

PseudoBioRes: a Bioinformatic Resource for Pseudomonas

Pseudomonas Migula 1894, includes bacterial species of medical relevant interest, phytopathogens of economical concern as well as and species of biotechnological and ecological interest.

Genome sequencing projects and gene sequence data increase rapidly as a demonstration of the interest of scientific community on this versatile bacterial genus. Nevertheless little information is available on strains, which genomes have not been sequenced yet. *Pseudomonas*-related data (gene and protein sequences, and metabolic pathways) despite being available for a large number of strains, are dispersed among many sources and information extraction could be accessed only via cross researches through either different web sources dedicated to specific class of enzymes or bacterial species or the GeneBank Database. (Fig. 1)

Database overview

We designed and projected a database, designated "PseudoBioRes", that aims to provide for researchers working on *Pseudomonas*, an integrated resource that collects information from the most accredited resources on genes and/or proteins on the basis of the metabolic pathway in which they are involved their potential applications, providing a user-friendly web interface.

The database has a tree-structure (Fig. 2). From the introduction page, showing the main characteristics of *Pseudomonas* species and their role in different fields, three branches were developed.

From the section "Pseudomonas Species" it is possible to access information on a single *Pseudomonas* species from a list based on the "List of Prokaryotic Names with Standing in Nomenclature" (LPNS), a brief description of the specie and its relevance, the genome sequence (when available), genes and proteins involved in a specific pathway and the relevant papers, as schematised in the fig. 3.

As a model, in this first version, a section dedicated to polyhydroxyalkanoates (PHA), biodegradable polymers produced by bacterial cells, has been developed.



Figura 1

Data sources

The database has been assembled by linking the collected data to their original sources. We organized the web data sources in three different categories:

1. Pseudomonas Genome Projects

PseudoBioRes access data on genome sequences from the section Gbrowse of Pseudomonas Genome Database v2 (PGDv2) that stores and integrates data from the project Pseudomonas Genome Project and from PseudoCAP (Pseudomonas aeruginosa Community Annotation Project).

2. Genes and proteins

GenBank was used as gene and protein data source using the searching engine of NCBI (National Center for Biotechnology Information).

3. Metabolic pathways and enzyme classes

The main source is the Japanese GenomeNet service, KEGG: Kyoto Encyclopedia of Genes and Genomes. KEGG integrates metabolic pathways (data on metabolic pathway and complex), genes (data on functional genes and their protein products) and ligands (Chemical compounds, drugs, glycans, and reactions). From here we extrapolated Pseudomonas polyhydroxyalkanoates metabolic pathway.

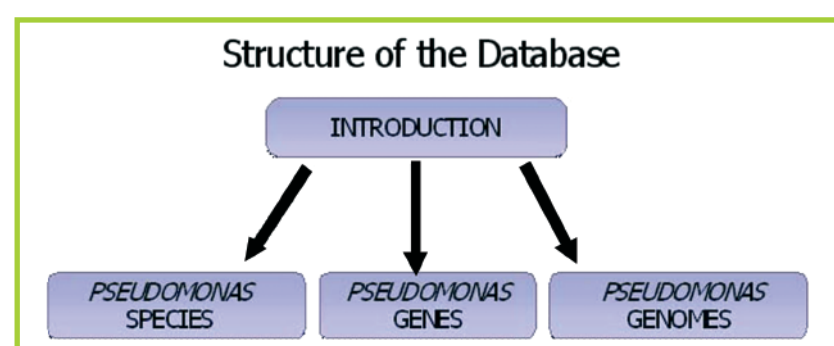


Figura 2

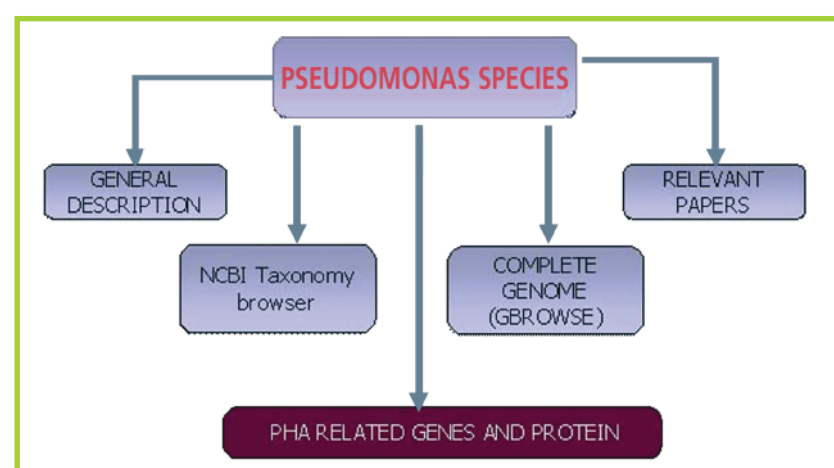


Figura 3

PHA dedicated section

PHAs, synthesized by many bacteria, are biodegradable polymers of great potential for industrial and medical applications. These microbial polymers are accumulated as inclusion bodies when nutrient supplies are imbalanced and are thus believed to play a role as a sink of carbon source and reducing equivalents (Fig. 5).

In the first version of this database all the protein involved in the metabolism of polyhydroxyalkanoate (PHA) isolated or deduced by the genome sequencing projects in species belonging to the *Pseudomonas* genus are present. The database consolidates information from public external sources (GenBank) and manually annotates them into a relational database. Nevertheless to complete the section we become aware that to include all the information it was necessary to articulate query terms and to manually implement data results for each single species. This took a lot of time but was essential to be sure to include all data. (Fig. 6)

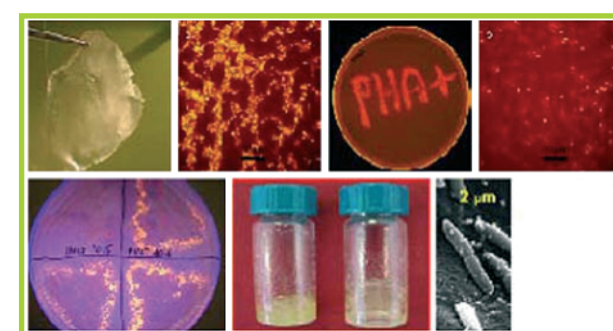


Figura 5

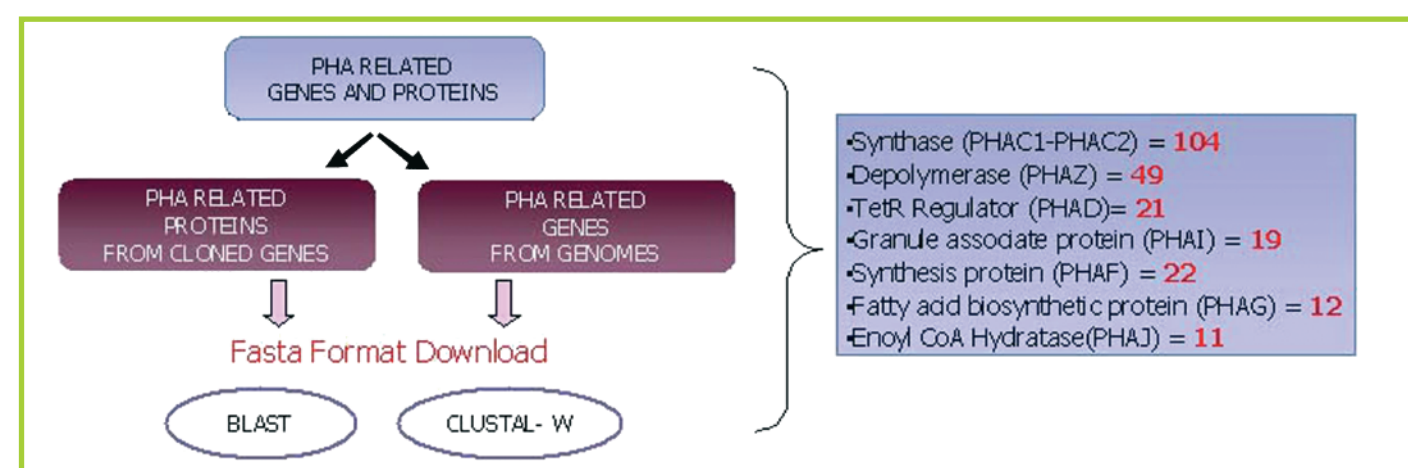


Figura 6



Figura 4

In this page are also provided links to:

- Genome sequencing projects (complete and in process) from the Gbrowse of the Pseudomonas genome database v2 web site;
- Pseudomonas species validated list by linking the List of Prokaryotic Names with Standing in Nomenclature" (LPNS) web site. (Fig. 4)

By the Pseudomonas Gene resource it is possible to access to specific sections dedicated to classes of gene with relevant interest. Gene chromosome location, sequence and structural information are extracted from the NCBI Taxonomy database, used also as reference for information on the biological sources of the protein sequenced providing links to the main important biological database (KEGG: Kyoto Encyclopedia of Genes and Genomes). This section is work in progress.

Through the web interface the user can browse PHA gene or protein sequences among all *Pseudomonas* spp. and perform multi-gene/protein comparisons (including BLAST and alignments).

The database is open source in order to maintain consistency with the new findings and can also be used as a guideline in order to create other sections for other relevant metabolites. At the moment, it is accessible at the following URL: www.ylichron.it/PHA_pseudomonas_DB. (Fig. 7)

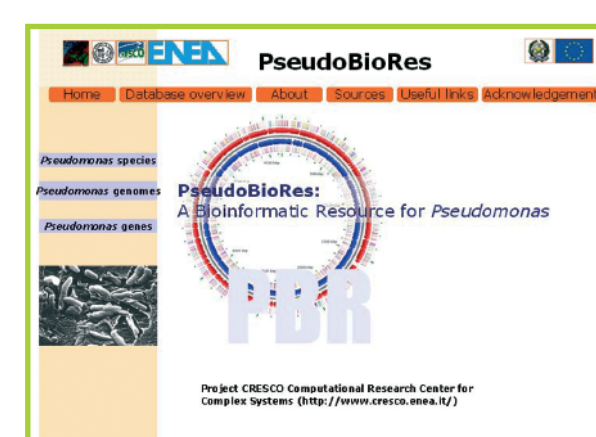


Figura 7

Autori: G. Licciardello^{(1) (2)}, V. Catara⁽²⁾, A. Mangtani⁽³⁾, R. Casilli⁽⁴⁾ and V. Rosato⁽³⁾

⁽¹⁾Parco Scientifico e Tecnologico della Sicilia, ⁽²⁾Università di Catania - Dip. di Scienze e Tecnologie Fitosanitarie, ⁽³⁾ ENEA, ⁽⁴⁾ Ylichron S.r.l.

e-mail contact: g.licciardello@unict.it ► www.cresco.enea.it