

DNA sequences need quality time too - guidelines for quality control published

September 5 2012



This is the cover for the latest *MycoKeys* issue. Credit: Pensoft Publishers

Like all sources of information, DNA sequences come in various degrees of quality and reliability. To identify, proof, and discard compromised molecular data has thus become a critical component of the scientific endeavor - one that everyone generating sequence data is assumed to carry out before using the sequences for research purposes.

"Many researchers find sequence [quality control](#) difficult, though", says Dr. Henrik Nilsson of the University of Gothenburg and the lead author

of a new article on sequence reliability, published in the Open Access journal *MycoKeys*. "There just isn't any straightforward document to put in their hands to give them a flying start. As a result, scientists differ in the degree to which they are aware of the need to exercise sequence quality control and in what measures they take." Previous studies have highlighted several shortcomings of publicly available [DNA sequences](#) - more than ten percent of the fungal DNA [sequences](#) may be misidentified at the species level, for example.

"A second complication", adds co-author Prof. Urmas Koljalg of the University of Tartu, "is that the software available for sequence quality management tend to be very complex and resource intensive. It borders on the unfair to expect everyone to have access to, and to master, such computer environments. Fortunately, a whole lot can be done towards quality control of DNA sequences using just manual means and a web browser. The current *MycoKeys* paper describes these means to help those [biologists](#) who do not have a strong background in computer science."

The article—"Five simple guidelines for establishing basic authenticity and reliability of newly generated fungal ITS sequences"—compiles principles and observations to assist the reader in the quality management of [sequence data](#). Although focusing on [fungi](#), the guidelines are general and apply to most groups of organisms and genes. The guidelines target traditional DNA sequencing and are broadly applicable to datasets used in systematics, taxonomy, and ecology.

Co-author Dr. Martin Hartmann of the Swiss Federal Research Institute WSL concludes, "We hope that our guidelines will assist the readers in sharpening their datasets so that, eventually, the trend of increasing noise in the public sequence databases can be arrested. Molecular data offer so much promise that we simply cannot afford to lose accuracy to bias and artifacts."

More information: Nilsson RH, Tedersoo L, Abarenkov K, Ryberg M, Kristiansson E, Hartmann M, Schoch CL, Nylander JAA, Bergsten J, Porter TM, Jumpponen A, Vaishampayan P, Ovaskainen O, Hallenberg N, Bengtsson-Palme J, Eriksson KM, Larsson K-H, Larsson E, Kõljalg U (2012) Five simple guidelines for establishing basic authenticity and reliability of newly generated fungal ITS sequences. *MycKeys* 4: 37-62. [doi: 10.3897/mycokeys.4.3606](https://doi.org/10.3897/mycokeys.4.3606)

Provided by Pensoft Publishers

Citation: DNA sequences need quality time too - guidelines for quality control published (2012, September 5) retrieved 27 April 2024 from <https://phys.org/news/2012-09-dna-sequences-quality-guidelines.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.