Saving our science from ourselves: the plight of biological classification

Malte C. Ebach[1], Marcelo R. de Carvalho[2] & Silvio S. Nihei[2]

[1]Evolution & Ecology Research Centre, School of Biological, Earth & Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia. mcebach@gmail.com (corresponding author)
[2]Departamento de Zoologia, Instituto de Biociências, Universidade de São Paulo, Rua do Matão, Trav. 14, no. 101, 05508–900 São Paulo-SP, Brazil. mrcarvalho@ib.usp.br; silvionihei@gmail.com

ABSTRACT. Saving our science from ourselves: the plight of biological classification. Biological classification (nomenclature, taxonomy, and systematics) is being sold short. The desire for new technologies, faster and cheaper taxonomic descriptions, identifications, and revisions is symptomatic of a lack of appreciation and understanding of classification. The problem of gadget-driven science, a lack of best practice and the inability to accept classification as a descriptive and empirical science are discussed. The worst cases scenario is a future in which classifications are purely artificial and uninformative.

KEYWORDS. Classification; DNA barcoding; morphology; nomenclature; systematic; taxonomy.

RESUMO. Salvando a nossa ciência de nós mesmos: a saga da classificação biológica. A classificação biológica (nomenclatura, taxonomia e sistemática) está sendo subestimada. O desejo por tecnologias novas e por descrições taxonômicas, identificações e revisões mais rápidas e baratas, é sintomático da falta de apreciação e entendimento sobre a ciência da classificação. O problema de uma ciência dirigida por máquinas, a ausência da 'melhor prática', e a inabilidade de se aceitar a classificação biológica como uma ciência descritiva e empírica são discutidos. O cenário mais pessimista é um futuro no qual as classificações serão inteiramente artificiais e não-informativas.

PALAVRAS-CHAVE. Classificação; código de barras DNA; morfologia; nomenclatura; sistemática; taxonomia.

Scientific fields that have made few technological and time efficient advances are slowly being phased out in many parts of the world, such as paleontology, taxonomy, and morphological systematics. Rather than promote these fields as essential to biological science, we instead fund new technologies that store taxonomic and morphological information, undercutting ourselves, to some degree, as systematists. Morphology and taxonomy are basic sciences that can be applied in a number of different ways within biology (ecology, co-speciation, etc.), but more is invested in applications and information technologies (databases, etc.) than in the general pursuit of knowledge (*e.g.* species discovery, relationships, classification). According to one reviewer (Flowers 2007: 5), database initiatives that have as primary goal the application of new technology to existing data, rather than generate new data, have "mopped up a not inconsiderable fraction of the available money during the Biodiversity Decade" (the 1990s). We agree with Flowers on that front, but more conceptual impediments to classification are what concern us here. We investigate some of these problems and identify present trends that are potentially damaging to systematics.

The Information Boom and the Knowledge Bust. A certain degree of confusion exists between what constitutes information and knowledge:

"*Our view is that it would be a major retrograde step, disconnecting future taxonomy from the wealth of knowledge that has been built up and indexed under the Linnaean system*" (Godfray *et al.* 2007: 954).

Classifications contain information in the form of data. Knowledge is a dynamic activity that involves observation and comparison, which over time produces informative statements about data. Not all information, however, is necessarily informative about knowledge. A diagnosis or taxonomic key may help a novice identify a beetle. However, a coleopterist is able to confirm whether the information is correct or not. A taxonomist provides knowledge, whereas a key or description merely records data. Information and knowledge are separate issues. The following examples highlight this confusion:

"*In a barcoded world, taxonomists will retain their leadership role in the association, integration, and interpretation of knowledge about the character state variation that delineates species and what this implies for higher level taxonomy*" (Hebert & Gregory 2005: 855).

Barcodes are simply data, which may or may not be informative. Actual knowledge about a taxon is not contained within DNA. This knowledge derives from direct observation of morphology, contra Hebert & Gregory (2005):

"*By contrast, DNA barcodes – by themselves – are never sufficient to describe new species. At some stage, clearly divergent DNA barcodes, in combination with other*

information, will be used as the basis for providing a new Linnaean name […] *and, as with any taxonomic hypothesis, this would be subject to ongoing reevaluation*" (Hebert & Gregory 2005: 853).

The first sentence confirms the limitations of information. Barcodes cannot describe and therefore cannot name species. In other words, naming a species is based on a description that allocates the new form to a place in the hierarchy of life. As such, descriptions are more than the accumulation of separate pieces of information – they reflect knowledge. This taxonomic procedure is misunderstood by Hebert & Gregory (2005).

The assumption that information = knowledge, however, is integral to bioinformatics. If knowledge cannot be recorded and stored, data-basing, for instance, looses its appeal somewhat as a taxonomic tool. Moreover, non-morphological pursuits, such as DNA barcoding, would appear less vital and less significant than their promoters claim. Information is simply data, which may or may not be informative. Knowledge is obtained from observation, comparison and experience (de Carvalho & Ebach 2009). Only the former can be stored, cataloged and retrieved with new emergent technologies. We doubt whether the latter can be replaced by technology. Information, however, drives technology and current trends in systematics and biogeography. We assess below three areas that we feel are compromising knowledge within systematics and biogeography.

Gadget Science: Technology and the Public Perception. The appeal of technology over science has spawned a new generation of scientists that shun biological classification (*e.g.* taxonomy) as the realm of dusty stamp collectors who refuse to accept emergent technologies. In an editorial in Systematic Biology, the then editor, Rod Page, claimed that:

"… *the Journal has published insightful investigations of the strengths and limitations of automated methods of delimiting taxa* […] *Phylogeneticists are tackling increasingly large amounts of data, often obtained from diverse sources, which raises issues about workflows to automate handling of large data sets* […] *and developing more sophisticated methods of linking data together* […]. *I expect that these areas will see further exciting developments*" (Page, in Page & Sullivan 2008: 2).

Imposing technologies such as "automated methods" creates the image of taxonomic technophobes. Taxonomy and systematics can function equally well without emergent technologies, as the last 250 years has successfully shown. It is obvious that technological innovations may be quite useful, but taxonomy and systematics are not dependent on those technologies to work.

Public perception of taxonomy and systematics differs remarkably from that of science funding bodies and end-users of taxonomic data, as shown by the recent discovery of Ida, the lemur-like primate *Darwinius masillae* (Franzen *et al.* 2009). The discovery of this wonderfully well-preserved fossil from Messel, Germany, and the possibility that it represents a "missing link" between primates and humans took the public by storm. The public appeal lies in the achievements of paleontology and paleontologists. Incredibly, while the public's interest in paleontology increases, the number of jobs for paleontologists decreases. One would consider the reverse to be true, given its recurring popularity.

Public perception does not necessarily influence science funding policy. Instead, technological initiatives drive science funding. Technology has greater appeal due to its increase in efficiency while at the same time reducing total cost. Public perception of museums filled with paleontologists hovering over the brink of discoveries runs contrary to what is actually happening. Rather than fund taxonomic initiatives, museums are attempting to compete with universities and applied research institutions. The result is a somewhat sterile world of molecular labs, computer algorithms, and a lack of taxonomic specialists-museums may quickly become mere service-providers to biotechnological industries at the expense of continued research in systematics.

Research institutions, which may include natural history museums, manipulate this false perception in order to gain public support, while at the same time radically altering their own internal policies (*i.e.*, shifting towards DNA barcoding, genomics, bioinformatics) in order to benefit financially. Yielding to funding bodies seems ironic for institutions such as public museums or even universities. Surely, researchers are capable of determining their own research priorities?

The reasons behind the above deception lie in how research priorities are met. Progress is usually confused in systematics and taxonomy as modernization. For example, very little progress has been made in understanding the systematic relationships and classification of asphasid trilobites. The dwindling numbers of trilobite researchers, especially those brave enough to tackle asphasids, have long to go before any real progress is made, both in trilobite classification and species discovery. Interpret 'progress' as 'modernization' (*i.e.*, technology), and we would only need to database existing classifications to meet this new standard. These new standards shift priorities, change our values and collectively influence funding bodies. Scientific standards are expected to evolve and are what regulate our research priorities. Taxonomists, preoccupied only with classifying their groups, do not live up to these standards, regardless of whether their work provides a stable foundation for these newly emerging applied fields such as DNA barcoding or genomics. Moreover, researchers from these applied fields have cast themselves as spokesmen for systematics and taxonomy – a practice usually reserved to end-users of a technology or service. Has the perception of systematics and taxonomy been reduced to a mere service by applied scientists?

Services are sought in times of need. Presently conservationists and ecologists are assessing what has been described as a 'biodiversity crisis'. Given the lack of taxonomists due

to the new priorities and skewed perception described above, researchers have opted for more technology to resolve a non-technological problem.

Best Practice in Biological Classification: the 'Clades in Grades' Problem. The rise of the Modern Synthesis demoted classification to a purely descriptive pursuit. 'Evolutionary systematists' discard natural classification as non-evolutionary and promote artificial groupings such as non-monophyletic groups. The reason for this departure from traditional classification is the lack of an evolutionary synthesis within taxonomy and systematics:

"*The point of systematics in an evolutionary world ought not to be the construction of classes, but the reconstruction of history* […] *and the analogy of systematics to classification is in fact a relict of the pre-evolutionary period, when living diversity was viewed ahistorically*" (O'Hara 1994: 14).

The problem of using an evolutionary synthesis in classifying taxa (and areas) is that something about evolution needs to be known prior to classification. Since natural classifications discover evolution in the form of monophyly, we are faced with a quandary, namely 'clades in grades'. Take for instance the classical example of dinosaurs and birds. Birds were united as a taxonomic group called Aves by Linneaus (1758) before we discovered that they are indeed monophyletic. In the original taxonomic description, no evolutionary synthesis was used. Linnaeus had just correctly inferred that birds are more closely related to each other than they are to any other taxon due to their morphological characteristics. Aves, the taxonomic group, were later shown to be monophyletic through systematic analysis. Further discoveries of 'primitive birds' and theropods (especially from China) changed what we know of bird and dinosaur systematics. This new information can now inform our taxonomy: Aves will need to be re-diagnosed as certain bird-like theropods may be included in this group. This rather common procedure in classification, however, seems to be ignored or misunderstood:

"*Despite the overwhelming evidence that birds are nested in theropods, major questions remain... The plethora of intermediates connecting dinosaurs and birds has shifted the question from whether birds are descended from dinosaurs, to where we shall draw the line between dinosaurs and birds. There is now a strong consensus that birds are an integral part of the dinosaurian radiation and must be classified as a subgroup of dinosaurs, in much the same way humans must be considered a subgroup of primate mammals*" (Lee *et al.* 2004: 463).

Birds may be nested in theropods, but this does not make them dinosaurs (characters that diagnose birds are sufficient to relate all 'primitive birds' and certain theropods, but exclude sauropods and other dinosaurs; in other words, birds are monophyletic and 'dinosaurs' are not). Similarly, recent phylogenies have supported a close relationship between Crustacea and Hexapoda. More than a close relationship, these phylogenies indicate that crustaceans are paraphyletic as part of Crustacea is more closely related to Hexapoda than to any other crustacean (Schram & Koenemann 2004). The crustacean taxon hypothesized as sister-group of Hexapoda is in dispute, and the list of candidates includes the branchiopods (Schram & Koenemann 2004), the Remipedia + Cephalocarida (Regier *et al.* 2010), among others.

Insects may be nested within branchiopods, but this does not make them crustaceans. Birds may be nested in theropods, but this does not make them dinosaurs. We know that 'invertebrates' are paraphyletic because they exclude vertebrates, a sub-group of descendants. There is a strong consensus that vertebrates are an integral part of the greater 'invertebrate' radiation (Deuterostomia) and must be classified as one of their subgroups. Thus, to say that birds are a subgroup of 'dinosaurs' is as informative as saying that birds are a sub-group of 'invertebrates'.

Another example of 'clades in grades' is the relationship among members of Dictyoptera. This higher-level taxon comprises the Blattaria (cockroaches), Mantodea (mantises) and Isoptera (termites). Although the monophyly of Dictyoptera is widely accepted today, the relationships among those three orders and even their monophyly are highly controversial. Some studies indicate that Isoptera is deeply nested in Blattaria, as Cryptocercidae (wood-feeding cockroaches) is more related to Isoptera than to any other taxon (Klass & Meier 2006).

In classification, dinosaurs can never be termed birds, and cockroaches cannot be called termites. In phylogenetics, however, it is a different matter. The 'clades in grades' problem is one of phylogenetic inference mistaken as classification. The source of this problem is a popular synthesis ignoring procedures or best practice in classification. Also referred to as 'tree thinking', a phylogenetic or synthetic approach sees ancestors as meaningful in determining what their descendants are – as in the case of the slogan 'dinosaurs are birds'.

For a long time, the Hemiptera was traditionally divided into Homoptera and Heteroptera. The monophyletic clade Heteroptera is now inferred to be descended from 'homopterans'. This means that by identifying the ancestor we can determine the genealogical lineage that leads to a descendant. Systematics has shown that some homopterans (Coleorrhyncha) are more closely related to Heteroptera than they are to any other group. Phylogenetic inference would immediately 'classify' heteropterans to be 'homopterans', because heteropteran characteristics can be linked to homopteran characteristics via a lineage or phylogeny. But what does this tell us about the classification of 'Homoptera'?

Grades, such as dinosaurs, crustaceans, homopterans, entognathans, among many other examples, have no meaning in a natural classification. This is because they belong to a taxon that shares closer relationships to other taxa than to each other. Such taxa exist in name only, because they have no shared characteristics unique to themselves.

Ignoring classification in favor of evolutionary scenarios is dangerous. Those that venture into 'Homoptera' or 'Entognatha' evolution will realize that homopterans and entognathans simply do not exist in any classification. Reptiles pose similar problems as they also represent a grade. As a taxonomic group they are no more than a name linked to a diagnosis, but yet 'reptiles' are commonly used in ecological and conservational research.

Best practice in systematics (the pursuit of natural groups), taxonomy (the classification of taxa) and nomenclature (rules governing the naming of names) needs to be observed before inferring phylogeny (ancestor-descendant lineages). These best practices are consistently compromised by 'tree thinking' and other syntheses that wish to 'classify' lineages rather than focus on natural classifications. Since natural classifications tell us whether taxa are based on homologies, they are evidence for evolution. Assuming that something is monophyletic prior to systematic analysis (testing for monophyly) is placing the cart before the horse.

Accepting Our Science: Classification not as Stamp Collecting. Various slogans used in science promote the importance of applied sciences by demeaning others. An infamous slogan attributed to the natural sciences is that of "stamp collecting":

"*The endeavour of natural history has often been ridiculed as "mere stamp collecting" by those unwilling to see anything scientific in naturalists' work*" (Johnson 2007: 172).

This mentality also exists within the natural sciences, with those considering their fields to be more experimental or 'scientific' and claiming superiority over others. Most notably, it is classification that bears the brunt of this sloganeering:

"*I have consequently announced that I have founded the fourth great school of classification, the It-Doesn't-Matter-Very-Much school. Actually, systematists 'voted with their feet' to establish this school, long before I announced its existence*" (Felsenstein 2004: 145).

"*It turns out that Joe is the founder of the "It-Doesn't-Matter-Very-Much" school – get the tree right and shut up – with a healthy scepticism towards the importance of a formal classification*" (Penny 2004: 669).

Classification has both a descriptive (taxonomy) and an empirical (systematics) component, both of which, together, form a highly dynamic field that many other disciplines rely on in order to make inferences about taxa. The recent discovery of a fossilized finger bone in Siberia has created some media attention. Mitochondrial DNA (mtDNA) extracted from the bone has revealed a molecular signature that is unlike any known for humans (Krause *et al.* 2010). The first reaction is that this is a new species. Unfortunately this is all that the molecular researchers can infer from one single individual bone. Without further morphological evidence, the

distinction between humans and other taxa cannot be made. Regardless of this lack of evidence, it is solely variation among molecular data that are heralded, providing science with new insights into human evolution. The same is true for genetic studies of fossil hominids and apes. Ironically, morphological data are consistently rejected in favor of molecular inferences, even by paleoanthropologists, as though sequence data is more meaningful than morphological data. In each case it is technology infused with new forms of data that appears to progress science, when in fact it is nothing more than a form of micro-morphological data that are (or should be) used in exactly the same way as their macro-morphological equivalent.

Setting the technology and varying forms of data to one side, classification is still an important part of natural science. In many cases traditional classifications do not provide the answers quick enough because of a lack of data, expertise or productivity. In its place many have suggested using quick-fix techniques that attempt to "fix" classification (DNA barcoding, DNA taxonomy, Phylocode, etc.), but which provide unreliable results.

Alternatives to traditional classification are not an option. Classification takes time, in the same way that deep-space probes take time to navigate between planets. Anyone claiming to cut times between vast distances would be met with extreme skepticism in the astronomical community. However, similar suggestions in biology are immediately embraced by those wishing to speed up the process of classification.

Saving Our Science from Ourselves. Classification is a vital part of comparative biology and without it we are unable to make evolutionary, ecological, behavioral or genetic inferences. Yet, classification is sold short through misconceptions about technology, an ignorance of best practices and attempts to find quick and cheap alternatives to classification. Re-educating students and academics of the importance of classification, 'best practice' and the need for time may improve the scientific and public perception of classification. Without such efforts from the biological community classification will degenerate into something akin to reading one-dimentional barcodes – saving classification from our own ignorance is paramount to saving evolutionary biology.

ACKNOWLEDGEMENTS

REFERENCES

de Carvalho, M. R. & M. C. Ebach. 2009. Death of the specialist, rise of the machinist. **History and Philosophy of the Life Sciences 31**: 467–469.

Felsenstein, J. 2004. **Inferring Phylogenies.** Sunderland, Sinauer Associates, xx+664 p.

Flowers, R. W. 2007. Taxonomy's unexamined impediment. **The Systematist 28**: 3–7.

Franzen, J. L.; P. D. Gingerich; J. Habersetzer; J. H. Hurum; W. Von Koenigswald & B. H. Smith. 2009. Complete primate skeleton from the middle Eocene of messel in Germany: Morphology and paleobiology. **PLoS ONE 4** (5): e5723.

Godfray, H. C. J.; B. R. Clark; I. J. Kitching; S. J. Mayo & M. J. Scoble. 2007. The web and the structure of taxonomy. **Systematic Biology 56**: 943–955.

Hebert, P. D. N. & T. R. Gregory. 2005. The promise of DNA barcoding for taxonomy. **Systematic Biology 54**: 852–859.

Johnson, K. 2007. Natural history as stamp collecting: A brief history. **Archives of Natural History 34**: 244–258.

Klass, K.-D. & R. Meier. 2006. A phylogenetic analysis of Dictyoptera (Insecta) based on morphological characters. **Entomologische Abhandlungen 63**: 3–50.

Krause, J.; Q. Fu; J. M. Good; B. Viola; M. V. Shunkov; A. P. Derevianko & S. Paabo. 2010. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. **Nature 464**: 894–897.

Lee, M. S. Y.; T. W. Reeder; J. B. Slowinski & R. Lawson. 2004. Resolving reptile relationships: Molecular and morphological markers, p. 451–467. *In*: J. Cracraft & M. J. Donoghue (eds.). **Assembling the Tree of Life.** New York, Oxford University Press, 576 p.

Linneaus, C. 1758. **Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis.** Tomus i. Editio decima, reformata.

O'Hara, R. J. 1994. Evolutionary history and the species problem. **American Zoologist 34**: 12–22.

Page, R. D. & J. Sullivan. 2008. The expanding contributions of systematic biology. **Systematic Biology 57**: 1–3.

Penny, D. 2004. Phylogeny in the comfort zone. **Systematic Biology 53**: 669–670.

Regier, J. C.; J. W. Shultz; A. Zwick; A. Hussey; B. Ball; R. Wetzer; J. W. Martin & C. W. Cunningham. 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. **Nature 463**: 1079–1084.

Schram, F. R. & S. Koenemann. 2004. Are the Crustaceans monophyletic? p. 319–329. *In*: J. Cracraft & M. J. Donoghue (eds.). **Assembling the tree of life.** New York, Oxford University Press, 576 p.