World Journal of Pharmaceutical Sciences
ISSN (Print): 2321-3310; ISSN (Online): 2321-3086
Published by Atom and Cell Publishers
Available online at: http://www.wjpsonline.com/
Research Article

PUTATIVE DRUG TARGET IDENTIFICATION FOR SEPTIC ARTHRITIS THROUGH DATA MINING APPROACH

Anup Tripathi, Sachidanand Singh, Atul Kumar*

Department of Bioinformatics, School of Biotechnology and Health Sciences, Karunya University, Karunya Nagar, Coimbatore, Tamil Nadu, India

Received: 10-10-2013 / Revised: 15-11-2013 / Accepted: 30-11-2013

ABSTRACT

Septic arthritis is the purulent invasion of a joint by an infectious agent which produces arthritis. The main organisms having great potential to infect human beings as well as other mammals are *Staphylococcus aureus, Streptococcus pneumoniae*and *Streptococcus pyrogens*. *In Silico* comparative analysis ofall the pathways of host *Homo sapiens* and pathogens was performed by using KEGG and Protein BLAST. 25, 20 and 16 unique pathways were identified for *Staphylococcus aureus*, *Streptococcus pyrogens*and *Streptococcus pneumonia* respectively. Out of these we identified 3 enzymes for *Staphylococcus aureus*, 4 for *Streptococcus pneumoniae* and 1 for *Streptococcus pyrogens*, which are non-homologous to *Homo sapiens* proteins. The enzymes essential for survival of the pathogens were found out by DEG database. Further CELLO analysis results showed that 50% enzymes are found to be Extracellular, 25% to be cytoplasmic and 25% to be membranous for *Staphylococcus aureus*. For *Streptococcus pneumoniae*, 50% enzymes are found to be Extracellular, 12% cytoplasmic, 13% membranous and 25% as cell wall proteins. 100% enzymes were found to be membranous for *Streptococcus pyrogens*. Finally, the enzymes from DEG were submitted in Drug Bank database to identify approve drug targets. This Data Mining approach found that mostly the enzymes which can act as targets belong to extracellular level in Staphylococcus aureus, Streptococcus pneumoniae and membranous in Streptococcus pyrogens. This finding gives an understanding of these enzymes' interaction with human protein protein interaction at extracellular and membrane level.

Keywords: Septic arthritis, *Staphylococcus aureus, Streptococcus pneumoniae, Streptococcus pyrogens*, essential enzymes, KEGG, DEG, PLSPred, CELLO and Drug Bank

INTRODUCTION

Septic arthritis is a serious disease in which an inflammatory reaction in a joint is caused by an infection. This disease requires emergency surgical intervention because it can cause irreversible and irreparable joint lesions and life threatening conditions in children and adults [1]. Most septic joints develop as a result of hematogenous seeding of the vascular synovial membrane due to a bacteremic episode [2, 3]. The most common infectious agents are *Staphylococcus aureus* [4], *Niesseriagonorrhea* [5], *Streptococcus pneumonia* [3], *Streptococcus pyrogens* [6] and *haemophilus influenza* [7]. Microorganisms may invade the joint via direct inoculation, contiguous spread from infected periarticular tissues, or the bloodstream. The major consequence of bacterial invasion is hyperaemia and oedema of the synovial membranes; the overproduced synovial fluid contains polymorphonuclear leukocytes, which can damage articular cartilage by releasing proteolytic enzymes [8]. The direct introduction of bacteria during joint surgery has increasingly been a source of bacterial arthritis, particularly in association with knee and hip arthroplasties.

In silico subtractive genomics approaches, based on the strategy that an essential survival gene non-homologous to any human host gene is a candidate drug target for a given pathogen, [9, 10] have been used to identify putative drug targets in *Staphylococcus aureus, Streptococcus pneumoniae* and *Streptococcus pyrogens*. In the present study, a similar data mining approach has been carried out to screen these organism's genome and proteome in

***Corresponding Author Address:** Mr. Atul Kumar, Department of Bioinformatics, School of Biotechnology and Health Sciences, Karunya University, Karunya Nagar, Coimbatore, Tamil Nadu, India, E-mail ID: atul.0298@gmail.com*

order to identify its essential genes and subsequent drug and vaccine targets from various metabolic pathways.

MATERIALS AND METHODS

Identification of potential drug targets: For metabolic pathway identification KEGG [11] (Kyoto Encyclopedia of Genes and Genome) (http://www.genome.jp/kegg/) was used. The metabolic pathway identification numbers of the host *Homosapiens* and the pathogens *Staphylococcus aureus, Streptococcus pneumoniae*and *Streptococcus pyrogens* were extracted from the KEGG database (Table 1). According to KEGG database annotation, the pathways which do not appear in host *Homosapiens* but are present in above pathogens have been identified as pathways unique to pathogens in comparison to the host. Enzymes present in these unique pathways as well as the enzymes involved in several metabolic pathways like Carbohydrate Metabolism, Energy Metabolism, Lipid Metabolism, Nucleotide Metabolism, Amino Acid Metabolism, Glycan Biosynthesis and Metabolism and Metabolism of Cofactors and Vitamins were also identified from KEGG database. The protein sequences corresponding to the unique pathways were retrieved from the KEGG database and a BLASTp [12] search against the non-redundant database of these protein sequences were performed. In this search the e-value inclusion threshold was set to 0.005 and the search was restricted to proteins from *Homo sapiens* by excluding *Homo sapiens* in BLASTp. This helps in finding only those targets which do not have detectable human homologues. Thus potential drug targets are obtained by selecting those enzymes which do not have hits below e-value threshold of 0.005 in BLASTp result.

Finding the essential targets: Essential genes are the genes which are indispensable for the survival of an organism and their functions are the foundation of life. The non-homologous enzymes obtained from the pBLAST result were subjected to DEG database [13] (http://tubic.tju.edu.cn/deg/) for analyzing the essentiality of these enzymes to pathogens.

Finding the biological significance of the targets: The non-homologous essential genes which do not show any similarity with any human sequence were considered as putative drug targets. The function and the cellular localization of each protein of identified targets were analyzed with Swiss-Prot protein database (http://us.expasy.org/sprot) and by using sub-cellular localization prediction tools: CELLO [14] and PSLpred [15]. Surface membrane proteins were selected for putative candidate vaccine targets.

RESULTS AND DISCUSSION

From KEGG server all the pathways associated with *S.aureus, Streptococcus pneumoniae* and *Streptococcus pyrogens* were analyzed and each enzyme of these pathways were compared with the proteins of the host *Homo sapiens* with the help of BLASTp search against the non-redundant database restricted to *Homosapiens*. To remove the homologous sequences from the BLASTp search the e-value inclusion threshold was set as 0.005. From KEGG, 15 strains and 25 unique pathways were identified for *Staphylococcus aureus*. Similarly for *Streptococcus pyrogens* 12 strains and 20 unique pathways and for *Streptococcus pneumonia* 14 strains and 16 unique pathways were identified. Finally by performing BLASTp search a total of 3 enzymes for *Staphylococcus aureus*, 4 for *Streptococcus pneumoniae* and 1 for *Streptococcus pyrogens*, which are non-homologous to *Homo sapiens* protein sequences were identified (Table 2). Inhibitors can be designed against these sequences in order to find better drugs.

Dispensable or non-essential enzymes or pathways are not considered to be a good drug target. Thus a DEG server analysis of all the 8 enzymes identified from KEGG server is necessary. To enhance the specificity of these enzymes in various pathogens the cutoff score was set to be > 100.From DEG, a total of 37 essential enzymes were found out and no essential enzymes were detected for SecE (SPY 2058), Sfb1 (SPT 1057) and Sfb1 (SNC 0837) (Table 3).

These enzymes are finally considered as potential drug targets. The subcellular localization analysis of all these enzymes was also performed by two tools: CELLO (http://cello.life.nctu.edu.tw/) and PLSpred(http://www.imtech.res.in/raghava/pslpred/). From the CELLO analysis (Figure 1), 50% enzymes are found to be Extracellular, 25% to be cytoplasmic and 25% to be membranous for *Staphylococcus aureus*. For *Streptococcus pneumoniae* 50% enzymes are found to be Extracellular, 12% cytoplasmic, 13% membranous and 25% as cell wall proteins. 100% enzymes were found to be membranous for *Streptococcus pyrogens.* And the PLSPred analysis (Figure 2) showed that for *Staphylococcus aureus* 33% enzymes are cytoplasmic, 33% enzymes are Peri-plasmic and 34% enzymes are Extracellular. For *Streptococcus pneumonia*, both the Extracellular and Peri-plasmic enzymes were found to be 50% each. 100% enzymes were found to be Inner-

membranous for *Streptococcus pyrogens*. The analysis of the essential enzymes of DEG database result against the drug bank database [16] was done and about 7 approved drug targets were identified. The results are shown in Table 4.

CONCLUSION

Septic arthritis is a serious problem worldwide and and can be life threatening with lethal development of sepsis, meningitis and other diseases, especially caused by *Staphylococcus aureus, Stretococcus pneumonia* and *Streptococcus pyrogens*. For the identification and analysis of the essential genes of these pathogens responsible for Septic arthritis, data mining approach was used. The unique pathways for *Staphylococcus aureus, Stretococcuspyrogens*and *Streptococcus pneumonia*were identified as Staphylococcus aureus infection, Bacterial secretion system and Bacterial invasion of epithelial cells respectively. From these pathways non-detectable human homologous enzymes were identified as TSST-1 (SA 1819), SCIN (SA 0221), SCIN (SA 1004) for *Staphylococcus aureus*; SecE (SPY 2058) for *Stretococcuspyrogens* and Sfb1 (SPJ 0350), Sfb1 (SPCG 0360), Sfb1 (SPT 1057), Sfb1 (SNC 0837) for *Streptococcus pneumonia*. These non-homologous enzymes showed 37 essential genes through DEG database, which confirmed that these enzymes won't perform their activity without the presence of essential genes. The subcellular localization of these enzymes was also found out with the help of CELLO and PLSPred. These enzymes were submitted in Drug Bank database against approve drug targets, which showed their relevance in septic arthritis. This approach gives us an insight to find putative target enzymes which play an important role in causing septic arthritis.

Table 1: DIFFERENT STRAINS AND UNIQUE PATHWAYS OF PATHOGENS IDENTIFIED FROM KEGG

Organisms	**Strains**	**Unique Pathways**
Staphylococcus aureus	Staphylococcus aureus N315 Staphylococcus aureus Mu50 Staphylococcus aureus Mu3 Staphylococcus aureus JH1 Staphylococcus aureus JH9 Staphylococcus aureus MW2 Staphylococcus aureus MSSA476 Staphylococcus aureus MRSA252 Staphylococcus aureus COL Staphylococcus aureus USA300_TCH1516 Staphylococcus aureus USA300_FPR3757 Staphylococcus aureus NCTC8325 Staphylococcus aureus Newman Staphylococcus aureus ED98 Staphylococcus aureus RF122	00660 C5-Branched dibasic acid metabolism 00680 Methane metabolism 00473 D-Alanine metabolism 00550 Peptidoglycan biosynthesis 00906 Carotenoid biosynthesis 00903 Limonene and pinene degradation 00281 Geraniol degradation 00312 beta-Lactam resistance 00521 Streptomycin biosynthesis 00401 Novobiocin biosynthesis 00362 Benzoate degradation 00627 Aminobenzoate degradation 00625 Chloroalkane and chloroalkene degradation 00642 Ethylbenzene degradation 00621 Dioxin degradation 00626 Naphthalene degradation 00624 Polycyclic aromatic hydrocarbon degradation 05150 Staphylococcus aureus infection 05100 Bacterial invasion of epithelial cells 00440 Phosphonate and phosphinate metabolism 00363 Bisphenol degradation 00623 Toluene degradation 00361 Chlorocyclohexane and chlorobenzene degradation 00791 Atrazine degradation 00121 Secondary bile acid biosynthesis
Streptococcus pyrogens	Streptococcus pyogenes SF370 Streptococcus pyogenes MGAS5005	00680 Methane metabolism 00473 D-Alanine metabolism

	Streptococcus pyogenes MGAS8232 Streptococcus pyogenes MGAS315 Streptococcus pyogenes SSI-1 Streptococcus pyogenes MGAS10270 Streptococcus pyogenes MGAS10750 Streptococcus pyogenes MGAS2096 Streptococcus pyogenes MGAS9429 Streptococcus pyogenesManfredo Streptococcus pyogenes MGAS10394 Streptococcus pyogenes MGAS6180 Streptococcus pyogenes NZ131	00550 Peptidoglycan biosynthesis 00903 Limonene and pinene degradation 00521 Streptomycin biosynthesis 00621 Dioxin degradation 00626 Naphthalene degradation 00624 Polycyclic aromatic hydrocarbon degradation 00362 Benzoate degradation 00627 Aminobenzoate degradation 00625 Chloroalkane and chloroalkene degradation 00363 Bisphenol degradation 00660 C5-Branched dibasic acid metabolism 05100 Bacterial invasion of epithelial cells 00440 Phosphonate and phosphinate metabolism 00523 Polyketide sugar unit biosynthesis 00643 Styrene degradation 02060 Phosphotransferase system (PTS) 03070 Bacterial secretion system 02020 Two-component system
Streptococcus pneumoniae	Streptococcus pneumoniae TIGR4 Streptococcus pneumoniae D39 Streptococcus pneumoniae R6 Streptococcus pneumoniae CGSP14 Streptococcus pneumoniae G54 Streptococcus pneumoniae ATCC 700669 Streptococcus pneumoniae Hungary19A 6 Streptococcus pneumoniae 70585 Streptococcus pneumoniae JJA Streptococcus pneumoniae P1031 Streptococcus pneumoniae Taiwan19F-14 Streptococcus pneumoniae TCH8431/19A Streptococcus pneumoniae 670-6B Streptococcus pneumoniae AP200	00791 Atrazine degradation 00680 Methane metabolism 00473 D-Alanine metabolism 00550 Peptidoglycan biosynthesis 00903 Limonene and pinene degradation 00621 Dioxin degradation 00626 Naphthalene degradation 00624 Polycyclic aromatic hydrocarbon degradation 00362 Benzoate degradation 00627 Aminobenzoate degradation 00625 Chloroalkane and chloroalkene degradation 00363 Bisphenol degradation 00660 C5-Branched dibasic acid metabolism 05100 Bacterial invasion of epithelial cells 00440 Phosphonate and phosphinate metabolism 00643 Styrene degradation

Table 2:ENZYMES WHICH ARE NON-HOMOLOGOUS TO THE HOST

Organism	**Unique Pathway**	**Non homologous enzymes**
Staphylococcus aureus	05150	TSST-1 (SA 1819) SCIN (SA 0221) SCIN (SA 1004)
Streptococcus pyrogens	03070	SecE (SPY 2058)
Streptococcus pneumoniae	05100	Sfb1 (SPJ 0350) Sfb1 (SPCG 0360) Sfb1 (SPT 1057) Sfb1 (SNC 0837)

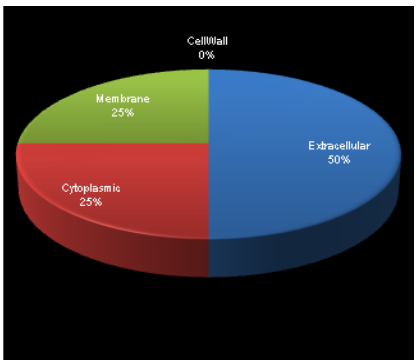
Table 3: ESSENTIAL ENZYMES OF PATHOGENS IDENTIFIED FROM DEG

Non homologous enzymes from KEGG	Essential enzymes from DEG	Gene name	Function Class	Description
toxic shock syndrome toxin-1	DEG10020122	Smc	DNA packaging and segregation	Chromosome segregation SMC protein
toxic shock syndrome toxin-1	DEG10020178	Alas	Protein synthesis	Alanyl-tRNAsynthetase
toxic shock syndrome toxin-1	DEG10020173	dnaK	Post-translational modification, protein turnover, chaperones	Molecular chaperone
toxic shock syndrome toxin-1	DEG10020268	rplN	Protein synthesis	50S ribosomal protein L14
toxic shock syndrome toxin-1	DEG10020103	mutS2	DNA replication/modification and repair	MutS-like protein
toxic shock syndrome toxin-1	DEG10020024	guaA	Nucleotide transport and metabolism	GMP synthase
toxic shock syndrome toxin-1	DEG10170336	SAOUHSC_02571	Function unknown	Secretary antigen precursor
toxic shock syndrome toxin-1	DEG10170214	dnaK	Protein folding	Molecular chaperone DnaK
toxic shock syndrome toxin-1	DEG10170324	rplN	Ribosomal proteins	50S ribosomal protein L14
toxic shock syndrome toxin-1	DEG10170020	guaA	Purine biosynthesis	GMP synthase
toxic shock syndrome toxin-1	DEG10170223	alaS	tRNAsynthetase	Alanyl-tRNAsynthetase
staphylococcal complement inhibitor SA0221	DEG10020145	SA1147	Function unknown	Hypothetical protein
staphylococcal complement inhibitor SA0221	DEG10020122	Smc	DNA packaging and segregation	Chromosome segregation SMC protein
staphylococcal complement inhibitor SA0221	DEG10020195	polA	DNA replication, recombination, and repair	DNA polymerase I
staphylococcal complement inhibitor SA0221	DEG10020297	SA2442	Protein secretion	PreproteintranslocaseSec A homolog
staphylococcal complement inhibitor SA0221	DEG10020039	lysS	Protein synthesis	Lysyl-tRNAsynthetase
staphylococcal complement inhibitor SA0221	DEG10020025	SA0422	Function unknown	Hypothetical protein
staphylococcal complement inhibitor SA0221	DEG10170190	SAOUHSC_01473	Lipids	BirAbifunctional protein
staphylococcal complement inhibitor SA0221	DEG10170159	SAOUHSC_01237	Peptidoglycan biosynthesis	Undecaprenyl pyrophosphate synthase
staphylococcal complement inhibitor SA0221	DEG10170035	lysS	tRNAsynthetase	Lysyl-tRNAsynthetase
staphylococcal complement inhibitor SA0221	DEG10170278	SAOUHSC_02123	DNA replication	ATP-dependent DNA helicase PcrA
staphylococcal complement inhibitor SA1004	DEG10020038	ftsH	Cell division	Cell-division protein
staphylococcal complement inhibitor SA1004	DEG10020195	polA	DNA replication, recombination, and repair (L)	DNA polymerase I
staphylococcal complement inhibitor SA1004	DEG10020207	ccpA	Inorganic ion transport and metabolism (P)/Signal transduction (T)	Catabolite control protein A
staphylococcal complement inhibitor SA1004	DEG10020066	uvrB	DNA replication, recombination, and repair (L)	Exonuclease ABC subunit B
staphylococcal complement inhibitor SA1004	DEG10020174	lepA	General function prediction only (R)	GTP-binding protein
staphylococcal complement inhibitor SA1004	DEG10020151	glcT	transcription antiterminator	Transcription antiterminator
staphylococcal complement inhibitor SA1004	DEG10020030	SA0447	Similar to unknown proteins	Conserved hypothetical protein
staphylococcal complement inhibitor SA1004	DEG10020197	pfk/pfkA	Metabolism of carbohydrates and related molecules	6-phosphofructokinase
staphylococcal complement inhibitor SA1004	DEG10020132	tsf	Protein synthesis	Homolog elongation factor TS
staphylococcal complement inhibitor SA1004	DEG10170129	SAOUHSC_01148	Cell division	Cell division protein

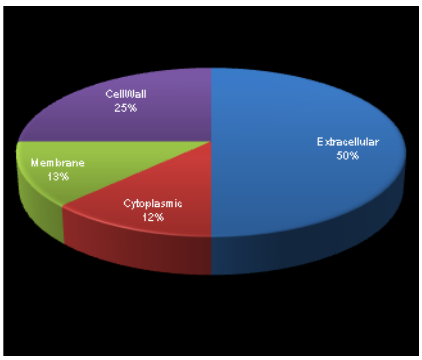
staphylococcal complement inhibitor SA1004	DEG10170157	pyrH	Pyrimidine biosynthesis	Uridylate kinase
staphylococcal complement inhibitor SA1004	DEG10170229	SAOUHSC_01739	Peptidoglycan biosynthesis	Hypothetical protein
staphylococcal complement inhibitor SA1004	DEG10170253	SAOUHSC_01807	Glycolysis	6-phosphofructokinase
fibronectin-binding protein 1 SPJ 0350	DEG10070096	SP_1737	-	DNA-directed RNA polymerase, omega subunit, putative
fibronectin-binding protein 1 SPCG 0360	DEG10070096	SP_1737	-	DNA-directed RNA polymerase, omega subunit, putative
fibronectin-binding protein 1 SPCG 0360	DEG10070127	sulC/folE	Coenzyme metabolism	GTP cyclohydrolase

Table 4: APPROVED DRUG TARGET IDENTIFICATION BY THE DRUG BANK SERVER

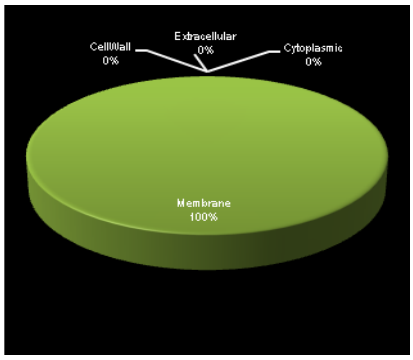
Target Enzyme	**Target Drug**	**Drug Bank ID**	**Target**
toxic shock syndrome toxin-1	Phenobarbital	DB01174	GABA receptor subunit alpha 1
staphylococcal complement inhibitor	Lincomycin	DB01627	50 S ribosomal protein L 10
preproteintranslocase subunit SecE (Secretory System)	Lincomycin	DB01627	50 S ribosomal protein L 10
preproteintranslocase subunit SecE (Secretory System)	Vancomycin	DB00512	DNA
preproteintranslocase subunit SecE (Secretory System)	Clarithromycin	DB01211	GlycosyltransferaseGtfA
fibronectin-binding protein 1	Cefditoren	DB01066	Penecillin binding proteins
fibronectin-binding protein 1	Lincomycin	DB01627	50 S ribosomal protein L 10



(A)

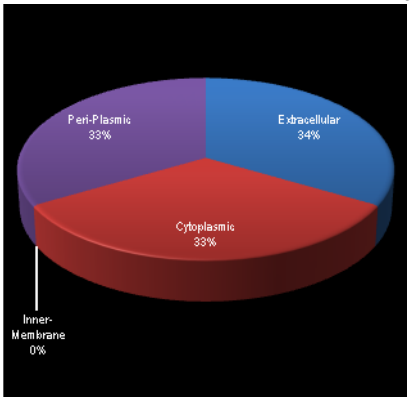


(B)

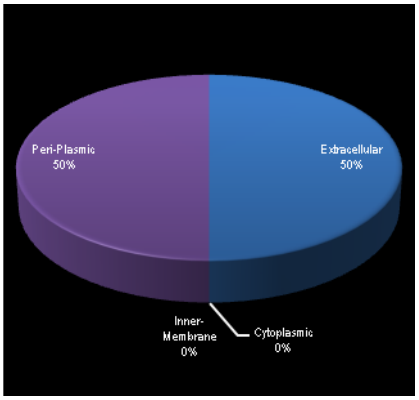


(C)

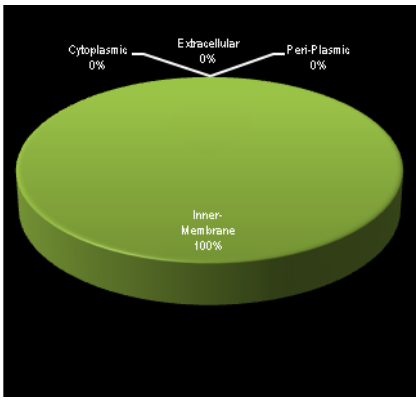
Figure 1: Pie diagram showing subcellular localization of enzymes of *Staphylococcus aureus*(A), *Streptococcus pneumoniae*(B), *Streptococcus pyrogens*(C)by CELLO



(A)



(B)



(C)

Figure 2: Pie diagram showing subcellular localization of enzymes of *Staphylococcus aureus*(A), *Streptococcus pneumoniae*(B), *Streptococcus pyrogens*(C)by PLSPred

REFERENCES

1. Cicak N. Ultrasound system to move. Medicins Ka Naklada 2003; 67-75.
2. Klein RS. Joint infection, with consideration of underlying disease and sources of bacteremia in hematogenous infection. Clinics in GeriatricMedicine 1988; 4(2): 375–394.
3. Morgan DS, Fisher D, Merianos A, Currie BJ. An 18 year clinical review of septic arthritis from tropical Australia. Epidemiology and Infection 1996; 117(3): 423–428.
4. Barton LL, Dunkle LM, Habib FH. Septic arthritis in childhood: A 13-year review.American Journal of Diseases of Children1987; 141(8): 898–900 .
5. O'Brien JP, Goldenberg DL, Rice PA. Disseminated gonococcal infection: a prospective analysis of 49 patients and a review of pathophysiology and immune mechanisms. Medicine (Baltimore) 1983; 62 (6): 395–406.
6. Ryan MJ, Kavanagh R, Wall PG, Hazleman BL. Bacterial joint infections in England and Wales: analysis of bacterial isolates over a four year period.British Journal of Rheumatology 1997; 36 (3): 370–373.
7. Shirtliff ME, Mader JT. Acute Septic Arthritis. Clinical Microbiology Reviews 2002; 15 (4): 527–544.
8. Bialik V, Volpin G, Jerushalmi J, Stein H. Sonography in the diagnosis of painful hips. International Orthopaedics 1991; 15 (2):155-159.
9. Allsop AE. New antibiotic discovery, novel screens, novel targets and impact of microbial genomics. Current Opinion Microbiology 1998; 1 (5): 530-534.
10. Schmid MB, Novel approaches to the discovery of antimicrobial agents. Current Opinionin Chemical Biology 1998; 2 (4): 529-534.
11. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M.KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research 1999; 27 (1): 29-34.
12. Altschul SF, Gish W, Miller W, E.W. Myers EW, Lipman DJ. Basic local alignment search tool. Journal of Molecular Biology 1990; 215 (3):403-410.
13. Zhang R, Ou HY, Zhang CT.DEG: A database of essential genes. Nucleic Acids Research 2004; 32 (1): D271-D272.
14. Yu CS, Chen YC, Lu CH, Hwang JK.Prediction of protein subcellular localization. Proteins: Structure and Function 2006; 64 (3):643-651.
15. Bhasin M, Garg A, Raghava GPS. PSLpred: prediction of subcellular localization of bacterial proteins. Bioinformatics 2005; 21 (10): 2522-2524.
16. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S,Tzur D, Gautam B, Hassanali M, DrugBank: a knowledgebase for drugs, drug actions and drug targets.Nucleic Acids Research 1991;3:901-906.