

A solution to integration and sharing data for phylogenetic research

Yanping Gao, Zhen Meng, Xing He, Yong Liu Yuanchun Zhou & Jianhui Li*

Scientific Data Center, Computer Network Information Center, CAS, Beijing, China,
zhenm99@cnic.cn

DarwinTree is a comprehensive platform on phylogenetic analysis, which organizes multi-source data and massive analysis tools, computing models as well. One of central functions of DarwinTree is to integrate and share data from international public databases and private data of users in several formats. In this paper, a solution to intergration and sharing data for phylogenetic research is presented. Some proverbial systems of taxonomy are presented as public references, such as the APG II system of plant classification and the NCBI Taxonomy; while the sequence data to reconstruct the phylogenetic trees for public users are screened from International public databases such as genBank, EMBL. At the same time, for users who are interested in ongoing phylogenetics projects, we have started a service called “My darwinTree”, which allows users to add their own evolutionary classification system as a reference system in addition to using public ones, to subscribe to the related taxon of their own to get the daily update information report. Meanwhile users are welcomed to submit their private sequence data for analysis and sharing with others. Furthermore, several data screening methods are implemented to keep quality control of the data both from International public databases and users.