

## ESTABLISHED DATABASES FOR AEROBIC AND ANAEROBIC KEY GENES AND PROTEINS INVOLVED IN AEROBIC AND ANAEROBIC DEGRADATION OF PAHS

### SUMMARY

Despite the overall increasing knowledge on aromatic degradation pathways, it is a matter of fact that this information is not clearly structured. The majority of protein sequences in public databases have not been experimentally characterized and the most common approach in use today continues to be the assignment of molecular function from the inference of homology followed by annotation transfer. Actually, misannotation in enzyme superfamilies containing multiple families that catalyze different reactions is a larger problem than has been recognized. To obtain a reliable frame of reactions and enzymes involved in PAH degradation, we developed databases on key catabolic enzymes.

### DESCRIPTION OF THE PRODUCT

The MAGICPAH Consortium gets ready databases on key catabolic enzymes:

- Rieske non-heme iron oxygenases
- Extradial dioxygenases of the vicinal chelate superfamily
- Extradial dioxygenases of the LigB superfamily
- Cupin dioxygenases
- Enzymes related to benzylsuccinate synthases activating methylsubstituted aromatics and alkanes
- Class I benzoyl-CoA reductases
- Class II benzoyl-CoA reductases
- (putative) anaerobic aromatic carboxylases

The distribution analysis of key catabolic enzymes was based on well-characterized sequences documented to be involved in

catabolism of aromatics as initial seeds to search for homologous sequences. Protein similarity searches were performed using the BLASTP program from the NCBI website. Translated protein sequences were aligned with MUSCLE using default values. Phylogenetic trees were constructed with MEGA5 and inspected for proteins with similarity to proteins of documented function. All databases are available in fasta format.

### INNOVATIONS AND ADVANTAGES OFFERED

- The databases can be used for: (1) to design PCR primers and microarray probes to survey the spread and activity in contaminated environments; (2) and to obtain a reliable annotation of genes observed in metagenomic surveys.

### POTENTIAL AREA OF UTILIZATION

Due to the exponential accumulation of missanotations on GenBank database, there is an increased interest to develop a fast and user-friendly web platform for drafting metabolic models, based on reliable knowledge on PAH metabolism.

Therefore, the established databases developed by the MAGICPAH Consortium might be of great value as a powerful tool to speed up the development of a web-tool for drafting metabolic models, based on reliable knowledge on PAH metabolism.

For further information contact:

[dpi@helmholtz.hzi.de](mailto:dpi@helmholtz.hzi.de)