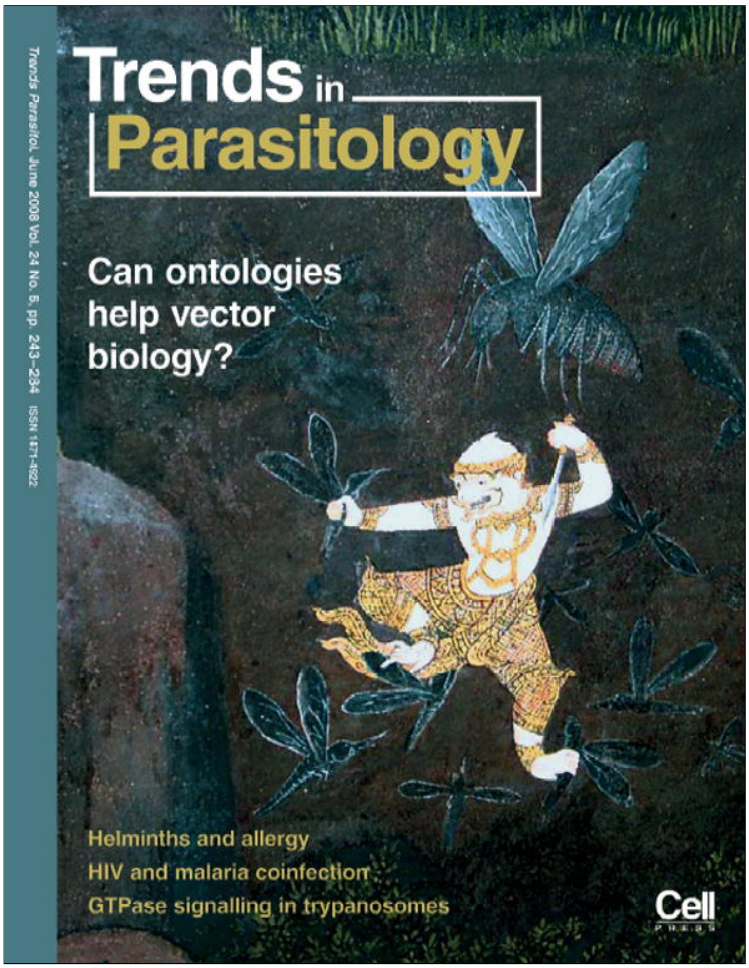

Letters

Parasite misidentifications in GenBank: how to minimize their number?

Gediminas Valkiūnas[1], Carter T. Atkinson[2], Staffan Bensch[3], Ravinder N.M. Sehgal[4] and Robert E. Ricklefs[5]

[1] Institute of Ecology, Vilnius University, Akademijos 2, Vilnius 21, LT-08412, Lithuania
[2] US Geological Survey, Pacific Island Ecosystems Research Center, Hawaii National Park, HI 96718, USA
[3] Department of Animal Ecology, Ecology Building, Lund University, SE-22362 Lund, Sweden
[4] Department of Biology, San Francisco State University, 1600 Holloway Avenue, San Francisco, CA 94132, USA
[5] Department of Biology, University of Missouri, St Louis, MO 63121-4499, USA

During the past ten years, new and remarkable information about the genetic diversity and phylogenetic relationships of parasites has been obtained through use of sensitive polymerase chain reaction-based techniques [1–8]. DNA sequences of numerous species have been deposited in GenBank, an invaluable resource for the study of parasites and other organisms. However, the majority of DNA sequences in GenBank, especially those obtained from wildlife parasites, have been identified only to the level of genus, and sometimes at even higher levels of classification. For example, as of 18 February 2008, GenBank contained sequence data for 40 named species and 400 unidentified lineages of *Plasmodium*, and 19 named species and 384 unidentified lineages of *Haemoproteus*. The low number of named species is unfortunate because linkage between DNA sequences and identifications based on traditional morphological species can provide important knowledge about basic life history strategies for parasitologists and evolutionary biologists studying the phylogenetic relationships of these organisms. There is an urgent need to remedy this because few experts possess the knowledge to identify parasite species in many branches of parasitology, and few people in the next generation of scientists are learning these taxonomic skills.

Unfortunately, the number of incorrectly identified species in GenBank is increasing. For example, the sequences of *Haemoproteus columbae* (AF069613), *Haemoproteus sylvae* (AY099040), *Plasmodium nucleophilum* (AF254962), *Plasmodium elongatum* (AF069611) and *Plasmodium relictum* (AY733088) were originally misidentified; in fact, they belong to *Plasmodium* sp. [9], *Haemoproteus payevskyi* [9], *Plasmodium ashfordi* [9], *P. relictum* [10] and *P. elongatum* [10], respectively. The sequence AY178904 does not belong to *Plasmodium rouxi,* as is referred to in GenBank [10]. The reader certainly could add to the list of such mistakes from their own fields of research. Phylogenetic analyses based on incorrect identifications are likely to be misleading and result in erroneous conclusions. It is difficult to determine how frequently these mistakes have occurred in the GenBank collection because details of voucher specimens of parasites and their place of deposition are rarely specified in recently published molecular studies.

If work on the comparison of morphospecies of organisms and their DNA lineages is to be continued without the expertise of taxonomists, there is a risk that GenBank data will include increasing numbers of misidentifications of species and even higher taxa. This would be misleading and would devalue the importance of GenBank data in the future. One way to avoid this situation might be for GenBank to require, or at least encourage, authors to deposit voucher specimens of parasites in recognized and accessible museum collections, so that they would be available for examination by experts [10]. Ideally, GenBank databases of accepted sequences, which are linked with morphospecies, should require information about the place of deposition, and accession numbers of voucher specimens of the parasites and other organisms from which the sequences were derived (Box 1). A larger issue is the continued existence of expertise and training in parasite taxonomy. Because molecular approaches are revealing vastly more parasite diversity than previously anticipated, the need for traditional systematic and biological information is becoming all the more crucial.

Box 1. Suggestions on how to optimize parasite identification in GenBank

1. For sequences that are linked with morphospecies, GenBank should require the deposition of voucher specimens and provision of the accession numbers of the organisms from which the sequences were derived. This would make molecular identifications of species repeatable, which is presently not the case. This would also link the rapidly developing field of molecular parasitology with classical parasite taxonomy and prominent parasite collections.
2. A publicly accessible ancillary database should be established for reference DNA sequences from positively identified parasite samples.
3. Projects which aim to link the knowledge of traditional parasite taxonomy with molecular parasitology should be encouraged.
4. The establishment of grants should be encouraged for long-term visits of prominent taxonomists to molecular parasitology groups and for workshops or short courses linking taxonomy and molecular parasitology.

Acknowledgements
The study was supported, in part, by the Lithuanian State Science and Studies Foundation, the Swedish Research Council, the joint NSF-NIH

Corresponding author: Valkiūnas, G. (gedvalk@ekoi.lt).

(USA) Ecology of Infectious Diseases Program award EF-0430146, the US Geological Survey Wildlife and Terrestrial Resources Program and NSF (USA) DEB-0542390. Any use of trade, product or firm names in this publication is for descriptive purposes only and does not imply endorsement by the US Government.

References

1 Bensch, S. *et al.* (2004) Linkage between nuclear and mitochondrial DNA sequences in avian malaria parasites: multiple cases of cryptic speciation? *Evolution Int. J. Org. Evolution* 58, 1617–1621

2 Fallon, S.M. *et al.* (2003) Island and taxon effects in parasitism revisited: avian malaria in the Lesser Antilles. *Evolution Int. J. Org. Evolution* 57, 606–615

3 Kimura, M. *et al.* (2006) Phylogeographic structuring of *Plasmodium* lineages across the North American range of the house finch (*Carpodacus mexicanus*). *J. Parasitol.* 92, 1043–1049

4 Križanauskienė, A. *et al.* (2006) Variation in host specificity between species of avian haemosporidian parasites: evidence from parasite morphology and cytochrome *b* gene sequences. *J. Parasitol.* 92, 1319–1324

5 Martinsen, E.S. *et al.* (2007) Morphologically defined subgenera of *Plasmodium* from avian hosts: test of monophyly by phylogenetic analysis of two mitochondrial genes. *Parasitology* 134, 483–490

6 Perkins, S.L. *et al.* (2007) The phylogeny of rodent malaria parasites: simultaneous analysis across three genomes. *Infect. Genet. Evol.* 7, 74–83

7 Sehgal, R.N. *et al.* (2005) Molecular evidence for host specificity of parasitic nematode microfilariae in some African rainforest birds. *Mol. Ecol.* 14, 3977–3988

8 Ricklefs, R.E. *et al.* (2004) Evolutionary relationships, cospeciation, and host switching in avian malaria parasites. *Syst. Biol.* 53, 111–119

9 Valkiūnas, G. *et al.* (2007) Linkage between mitochondrial cytochrome *b* lineages and morphospecies of two avian malaria parasites, with a description of *Plasmodium (Novyella) ashfordi* sp. nov. *Parasitol. Res.* 100, 1311–1322

10 Valkiūnas, G. *et al.* (2008) Polymerase chain reaction-based identification of *Plasmodium (Huffia) elongatum,* with remarks on species identity of haemosporidian lineages deposited in GenBank. *Parasitol. Res.* 102, 1185–1193

1471-4922/$ – see front matter
doi:10.1016/j.pt.2008.03.004 Available online 25 April 2008