Title: FunCatTM, a controlled vocabulary encompassing the biology of prokaryotes, plants and animals from cellular to systemic level

Authors: Dieter Maier1, Alfred Zollner1, Kaj Albermann1, Jean Hani1, Reinfried Gotz1, Armin Ehrenreich1, Klaus Mayer2, Heiko Schoof2 and Hans-Werner Mewes2
Address: 1 Biomax Informatics AG, Lochhammer Str. 11, 82152 Martinsried, Germany
2 Institut fur Bioinformatik, GSF - Forschungszentrum fur Umwelt und Gesundheit, GmbH, Ingolstadter Landstr. 1, 85764 Neuherberg, Germany

Abstract: The FunCat vocabulary starts with general terms describing biological categories, such as metabolism, transcription or development and becomes increasingly more specific, with entries such as amino acid biosynthesis, mRNA-capping or gastrulation. In this way, the FunCat vocabulary covers the biological processes involving a protein, the protein's molecular functions (such as kinase or protease) and the protein localization from the subcellular to the organ level.
More than 1100 specifications in 26 categories and up to six different levels enable the description of prokaryotes, unicellular eukaryotes, plants and animals.
The FunCat has gradually been extended to described biology from the cellular to the systemic level, covering the complete Phenotype of an organism, from Development to Physiology.
Grouping genes into meaningful categories using the multidimensional annotation scheme of the FunCat enables automatic clustering and data mining across diverse datasets as well as transfer and integration of information between different bioinformatics applications.
Much information about genes has been gathered describing not their normal biological function but the effects they cause if mutated. As we do not want to mix those informations we have created separate Phenotype catalogues to describe the pathology caused by mutations in annotated genes.