

Abstract

Anna L Weitzman & Christopher H C Lyal

Integrating DNA Barcoding and taxonomic data. INOTAXA: how new technology can facilitate Open Access to 300 years of vitally important information

Informatics Branch & Dept. of Botany, National Museum of Natural History, Smithsonian Institution, Washington DC 20560, USA, Weitzman@si.edu

Dept. of Entomology, Natural History Museum, London SW7 5BD, UK, c.lyal@nhm.ac.uk

A major problem for access to biodiversity information is the many forms and places where it is stored. While new technology such as DNA barcoding may speed the process of identification of organisms, it does not obviate, and in fact increases the need for ready access to the 300 years of data about earth's plants and animals.

Biological taxonomy is the study of the names and relationships of organisms. Taxonomists have been accumulating data in the form of names, concepts, specimens, descriptions and geographic distributions for some 300 years, much of which is published, but not easily retrievable. More recently, molecular sequence (DNA) data have been added to the mix. Standards for storage and exchange of taxonomic data are being actively developed with international cooperation through several groups.

The fundamental assumption behind DNA barcoding is that a short sequence of DNA can be used to quickly and easily identify any organism. However, simply having that sequence does not tell someone what name to attach to that organism; that requires an understanding of taxonomy which in turn requires immediate access to millions of volumes of literature. Taxonomic literature is generally not accessible electronically or, if it is, the format is not one that enables interoperability with other data types. A wide variety of kinds of research and products are hindered by the lack of easy access to taxonomic literature. The INOTAXA project aims to create a model to resolve these issues. The project will make major taxonomic resources for Mesoamerica, including the 58 biological volumes of *Biologia Centrali-America* and modern works such as *Flora Mesoamericana*, available as a fully searchable resource. The model will be appropriate for all taxonomic literature and will enable interoperability with specimen databases, taxonomic authority files, and other datasets made available on the web. In addition to integrating literature and other data sources, the method opens possibilities of generating further web products including checklists from multiple publications available in this form and a way of speeding preparation and delivery of taxonomic products.