

StreptomycesInforSys: A web-enabled information repository

Chakresh Kumar Jain^{1*}, Vidhi Gupta¹, Ashvarya Gupta¹, Sanjay Gupta¹, Gulshan Wadhwa², Sanjeev Kumar Sharma¹ & Indira P Sarethy¹

¹Department of Biotechnology, Jaypee Institute of Information Technology, A-10, Sector 62, Noida- 201307, India; ²Department of Biotechnology (DBT), Ministry of Science & Technology, New Delhi-110003, India; Chakresh Kumar Jain - Email: ckj522@yahoo.com; Phone: +91-120 2400973; Fax: +91-120 2400986; *Corresponding author

Received May 11, 2012; Revised December 01, 2012; Accepted December 04, 2012; Published December 19, 2012

Abstract:

Members of *Streptomyces* produce 70% of natural bioactive products. There is considerable amount of information available based on polyphasic approach for classification of *Streptomyces*. However, this information based on phenotypic, genotypic and bioactive component production profiles is crucial for pharmacological screening programmes. This is scattered across various journals, books and other resources, many of which are not freely accessible. The designed database incorporates polyphasic typing information using combinations of search options to aid in efficient screening of new isolates. This will help in the preliminary categorization of appropriate groups. It is a free relational database compatible with existing operating systems. A cross platform technology with XAMPP Web server has been used to develop, manage, and facilitate the user query effectively with database support. Employment of PHP, a platform-independent scripting language, embedded in HTML and the database management software MySQL will facilitate dynamic information storage and retrieval. The user-friendly, open and flexible freeware (PHP, MySQL and Apache) is foreseen to reduce running and maintenance cost.

Availability: www.sis.biowaves.org

Keywords: Streptomyces, PHP, MySQL, Literature, Molecular

Background:

Members of the genus *Streptomyces* are known for production of structurally diverse bioactive compounds, especially antibiotics and anti-tumour compounds and are responsible for the production of 70% of known microbial natural products [1]. Since their discovery in the 1940s, these organisms have been of immense interest to industry and academia. Emergence of multi-drug resistant pathogens has caused renewed emphasis on discovery of newer bioactive molecules. There is considerable information available on *Streptomyces* including their genotypic and phenotypic traits for classification purposes. This type of polyphasic information provides conclusive evidence that isolates belonging to specific clusters

such as the *S. griseus* clade are more prolific producers of bioactive compounds. Pharmacological screening programmes have to be based on logical parameters such as traits displayed by the isolates since there may be a coupling between taxonomic and chemical diversity at the level of strains [2]. However, the relevant information is scattered across various resources and not accessible from a single point source or even freely available to researchers. Easy access to such data will facilitate quicker initial screening studies.

A search of the 2012 Nucleic Acids Research Database [3] reveals that such a comprehensive database has not been created till date. Currently available databases [4-14] focus on

specific aspects such as degradation of xenobiotics or of environmental importance and include some information pertaining to *Streptomyces* but are not exhaustive. The designed database collates information about *Streptomyces* from different platforms – publications (journals, books, theses, etc.), internet, public information repositories (GenBank, RDP) web-servers (StrepDB, PubMLST) and integrates them. The webpage is facilitated with various search options and results are displayed in tabular form. Currently, the website has been launched using a free web portal.

Methodology:

Data collection: Database was manually curated and collated from various sources like journals, books, theses and websites. Database feature and design: The database is comprised of four tables viz. Literature, Molecular, Morphological, and Physiological-biochemical and follows RDBMS architecture for database modeling and designing. The RDBMS technology will be compatible with other existing platforms and can facilitate information exchange for easy retrieval, modification and analysis. The detailed architecture of database design component is described in (Figure 1). We have employed the apache 2.2.11, PHP 5.2.9, and MySQL client version 5.0.51a for development of our database. In designing the database, normalization eliminates data redundancy to improve consistency, maximizes the flexibility for application of search system by keeping tables corresponding to object types in their simplified forms. MySQL is extremely fast for small to medium size database like StreptomycesInforSys. MySQL utilizes hash based indexing structure for search; therefore search output is received in milliseconds. For efficient database access, elimination of anomalies was done by decomposition and normalization. Care has been taken to reduce redundancy and maintain structure of database. Normalization eliminates data redundancy to improve consistency, maximizes the flexibility for application of search system by keeping tables corresponding to object types in their simplified forms. Referential integrity was sensitized to reject redundant data. Many-to-many relationship was directed to build the whole rational structure and provide flexibility to incorporate any number of contacts. Screenshots of input tables and search options are shown in (Figure 2).

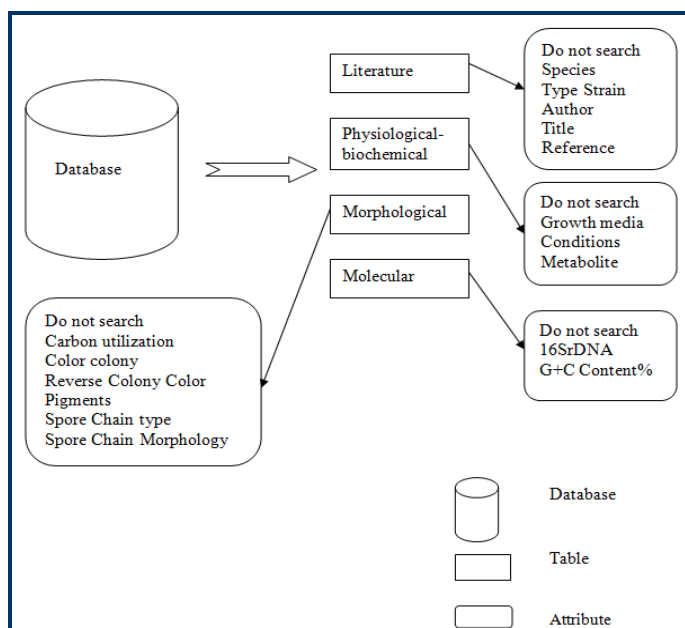


Figure 1: Representation of database architecture

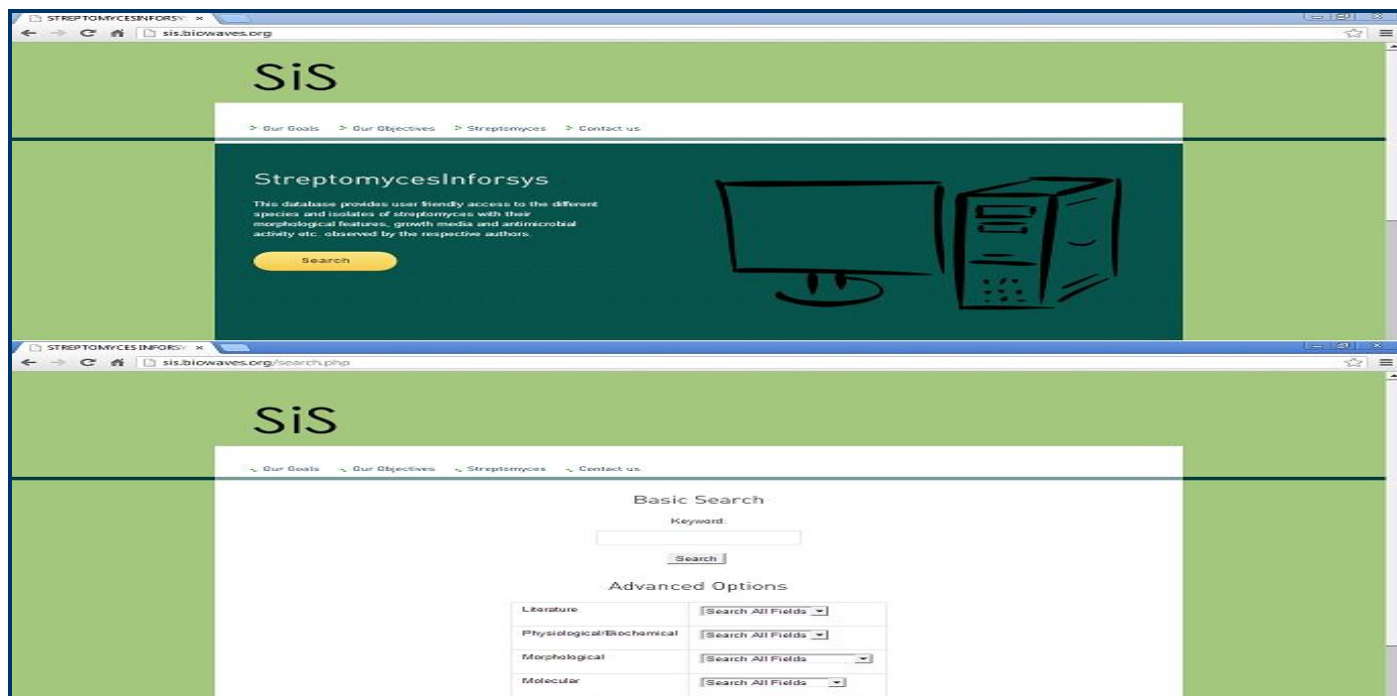


Figure 2: Screenshots of StreptomycesInforSys homepage and search options.

Utility:

The database will serve as a single-point free reference resource to facilitate search for *Streptomyces*-specific

information. It provides quality data set with non-redundancy to help in data mining. It supports multi-user, multitasking and multi-threading approaches for end-user.

The database is free for academicians and industry. It provides detailed information in simple downloadable format.

Future Developments:

There is also future provision for linking to other available databases containing specific information on *Streptomyces/Actinomycetes* like the Multi-Locus Sequence Typing (MLST) database and The Streptomyces Annotation Server (StrepDB) for easy integration. It is being updated to improve the data size and number of records.

Acknowledgement:

The authors are thankful to Jaypee Institute of Information Technology, Noida for providing infrastructure facilities for carrying out the work.

References:

- [1] Laureti L *et al.* *Proc Natl Acad Sci USA*. 2011 **108**: 6258 [PMID: 21444795]
- [2] Antony-Babu S *et al.* *Antonie Van Leeuwenhoek*. 2010 **97**: 231 [PMID: 20033289]
- [3] <http://nar.oxfordjournals.org>
- [4] <http://www.straininfo.net>
- [5] <http://pubmlst.org/>
- [6] <http://bsd.cme.msu.edu>
- [7] <http://www.dzumenvis.nic.in>
- [8] <http://www.ohsu.edu>
- [9] <http://www.microbes.info>: Pathological strains
- [10] <http://www.sciencedirect.com/science>
- [11] <http://www.microbes.info>
- [12] <http://mbgd.genome.ad.jp>
- [13] <http://www.microbialrosettastone.com>
- [14] <http://www.microbesonline.org/>

Edited by P Kanguane

Citation: Jain *et al.* *Bioinformation* 8(25): 1283-1285 (2012)

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