

## Poster B-12

### Integr8 – Access to Complete Genomes and Proteomes



#### Authors:

Manuela Pruess (*EMBL Outstation - European Bioinformatics Institute (EBI)*)  
Lawrence Bower (*EMBL Outstation - European Bioinformatics Institute (EBI)*)  
Paul Kersey (*EMBL Outstation - European Bioinformatics Institute (EBI)*)  
Rajesh Radhakrishnan (*EMBL Outstation - European Bioinformatics Institute (EBI)*)  
Rolf Apweiler (*EMBL Outstation - European Bioinformatics Institute (EBI)*)

**Short Abstract:** Integr8 is a web portal for exploring the biology of organisms with completely deciphered genomes. For 250 species, Integr8 provides access to general information, publications, and a statistical overview of the genome and proteome of the organism constructed using data resources such as UniProtKB, GenomeReviews, IPI, InterPro, CluSTr and GOA.

#### Long Abstract:

The development of large scale, high throughput technologies capable of determining the nature of the genome, transcriptome and proteome of living organisms has led to the concomitant development of new ways of interpreting information about individual genes and proteins in a wider context. However, the enormous quantities of data generated through such experimental approaches have increasingly been spread over many different primary resources, making it harder to draw scientific conclusions. Coherently integrating such data, and offering access to it, has thus emerged as one of the most important challenges in bioinformatics. The Integr8 web portal has been developed to provide a single point of access to data obtained from many different primary sources, integrated in a model reflecting the central dogma of biology, in order to describe the current state of knowledge about the biology of individual species. The focus of Integr8 is species with completely deciphered genomes, in which it is possible to interpret information about particular genes in the wider of the genome as a whole. Integr8 is built by taking reference genome sequences (selected in a species-specific fashion from resources including as the EMBL Nucleotide Sequence Database, SGD, and Ensembl) and incorporating data from protein-centric resources such as the UniProt Knowledgebase and IPI into this framework. The Integr8 portal provides public access to this data, and offers tools for its interpretation and analysis. For each species, a short descriptive overview is provided, as well as a list of recent publications. A detailed statistical analysis of its genome and proteome is also presented via a combination of tabular and graphical displays. Key tools used in this analysis include InterPro, CluSTr, and GO, which each provide ways proteomes by classifying their constituent proteins. Comparative analysis between genomes is also supported, while additional tools allow users to customise their own analyses. Information about individual genes can be accessed through the "Integr8or" component of the Integr8 portal. Access to the Integr8or is provided through a graphical search interface that allows the user to direct their search to a specified portion of the taxonomic tree. Once a gene has been selected, the Integr8or provides a clear view of the products (transcripts and proteins) of the gene, and the annotations (and cross-referencing entries in external resources) that have been associated with each. The

Integr8or also provides a schematic representation of the location of each gene in the context of its neighbours, and allows for the detection of syntenic regions by supporting the customisable expansion of this view into a comparative display of multiple genomes centred on putative orthologues in each species. In this display, each gene is colour-coded according to the domain architecture of its protein product, enabling users to quickly compare genes not yet assigned a descriptive name. Another component of Integr8, the “Inquisitor”, provides an expert system for protein sequence analysis, running a number of different sequence analysis tools to identify each sequence in Integr8, or, if the sequence is novel, to classify it as fully as possible. Integr8 also serves as a central hub for downloading information. Access to EMBL and Genome Reviews files, non-redundant protein sets (derived from the UniProt KB) in a variety of file formats, and GO/InterPro annotation for each proteome is available for download. The Integr8 data is also available through the BioMart data warehousing system, which allows users to selectively identify genes and proteins matching their own criteria for download and analysis. Integr8 is updated fortnightly, and can be accessed at <http://www.ebi.ac.uk/integr8>.