Correspondence

Automating the "Material examined" section of taxonomic papers to speed up species descriptions

BRIAN V. BROWN

Natural History Museum of Los Angeles County, 900 Exposition Blvd., Los Angeles, CA, 90007, USA

There has been much discussion of the "taxonomic impediment" and of speeding up descriptive taxonomy (e.g., Hoagland 1996; Wheeler *et al.* 2004; Godfray 2007; Ebach *et al.* 2011). It is claimed that the rate of species description is being impeded by the lack of cyber infrastructure resources (Wheeler 2008, for example). Other factors, such as lack of funding, lack of recognition for taxonomic work, the small number of researchers, and distractions from research (committee work, teaching, administration) are also likely sources of slow rates of species description, but some aspects of traditional species description could be improved upon.

Some of the suggested cyber infrastructure resources, although excellent in quality, are expensive to purchase or have a fairly steep learning curve, or both (e.g., LUCID—*www.lucidcentral.org*). Their many advantages are exploited by relatively few workers. These tools are unlikely to help beleaguered taxonomists, many of whom pursue their studies in their spare time, or who are already underfunded and lacking in time and energy. What can more easily be done is to examine the existing species description process and determine where taxonomists can get bogged down.

In researching a group (at least a group of insects, my focal taxon), a taxonomist sorts the material, does necessary dissections and comparisons, eventually ending up in a series of specimens of each putative species. He or she then prepares descriptions, including a list of characteristics of the specimens, a definition or diagnosis of its differences from other species, a summary of its natural history, its distribution, and an all-important list of primary specimen data called "material examined". All must be formatted in an appropriate manner for both the discipline and the publication outlet.

Species descriptions in large papers can be more efficiently produced using a program like MANTIS (Naskrecki 2008), DELTA (Dallwitz *et al.* 2000), or LUCID, if one wishes to climb the learning curve or pay for the program (in the case of LUCID). On the other hand, the "material examined" section of revisions and monographs can require a large amount of time in formatting species records. Especially for large papers where species are represented by many specimens, hours of time are necessary merely to prepare this part of the manuscript. Two issues are at stake: 1) getting species records into a material examined section, and 2) getting them into the particular format of the journal at hand.

Formatting records. Most taxonomists today record specimen label data digitally, or have the capability of doing so. Usually, researchers come to collections with a laptop computer and enter data into a spreadsheet or database program. Often the program is Microsoft Excel, whose simple format is a *lingua franca* for many other programs. If not Microsoft Excel, then most programs can produce comma- or tab-delimited records that can be read by Excel.

To translate these records into a material examined section, the records must be organized by country, then state, collecting locality, date, etc. If done manually, especially in large revisions with hundreds or thousands of specimens, this can consume a significant part of the time invested in the paper. In order to circumvent this suboptimal use of time, I wrote a macro that translates Excel records into material examined sections. It is available without charge from the website *www.phorid.net/software*. Input can be cut and pasted from other Excel spreadsheets, imported from comma delimited files, or typed in directly. The fields are common ones, found in or easily extracted from most taxonomic databases (Appendix 1). Once in the spreadsheet, a single click of a button provides completely formatted output that can be cut and pasted directly into a manuscript for Zootaxa. An example prepared by the spreadsheet is given below:

Material examined. ARGENTINA: La Rioja: Santa Cruz, 28.57°S, 67.6°W, 1♂, 15.ix–4.x.2002, P.Fildago, Malaise trap (LACM); **Mendoza:** La Consulta, 33.73°S, 69.12°W, 2♂, 15–19.i.2007, S.Lanati, Malaise trap (LACM), 1♂, 19–22.ii.2007, S.Lanati, Malaise trap (LACM). **COLOMBIA: Cauca:** Isla Gorgona, 2.97°N, 78.18°W, 1♂, 1–5.iii.2000, B.Brown,G.Kung,M.Sharkey, Malaise trap #10 (LACM). **COSTA RICA: Alajuela:** Estación San Ramón Oeste, 620m, 1♂, 11–15.iv.1994, C.Moraga (INBC); **Guanacaste:** 3km SE Rio Naranjo, Finca Montezuma, 10.67°N, 85.06°W, 600m, 1♀, 21.ii.2002, Brown,Walker,Gonzalez, Malaise trap (LACM).

Specific journal formats. A further question is whether different formats should be provided for other journals. The material examined section for various journals can vary markedly. For instance, a record for the journal Zootaxa might read:

Material examined. BELGIUM: Gent: Bourgooien-Ossemeersen, 51.05°N, 3.72°E, 1♂, 7.viii–4.ix.1993, M.Pollet, white pan traps (LACM),

whereas the same information published in the journal Systematic Entomology would look like this:

Material examined, BELGIUM: 1♂, Gent, Bourgooien-Ossemeersen, 51.05°N, 3.72°E, 7.viii–4.ix.1993, white pan traps *(Pollet)* (LACM).

Getting these fields just right for the journal expends a large amount of the author's time, the journal's copy editor's time, or both. The format of the spreadsheet program described here is that of Zootaxa, which is relatively free of idiosyncrasies (other than the "#", which can be easily stripped out), easily readable, with full access to all data. Other formats might be supported in future versions of the program.

Another concern for marerial examined lists is their accuracy. A missed "-" sign can place a South American specimen in the Northern Hemisphere, for example. A function in AutoMatEx, called MAPR, prints out an online distribution map for the AutoMatEx data so that records can be visually skimmed and verified. This program uses the web mapping site SIMPLEMAPPR, and requires an internet connection to function.

Summary. An EXCEL spreadsheet that organizes specimen records for the material examined section of Zootaxa articles is described. An easily used, standalone utility, it speeds the process of species description for those who do not wish to invest the time or money in a more comprehensive specimen management software package.

Acknowledgements

Literature cited

Dallwitz, M.J., Paine, T.A. & Zurcher, E.J. (2000) Principles of interactive keys. Available from: http://delta-intkey.com (Accessed 5 July 2012).

Ebach, M., Valdecasas, A. & Wheeler, Q. (2011) Impediments to taxonomy and users of taxonomy: accessibility and impact evaluation. *Cladistics,* 27, 550–557.

Godfray, H. (2007) Linnaeus in the information age. *Nature, 446*, 259–260.

Hoagland, K. (1996) The taxonomic impediment and the Convention on Biodiversity. *Association of Systematics Collections Newsletter,* 24, 61–62, 66–67.

Naskrecki, P. (2008) Mantis v. 2.0 - A Manager of Taxonomic Information and Specimens. Available from: http://insects.oeb.harvard.edu/mantis (Accessed 30 July 2012).

Wheeler, Q. (2008) The new taxonomy. *In:* A. Warren (Ed), *The Systematics Association Special Volume Series*. CRC Press, Boca Raton, London, New York, p. xi + 237.

Wheeler, Q., Raven, P. & Wilson, E. (2004) Taxonomy: impediment or expedient? *Science,* 303, 285.

APPENDIX 1. The AutoMatEx spreadsheet (available at http://phorid.net/software)

System requirements. Microsoft Excel 2007, 2010, 2013. Works on Windows or MacOS platforms.

Terms of use. Any publication resulting from use of this spreadsheet should cite in the methods section "Material examined sections of species descriptions were prepared using AUTOMATEX (Brown 2013)", or something similar.

The literature citation is:

Brown, B.V. (2013) Automating the "Material examined" section of taxonomic papers to speed up species descriptions. *Zootaxa*, 3683 (3), 297–299.

Instructions
Download the spreadsheet. Make a backup copy. The fields are listed across the top of the first worksheet ("Input"), and consist of:

Barcode. This is the place for globally unique number for the specimen such as "LACM ENT 134115". If this field is blank, AutoMatEx will fill it with sequential numbers ("specimen 1, specimen 2, etc.').

Gender. Either "m" (male), "f" (female), or "?". REQUIRED. If this field is blank, the program will not work.

Country. The program will both capitalize and bold the country names for easier viewing in long lists.

State/province. If not given, then "[no state]" is inserted. The program will bold this field for easier viewing in long lists. I recommend that the dataset be edited to replace blank "state" fields with their correct values and AutoMatEx be rerun.

Locality. Can include County, and any other directional information. Those who wish to organize their records by County should put the county information first, so it is alphabetized. No limit on length.

Latitude. Any format; include degrees symbol and "N" or "S" (for North or South) for AutoMatEx; decimal degrees required for MAPR (i.e., 3.5°N, not 3°30'N).

Longitude. Any format; include degree symbol and "W" or "E" (for West or East) for AutoMatEx; decimal degrees required for MAPR (i.e., 3.5°E, not 3°30'E).

Elevation. Units are meters.

Date. A single field that can hold ranges, such as "1–10.vii.2010". If not given, then "[no date]" is inserted.

Collector(s). Any format.

Trap. Any collecting method used.

Ownership. Place for the coden of the collection to which the specimen belongs, e.g., "LACM" or "AMNH".

Name. Genus, species, Author. Optional¸ unless more than one species is included.

Holotype. Optional. If the text "HT" (upper case) is entered in this column for one of the records, that record is treated as a holotype specimen and all other specimens are treated as paratypes. If you don't want this, leave out the HT and just format as "Material examined". Currently, there is no provision to have holotype, paratypes, and other material examined for a single species (although I might add this if it is requested).

Running the program.
Windows: If prompted to allow macros, select allow.
Before entering data, I recommend running a built-in macro, "Ctr-k" (that is, hold down the Control key and press the "k" key at the same time) to make sure the sheet is clean. Do this while on "Input."
Still on Input, enter data either manually or by cutting and pasting from another spreadsheet. Start at cell A2.
Once the data are in place, enter "Ctr-m". A few seconds later, a popup will appear, stating "Program done". Click on "ok" and you will be on a sheet called "output" with a field highlighted containing your material examined section. Copy this cell (Ctr-C) and paste it into whatever document you wish to use.
If the macro doesn't seem to work, make sure you are not editing a field in EXCEL, as nothing can happen while an edit is open. Click once on an empty cell to be sure.

MacOS: If prompted to allow macros click "allow".
Click on "Tools", "Macros", select "Automatex" from the list, then "Run".

For comments, questions, or to report bugs, send email to AutoMatEx2012@gmail.com.

MAPR

This is a quick mapping utility using SIMPLEMAPPR (www.simplemappr.net) and the same data that you used for AutoMatEx. It is good for visually checking to make sure your coordinates are correct (for instance, that no dots for terrestrial taxa appear in the mid-Atlantic). Currently intended for one species only, but multiple species data sets will display all localities.

Instructions
Windows: As for AutoMatEx, but instead of using "ctr-m", type 'ctr-p". Microsoft Internet Explorer required.
MacOS: Not currently available.