Correspondence

Widespread mistaken identity in tropical plant collections

Zoë A. Goodwin[1], David J. Harris[2], Denis Filer[1], John R.I. Wood[1], and Robert W. Scotland[1,*]

Specimens of plants and animals preserved in museums are the primary source of verifiable data on the geographical and temporal distribution of organisms. Museum datasets are increasingly being uploaded to aggregated regional and global databases (e.g. the Global Biodiversity Information Facility; GBIF) for use in a wide range of analyses [1]. Thus, digitisation of natural history collections is providing unprecedented information to facilitate the study of the natural world on a global scale. The digitisation of this information utilises information provided on specimen labels, and assumes they are correctly identified. Here we evaluate the accuracy of names associated with 4,500 specimens of African gingers from 40 herbaria in 21 countries. Our data show that at least 58% of the specimens had the wrong name prior to a recent taxonomic study. A similar pattern of wrongly named specimens is also shown for Dipterocarps and *Ipomoea* (morning glory). We also examine the number of available plant specimens worldwide. Our data demonstrate that, while the world's collections have more than doubled since 1970, more than 50% of tropical specimens, on average, are likely to be incorrectly named. This finding has serious implications for the uncritical use of specimen data from natural history collections.

Our study examined the history of all names associated with more than 4,500 specimens of *Aframomum*. We extracted the full determination history of each specimen from the time it was collected until the present (Figure 1A). Subsequently, we evaluated the accuracy of names associated with each specimen over time relative to the current name, as determined in the recent monograph [2]. Additionally, we measured the inconsistency of names associated with more than 21,000 specimens of the tree family Dipterocarpaceae, basing this work on material duplicated in different herbaria. We also assessed the percentage of synonyms and invalid names relative to the total number of names associated with 49,500 specimens of *Ipomoea* stored in GBIF (Supplemental Figure S1A). Finally, we documented the increase in the number of tropical herbaria and the accumulation of specimens for several taxa and geographical regions (Supplemental Figure S1B–CI-X).

Figure 1B charts the complete determination history of all names for all specimens of *Aframomum*. Before the current monograph, more than 58% of specimens were misidentified or only identified to genus or family (indeterminate), or given a name that was a synonym of the correct name (Supplemental Table S1). Figure 1B demonstrates two important facts. First, before a taxonomic revision, a large percentage of specimens have the wrong name. Second, the number of specimens doubled between 1969 and 2000 (Figure 1B).

To explore the generality of our results, we assembled a number of other datasets. From a total of 58,860 specimens of Dipterocarpaceae from nine herbaria, we identified 9,222 collections, each represented by at least two duplicate specimens held at different herbaria, making a total of 21,075 specimens. Of these collections, 29.1% had different names in different herbaria. We also investigated names in the aggregator database GBIF. For this, we measured the accuracy of names rather than specimens. Examination of the 560 *Ipomoea* names

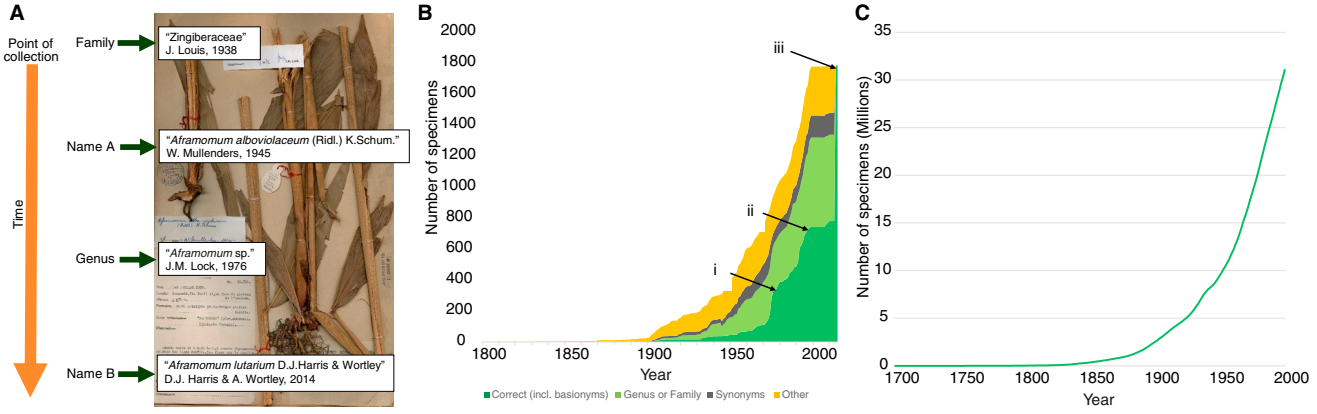

Figure 1. Determination history, accurate identification and growth in the number of plant specimens.
(A) Visualization of determination slips on a specimen of *Aframomum*, J. Louis 10158 (BR). In 1938 when the plant was collected, it was identified to family, as Zingiberaceae, by the collector J. Louis. In 1945, it was identified as *Aframomum alboviolaceum* (Ridl.) K.Schum. by W. Mullenders. Thirty two years later, in 1976, it was recognized as not fitting any existing species concepts and re-identified as *Aframomum* sp. by J.M. Lock. In 2014 the plant was recognized as a new species, *Aframomum lutarium* D.J.Harris & Wortley, by Harris & Wortley as part of the revision of the genus *Aframomum*. (B) Accumulation of *Aframomum* specimens and their determinations between 1800 & 2014 (N = 1,779), rapid increases in quality of determinations due to (i) J.M. Lock in the mid-1970's, (ii) M.M. Dhetchuvi in the mid-1990's and (iii) the current monograph. (C) Accumulation of 31,068,510 plant specimens in herbaria between 1700 and 2000, which are now present in GBIF; 50% of specimens collected since 1969.

associated with 49,500 specimens in GBIF (Figure S1A) revealed a large proportion of the names to be nomenclatural and taxonomic synonyms (40%), invalid, erroneous or unrecognised names (16%, 'invalid' in Figure S1A). In addition, 11% of the specimens in GBIF were unidentified to species.

The number of herbaria in the world has increased enormously during the late 20th Century (Figure S1B), doubling between 1957 and 2000.This partly reflects the enormous growth in the number of tropical plant specimens in the latter part of the 20th century, with a doubling of the number of collected specimens stored in herbaria between 1970 and 2000. This pattern can be seen in a range of taxa (Figure 1B, Figure S1CI–VI), geographical regions (Figure S1CVII–X) and in the 31 million plant specimens currently available in GBIF (Figure 1C). Therefore, our results suggest at least three reasons why so many specimens have inaccurate names. First and most important, too few taxonomic revisions across the entire geographical distribution of taxa in recent times means that the taxonomy and nomenclature of these groups are provisional and many specimens remain wrongly named, unrecognised and/or not determined for decades (Figure 1B). Second, the number of available specimens for any sizeable group is considerable (Figure 1C, Figure S1CI–VI). Third, the number of herbaria has greatly increased (Figure S1B), which means that there are too many herbaria for a given expert to visit or request loans from. Rapidly increasing numbers of specimens in increasing numbers of herbaria are not being revised because there are too few taxonomists.

We consider our results are representative of other tropical collections for several reasons. First, most tropical taxa have not been monographed in the last 50 years [3]. This means that the patterns we document for *Aframomum* can be predicted for many other taxa. In addition, many other published measures demonstrate levels of uncertainty in flowering plant taxonomy and nomenclature. These include an estimated synonymy rate of 65% [4]; general recognition that most described species are poorly understood [5]; and the facts that 90% of tropical species have been collected so infrequently that they are effectively unavailable for climatic modelling [6] and that less than 20,000 species of flowering plant have had an IUCN Red List conservation assessment [7]. The taxonomic status of only 39.4% of 951,140 published species names of angiosperms are considered by the Plant List as having been determined with 'high confidence' [8].

These reported uncertainties combined with our research provide further support for our main result that more than half of all tropical plant collections may be wrongly named — a result that could be true for temperate species as well. Our results highlight a serious disconnection between the rapid accumulation of tropical plant collections and the capacity to accurately identify those collections. Even when an economically important group such as the Dipterocarps has been recently revised [9], this knowledge is not necessarily transferred to accurate names in herbaria. We assume that the pattern we document for flowering plants in this paper is also true and possibly worse for insects, given that the number of described insects is three times that of flowering plants.

The *Aframomum* results (Figure 1B) illustrate the dynamic and interdependent nature of species discovery, specimen accumulation and accuracy of specimen names. The reason why more than 58% of the specimens were incorrectly named is because the taxonomy and species delimitation had advanced piecemeal over a century or so using a limited sample of specimens (see Supplemental Experimental Procedures). As specimens have accumulated in an increasing number of herbaria [10] over the last 40 years, the traditional approach of expert taxonomists visiting relevant herbaria to identify specimens is no longer tenable. Digitised specimens, remotely accessed and integrated into species-level taxonomy, are essential to improve the names associated with the world's natural history collections. Specimen data have huge potential to address global environmental problems, but the rate of increase in natural history collections across the world has greatly outpaced the ability to process, evaluate and name them correctly.

SUPPLEMENTAL INFORMATION

Supplemental Information includes one supplemental figure, one table, and experimental procedures and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2015.10.002.

ACKNOWLEDGEMENTS

This research was funded by an NERC studentship and Systematics Research Fund award from the Systematics Association and Linnean Society to Z.A.G. and a Leverhulme Trust & SynTax award from BBSRC, NERC, Systematics Association and the Linnean Society to R.W.S.

REFERENCES

1. Khoury, C.K., Heider, B., Castañeda-Alvarez, N.P., Achicanoy, H.A., Sosa, C.C., Miller, R.E., Scotland, R.W., Wood, J.R.I., Rossel, G., Eserman, L.A., *et al*. (2015). Distributions, ex situ conservation priorities, and genetic resource potential of crop wild relatives of sweetpotato [*Ipomoea batatas* (L.) Lam., I. series Batatas]. Front. Plant. Sci. *6*, 251.
2. Harris, D.J., and Wortley, A.H. (2015). Monograph of *Aframomum* (Zingiberaceae). Syst. Bot. Monogr., in press.
3. Wayt Thomas, W. (1999). Conservation and monographic research on the flora of Tropical America. Biodivers. Conserv. *8*, 1007–1015.
4. Wortley, A.H., and Scotland, R.W. (2004). Synonymy, sampling and seed plant numbers. Taxon *53*, 478–480.
5. Heywood, V. (2001). Floristics and monography-an uncertain future? Taxon *50*, 361–380.
6. Feeley, K.J., and Silman, M.R. (2011). The data void in modeling current and future distributions of tropical species. Global Change Biol. *17*, 626–630.
7. IUCN (2014). The IUCN Red list of threatened species. Version 2014-3. http://www.iucnredlist.org/ Accessed 17 February 2015.
8. The plant list (2013). Version 1.1. http://www.theplantlist.org/ Accessed 1st January 2015.
9. Ashton, P.S. (2004). Dipterocarpaceae. In tree flora of Sabah and Sarawak, Volume 5, E. Soepadmo, L.G. Saw, and R.C.K. Chung, (eds.) (Kuala Lumpur, Malaysia: Government of Malaysia), pp. 63–388.
10. Thiers, B.M. (2015). Index herbariorum: A global directory of public herbaria and associated staff. (New York Botanical Garden's Virtual Herbarium). http://sweetgum.nybg.org/ih/ Accessed 20 February 2015.

[1]Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, UK. [2]Royal Botanic Garden Edinburgh, Edinburgh EH3 5LR, UK.
*E-mail: robert.scotland@plants.ox.ac.uk