



Published in final edited form as:

Mol Phylogenet Evol. 2009 October ; 53(1): 357–358. doi:10.1016/j.ympev.2009.04.016.

Comments on the paper by : vouching for GenBank^{Pleijel et al. (2008)}

Scott Federhen, Carol Hotton, and Ilene Mizrahi

NCBI, NLM, NIH, 8600 Rockville Pike, Bethesda MD, USA

Scott Federhen: federhen@ncbi.nlm.nih.gov; Carol Hotton: ; Ilene Mizrahi:

We commend Pleijel et al. (2008) for advocating forcefully for inclusion of specimen vouchers in all molecular systematic studies. Unfortunately, the paper contains significant factual errors that have the potential of spreading serious misinformation about GenBank among the scientific community. [GenBank is used here to denote the International Nucleotide Sequence Database, comprising GenBank, EMBL & DDBJ].

GenBank is an archival database, and our submitters are responsible for providing the taxonomic identifications for their entries – as a result, there are certainly misidentified sequence entries in GenBank. We also agree that “vouchers constitute an essential link between data and taxa”, and we have worked very hard to support and encourage that kind of annotation in GenBank entries. But they go on to state: “*Furthermore, at GenBank there is currently no dedicated field for specification of vouchers.*” This is incorrect. GenBank introduced the /specimen_voucher qualifier in 1998, and have actively promoted its use since then – GenBank currently has more than 600,000 entries from systematic and phylogenetic studies annotated with specimen vouchers. We also note that one of the authors (Oxelman) actually used the /specimen_voucher field in one of their own entries (EF061375) which appears in Table 1 of the paper.

To provide the background overlooked by Pleijel et al. (2008), specimen vouchers may be simply annotated as follows:

/specimen_voucher="Oxelman 1234 (K)" [for specimens deposited in curated natural history collections]

When we get large submissions (>10 entries) that lack adequate specimen annotation, we routinely send a form letter asking for it. We don't (and can't) require it. As Pleijel et al. point out, most journals don't require voucher annotation, though most do require GenBank accessions. We can't set the bar higher for GenBank submissions without jeopardizing this arrangement.

A free-text formatted field for specimen vouchers was a significant improvement in our annotation palette, but after several years it became clear that we could do better by adding some structure to this qualifier. We adopted the Darwin Core triplet format of <institution-code>:<OPTIONAL collection-code>:<specimen-id> and introduced two new qualifiers, so as not to dilute the meaning of already popular /specimen_voucher.

/specimen_voucher – for specimens vouchered in natural history collections

/culture_collection – for live cultures, cell lines &c.

/bio_material – for other kinds of collections: stock centers, seed banks, zoos & aquaria, DNA banks &c.

In the best cases, when a collection is digitized and accessible on the web, this structured format allows us to generate hotlinks directly from the qualifier in the sequence entry to the specimen record itself. Many culture collections have this capability, but very few museums or herbaria

to date – most notably the Museum of Vertebrate Zoology at Berkeley, and the University of Alaska Museum of the North. For example:

```
FJ151112 – /specimen_voucher="MVZ:Herp:244898"
```

This entry is hotlinked to the corresponding specimen page at the MVZ, and was published in *Mol. Phylogenet. Evol.* 49 (3), 806–826, within a month of Pleijel et al.

Table 1 deserves more discussion – each of the accessions listed there were submitted by one of the authors of Pleijel et al. (Thollesson, Jondelius & Oxelman), and each is instructive in its own way. Thollesson did not see fit to submit any specimen data with the first two entries in Table 1. Accession AJ225185, from *Dendronotus frondosus*, is linked to *Mol. Phylogenet. Evol.* 16 (2), 161–172 (2000). There is no voucher information published in this paper, nor in the previous paper, which reported the 16S sequences from the same set of specimens.

The accession cited in the second entry in Table 1 is apparently incorrect, and the genus is misspelled. Thollesson has submitted four entries from *Protopelagonemertes* sp. 544 (AJ436817, AJ436872, AJ436927, AJ436975). Neither these entries nor the associated publication [*Proc. Biol. Sci.* 270 (1513), 407–415] contain any voucher data. This is not GenBank's fault.

Likewise, Jondelius did not see fit to submit any specimen data with the third entry, AF167423. If he had, we would have annotated this with:

```
/bio_material="SMNH<SWE>:99999"
```

```
/note="SMNH 99999: illustrations"
```

The annotation 'SMNH<SWE>' deserves an aside. We were surprised to find that a catalog of natural history collection codes did not exist – so we built one, starting with the *Index Herbariorum*, the World Data Centre for Microorganisms, and specialty lists associated with resources like the Catalog of Fishes, the World Spider Catalog &c. This resource allows us to recognize that SMNH (also from Table 1) is used by three natural history museums.

SMNH – Swedish Museum of Natural History

SMNH – Saskatchewan Museum of Natural History (aka Royal Saskatchewan Museum)

SMNH – Schmidt Museum of Natural History, Emporia State University

The tag '<SWE>' indicates that SMNH in this case refers to the Swedish Museum of Natural History.

The fourth entry, from Oxelman, is properly annotated with a specimen voucher, and we commend him for this. His only fault was to put his name on a paper that claimed that it was not possible to do this.

The final entry, AJ511670, also from Thollesson is also well annotated, with `/strain="ATCC 51973 = CCUG 35103"`, which was the best that we could do in 2004. We could enhance this annotation today with `/culture_collection="CCUG:35103"`, which would hotlink directly to the CCUG page for this strain.

Pleijel et al. close by contrasting the poor example of GenBank with the Barcoding of Life Database “which demands not just voucher specimens, but also trace files from the actual sequencing.” GenBank pioneered the archiving of trace files to support the Human Genome Project, and we have worked to support voucher annotation and linkages to specimen records in natural history collections long before the Barcoding initiative existed. The GenBank tools exist, but in the end the onus is on the submitters to include relevant sequence annotation. It is

never too late to update your entries, and we encourage authors everywhere to update theirs – in most cases a simple table of accessions and field values is all it takes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- Pliejel F, Jondelius U, Norlinder E, Nygren A, Oxelman B, Schander C, Sundberg P, Thollessen M. Phylogenies without roots? A plea for the use of vouchers in molecular phylogenetic studies. *Mol. Phylogenet. Evol* 2008;48:369–371. [PubMed: 18424089]