Management of DNA Barcoding data

Di Liu, Guanghui Guo, Linhuan Wu and Juncai Ma

Network Information Center, Institute of Microbiology, Chinese Academy of Sciences

Beichen West Road, Chaoyang District, Beijing 100101, China P.R.

Abstract:

Barcode of Life is an international collaborative program aiming to collect and identify as much as possible living creatures via DNA barcode, a relatively short DNA segment that can be used to distinguish species by sequence. To fulfill these goals, field works, wet-lab experiments as well as informatics and bioinformatics efforts are necessary. China, as a central node, is responsible for the collection and distribution of DNA barcode data, so that a data management system is required. In addition to the data distribution system, which is designed as the barcode of life data mirror system, the data collection system is also constructed as the system to manage data mainly submitted by the Chinese scientists. To ensure the data security, the data management system (data submission system) is accessible only to the registered users. To acquire a valid user ID, the registration and approval procedure is required. The data stored in the system is project-based, and each user may view only his/her own data and the public data. To upload data, user may use single or batch submission. For the single submission, there are four kinds of forms to be filled, the sample details, pictures, sequence files and DNA chromatogram (trace files). To upload a set of data, “quick submission” mode can be used. The users are required to prepare four files for batch submission. One is an Excel file, which describes project information, sample information, collection information, sequencing information, etc. The second file is a zipped sequence file (in FASTA format). The third is a zipped DNA trace and score file, and the fourth is a zipped figures (in JPEG, PNG or GIF format). The

template of the Excel file can be downloaded from the quick submission page. When data is uploaded, each record is viewable and editable through the web pages. Users can check the correctness or make modifications to their data. Besides, a BLAST search is provided for the sequence comparison with all the public DNA barcode data.

Keywords: DNA Barcode, Management System, Data Submission