e-ISSN - 2348-2184
Print ISSN - 2348-2176

AMERICAN JOURNAL OF BIOLOGICAL AND PHARMACEUTICAL RESEARCH

Journal homepage: www.mcmed.us/journal/ajbpr

COMPARATIVE GENOMICS OF HEAT SHOCK PROTEIN IN PATHOGENIC BACTERIA

S. Jenifer[1]* and S. Mannivannan[2]

[1]Assistant Profesor, PG and Research Department of Biotechnology, Marudupandiyar Institutions, Thanjavur, Tamil Nadu, India.
[2]Assistant Professor, PG Research Department of Biotechnology, Bharath College of Science and Management, Thanjavur, Tamilnadu, India.

Article Info

Received 29/01/2017
Revised 16/02/2017
Accepted 19/03/2017

Key words: - HSP genes, FASTA, NCBI/Gen Bank Database , Clastal W, B-galactosidase.

ABSTRACT

Genome comparison has recently become a common research topic in computer science. The development of computer-assisted mathematics (using products such as Mathematica or Matlab) has helped engineers, mathematicians and computer scientists to start operating in this domain, and a public collection of case studies and demonstrations is growing, ranging from whole genome comparisons to gene expression analysis. This has increased the introduction of different ideas, including concepts from systems and control, information theory, strings analysis and data mining. It is anticipated that computational approaches will become and remain a standard topic for research and teaching, while students fluent in both topics start being formed in the multiple courses created in the past few years. In the present investigation, four pathogenic microorganisms have been selected. They are *1.Salmonella typhimurium, 2.Bacillus subtilis, 3.Mycobacterium tuberculosis and, 4.Thermoplasma acidophilum.* In the above organisms, heat shock protein genes (HSP genes) sequences have been retrived from the Bioinformatics databases.

INTRODUCTION

Comparative genomics is the study of relationships between the genomes of different species or strains. Comparative genomics is an attempt to take advantage of the information provided by the signatures of selection to understand the function and evolutionary processes that act on genomes. While it is still a young field, it holds great promise to yield insights into many aspects of the evolution of modern species [1, 2]. The sheer amount of information contained in modern genomes (750 megabytes in the case of humans) necessitates that the methods of comparative genomics are automated. Gene finding is an important application of comparative genomics, as is discovery of new, non-coding functional elements of the genome [3].

Comparative genomics exploits both similarities and differences in the proteins, RNA, and regulatory regions of different organisms to infer how selection has acted upon these elements. Those elements that are responsible for similarities between different species should be conserved through time (stabilizing selection), while those elements responsible for differences among species should be divergent (positive selection). Finally, those elements that are unimportant to the evolutionary success of the organism will be unconserved [4].

Corresponding Author

S. Jenifer
Email:- jsdrjsdr@gmail.com

In the present investigation, four pathogenic microorganisms have been selected. They are

1. Salmonella typhimurium
2. Bacillus subtilis
3. Mycobacterium tuberculosis
4. Thermoplasma acidophilum.

In the above organisms, heat shock protein genes (HSP genes) sequences have been retrived from the Bioinformatics databases.

Heat shock proteins

Heat shock proteins (HSP) are a group of proteins whose expression is increased when the cells are exposed to elevated temperatures or other stress. This increase in expression is transcriptionally regulated. This dramatic upregulation of the heat shock proteins induced mostly by Heat Shock Factor (HSF) is a key part of the heat shock response [5-7].

Bacillus subtilis is a Gram-positive, catalase-positive bacterium commonly found in soil. A member of the genus *Bacillus*, *B. subtilis* has the ability to form a tough, protective endospore, allowing the organism to tolerate extreme environmental conditions [8].

Salmonella typhimurium multiplies in the gastrointestinal tract of many animal species where it usually causes no disease, but in humans its growth causes gastroenteritis. Six to 48 hours after ingestion of contaminated water or food (usually poultry or beef), illness may begin with nausea and vomiting, often followed by diarrhea.

Thermoplasma is a genus of archaea. It belongs to the Thermoplasmata, which thrive in acidic and high-temperature environments. *Thermoplasma* are facultative anaerobes and respire using sulfur and organic carbon. They do not contain a cell wall but instead contain a unique membrane composed mainly of a tetraether lipoglycan containing atypical archaeal tetraether lipid attached to a glucose- and mannose-containing oligosaccharide.

Mycobacterium tuberculosis is the bacterium that causes most cases of tuberculosis [8-10]. *M. tuberculosis*, then known as the "tubercle bacillus," was first described on March 24, 1882 by Robert Koch, who subsequently received the Nobel Prize in physiology or medicine for this discovery in 1905; the bacterium is also known as *Koch's bacillus*.

MATERIALS AND METHODS

Materials

Hardware Processor
Intel Pentium 3928MHz
Ram: 128MB
HardDisk: 40GB
Modem: 56KBPS
CDDrive: 40X Max
FloppyDrive: 3.5”Floppy
Controls: Keyboard/Mouse
Monitor: VGA or Higher resolution
Server: LAN Server with 8 nodes
Software
Windows 98/NT
MS Office

Bioinformatic Tools

Search Engines
(i) Google
(ii) Nucleotide databases
(a) European Molecular Biology Laboratory(EMBL)
(b)NCBI/Gen Bank Database
(c) DNA Database of Japan (DDBJ).
(iii)Methodologies for searching databases:

FASTA and Basic Local Alignment Search Tool (BLAST)

Sequence Analysis

Sequence Database Searching
(i) Keyword Searching
Sequence Retrieval System (SRS) Entrez
(ii) Database Scanning
Basic Local Alignments Searching Tool (BALAST)
(iii)Multiple Sequence Alignments (Clustal w)
(iv)Phylogenetic tree

Literature Databases

Medline and premedline citations (NCBIs Pubmed) Bibliographic services for the UK higher education community (BIDS)
Center for microbial Ecology (CME)
Applied Environmental Microbiology (AEM)
Journal of Bacteriology (JB)
Ribosomal ribonucleic acid databases (dRRND)

European Bioinformatics Institute (EBI)

Traditionally, molecular biology research was carried out entirely at the experimental laboratory bench but the huge increase in the scale of data being produced in this genomic era has seen a need to incorporate computers into this research process.
There are three central biological processes around which bioinformatics tools must be developed.
DNA sequence determines protein sequence.
Protein sequence determines protein sequence.
Protein structure determines protein sequence.

Biological Databases

Biological databases are archives of consistent data that are stored in a uniform and efficient manner. These databases contain data from a broad spectrum of molecular biology areas. Primary or archived databases contain information and annotation of DNA protein sequences, DNA and protein structure and DNA and

protein expression profiles. Secondary or derived databases are so called because they contain the results of analysis on the primary resources including information on sequence patterns or motifs, variants and mutation and evolutionary relationships resources

Tools for Web Search

As information on the web is huge, there are numerous search engines to aid in information search. Several general search engines like GOOGLE, ALTAVISTA, INFOSEEK, HOTBOT, etc are widely used by the internet users [11].

NCBI

Understanding nature's mute but elegant language of living cell is the quest of modern molecular biology. From an alphabet of only four letters representing the chemical subunits of DNA emerges a syntax of life processes whose most complex expression is man. The unraveling and use of this "alphabet" to form new "words and phrases" is a central focus of the field of molecular biology. The staggering volume of molecular data and its cryptic and subtle patterns have le to an absolute requirement for computerized databases and analysis tools. The challenge is in finding new approaches to deal with the volume and complexity of data and in providing researches with better access to analyze and computing tools to advance understanding of our genetic legacy and its role in health and disease [12].

EBI

The European Bioinformatics Institute is a non-profit academic organization that forma part of the European Molecular Biology Laboratory (EMBL). The EMBL is an international network of research institute funded by contribution from seventeen countries and dedicated to research in molecular biology.

SOFTWARE USED

PSI BLAST

Position specific iterative BLAST (PSI-BLAST) refers to a feature of BLAST 2.0 in which a profile (or position specific scoring matrix, PSSM) is constructed (automatically) from a multiple alignment of the highest scoring hits in an initial BLAST search. The PSSM is generated by calculating position-specific scores for each position in the alignment. Highly conserved positions receive high scores and weakly conserved positions receive scores near zero. The profile is used to perform a second (etc.) BLAST search and the results of each "iteration" used to refine the profile. This iterative searching strategy results in increased sensitivity [13].

CLUSTAL W

Clustal W is a general purpose multiple sequence alignment program for DNA or proteins. It produces biologically meaningful multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be seen. Evolutionary relationships can be seen via viewing Cladograms or Phylograms [14].

MATRILS

Salmonella typhimurium

CGCGACCGTACTGGCGCAGTCCATCATTACCGAAG
GCTTGAAAGCCGTTGCTGCGGGCATGAACCCGATG
GACCTGAAACGTGGTATCGACAAAGCGGTTGCTGC
GGCGGTTGAAGAGCTGAAAGCCCTGTCCGTACCGT
GCTCCGACTCTAAAGCGATTGCTCAGGTAGGTACT
ATCTCCGCTAACTCCGACGAAACCGTAGGTAAACT
GATCGCGGAAGCGATGGATAAAGTCGGTAAAGAA
GGCGTCATCACCGTTGAAGACGGTACCGGTCTGCA
G
GACGAACTGGACGTGGTTGAAGGTATGCAGTTTGA
CCGCGGCTACCTGTCTCCTTACTTCATCAACAAGC
CGGAAACTGGCGCAGTAGAGCTGGAAAGCCCGTT
CATCCTGCTGGCTGATAAGAAAATCTCCAACATCC
G
CGAAATGCTGCCGGTTCTGGAAGCCGTTGCAAAAG
CAGGCAAACCGCTGCTGATCATCGCTGAAGATGTT
GAAGGCGAAGCGCTGGCTACCCTGGTTGTTAACAC
CATGCGTGGCATCGTGAAAGTGGCTGCTGTT

Mycobacterium tuberculosis

ATGGTGTGTCCATCGCCAAGGAGATCGAGCTGGA
GGATCCGTACGAGAAGATCGGCGCCGAGCTGGTC
AA
AGAGGTAGCCAAGAAGACCGATGACGTCGCCGGT
GACGGCACCACGACGGCCACCGTGCTGGCCCAGG
CG
TTGGTTCGCGAGGGCCTGCGCAACGTCGCGGCCGG
CGCCAACCCGCTCGGTCTCAAACGCGGCATCGAAA
AGGCCGTGGAGAAGGTCACCGAGACCCTGCTCAA
GGGCGCCAAGGAGGTCGAGACCAAGGAGCAGATT
GC
GGCCACCGCAGCGATTTCGGCGGGTGACCAGTCCA
TCGGTGACCTGATCGCCGAGGCGATGGACAAGGT
G
GGCAACGAGGGCGTCATCACCGTCGAGGAGTCCA
ACACCTTTGGGCTGCAGCTCGAGCTCACCGAGGGT
A
TGCG

Bacillus subtilis

AGGTGTAGGTAAAACGGCTATCGCAGAAGGTTTG
GCACAGCAAATTATCAATAATGAAGTACCCGAAA
TT

TTGCGTGATAAACGTGTGATGACATTAGACATGGG
AACAGTTGTTGCCGGCACAAAATACCGCGGAGAA
T
TTGAGGATCGCCTGAAGAAGGTCATGGATGAAATT
CGCCAGGCAGGAAATATCATTCTATTCATCGATGA
GCTCCATACATTAATCGGGGCAGGCGGAGCAGAA
GGTGCTATTGATGCATCTAATATTTTAAAACCTTC
A
CTTGCTCGTGGCGAACTCCAATGTATTGGTGCAAC
GACTCTTGATGAGTACCGTAAATATATTGAAAAAG
ATGCAGCACTGGAACGCCGTTTTCAGCCGATTCAG
GTTGATCAGCCATCTGTAGATGAAAGTATTCAAAT
TTTACAAGGTCTGCGTGACAGATACGAAGCCCACC
ACCGCGTTTCTATCACTGATGATGCCATTGAAGCT
GCGGTTAAGCTTTCTGACAGATATATTTCTGACCG
CTTCCTT

Thermoplasma acidophilum
CCCTTAACGATGGCTGTCCGCACTTCGAAGGTAAT
CGTTGTCAAAAGGTGATGATGTATGTCAAAGATAA
TTGGTATTGATCTGGGTACAAGCAATTCTGCTGCT
GCAGTTGTGATATCGGGGAAGCCAACCGTGATCCC
AAGCTCGGAGGGGGTATCGATAGGAGGCAAGGCT
TTCCCCAGCTATGTCGCATTTACGAAGGATGGGCA
G
ATGCTTGTGGGCGAACCTGCGAGGAGACAGGCTCT
ACTCAATCCAGAAGGCACCATATTCGCAGCAAAG
A
GAAAGATGGGTACAGATTACAAGTTCAAGGTTTTT
GATAAGGAGTTCACGCCTCAGCAGATCTCTGCATT
CATACTTCAGAAGATAAAAAAGGATGCTGAGGCC
TTCCTCGGCGAACCAGTGAATGAAGCTGTTATAAC
T
GTGCCCGCTTATTTCAATGATAATCAGAGGCAGGC
AACCAAAGATGCCGGTACAATAGCTGGCTTCGATG
TTAAGAGAATAATAAATGAACCAACAGCCGCTGC
ACTCGCCTATGGTGTAGATAAGAGCGGGAAATCC
GA
AAAGATCCTCGTTTTCGATCTCGGAGGGGGAACTC
TGGATGTTACGATAATGGATTTCGGTGATGCCGTT
TTCCAGGTGCTTTCAACATCCGGCGACACAAGGCT
TGGAGGTACTGACATGGACGAGGCCATCGTCAACT
ATATAGCCGATGACTTCCAGAAGAAGGAGGGTAT
AGACCTCAGAAAGGATCGATCCGCCTACATAAGG
CT
GAGGGATGCGGCTGAAAAGGCCAAGATAGAACTT
TCAACTACGCTCTCAACAGATATCGATCTGCCGTA
C
ATAACGGTAACAAACAGCGGGCCAAAACACATAA
AGATGACGCTCACAAGGGCAAAGCTAGAAGAGCT
AT
ATTCTCCAATAGTTGAGAGGGTGAAAGGCCCGATA
GACAAGGCTCTTGAAGGCGCAAAGCTCAAGAAAA
C
CGAGATCACAAAGCTGCTATTCGTGGGCGGGCCG
ACCAGGATACCATATGTTAGGAAATATGTTGAGGA
T
TACCTTGGCATAAAGTCGCCGGAGGGTGGAGTGG
ACCCGATGGAAGCTGTGGCCATCGGTGCTGCAATA
C
AGGGCGCAGTCCTAAAGGGAGAGATAAAAGACAT
CGTTCTGCTGGATGTGACCCCTGTCACGCTCAGCG
T
TGAAACGCTTGGTGGCATCGCAACCCCGATAATTC
CTGCAAACACCACCATACCGGTGAGAAAGAGCCA
G
ATATTTACGACAGCTGAGGACATGCAGACAACGG
TCACCATACACGTGGTGCAGGGTGAGAGGCCGCTC
G
CGAAGGATAACGTTTCGCTGGGTATGTTCAATCTT
ACCGGAATAGCGCCGGCGCCAAGGGGCGTTCCAC
A
GATAGAGGTTACGTTCGATATGCACTCAAACGGCA
TTCTGAACGTGACCGCGGTTGACAAGGCTACTGGA
AAGAAGCAGGGTATAACGATAACGGCTTCCACGA
AGCTCTCCAAAGAGGAGATAGAGAGGATGAAGAA
AG
AAGCCGAGCAATACGCTGAGCAGGACAGAAAGGC
GAAGGAACAGATAGAACTGCTAAACAATGCAGAG
TC
TTTAGCTTACAGTGTTGAGAAGAGCCTGAAGCATG
CTGGAGACAAGGTGGACAAGGAGACTAAGGAAAG
G
CTGACCAACGAGGTAAAGGATCTGAGAAAGGCCA
TAGAGGAGAAGAACACGGAGAACGTAAAGACGCT
GA
TGGACAAGCTGTCAAAGGACATACAGGAAGTCGG
GGCCAAGATGTACCAGCAGGCTTCAGCGAACACC
CA
GCAGAGTGCACAGTCAAACAGCCA.

RESULT and DISCUSSION

The present investigation was carriedout in pathogenic bacteria and Thermophilic bacteria. In all the four bacteria, stress protein called heat shock proteins producing genes were taken in to consideration for analysis.

In the genomic analysis gene annotation, GC contents, pairs of nucleotide contents, multiple sequence alignment, pairwise alignment using Needle Wunch algoritum, pairwise alignment using smith Waterman Algoritum.

The gene annotation gives the four nitrogenous bases namely adinine, guanine, cytosine, and thiamin in heat shock protein(HPS)gene for the four pathogen.

Percentage of GC contents:

In the presence study for the analysis of GC content in various genera like *Mycobacterium, Salmonella, Bacillus* and *Thermoplasma* were selected. The GC contents always helpful in delimiting the genera of microbes. In *Mycobacterium* the percentage of GC content was 65% in *Salmonella* 54% in *Thermoplasma* 50% and in *Bacillus* it was only 44%.

Multiple Sequence Alignment

Clastal W is a typical alignment tool used to align the sequences of various organisms. Using this alignment programme phylogenetic tree through cadogram could be constructed.

Multiple sequence alignment through clastalW showed the close relarionship between *Salmonella typimurium* and *Mycobacterium tuberclosis*.

Pair Wise Alignment:

Pair wise alignment using Needle Wunch algoritum showed 45%similarity among the two microbes Salmonella and *Mycobacterium tuberclosis*.

Pair wise alignment using Smith Waterman algorithms showed 48.7% among the above two organisms

Phylogenetic Tree View of Pathogenic Microbes

In the clastalW program cladogram was contracted based on the percentage of homology found among the three microorganisms. In the phylogenetic tree, the short branches indicated more percentage of similarity while the long branches indicated less percentage of homology in the phylogenitic tree.

Salmonella typimurium and *Mycobacteriam tuberclosis* showed shout branches in the phylogram were as Bacillus indicated distant relationship having Long branch [15].

Bacteria respond to environmental stimuli. It was originally recognized by monitoring gross phenotypic, biochemical and behavioral changes. Much of the genetic basis for how bacteria change their phenotypic was established 1960s and 1870s for B-galactosidase regulation in *E.coli* by Jacob and monad.

The bacterial responses became apparent through the use of global analytical approaches. At the translational level, the use of two-dimensional gel electrophoresis has established the proteomic approach. This technique reveals and separates most of the several hundred proteins which are being synthesized by the bacteria. The different proteins detected represent those proteins that the organisms require to function in the circumstances from which the sample was drawn. The different types of protein synthesized are in response to all most any environment change.

This finding helps to recognize how different phenotype of a single organism can be in different physiological states. More recently the development of DNA-arrays has enabled global analysis of responses at the transcriptional level by detecting mRNA molecules relating to every gene in the organism in a single analysis.

Specific sublethal and noxious stimuli a gene expression are the subject of intense current study. Each different stimulus leads to an adoptive stress response heat shock protein (the effects of the rising temperatures of 45C and above for a few minutes) has been studied most extensively. The newly synthesized proteins are called as heat shock proteins (HSPs).

Environmental factors induce microbial organisms to produce heat shock proteins by their HSP gene, when they are subjected to extreme conditions [16]. The length of the HSP protein gene is 424 in *Mycobacterium tuberculosis*. This organism is having the HSP gene shorterthan any other organisms. The phylogram of clastalw analysis indicates a close relationship between salmonella and mycobacterium, eventhough the length of the HSP protein is different. *Bacillus* eventhough having similar length of gene to that of the *Salmonella*, it differs in the protein production. *Thermoplasma acitophilum* an organisms living in an extreme condition of heat. In order to suit the extreme environment, the length of the heat shock protein gene is long when c to other three microorganisms. The clustalw analysis also indicates that *Thermoplasma* is an outstanding organism among the four. The clustalw analysis, on the whole gives a clear picture stating that salmonella and *Mycobacterium* are closely related while that of *Bacillus* and *Thermoplasma* forming a separate group.

In molecular taxonomy of microorganisms, the GC content of the particular gene place anImportant role in delimiting the various genera of microbes. In the present study also the GC content of the heat shock protein gene also show distinct percentage of similarity for every genus (*Bacillus*=44, *Thermoplasma*=50. *Salmonella*=54, *Mycobacterium*=65)[fig:4.2].

CLUSTAL 2.0.8 multiple sequence alignment

```
Salmonella      ------------------------------------------------------------
Mycobacterium   ------------------------------------------------------------
Thermoplasma    CCCTTAACGATGGCTGTCCGCACTTCGAAGGTAATCGTTGTCAAAAGGTGATGATGTATG 60
Bacillus        ------------------------------------------------------------

Salmonella      ------------------------------------------------------------
Mycobacterium   ------------------------------------------------------------
Thermoplasma    TCAAAGATAATTGGTATTGATCTGGGTACAAGCAATTCTGCTGCTGCAGTTGTGATATCG 120
Bacillus        ------------------------------------------------------------
```

```
Salmonella      ------------------------------------------------------------
Mycobacterium   ------------------------------------------------------------
Thermoplasma    GGGAAGCCAACCGTGATCCCAAGCTCGGAGGGGGTATCGATAGGAGGCAAGGCTTTCCCC 180
Bacillus        ------------------------------------------------------------
Salmonella      ------------------------------------------------------------
Mycobacterium   ------------------------------------------------------------
Thermoplasma    AGCTATGTCGCATTTACGAAGGATGGGCAGATGCTTGTGGGCGAACCTGCGAGGAGACAG 240
Bacillus        ------------------------------------------------------------
Salmonella      ------------------------------------------------------------
Mycobacterium   ------------------------------------------------------------
Thermoplasma    GCTCTACTCAATCCAGAAGGCACCATATTCGCAGCAAAGAGAAAGATGGGTACAGATTAC 300
Bacillus        ------------------------------------------------------------
Salmonella      ------------------------------------------------------------
Mycobacterium   ------------------------------------------------------------
Thermoplasma    AAGTTCAAGGTTTTTGATAAGGAGTTCACGCCTCAGCAGATCTCTGCATTCATACTTCAG 360
Bacillus        ------------------------------------------------------------
Salmonella      ------------------------------------------------------------
Mycobacterium   ------------------------------------------------------------
Thermoplasma    AAGATAAAAAAGGATGCTGAGGCCTTCCTCGGCGAACCAGTGAATGAAGCTGTTATAACT 420
Bacillus        ------------------------------------------------------------
Salmonella      ------------------------------------------------------------
Mycobacterium   ------------------------------------------------------------
Thermoplasma    GTGCCCGCTTATTTCAATGATAATCAGAGGCAGGCAACCAAAGATGCCGGTACAATAGCT 480
Bacillus        ------------------------------------------------------------
Salmonella      ------------------------------------------------------------
Mycobacterium   ------------------------------------------------------------
Thermoplasma    GGCTTCGATGTTAAGAGAATAATAAATGAACCAACAGCCGCTGCACTCGCCTATGGTGTA 540
Bacillus        ------------------------------------------------------------
Salmonella      ------------------------------------------------------------
Mycobacterium   ------------------------------------------------------------
Thermoplasma    GATAAGAGCGGGAAATCCGAAAAGATCCTCGTTTTCGATCTCGGAGGGGGAACTCTGGAT 600
Bacillus        ------------------------------------------------------------
Salmonella      ----------CGCGACCGTACTGGCGCAGTCCATCATTAC-CGAAGGCTT 39
Mycobacterium   ------------------------------------------------------------
Thermoplasma    GTTACGATAATGGATTTCGGTGATGCCGTTTTCCAGGTGCTTTCAACATC-CGGCGACAC 659
Bacillus        ----------AGGTGTAGGTAAAACGGCTATCGCAG 26
Salmonella      GAAAGCCGTTGCTGCGGGC-ATGAACCCGATGGACCTGAAACGTGGTATCGACAAAGCGG 98
Mycobacterium   ----------ATGGTGTGT 9
Thermoplasma    AAGGCTTGGAGGTACTGAC-ATGGACGAGGCCATCGTCAACTATATAGCCGATGACTTCC 718
Bacillus        AAGGTTTGGCACAGCAAATTATCAATAATGAAGTACCCGAAATTTTGCGTGATAAACGTG 86
Salmonella      TTGCTGCGGCGGTTGAAGAGCT---GAAAGCCCTGTCCG--TACCGTGCTCCGACTCTAA 153
Mycobacterium   CCATCGCCAAGGAGATCGAGCT---GGAGGATCCGTACG--AGAAGA---TCGGCGCCGA 61
Thermoplasma    AGAAGAAGGAGGGTATAGACCTCAGAAAGGATCGATCCGCCTACATAAGGCTGAGG--GA 776
Bacillus        TGATGACATTAGACATGGGAAC-----AGTTGTTGCCGGCACAAAATACCGCGGAGAATT 141
                  *   *       *       *         *
Salmonella      AGCGATTGCTCAGGT--AGGTACTATCTCCGCTAACTCCGACGA-AACCGTAGGTAAACT 210
Mycobacterium   GCTGGTCAAAGAGGT--AGCCAAGAAGACCGATGACGTCGCCGG-TGACGGCACCACGAC 118
Thermoplasma    TGCGGCTGAAAAGGCCAAGATAGAACTTTCAACTACGCTCTCAACAGATATCGATCTGCC 836
Bacillus        TGAGGATCGCCTGAAGAAGGTCATGGATGAAATTCGCCAGGCAGGAAATATCATTCTATT 201
              *     *   **              *
Salmonella      GATCGCGGAAGCGATGGATA--------AAGTCGGTAAAGAAGGCGTCATCACCGTTGAA 262
Mycobacterium   GGCCACCGTGCTGGCCCAGG--------CGTTGGTTCGCGAGGGCCTGCGCAACGTCGCG 170
Thermoplasma    GTACATAACGGTAACAAACAGCGGGCCAAAACACATAAAGATGACGCTCACAAGGGCAAA 896
Bacillus        CATCGATGAGCTC-CATACATTAA-TCGGGGCAGGCGGAGCAGAAGGTGCTATTGATGCA 259
              *         *             * *      * *
```

```
Salmonella      GACGGTACCGGTCTGCAGGA---CGAACTGGACGTGGTTGAAGGT-ATGC-AGTTTGACC 317
Mycobacterium   GCCGGCGCCAACCCGCTCGGTCTCAAACGCGGCATCGAAAAGGCC-GTGG-AGAAGGTCA 228
Thermoplasma    GCTAGAAGAGCTATATTCTCCAATAGTTGAGAGGGTGAAAGGCCCGATAG-ACAAGGCTC 955
Bacillus        TCTAATATTTTAAAACCTTCACTTGCTCGTGGCGAACTCCAATGTATTGGTGCAACGACT 319
                          *           *       *

Salmonella      GCGGCTACCTGTCTCCTTACTTCATCAA-----CAAGCCGGAAACTGGCGCAGTAGAGC- 371
Mycobacterium   CCGAGACCCTG---CTCAAGGGCGCCAA-----GGAGGTCGAGACCAA-GGAGCAGATTG 279
Thermoplasma    TTGAAGGCGCAAAGCTCAAGAAAACCGAGATCACAAAGCTGCTATTCGTGGGCGGGCCGA 1015
Bacillus        CTTGATGAGTAC--CGTAAATATATTGAAA---AAGATGCAGCACTGGAACGCCGTTTTC 374
                    *  *      *           *

Salmonella      TGGAAAGCCCGTTCATCCTGCTGGCTGATAAGAAAATCTCCAACAT-------CCGCGAA 424
Mycobacterium   CGGCCACCGCAGCGATTTCGGCGGGTGACCAG----TC-----CAT-------CGGTGAC 323
Thermoplasma    CCAGGATACCATATGTTAGGAAATATGTTGAGGATTACCTTGGCATAAAGTCGCCGGAGG 1075
Bacillus        AGCCGATTCAGGTTGATCAGCCATCTGTAGATGAAAGTATTCAAATT------TTACAAG 428
                *         *    **  *          **

Salmonella      ATGCTGCCGGTTC--TGGAAGCCGTTGCAAAAGCAG--GCAA-ACCGCTGCTGATCATCG 479
Mycobacterium   ------CTGATCG--CCGAGGCGATGGACAAGGTGG--GCAA-CGAGGGCGTCATCACCG 372
Thermoplasma    GTGGAGTGGACCCGATGGAAGCTGTGGCCATCGGTGCTGCAATACAGGGCGCAGTCCTAA 1135
Bacillus        GTCTGCGTGACAGATACGAAGCC--CACCACCGCGTTTCTATCACTGATGAT-GCCATTG 485
                  *      ** **     * *      *    *      *

Salmonella      CTGAAGATGTTGAAGGCGAAGCGCTGGCTACCCTGGTTGTTAACACCATGCGTGGCATCG 539
Mycobacterium   TCGAGGAGTCCAACACCTTTGGGCTGCAGCTCGAGCTCACCGAGGGTATGCG-------- 424
Thermoplasma    AGGGAGAGATAAAAGACATCGTTCTGCTGGATGTGACCCCTG---TCACGCTCAGCGTTG 1192
Bacillus        AAGCTGCGGTTAAG-----CTTTCTGACAGATAT-ATTTCTGACCGCTTCCTT------- 532
              * *    *       ***              *

Salmonella      TGAAAGTGGCTGCTGTT---------------------------------------- 556
Mycobacterium   ----------------------------------------------------------
Thermoplasma    AAACGCTTGGTGGCATCGCAACCCCGATAATTCCTGCAAACACCACCATACCGGTGAGAA 1252
Bacillus        ----------------------------------------------------------
Salmonella      ----------------------------------------------------------
Mycobacterium   ----------------------------------------------------------
Thermoplasma    AGAGCCAGATATTTACGACAGCTGAGGACATGCAGACAACGGTCACCATACACGTGGTGC 1312
Bacillus        ----------------------------------------------------------
Salmonella      ----------------------------------------------------------
Mycobacterium   ----------------------------------------------------------
Thermoplasma    AGGGTGAGAGGCCGCTCGCGAAGGATAACGTTTCGCTGGGTATGTTCAATCTTACCGGAA 1372
Bacillus        ----------------------------------------------------------
Salmonella      ----------------------------------------------------------
Mycobacterium   ----------------------------------------------------------
Thermoplasma    TAGCGCCGGCGCCAAGGGGCGTTCCACAGATAGAGGTTACGTTCGATATGCACTCAAACG 1432
Bacillus        ----------------------------------------------------------
Salmonella      ----------------------------------------------------------
Mycobacterium   ----------------------------------------------------------
Thermoplasma    GCATTCTGAACGTGACCGCGGTTGACAAGGCTACTGGAAAGAAGCAGGGTATAACGATAA 1492
Bacillus        ----------------------------------------------------------
Salmonella      ----------------------------------------------------------
Mycobacterium   ----------------------------------------------------------
Thermoplasma    CGGCTTCCACGAAGCTCTCCAAAGAGGAGATAGAGAGGATGAAGAAAGAAGCCGAGCAAT
1552
Bacillus        ----------------------------------------------------------
Salmonella      ----------------------------------------------------------
Mycobacterium   ----------------------------------------------------------
Thermoplasma    ACGCTGAGCAGGACAGAAAGGCGAAGGAACAGATAGAACTGCTAAACAATGCAGAGTCTT
1612
Bacillus        ----------------------------------------------------------
```

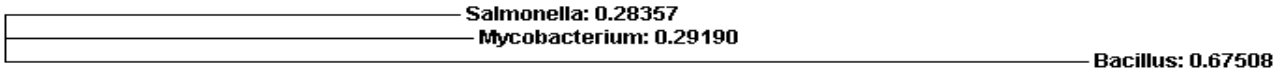
```
Salmonella      ----------------------------------------------------------
Mycobacterium   ----------------------------------------------------------
Thermoplasma    TAGCTTACAGTGTTGAGAAGAGCCTGAAGCATGCTGGAGACAAGGTGGACAAGGAGACTA 1672
Bacillus        ----------------------------------------------------------
Salmonella      ----------------------------------------------------------
Mycobacterium   ----------------------------------------------------------
Thermoplasma    AGGAAAGGCTGACCAACGAGGTAAAGGATCTGAGAAAGGCCATAGAGGAGAAGAACACGG
1732
Bacillus        ----------------------------------------------------------
Salmonella      ----------------------------------------------------------
Mycobacterium   ----------------------------------------------------------
Thermoplasma    AGAACGTAAAGACGCTGATGGACAAGCTGTCAAAGGACATACAGGAAGTCGGGGCCAAGA
1792
Bacillus        ----------------------------------------------------------
Salmonella      ------------------------------------------------
Mycobacterium   ------------------------------------------------
Thermoplasma    TGTACCAGCAGGCTTCAGCGAACACCCAGCAGAGTGCACAGTCAAACAGCCA 1844
Bacillus        ------------------------------------------------
```

Guide Tree

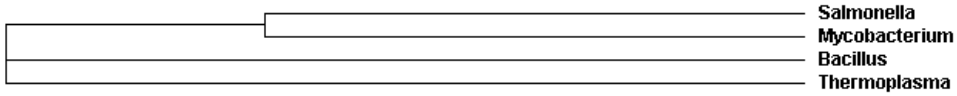
Show as Cladogram Tree | Show Distances | View DND File

```
(
Salmonella:0.28357,
Mycobacterium:0.29190,
Bacillus:0.67508);
```

Phylogram



Cladogram



Pair Wise Alignment using Needle Wunch Algorithm

Aligned_sequences: 2

- Mycobacterium
- Salmonella

Matrix: EBLOSUM62
Gap_penalty: 10.0
Extend_penalty: 0.5
Length: 587
Identity: 272/587 (46.3%)
Similarity: 272/587 (46.3%)

```
Gaps:        194/587 (33.0%)
Score: 1108.0
Mycobacterium     1          ATGGTG-TGTCCATC---GCCAAGGAGATCGAGCTGGAGG     36
                             .|||.| .|||||||   .||.|.|       |||.|| .
Salmonella        1 CGCGACCGTACTGGCGCAGTCCATCATTACCGAAG-------GCTTGA-A     42

Mycobacterium     37 ATCCG----TACGAG-AAGATCGGCGCCGAGCTGGTC----AAAGAGGTA     77
                     |.|||    |.||.| |.||.|    ||||  |||.|    ||.|.||||
Salmonella        43 AGCCGTTGCTGCGGGCATGAAC----CCGA--TGGACCTGAAACGTGGTA     86

Mycobacterium     78 GCCAAGAAGACCGATGACGTCGCCGG-TGACG-GCACCACGAC--GGCC-    122
                     .|.|..|||  ||.|..|..||.||| |||.| ||...|.|.|  |.||
Salmonella        87 TCGACAAAG--CGGTTGCTGCGGCGGTTGAAGAGCTGAAAGCCCTGTCCG    134

Mycobacterium    123 -ACCGTGCT--GGCCC--AGGCGTTGGTTCGCGAGGG--C---CTGCGCA    162
                      ||||||||  |.|.|  |.|||.|.|.||..|..||  |   ||.|||.
Salmonella       135 TACCGTGCTCCGACTCTAAAGCGATTGCTCAGGTAGGTACTATCTCCGCT    184

Mycobacterium    163 ACGTCGCGGCCGGCGCCAACCCGCTCGGT---CTCAAACGCGGCATCGA-    208
                     |..|    |||.||  ||.||| |.|||   || .|.|||||.|.|||
Salmonella       185 AACT-----CCGACG--AAACCG-TAGGTAAACT-GATCGCGGAAGCGAT    225

Mycobacterium    209 ----AAAGGCCGTGGAGAAGG-----TCACCG----AGAC----CCTG-C    240
                     ||||.|.||..||||||     ||||||    ||||    ||.| |
Salmonella       226 GGATAAAGTCGGTAAAGAAGGCGTCATCACCGTTGAAGACGGTACCGGTC    275

Mycobacterium    241 TCAAGGGCGCCAAGGAGGTCGAGACCAAGG--AGCAGATT----GCGGCC    284
                     |..|||.||....|||.||  .|...||||  .||||.||    |||||.
Salmonella       276 TGCAGGACGAACTGGACGT--GGTTGAAGGTATGCAGTTTGACCGCGGCT    323

Mycobacterium    285 ACC----------------------GC--------AGCGATTTCG-GCGG    303
                     |||                    ||       .|||...|.| ||.|
Salmonella       324 ACCTGTCTCCTTACTTCATCAACAAGCCGGAAACTGGCGCAGTAGAGCTG    373

Mycobacterium    304 GTGA-CCAGTCCATC------GGTGA-----------CCTG-ATC-----    329
                     |..| ||.||.||||      |.|||          ||.. |||
Salmonella       374 GAAAGCCCGTTCATCCTGCTGGCTGATAAGAAAATCTCCAACATCCGCGA    423

Mycobacterium    330 ------GCC-------GAGGCGATGGACAAGGTGGGCAA--CGAGGGCG-    363
                       |||      ||.||..|.|..||.|..|||||  ||..|..|
Salmonella       424 AATGCTGCCGGTTCTGGAAGCCGTTGCAAAAGCAGGCAAACCGCTGCTGA    473

Mycobacterium    364 TCATCACCGTCGA----GGAGTCCAACACCTTTGGGCTG--CAGCTCGAG    407
                     |||||.|.|..||    |.||.|.||...||  ||||.  |.|.|.|..
Salmonella       474 TCATCGCTGAAGATGTTGAAGGCGAAGCGCT---GGCTACCCTGGTTGTT    520

Mycobacterium    408 CTCACCGAG---GGTATGCG                424
                     ..|||..|   ||.|| ||
Salmonella       521 AACACCATGCGTGGCAT-CGTGAAAGTGGCTGCTGTT    556
```

Pair Wise Alignment using Smith waterman Algorithm

Aligned_sequences: 2
1: Mycobacterium
2: Salmonella
Matrix: EBLOSUM62

Gap_penalty: 10.0
Extend_penalty: 0.5
Length: 559
Identity: 272/559 (48.7%)
Similarity: 272/559 (48.7%)
Gaps: 167/559 (29.9%)
Score: 1108.0

```
Mycobacterium     2 TGGTG-TGTCCATC---GCCAAGGAGATCGAGCTGGAGGATCCG----TA     43
                    |||.| .|||||||   .||.|.|       |||.|| .|.|||    |.
Salmonella       12 TGGCGCAGTCCATCATTACCGAAG-------GCTTGA-AAGCCGTTGCTG     53

Mycobacterium    44 CGAG-AAGATCGGCGCCGAGCTGGTC----AAAGAGGTAGCCAAGAAGAC     88
                    ||.| |.||.|    ||||  |||.|    ||.|.||||.|.|..|||
Salmonella       54 CGGGCATGAAC----CCGA--TGGACCTGAAACGTGGTATCGACAAAG--     95

Mycobacterium    89 CGATGACGTCGCCGG-TGACG-GCACCACGAC--GGCC--ACCGTGCT--    130
                    ||.|..|..||.||| |||.| ||...|.|.|  |.||  ||||||||
Salmonella       96 CGGTTGCTGCGGCGGTTGAAGAGCTGAAAGCCCTGTCCGTACCGTGCTCC    145

Mycobacterium   131 GGCCC--AGGCGTTGGTTCGCGAGGG--C---CTGCGCAACGTCGCGGCC    173
                    |.|.|  |.|||.|.|.||..|..||  |   ||.|||.|..|     ||
Salmonella      146 GACTCTAAAGCGATTGCTCAGGTAGGTACTATCTCCGCTAACT-----CC    190

Mycobacterium   174 GGCGCCAACCCGCTCGGT---CTCAAACGCGGCATCGA-----AAAGGCC    215
                    |.||  ||.||| |.|||   || .|.|||||.|.|||     ||||.|.
Salmonella      191 GACG--AAACCG-TAGGTAAACT-GATCGCGGAAGCGATGGATAAAGTCG    236

Mycobacterium   216 GTGGAGAAGG-----TCACCG----AGAC----CCTG-CTCAAGGGCGCC    251
                    ||..||||||     ||||||    ||||    ||.| ||..|||.||..
Salmonella      237 GTAAAGAAGGCGTCATCACCGTTGAAGACGGTACCGGTCTGCAGGACGAA    286

Mycobacterium   252 AAGGAGGTCGAGACCAAGG--AGCAGATT----GCGGCCACC--------    287
                    ..|||.||  .|...||||  .||||.||    |||||.|||
Salmonella      287 CTGGACGT--GGTTGAAGGTATGCAGTTTGACCGCGGCTACCTGTCTCCT    334

Mycobacterium   288 --------------GC---------AGCGATTTCG-GCGGGTGA-CCAGTC    313
                                  ||        .|||...|.| ||.||..| ||.||.
Salmonella      335 TACTTCATCAACAAGCCGGAAACTGGCGCAGTAGAGCTGGAAAGCCCGTT    384

Mycobacterium   314 CATC------GGTGA-----------CCTG-ATC-----------GCC--    332
                    ||||      |.|||          ||.. |||          |||
Salmonella      385 CATCCTGCTGGCTGATAAGAAAATCTCCAACATCCGCGAAATGCTGCCGG    434

Mycobacterium   333 -----GAGGCGATGGACAAGGTGGGCAA--CGAGGGCG-TCATCACCGTC    374
                     ||.||..|.|..||.|..|||||  ||..|..| |||||.|.|..
Salmonella      435 TTCTGGAAGCCGTTGCAAAAGCAGGCAAACCGCTGCTGATCATCGCTGAA    484

Mycobacterium   375 GA----GGAGTCCAACACCTTTGGGCTG--CAGCTCGAGCTCACCGAG--    416
                    ||   |.||.|.||...||   ||||.  |.|.|.|....||||..|
Salmonella      485 GATGTTGAAGGCGAAGCGCT---GGCTACCCTGGTTGTTAACACCATGCG    531

Mycobacterium   417 -GGTATGCG    424
                     ||.|| ||
Salmonella      532 TGGCAT-CG    539
```

Fig 1. Base Pair Of Thermoplasma, Salmonella, Mycobacteria Thermoplasma.

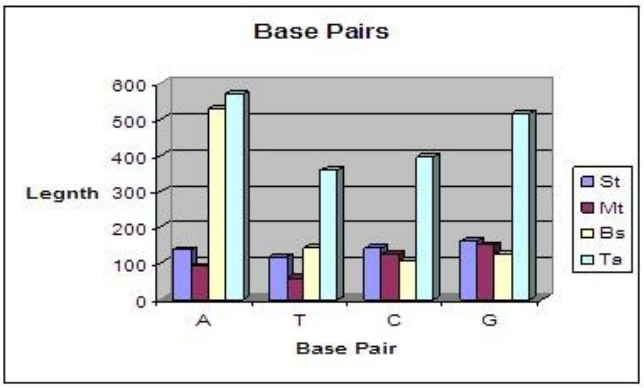


Fig 2. Gc Content Of Thermoplasma, Salmonella, Mycobacteria Thermoplasma.

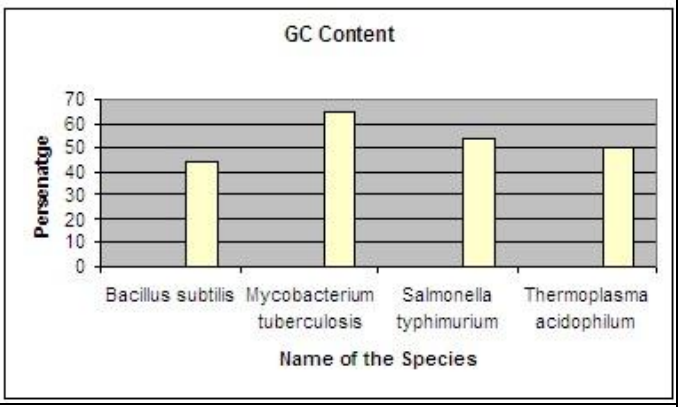


Fig 3. Pairs Of Nucleotides-Bacillus Subtills

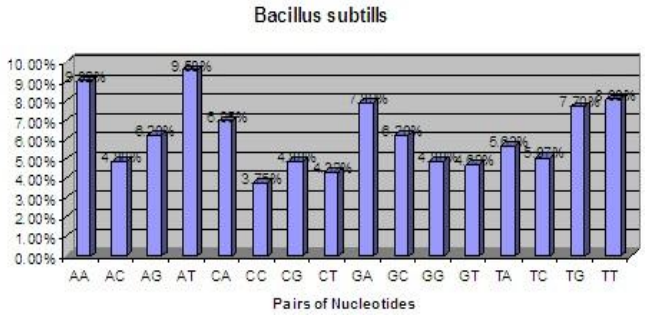


Fig 4. Pairs Of Nucleotides-Mycobacterium Tuberclosis

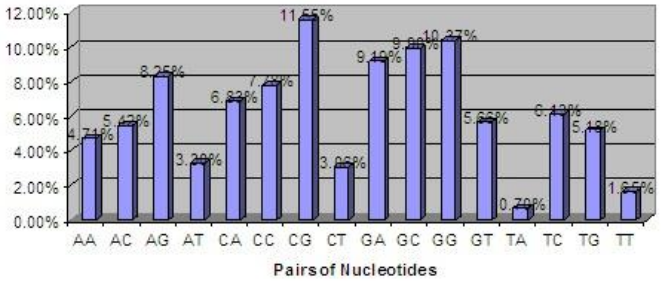


Fig 5. Pairs Of Nucleotides-Salmonella Typhimurium

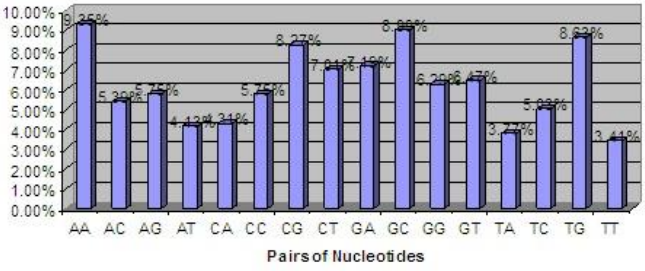


Fig 6. Pairs Of Nucleotides-Thermoplasma Acidophilum

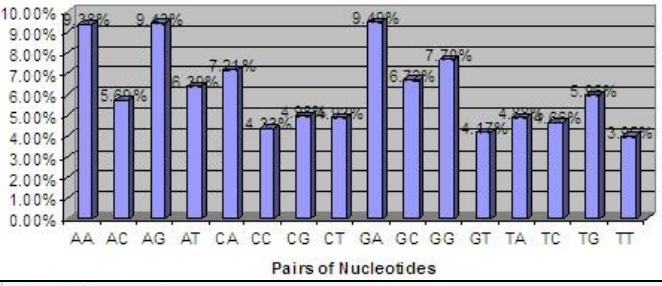
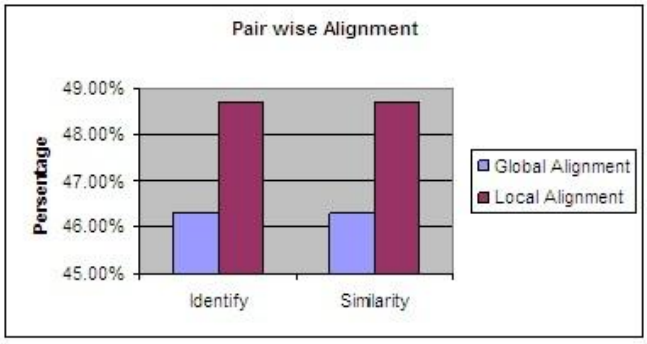


Fig 7. Pair Wise Alignment



ClustalW2 Results

Results of search	
Number of sequences	4
Alignment score	4328
Sequence format	Pearson
Sequence type	nt
JalView	Start Jalview
Output file	clustalw2-20080823-12132736.output
Alignment file	clustalw2-20080823-12132736.aln
Guide tree file	clustalw2-20080823-12132736.dnd
Your input file	clustalw2-20080823-12132736.input

SUBMIT ANOTHER JOB

Scores Table

Sort by | Sequence Number | View Output File

SeqA	Name	Len(nt)	SeqB	Name	Len(nt)	Score
1	Salmonella	556	2	Mycobacterium	424	42
1	Salmonella	556	3	Thermoplasma	1844	11
1	Salmonella	556	4	Bacillus	532	4
2	Mycobacterium	424	3	Thermoplasma	1844	8
2	Mycobacterium	424	4	Bacillus	532	3
3	Thermoplasma	1844	4	Bacillus	532	4

PLEASE NOTE: Some scores may be missing from the above table if the alignment was done

Sort by | Sequence Number | View Output File

CONCLUSION

Pathogenic bacteria and *Thermophilic* bacteria have produced heat shock proteins due to environment stress.

Heat Shock proteins are synthesized by HSP genes found in the genome of bacteria. Number of nucleotides found in *Salmonella* and *Bacillus* was 556 and 532 respectively in *Mycobacterium* it was 424 nucleotide basepairs.However the base pair of *Thermoplasma acidophilum* was 1844 which is a lengthy gene [17]. The phylogram indicates a close relationship between *Salmonella* and *Mycobacterium.* Whereas *Bacillus* showed a distant relationship.

Global pair wise alignment indicates 46% similarity in *Salmonella* and *Mycobacterium*. Logal pair wise alignment also showed close relationship by having 48% alignment between the two organisms. In the other to showed distant relationship.

The GC content indicates the four bacteria belonging to four distinct groups *Bacillus* with 44% *Mycobacterium* with 65%, Salmonella with 54% and *Thermoplasma* with 50%.

ACKNOWLEDGEMENT

None

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

STATEMENT OF HUMAN AND ANIMAL RIGHTS

All procedures performed in human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

REFERENCES

1. Jiuzhou S, Tony W, Shu LL. (2003). Comparative Genomics via Wavelet Analysis for Closely Related Bacteria Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Calgary, 3330.
2. Chen B, Piel WH, Gui L, Bruford E, Monteiro A. (2005). The HSP90 family of genes in the human genome: insights into their divergence and evolution. *Genomics*, 86(6), 627-37.
3. Fukuda D, Watanabe M, Sonezaki S, Sugimoto S, Sonomoto K, Ishizaki A. (2002). Molecular characterization and regulatory analysis of dnaK operon of halophilic lactic acid bacterium Tetragenococcus halophila. J Biosci Bioeng, 93(4), 388-94.
4. Rungeling E, Laufen T, Bahl H. (1999). Functional characterisation of the chaperones DnaK, DnaJ, and GrpE from Clostridium acetobutylicum. FEMS Microbiol Lett, 170(1), 119-23.
5. Kim H, Kim SH, Shim TS, Kim MN, Bai GH, Park YG. (2005). Differentiation of Mycobacterium species by analysis of the heat-shock protein 65 gene. Int J Syst Evol Microbiol., 55(4), 1649-56.

6. Dong CW, Zhang YB, Zhang QY, Gui JF. (2005). Differential expression of three Paralichthys olivaceus Hsp40 genes in responses to virus infection and heat shock. *Fish Shellfish Immunol*, 21(2), 146-58.
7. Gallegos Ruiz MI, Floor K, Roepman P. (2008). Integration of gene dosage and gene expression in non-small cell lung cancer, identification of HSP90 as potential target. *PLoS One*, 3(3), 1722.
8. Aavindhan V, Sulochana S, Narayanan S, Paramasivam CN, Narayanan PR. (2007). Identification & differentiation of Mycobacterium avium & M. intracellulare by PCR- RFLP assay using the groES gene. *Indian J Med Res*, 126(6), 575-9.
9. Shah MM, Iihara H, Noda M, Song SX, Nhung. (2007). DNAJ gene sequence-based assay for species identification and phylogenetic grouping in the genus Staphylococcus. *Int J Syst Evol Microbiol*, 57(1), 25-30.
10. Qiu Z, Bossier P, Wang X, Bojikova-Fournier S, MacRae TH. (2006). Diversity, structure, and expression of the gene for p26, a small heat shock protein from Artemia. *Genomics*, 88(2), 230-40.
11. http://www.ncbi. nlm.nih.gov
12. http/www.embl_heidelberg.de/
13. http://ncbi.nlm.nih.gov/BLAST
14. http://www.ebi.ac.uk/clustal W
15. *McClelland M, M Sanderson KE, et al.* (2001).Complete genome sequence of Salmonella enterica serovar Typhimurium LT2. *Nature*, 25, 852-6.
16. Lemos JA, Chen YY, Burne RA. (2001). Genetic and physiologic analysis of the gro E operon and role of the HrcA repressor in stress gene regulation and acid tolerance in Streptococcus mutans. *J Bacteriol*, 183(20), 6074-84.
17. Wu Y, Wan T, Zhou X, Wang B, *et al.* (2007). Identification & differentiation of Mycobacterium avium & M. intracellulare by PCR- RFLP assay using the groES gene. *Indian J Med Res*, 126(6), 575-9.