

facets of the biotic response to extreme events, and provide a timely call for increased study of this challenging topic.

David Ackerly

Department of Biological Sciences, Stanford University,
Stanford CA, USA 94305
(tel +1 650 723 0176; fax +1 650 723 6132;
email dackerly@stanford.edu)

References

- Ackerly DD, Dudley S, Sultan S, Schmitt J, Coleman JS, Linder R, Sandquist D, Geber M, Evans A, Dawson T, Lechowicz M. 2000. The evolution of plant ecophysiological traits: recent advances and future directions. *Bioscience* 50: 979–995.
- Ackerly DD. 2003. Community assembly, niche conservatism and adaptive evolution in changing environments. *International Journal of Plant Sciences* 164: S165–S184.
- Ball MC, Hodges VS, Laughlin GP. 1991. Cold-induced photoinhibition limits regeneration of snow gum at tree line. *Functional Ecology* 5: 663–668.
- Bazzaz FA. 1996. *Plants in changing environments*. Cambridge, UK: Cambridge University Press.
- Denny M, Gaines S. 2000. *Chance in biology: using probability to explore nature*. Princeton, NJ, USA: Princeton University Press.
- Easterling DR, Meehl GA, Parmesan C, Changnon SA, Karl TR, Mearns LO. 2000. Climate extremes: observations, modeling, and impacts. *Science* 289: 2068–2074.
- Gutschick VP, BassiRad H. 2003. Extreme events as shaping physiology, ecology, and evolution of plants: toward a unified definition and evaluation of their consequences. *New Phytologist* 160: 21–42.
- Hoffman AA, Parson PA. 1991. *Evolutionary genetics and environmental stress*. Oxford, UK: Oxford University Press.
- Jackson ST, Overpeck JT. 2000. Responses of plant populations and communities to environmental changes of the late Quaternary. *Paleobiology* 26: 194–220.
- Langan SJ, Ewers FW, Davis SD. 1997. Xylem dysfunction caused by water stress and freezing in two species of co-occurring chaparral shrubs. *Plant, Cell & Environment* 20: 425–437.
- Schmaulhausen II. 1949. Factors of evolution. The theory of stabilizing selection. Philadelphia, PA, USA: Blakiston.
- Williams CD, Shuman BN, Webb T III. 2001. Dissimilarity analyses of Late-Quaternary vegetation and climate in eastern North America. *Ecology* 82: 3346–3362.

Key words: global change, climate extremes, realized environment, acclimation, genetic variation.

Taxonomic misidentification in public DNA databases

There is a growing problem of taxonomic misidentification in public DNA databases, and this issue is highlighted by

Bridge *et al.* (pp. 43–48). DNA sequences are becoming the primary currency by which we measure and study microbial biodiversity (Tautz *et al.*, 2003). The sheer volume of sequence data places enormous pressure on public sequence databases (such as GenBank and EMBL), which must curate and annotate an ever-growing catalogue of genomic and environmental sequences. Like a library in which books are sometimes mistakenly assigned a wrong call number, sequence errors inevitably end up in public databases. A significant portion of the data in public databases is known to contain random as well as systematic sequencing errors (Clark & Whittam, 1992; Harris, 2003). The exponential proliferation of environmental sequences (e.g. directly from soil and other environmental clone banks) has also resulted in the proliferation of an important class of data for which voucher specimens are not available. While environmental sequences can provide useful information about microbial diversity (Vandenkoornhuys *et al.*, 2002), their relevance in DNA databases depends on comparison with reference material from voucher-based taxonomic studies.

'Upwards of 20% of the named sequences in public databases may be misidentified'

Bridge *et al.* retrieved multiple sequences for ribosomal DNA sequences from the EMBL server for three relatively well-studied groups of fungi (*Phoma*, *Amanita* and the *Helotiales*). Using a combination of standard bioinformatics search tools (FASTA, BLAST) as well as phylogenetic analyses of aligned sequences, they identified numerous cases of obvious or apparent mistaken identity in each group of fungi. The problems in each data set are different. In the case of *Amanita*, most problems are attributable to the use of misidentified cultures (presumably from well reputed culture collections). In the case of *Helotiales*, the problem stems from the diversity of investigators who work with these groups, and might also reflect differences in taxonomy between specialists. The most common causes of errors they cite include misidentification or mislabeling of original materials, contamination by other fungi during culture, or other PCR-based errors including chimaeric sequences. Although their study primarily addresses ribosomal DNA sequences (the most common type of data used for molecular systematics), it applies to any genes for which comparative sequence data may be collected. By their estimate, upwards of 20% of the named sequences in public databases may be misidentified in some fungi. This is a

serious problem, which threatens the utility of public sequence databases as archives of biodiversity.

By their inclusivity, public databases sometimes become a home for junk data (like the rest of the internet). Unless curated and annotated, mistakes can proliferate and reduce the ability of correct entries to serve as references for new data. For this reason, Bridge *et al.* recommend that protocols be adopted by public databases similar to those used by taxonomic and culture collections. Like forensic biologists who must adhere to fixed standards for handling of DNA evidence, natural historians who study DNA must also apply appropriate controls to guarantee the identity of strains, collections, and even DNA samples. One solution they propose is the annotation of questionable database sequences by taxonomic experts who work with the public databases.

How serious is the threat of misidentified sequences? As DNA databases grow, the most blatant taxonomic misidentifications are often sorted out by nonspecialists (e.g. a basidiomycete sequence that erroneously groups with ascomycete sequences). The inclusivity of public databases may also be their greatest asset. The prospects for better taxonomic accuracy of databases may not be so bad. In most instances, it is still too hard to tell if a sequence has been misidentified with certainty. For most fungi, taxonomic coverage in public databases is still relatively poor (< 1% of the estimated 1.5 million fungal species are represented in public databases). Although a complete taxonomic scaffolding is not yet in place, programs have recently been established to populate the fungal 'tree of life' with useful sequences (e.g. the Deep Hypha research coordination network supported by the U.S. National Science Foundation, <http://ocid.NACSE.ORG/research/deephyphae/>). By working together, the greater community of systematists can help plug the larger gaping holes that exist in our taxonomic coverage of known fungi, and also help in the discovery of unknown groups (Vandenkoornhuysen *et al.*, 2002). Another solution is for the systematics community to develop special-purpose databases for taxonomic identification, such as the Ribosomal Database Project (<http://rdp.cme.msu.edu>), with tools necessary for accurate sequence identification and analysis. New sequences will always need to be compared against standard reference data sets as well as special aligned sequence data sets for ribosomal DNA-based identification in mycorrhizal fungi (Bruns *et al.*, 1998), yeasts (Kurtzman & Robnett, 1997; Scorzetti *et al.*, 2002), and agarics (Moncalvo *et al.*, 2002).

It will always be the responsibility of users to check the identity of specimens and the integrity of their sequence data. As with all systematics research, responsible vouchering is also essential (Agerer *et al.*, 2000).

Rytas Vilgalys

Department of Biology, Duke University, Durham,
NC 27708–0338 USA
(email fungi@duke.edu)

References

- Agerer R, Ammirati J, Blanz P, Courtecuisse R, Desjardin DE, Gams W, Hallenberg N, *et al.* 2000. Always deposit vouchers. *Mycological Research* 104: 642–644.
- Bridge PD, Spooner BM, Roberts PJ, Panchal G. 2003. On the unreliability of published DNA sequences. *New Phytologist* 160: 43–48.
- Bruns TD, Szaro TM, Gardes M, Cullings KW, Pan JJ, Taylor DL, Horton TR, Kretzer A, Garbelotto M, Li Y. 1998. A sequence database for the identification of ectomycorrhizal basidiomycetes by phylogenetic analysis. *Molecular Ecology* 7: 257–272.
- Clark AG, Whittam TS. 1992. Sequencing Errors and Molecular Evolutionary Analysis. *Molecular Biology and Evolution* 9: 744–752.
- Harris DJ. 2003. Can you bank on GenBank? *Trends in Ecology and Evolution* 18: 317–319.
- Kurtzman CP, Robnett CJ. 1997. Identification of clinically important ascomycetous yeasts based on nucleotide divergence in the 5' end of the large-subunit (26S) ribosomal DNA gene. *Journal of Clinical Microbiology* 35: 1216–1223.
- Moncalvo JM, Vilgalys R, Redhead SA, Johnson JE, James TY, Aime MC, Hofstetter V, Verduin SJW, Larsson E, Baroni TJ, Thorn RG, Jacobsson S, Clemençon H, Miller Jr OK. 2002. One hundred and seventeen clades of euagarics. *Molecular Phylogenetics and Evolution* 23: 357–400.
- Scorzetti G, Fell JW, Fonseca A, Statzell-Tallman A. 2002. Systematics of basidiomycetous yeasts: a comparison of large subunit D1/D2 and internal transcribed spacer rDNA regions. *Fems Yeast Research* 2: 495–517.
- Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP. 2003. A plea for DNA taxonomy. *Trends in Ecology and Evolution* 18: 70–74.
- Vandenkoornhuysen P, Baldauf SL, Leyval C, Straczek J, Young JP. 2002. Extensive fungal diversity in plant roots. *Science* 295: 2051.
- Key words:** DNA sequences, public sequence databases, GenBank, EMBL, voucher-based taxonomic studies.