3	11
	Thermofilaceae	2	0	0
Unclassified	*Nostocaceae*	11	2	8

	Unclassified	17	12	
	Prochlorococcaceae	9	5	7
Deinococci	*Deinococcaceae*	6	4	13
	Thermaceae	5	2	8
Archaeoglobi	*Archaeoglobaceae*	1	1	4
Halobacteria	*Halobacteriaceae*	17	4	10
Methanobacteria	*Methanobacteriaceae*	3	2	2
Methanococci	*Methanocaldococcaceae*	3	0	0
	Methanococcaceae	1	0	0
Methanomicrobia	*Methanosarcinaceae*	5	3	7
	Methanocorpusculaceae	1	1	2
	Methanomicrobiaceae	1	1	3
	Methanosaetaceae	1	1	2
	Methanospirillaceae	1	1	6
Methanopyri	*Methanopyraceae*	1	1	2
Thermococci	*Thermococcaceae*	5	4	10
Thermoplasmata	*Picrophilaceae*	1	1	2
	Thermoplasmataceae	2	2	3
Bacilli	*Bacillaceae*	27	16	281
	Enterococcaceae	4	1	4
	Lactobacillaceae	20	11	54
	Streptococcaceae	32	25	65
	Leuconostocaceae	3	2	2
	Listeriaceae	5	4	15
	Staphylococcaceae	29	17	67
Clostridia	*Peptococcaceae*	2	2	9
	Clostridiaceae	12	7	21
	Thermoanaerobacteriaceae	2	2	6
	Syntrophomonadaceae	1	1	6
Mollicutes	*Entomoplasmataceae*	1	1	1
	Mycoplasmataceae	13	0	0
	Acholeplasmataceae	1	0	0
Fusobacteria	*Fusobacteriaceae*	1	1	1
Planctomycetacia	*Planctomycetaceae*	1	1	8
Alphaproteobacteria	*Rhizobiaceae*	24	13	37
	Anaplasmataceae	8	0	0
	Bartonellaceae	3	0	0
	Bradyrhizobiaceae	12	10	86
	Brucellaceae	8	8	16

	SAR11 cluster	1	0	0
	Caulobacteraceae	1	1	15
	Erythrobacteraceae	1	1	12
	Acetobacteraceae	7	2	6
	Hyphomonadaceae	2	2	34
	Rhodobacteraceae	22	10	30
	Rhodospirillaceae	3	2	6
	Phyllobacteriaceae	7	2	16
	Sphingomonadaceae	4	3	27
	Rickettsiaceae	9	5	11
Betaproteobacteria	*Comamonadaceae*	20	6	21
	Rhodocyclaceae	5	3	15
	Alcaligenaceae	3	3	7
	Burkholderiaceae	41	34	109
	Neisseriaceae	5	5	10
	Methylophilaceae	1	1	5
	Nitrosomonadaceae	8	3	5
	Hydrogenophilaceae	1	1	4
Deltaproteobacteria	*Myxococcaceae*	2	2	22
	Bdellovibrionaceae	1	1	2
	Desulfobulbaceae	3	1	2
	Desulfovibrionaceae	9	1	1
	Geobacteraceae	3	2	12
	Pelobacteraceae	4	2	6
	Syntrophobacteraceae	1	1	8
	Syntrophaceae	1	1	7
Epsilonproteobacteria	*Campylobacteraceae*	6	3	5
	Helicobacteraceae	9	1	4
Gammaproteobacteria	*Moraxellaceae*	4	2	9
	Pasteurellaceae	13	4	5
	Aeromonadaceae	1	1	3
	Alcanivoracaceae	1	1	2
	Ectothiorhodospiraceae	2	1	1
	Enterobacteriaceae	72	30	112
	Halomonadaceae	1	1	1
	Colwelliaceae	1	1	11
	Coxiellaceae	2	0	0
	Francisellaceae	5	5	10
	Hahellaceae	1	1	7
	Idiomarinaceae	1	1	4
	Legionellaceae	5	4	20
	Alteromonadaceae	4	2	7
	Methylococcaceae	1	1	2
	Chromatiaceae	2	1	2
	Vibrionaceae	15	9	22
	Pseudoalteromonadaceae	3	2	14
	Pseudomonadaceae	13	9	54
	Psychromonadaceae	1	1	2
	Shewanellaceae	16	9	39
	Piscirickettsiaceae	1	1	8

	Xanthomonadaceae	17	8	32
Spirochaetes	*Spirochaetaceae*	36	0	0
	Leptospiraceae	8	4	8
Thermotogae	*Thermotogaceae*	1	1	1

Ecophysiological distribution

Eco-physiological details of the sequenced organisms were extracted from the Microbial Genome Resources available at NCBI (http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi?view=1). We found that lactamase like genes were most prevalent in organisms which are in close association with humans or mesophile soil inhabitants. Very few or no genes were found in free-living extremophiles indicating that the use of antibiotics has facilitated the spread of lactamase genes among selected taxonomic groups and selected for their rapid evolution [22]. The genes were present mainly among pathogenic organisms. Non pathogenic strains that contain lactamase genes were from families that contain pathogenic organisms. For example, *Ralstonia metallidurans CH34* from family *Burkholderiaceae* is non-pathogenic and contain lactamase gene. *Ralstonia solanacearum* is a devastating, soil-borne plant pathogen (26).

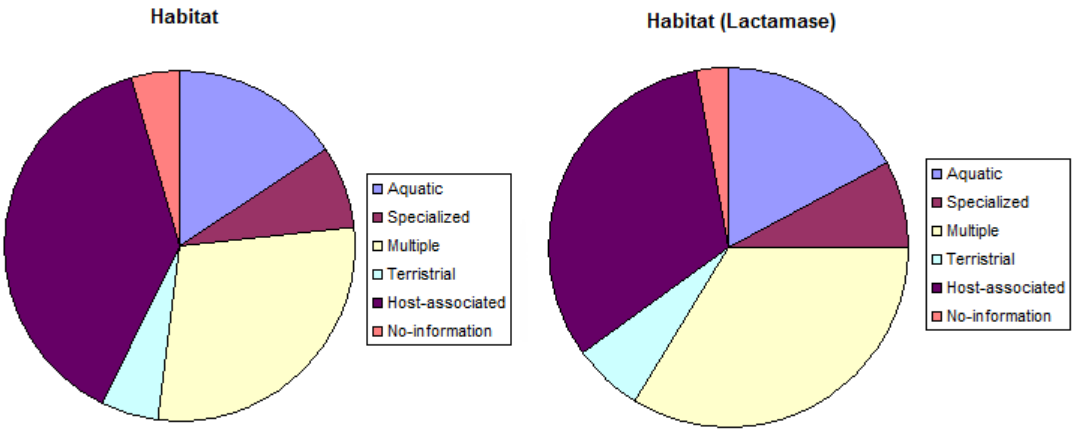

Figure 1: Habitat distribution of lactamase like genes.

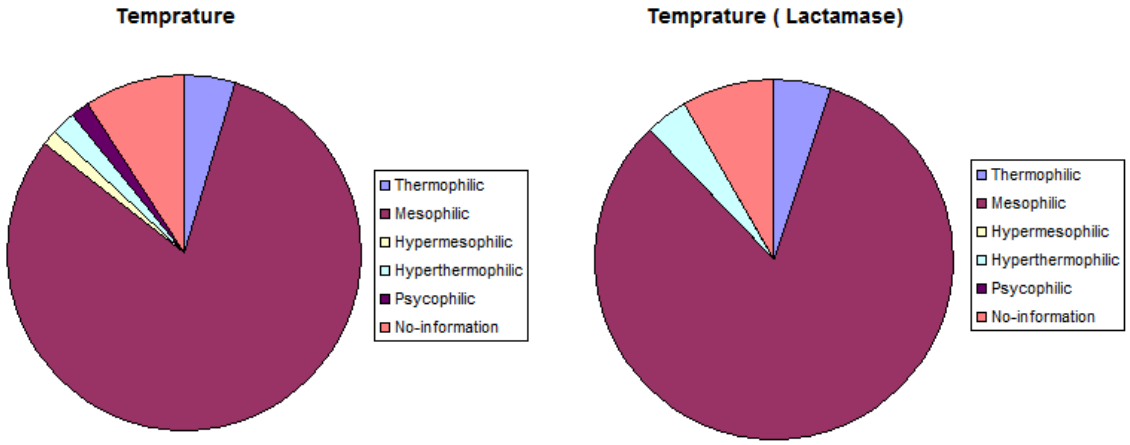

Figure 2: Distribution of lactamase like genes at the temperature level.

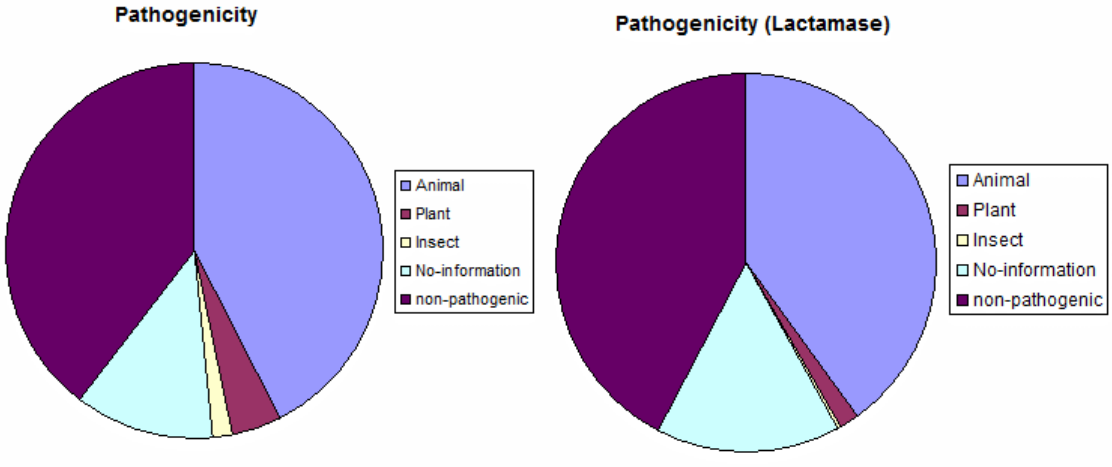

Figure 3: Distribution of lactamase like genes according to pathogenicity.

Domain length diversity within the Lactamase genes

The lactamase domain was identified using Interproscan [27]. Figure 4 shows the cumulative frequency distribution for the average domain length of the lactamase genes. Majority of lactamase genes in our database were from 200-400 residues long.

The distribution indicates at least two separate classes (200-250 and 325-350) of lactamase based on the domain length.

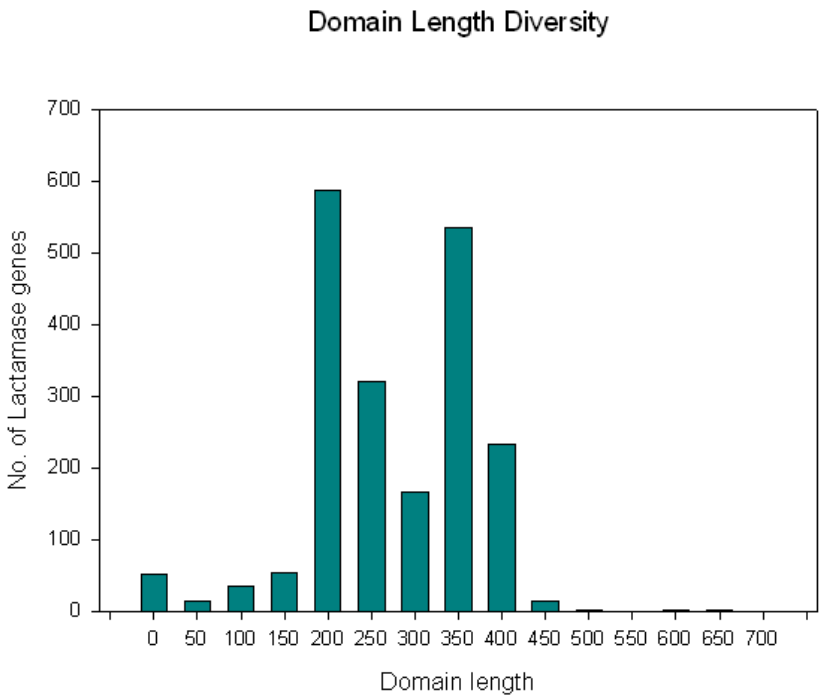

Figure 4: Domain length diversity bar graph.

We have found four proteins containing two lactamase like domains (as identified by InterProscan) (Table 2). Presence of double domain in a lactamase gene may be a case of gene fusion. Gene fusion has been observed for a number of bacterial proteins e.g. *Pseudomonas aeruginosa*. However the double domain found in those genes which are not associated with any human disease. The lactamase genes containing double domain may be a reservoir of generator of domain or sequence diversity.

Table 2: Occurrence of double domain lactamase genes in bacteria.

Organism Name	Taxon	Gi_number	Genome_accession	Start-End
Bacillus thuringensis	*Bacillales*	118478619	NC_008600	40 - 367 386 - 720
Pectobacterium atrosepticum	*Proteobacteria,Gammaproteobacteria; Enterobacteriales*	52142220	NC_004547	40 – 367 386 - 720
Bacillus cereus E33L	*Bacilli, Bacillales, Bacillaceae*	71280521	NC_006274	112 - 124 172 - 265
Colwellia psychrerythraea	*Gammaproteobacteria,Alteromonadales Colwelliaceae*	50121773	NC_003910	99 - 235 194 - 271

Sequence diversity within the Lactamase gene

We have used needle to find out pairwise sequence identity between the lactamase genes. Figure 5 shows the cumulative frequency distribution for pairwise percentage sequence identity between the lactamase genes. Pairwise sequence identity in lactamase genes ranges from 0.0% to 100 %.It has been noted that 86% lactamase genes has the sequence identity in the range of 10-20% and 35% in the range of 0-10%.It means lactamase genes are highly diverged.

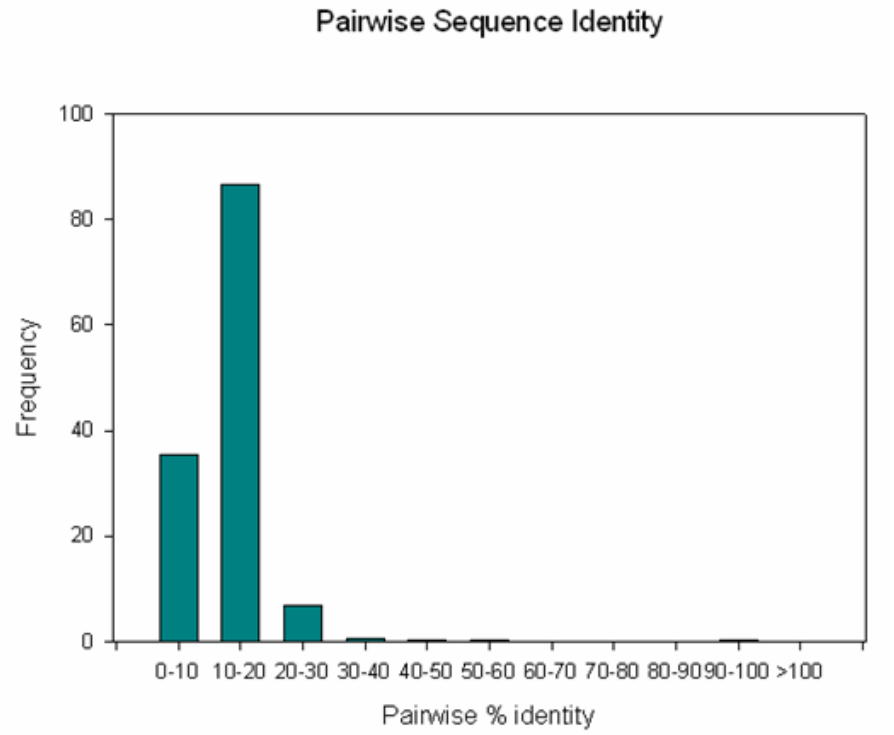

Figure 5: Distribution of pair wise sequence identity for lactamase gene.

Web Interface

DLact was constructed with MySql, and the web interface was implemented with PHP. In contrast to Lahey.org where sequence based search cannot be made, our database can be searched by using text or sequence queries. The text search window (Fig.6) allows searching by element of the gene-name, strand polarity OR via selecting taxonomic classification. The user also has the option to refine the search using a combination of gene-name, strand polarity and taxonomic listing. Search results are presented in table form as shown in Figure 7.In this case, a search for gene blaZ yielded three records. Each record includes gene-name, synonym, gi-number, start, end, strand, protein name, strain and taxonomic information. Links to NCBI database are provided for retrieval of full GenBank records.

Figure 6: The DLact text query page. A screenshot shows different fields by which a user can effectively search the database.

Sequence based search can be conducted by using nucleotide or protein queries. BLAST against lactamase genes has been made available with both DNA and protein blast, through blastp and blastn codes. User has the privilege to customize the blast by changing any of the listed parameter like e-value, word-size, Gap-Penalties or by selecting the available matrices. Filtering the low complexity regions of DNA/ protein can be done using the corresponding DUST/SEG filter. The blast options are standard as given at NCBI.

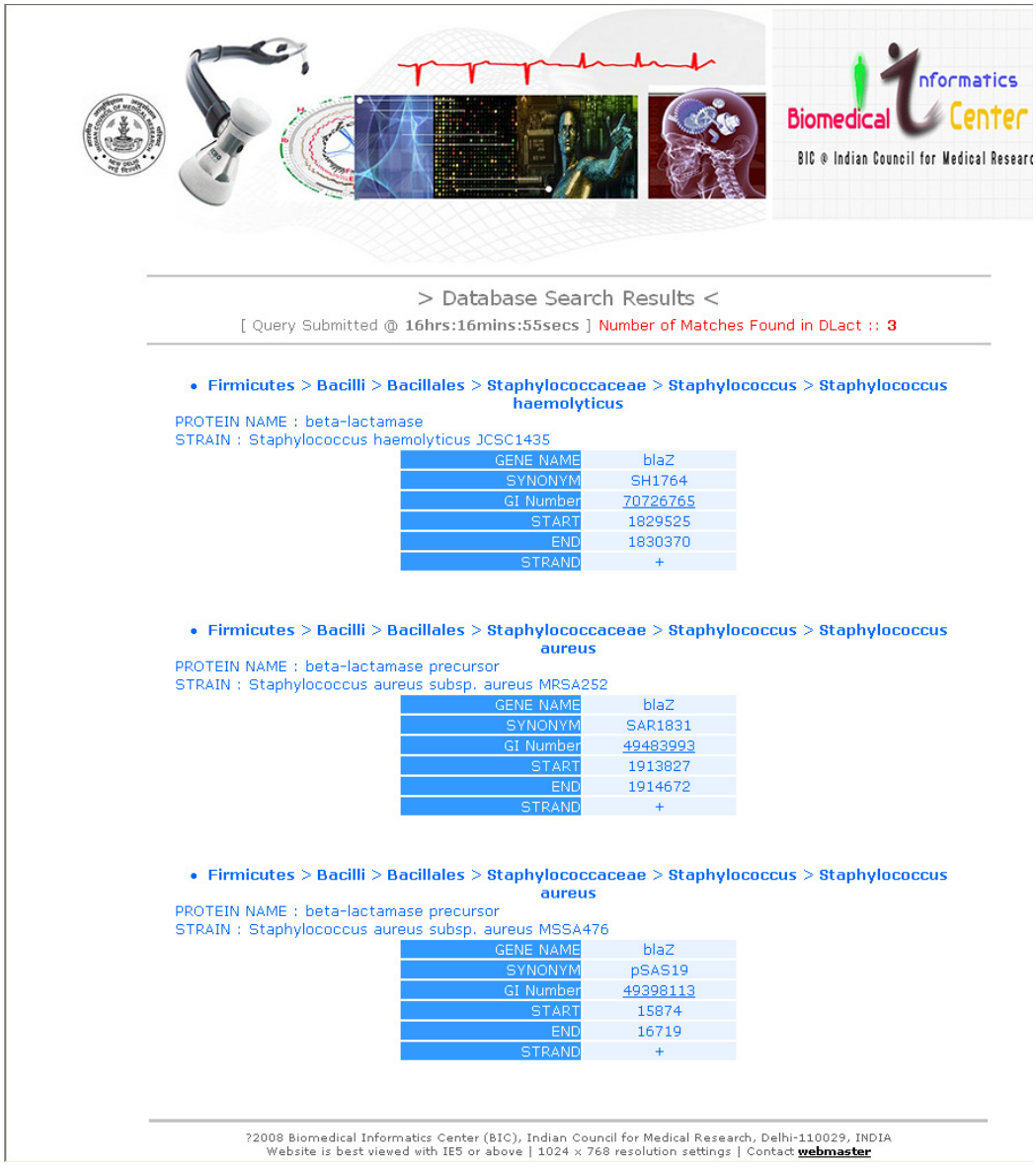

Figure 7: DLact record details. Records include the name of gene, synonym, gi number, star, end, strand and taxonomic information.

Methods

Data Collection

Chromosomal and Plasmid DNA sequences of 457 sequenced bacterial strains (as available on 1st March, 2007) were downloaded from NCBI genome repository. Sequences of 158 lactamase proteins belonging to well characterized lactamase families like Serine and Metello-beta lactamase etc were selected from literature and downloaded from SwissProt/UniProt database (The UniProt Consortium, 2007).

Identification of putative lactamase genes

Using each of the initial 158 sequences as a query set, blastx [28] search was carried out in order to identify regions of chromosome and plasmid DNA sequences homologous to the 158 lactamase proteins. Hits were merged and aligned regions were mapped onto the protein translation table (ptt) files (downloaded from the NCBI web site) in order to identify the gene boundaries of the lactamase genes.

Annotation of putative lactamase genes

Putative lactamase genes were annotated by characterizing domains in protein sequences using InterProScan [27] and rpsBlast [29]. Both these methods are very sensitive in detecting remote homologies and annotating sequences that may have diverged significantly. InterPro is an integrated resourse for the identification of protein families, domains and functional sites and was created by integrating several major protein signature databases [30]. Currently it includes PROSITE [31], Pfam [32], PRINTS [33] , ProDom [34], SMART [35], TIGR-FAMs [36], PIRSF [37] , Gene3D [38], Panther [39] and SUPERFAMILY. A common query can be used to interrogate the various databases and the outputs from all of them are integrated in a common format. Specific protein family signatures are integrated into InterPro entries that are expertly curated to provide significant biological and functional information. In our case, InterPro searches were carried out in batch mode using the perl script InterProScan.pl available at http://www.ebi.ac.uk/Tools/webservices/downloads/perl/interproscan.pl . RpsBLAST is another powerful heuristic search tool used for comparing a protein sequence against Position Specific Scoring Matrices (PSSMs) generated from conserved protein domain families. RpsBLAST searches of the putative lactamase proteins were carried out against PSSMs generated from protein domain families available in the CDD [40], SMART [35], COG [41] and Pfam [32] databases. Based on InterProScan and rpsBLAST searches, lactamase genes were selected. A database has been designed using MySql Version 4.001 on Red Hat Linux ES4.

References

[1] Wise,R., Hart,T., Cars,O., Streulens,M., Helmuth,R., Huovinen,P., and Sprenger,M .,1998, “Antimicrobial resistance: Is a major threat to public health”, BMJ: British Medical Journal *317*, 609.

[2] Welch,T.J., Fricke,W.F., McDermott,P.F., White,D.G., Rosso,M.L., Rasko,D.A., Mammel,M.K., Eppinger,M., Rosovitz,M.J., and Wagner,D. (2007). Multiple Antimicrobial Resistance in Plague: An Emerging Public Health Risk. PLoS ONE *2*.

[3] Livermore,D.M. and Woodford,N. ,2006, “The β-lactamase threat in Enterobacteriaceae, Pseudomonas and Acinetobacter”, Trends in Microbiology *14*, pp. 413-420.

[4] Pitout,J.D.D. and Laupland,K.B., 2008, “Extended-spectrum β-lactamase-producing Enterobacteriaceae: an emerging public-health concern”, The Lancet Infectious Diseases *8*, pp. 159-166.

[5] Zhang,R., Eggleston,K., Rotimi,V., and Zeckhauser,R.J. ,2006," Antibiotic resistance as a global threat: evidence from China, Kuwait and the United States", Global. Health *2*, 6.

[6] Spratt,B.G.,1988, "Hybrid penicillin-binding proteins in penicillin-resistant strains of Neisseria gonorrhoeae", Nature *332*, pp. 173-176.

[7] Jacoby,G.A. and Munoz-Price,L.S., 2005," The new beta-lactamases", N. Engl. J. Med. *352*, pp. 380-391.

[8] Weldhagen,G.F., 2004, "Integrons and beta-lactamases--a novel perspective on resistance",Int. J. Antimicrob. Agents *23*, pp. 556-562.

[9] Arthur,M. and Courvalin,P. ,1993, "Genetics and mechanisms of glycopeptide resistance in enterococci", Antimicrob. Agents Chemother. *37*, pp. 1563-1571.

[10] Cosgrove,S.E. and Carmeli,Y. ,2003," The impact of antimicrobial resistance on health and economic outcomes", Clin. Infect. Dis *36*, pp. 1433-1437.

[11] Bush,K., Jacoby,G.A., and Medeiros,A.A.,1995, "A functional classification scheme for beta-lactamases and its correlation with molecular structure", Antimicrob. Agents Chemother. *39*, pp. 1211-1233.

[12] Ambler,R.P., Coulson,A.F., Frere,J.M., Ghuysen,J.M., Joris,B., Forsman,M., Levesque,R.C., Tiraby,G., and Waley,S.G.,1991, "A standard numbering scheme for the class A beta-lactamases", Biochem. J. *276 (Pt 1)*, pp. 269-270.

[13] Ambler, R.P., 1980, "The structure of beta-lactamases", Philos. Trans. R. Soc. Lond B Biol. Sci. *289*, pp. 321-331.

[14] Jaurin,B. and Grundstrom,T.,1981, "ampC cephalosporinase of Escherichia coli K-12 has a different evolutionary origin from that of beta-lactamases of the penicillinase type", Proc. Natl. Acad. Sci. U. S. A *78*, pp. 4897-4901.

[15] Ouellette,M., Bissonnette,L., and Roy,P.H.,1987, "Precise insertion of antibiotic resistance determinants into Tn21-like transposons: nucleotide sequence of the OXA-1 beta-lactamase gene", Proc. Natl. Acad. Sci. U. S. A *84*, pp. 7378-7382.

[16] Philippon,A., Labia,R., and Jacoby,G.,1989, "Extended-spectrum beta-lactamases", Antimicrob. Agents Chemother. *33*, pp. 1131-1136.

[17] Koch,A.L.,2000, "Penicillin binding proteins, beta-lactams, and lactamases: offensives, attacks, and defensive countermeasures", Crit Rev Microbiol. *26*, pp. 205-220.

[18] Waxman,D.J. and Strominger,J.L.,1983, "Penicillin-binding proteins and the mechanism of action of beta-lactam antibiotics", Annu. Rev Biochem. *52*, pp. 825-869.

[19] Medeiros,A.A.,1997, "Evolution and dissemination of beta-lactamases accelerated by generations of beta-lactam antibiotics", Clin. Infect. Dis *24 Suppl 1*, S19-S45.

[20] Martin,J.F. and Gutierrez,S.,1995, "Genes for beta-lactam antibiotic biosynthesis", Antonie Van Leeuwenhoek *67*, pp. 181-200.

[21] Coque,J.J., Liras,P., and Martin,J.F.,1993, "Genes for a beta-lactamase, a penicillin-binding protein and a transmembrane protein are clustered with the cephamycin biosynthetic genes in Nocardia lactamdurans", EMBO J. *12*, pp. 631-639.

[22] Normark,B.H. and Normark,S., 2002, "Evolution and spread of antibiotic resistance", J. Intern. Med. *252*, pp. 91-106.

[23] Rowe-Magnus,A.D., Davies,J., and Mazel,D., 2002, "Impact of integrons and transposons on the evolution of resistance and virulence", Curr. Top. Microbiol. Immunol. *264*, pp. 167-188.

[24] Moellering,R.C., Jr.,1990, "Interaction between antimicrobial consumption and selection of resistant bacterial strains", Scand. J. Infect. Dis Suppl *70*, pp. 18-24.

[25] Falagas,M.E. and Karveli,E.A.,2006, "World Wide Web resources on antimicrobial resistance", Clin. Infect. Dis *43*, pp. 630-633.

[26] Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S., Arlat,M., Billault,A., Brottier,P., Camus,J.C., and Cattolico,L., 2002, "Genome sequence of the plant pathogen Ralstonia solanacearum", Nature *415*, pp. 497-502.

[27] Quevillon,E., Silventoinen,V., Pillai,S., Harte,N., Mulder,N., Apweiler,R., and Lopez,R., 2005, "InterProScan: protein domains identifier", Nucleic Acids Res. *33*, W116-W120.

[28] Altschul,S.F., Gish,W., Miller,W., Myers,E.W., and Lipman,D.J., 1990, "Basic local alignment search tool", J. Mol. Biol. *215*, pp. 403-410.

[29] Marchler-Bauer,A., Anderson,J.B., Weese-Scott,C., Fedorova,N.D., Geer,L.Y., He,S., Hurwitz,D.I., Jackson,J.D., Jacobs,A.R., Lanczycki,C.J., Liebert,C.A., Liu,C., Madej,T., Marchler,G.H., Mazumder,R., Nikolskaya,A.N., Panchenko,A.R., Rao,B.S., Shoemaker,B.A., Simonyan,V., Song,J.S., Thiessen,P.A., Vasudevan,S., Wang,Y., Yamashita,R.A., Yin,J.J., and Bryant,S.H.,2003, "CDD: a curated Entrez database of conserved domain alignments", Nucleic Acids Res. *31*, pp. 383-387.

[30] Mulder,N.J., Apweiler,R., Attwood,T.K., Bairoch,A., Bateman,A., Binns,D., Bork,P., Buillard,V., Cerutti,L., Copley,R., Courcelle,E., Das,U., Daugherty,L., Dibley,M., Finn,R., Fleischmann,W., Gough,J., Haft,D., Hulo,N., Hunter,S., Kahn,D., Kanapin,A., Kejariwal,A., Labarga,A., Langendijk-Genevaux,P.S., Lonsdale,D., Lopez,R., Letunic,I., Madera,M., Maslen,J., McAnulla,C., McDowall,J., Mistry,J., Mitchell,A., Nikolskaya,A.N., Orchard,S., Orengo,C., Petryszak,R., Selengut,J.D., Sigrist,C.J., Thomas,P.D., Valentin,F., Wilson,D., Wu,C.H., and Yeats,C., 2007, "New developments in the InterPro database", Nucleic Acids Res. *35*, D224-D228.

[31] Hulo,N., Bairoch,A., Bulliard,V., Cerutti,L., de,C.E., Langendijk-Genevaux,P.S., Pagni,M., and Sigrist,C.J., 2006, "The PROSITE database", Nucleic Acids Res. *34*, D227-D230.

[32] Bateman,A., Birney,E., Cerruti,L., Durbin,R., Etwiller,L., Eddy,S.R., Griffiths-Jones,S., Howe,K.L., Marshall,M., and Sonnhammer,E.L. (2002). The Pfam protein families database. Nucleic Acids Res. *30*, 276-280.

[33] Attwood,T.K., Bradley,P., Flower,D.R., Gaulton,A., Maudling,N., Mitchell,A.L., Moulton,G., Nordle,A., Paine,K., Taylor,P., Uddin,A., and

Zygouri,C., 2003, "PRINTS and its automatic supplement, prePRINTS", Nucleic Acids Res. *31*, pp. 400-402.

[34] Bru,C., Courcelle,E., Carrere,S., Beausse,Y., Dalmar,S., and Kahn,D.,2005, "The ProDom database of protein domain families: more emphasis on 3D", Nucleic Acids Res. *33*, D212-D215.

[35] Letunic,I., Copley,R.R., Pils,B., Pinkert,S., Schultz,J., and Bork,P.,2006, "SMART 5: domains in the context of genomes and networks", Nucleic Acids Res. *34*, D257-D260.

[36] Selengut,J.D., Haft,D.H., Davidsen,T., Ganapathy,A., Gwinn-Giglio,M., Nelson,W.C., Richter,A.R., and White,O.,2007, "TIGRFAMs and Genome Properties: tools for the assignment of molecular function and biological process in prokaryotic genomes", Nucleic Acids Res. *35*, D260-D264.

[37] Wu,C.H., Nikolskaya,A., Huang,H., Yeh,L.S., Natale,D.A., Vinayaka,C.R., Hu,Z.Z., Mazumder,R., Kumar,S., Kourtesis,P., Ledley,R.S., Suzek,B.E., Arminski,L., Chen,Y., Zhang,J., Cardenas,J.L., Chung,S., Castro-Alvear,J., Dinkov,G., and Barker,W.C. ,2004, "PIRSF: family classification system at the Protein Information Resource", Nucleic Acids Res. *32*, D112-D114.

[38] Yeats,C., Maibaum,M., Marsden,R., Dibley,M., Lee,D., Addou,S., and Orengo,C.A., 2006, "Gene3D: modelling protein structure, function and evolution", Nucleic Acids Res. *34*, D281-D284.

[39] Mi,H., Guo,N., Kejariwal,A., and Thomas,P.D.,2007, "PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways", Nucleic Acids Res. *35*, D247-D252

[40] Marchler-Bauer,A., Anderson,J.B., Derbyshire,M.K., Weese-Scott,C., Gonzales,N.R., Gwadz,M., Hao,L., He,S., Hurwitz,D.I., Jackson,J.D., Ke,Z., Krylov,D., Lanczycki,C.J., Liebert,C.A., Liu,C., Lu,F., Lu,S., Marchler,G.H., Mullokandov,M., Song,J.S., Thanki,N., Yamashita,R.A., Yin,J.J., Zhang,D., and Bryant,S.H., 2007, "CDD: a conserved domain database for interactive domain family analysis" Nucleic Acids Res. *35*, D237-D240.

[41] Tatusov,R.L., Fedorova,N.D., Jackson,J.D., Jacobs,A.R., Kiryutin,B., Koonin,E.V., Krylov,D.M., Mazumder,R., Mekhedov,S.L., Nikolskaya,A.N., Rao,B.S., Smirnov,S., Sverdlov,A.V., Vasudevan,S., Wolf,Y.I., Yin,J.J., and Natale,D.A., 2003, "The COG database: an updated version includes eukaryotes" BMC. Bioinformatics. *4*, 41