

# Enzyme Detector – integrated genomic enzyme function prediction

Susanne Quester and Dietmar Schomburg

*Department of Bioinformatics and Biochemistry , Technische Universität Carolo-Wilhelmina zu Braunschweig*  
[s.quester@tu-bs.de](mailto:s.quester@tu-bs.de)

The main biological genome annotation servers often provide inconsistent and incomplete genome annotations of enzyme functions. To cope with that fact we developed the program EnzymeDetector. This tool provides an integrated rated view of the available annotations for every gene of the organism of interest. The information of the databases NCBI [1], PEDANT [2], BRENDA [3] and SwissProt [4] are integrated, completed by the information of an up-to-date BLAST-search [5] and the pattern search program BrEPS [6].

The combined data are presented on a web interface [<http://enzymedetector.tu-bs.de>]. After selecting the organism of interest, the user can choose, which databases to be integrated and the relative rating of these annotation sources. As a help a default rating scheme is given. A summarised relevance score is provided.

Additionally to the different sources that can be integrated, the user can upload an own annotation file in a preset format and thus integrate his own data in the analysis.

The web interface presents the results of EnzymeDetector in four different views. For all views the data can be filtered with two different cut-offs, the maximal E-value the BLAST hits are allowed to have, to be considered and the minimal relevance score the results should have.

In the tabular view the annotations are listed sorted by GI (gene identifier). Optionally the results can be sorted by EC-number or recommended name of the enzyme as well. For every gene all possible annotation candidates are listed. It is noted which annotation sources contain the candidates. If several candidates are given, the user can decide based on the given E-value of the BLAST-search and the assigned overall relevance score. It is possible to search by GI, by gene position, by the common name of the enzyme annotated and by the sources that have to support an annotation. If wanted, it is possible to activate the 'show full entries'-Button and thereby not only the annotations that fulfill the search criteria are shown, but the other annotation candidates for that gene. Thereby a decision can be easier made.

The statistical view shows a statistical evaluation of the results. Important key values are shown. This site is splitted in half. On the left side the statical values are shown for the default cut-off values. On the right side the user has the possibility to assign own cut-off values to see which difference the chosen cut-offs make.

With the comparison view the user has the possibility to compare the list of enzymes of the organism of interest to up to two other organisms. The user gets a sorted list of all present enzymes and the information of the best E-value and the highest relevance score, with which the enzyme was found in the organism. By choosing one or two other organism the genome annotations of those organism are easily compared.

The pathway view provides an overview of all Metacyc [7] and KEGG pathways and the coverage of those by the given organism genome. If the coverage is not 100 %, the missing enzymes are given. This result table can be sorted by pathway name, source of the pathway or by coverage. It is possible to show only pathways with a certain coverage.

With the different views the web interface provides a comprehensive overview and analysis of the available annotations of an organism. It facilitates the identification of a reasonable enzyme stock and thereby lay a solid foundation for the construction of a metabolic model of the organism.

1. Pruitt KD, Tatusova T, Maglott DR: **NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins**. *Nucleic Acids Res* 2007, **35**:D61-65.
2. Riley ML, Schmidt T, Artamonova II, Wagner C, Volz A, Heumann K, Mewes H-W, Frishman D: **PEDANT genome database: 10 years online**. *Nucleic Acids Res* 2007, **35**:D354-357.
3. Chang A, Scheer M, Grote A, Schomburg I, Schomburg D: **BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009**. *Nucleic Acids Res* 2009, **37**:D588-592.
4. Bairoch A, Boeckmann B, Ferro S, Gasteiger E: **Swiss-Prot: juggling between evolution and stability**. *Brief. Bioinformatics* 2004, **5**:39-55.
5. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL: **BLAST+: architecture and applications**. *BMC Bioinformatics* 2009, **10**:421.
6. Bannert C, Welfle A, Aus dem Spring C, Schomburg D: **BrEPS: A flexible and automatic protocol to compute enzyme-specific sequence profiles for functional annotation**. *BMC Bioinformatics* 2010, **11**:589.
7. Caspi R, Altman T, Dale JM, Dreher K, Fulcher CA, Gilham F, Kaipa P, Karthikeyan AS, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Paley S, Popescu L, Pujar A, Shearer AG, Zhang P, Karp PD: **The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases**. *Nucleic Acids Res* 2010, **38**:D473-479.