

Integrated Information Platform for Biological Resource Centers

Author(s) Di Liu, Linhuan Wu, Quanhe Liu, Guanghui Guo, Juncai Ma

Institution(s) 1. IMCAS, Institute of Microbiology, Chinese Academy of Sciences, Beichen West Road, Chaoyang District, Beijing, China

Abstract:

Biological resource centers (BRCs) are important organizations for the long-term preservation and distribution of the cultures, especially microorganisms, for the scientific society. To facilitate the management of the preserved cultures, an information platform is essential. In this paper, we described the design and implementation of an information platform for BRCs, integrating not only the information and knowledge of bio-resource, species and taxonomy but also bioinformatic database, software, and DNA barcodes. The integrated information platform is designed as a culture collection centric system. Differed with bioinformatics platforms, the main entry point is to search or browse the bio-resources and preserved species. From species on, the hierarchy of NCBI and/or Species 2000 taxonomic trees can be retrieved. Meanwhile, the DNA/RNA sequences corresponding to each species were listed through the web services of Genbank and RefSeq database. From those nucleotide sequences on, more gene and function-related resources were linked, e.g. entrez-Gene, Gene Ontology, KEGG-Pathway. In addition, the DNA barcoding data from BOLD were integrated as well, especially for fungi and yeasts. Based on the distributed computing technology, the general sequence analysis tools, including BLAST, CLUSTAL_W were loosely coupled. Retrieved sequences and user's own data can be submitted to the bioinformatics-computing platform and scheduled to run. The implementation of the integrated information platform is mainly relying on LAMP (Linux + Apache + MySQL + PHP) technology. BRC data of CGMCC were directly accessed via MySQL connections. As for other data integration, locally installed MRS indexing system and NCBI web services were applied. The Condor scheduling system was used as the scheduler for bioinformatics programs running on high throughput computing clusters. In summary, we designed and implemented a culture collection centric integrated information system, by utilizing web and computing technology, as well as biological database and software.

Key words: BRC, Integrated Information Platform, Database, Bioinformatics