Taxonomic Epistemology, Observer Bias, and Biodiversity Error

Jeffrey C. Nekola, Barbora Winterová & Michal Horsák
Department of Botany and Zoology, Faculty of Science, Masaryk University, Kotlářská 2, CZ-61137 Brno, Czech Republic

---

"Science is what we know, and philosophy is what we don't know." – Bertrand Russell, 1950, *Unpopular Essays*

Before biodiversity can be protected the things making it up must be known. For this reason taxonomists rightly say that their field is foundational to conservation as it though their efforts that the biological universe can be known. But there is a problem: the rules governing the naming of life were drafted early on in the Scientific Revolution, and as a result do not require that proposed names be confronted against empirical data before being accepted. Rather, their validity is often justified based on trusting that statements made by the author were accurate. And therein lies the problem: psychological factors can easily make observations inaccurate. Many taxonomists will scoff at this assertion, and contend that only through expert opinion based on a lifetime of observation can life truly be categorized. But is this true?

This issue falls within the realm of epistemology, the branch of philosophical inquiry which considers the conditions required for a belief to constitute knowledge. In the case of taxonomy, this means considering the implications of allowing "knowledge" to be based on expert-opinion rather than on limiting it only to things which have survived empirical vetting. How often do these two types of "proof" disagree, and why? And what are their implications for our ability to accurately measure and protect biodiversity?

This issue can occur in any scientific discipline when observations cannot be independently verified. To help illustrate how this happens, we will start with (hopefully) an uncontroversial example from astronomy that shows how psychology can taint "objective" observations in the absence of replicable data. We then illustrate an eerily similar situation from land snail taxonomy and show how similar issues gave rise to both.

Percival Lowell and the canals of Mars

"The scientist does not lie outside of the natural world. Rather, the scientist is entirely part of that world and is subject to ... laws of perception and cognition and to the laws of related areas of experimental psychology. It is important for all scientists, in all disciplines, to be aware of these essential facts and to use them to exert caution in the interpretation of what might otherwise be interpreted as purely objective observations." – Matthew Sharps, 2018, *Skeptical Inquirer*. 42:41-46.

Percival Lowell graduated with distinction in mathematics from Harvard in 1876. By the 1890s, using wealth from his family's cotton mills, he had retired and dedicated his life to astronomy. He became fascinated by Mars, and when he trained his 24-inch telescope on this planet he saw many "non-natural features", including “single and double canals” and "oases" where the canals intersected. He concluded that these had been made by a desperate alien race tapping Martian polar ice to continue life on a now desert world.

The problem was that not every astronomer could see Martian canals, and without the existence of high quality astronomical photographs there was no way to independently assess whether Lowell was inventing features that did not exist or whether others were not able to see things that did. Because no empirical data existed to mediate debates, Lowell saw all disagreements as personal attacks; he fired his trusted assistant A.E. Douglas after being questioned by him about his conclusions. It was not until larger telescopes with photographic abilities became available that it became clear that large canals did not exist. But Martian maps well into the 1960s generally included them until space probes established beyond all doubt that canals of any size were not present (Figure 1).

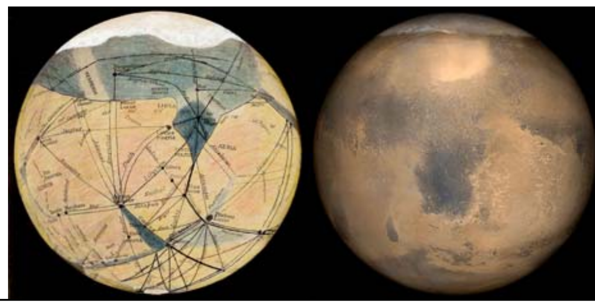

Figure 1: Comparison of Percival Lowell’s hand-drawn 1905 Martian map (left) with 2003 Mars Global Surveyor telemetry (right) taken from approximately the same perspective centered on Syrtis Major.

How could a highly trained and reputable astronomer like Lowell see things that simply did not exist? The answer likely lies in a combination of psychological factors, including personal history (his family’s wealth was partly due to New England canals with the immense engineering feats of the Suez and Panama canals having just been completed), Gestalt reconfiguration (perceiving multiple independent features as an artificial single unit), and various sociocognitive factors, like *cognitive dissonance*, where faulty ideas are defended, using increasingly incoherent rationales, when someone is sufficiently invested in them (Festinger 1957, Sharps 2018).

Clearly it is possible for apparently objective scientific observations to be faulty from human psychology. And only through the use of replicable, empirical data can such spurious observations be filtered out and excluded from further scientific discussion.

Terry Frest and Iowa’s Endemic Algific Talus Slope Snails

> Human beings have a demonstrated talent for self-deception when their emotions are stirred. Carl Sagan, 1980, *Cosmos*

Terrence Frest received his PhD from the University of Iowa in 1983, where his thesis considered the paleontology of Silurian Echinoderms. He quickly applied this training to modern invertebrates, becoming the regional expert in the (as of then) poorly known land snail fauna of the Upper Midwestern USA. Among his most important contributions was determination of the geological conditions required to generate 'algific' (cold-producing) talus slopes, in which buried ice caves expel a constant flow of near-freezing air through rock rubble, creating small patches of late-glacial maximum microclimates within the modern landscape (Figure 2). His discovery of more than 300 of these sites also documented the presence of a diverse relict biota, including the globally endangered land snail *Discus macclintocki*, globally threatened vascular plant *Aconitum noveboracense*, plus a large number of other highly disjunct plant and invertebrate species from the north (Nekola 1999). Because these sites were largely unknown prior to Frest's work, few had been afforded any form of protection. Their steep, unconsolidated rock rubble made them extremely fragile and easily degraded by livestock pasturing and even scientific research. Yet, they were among the most charismatic habitats of the central USA, with their ground surface often being bathed in cold fog even on hot summer days. Because only two of the taxa restricted to them were legally protected, money was not available to preserve most sites. This changed when Terry reported to various government agencies (Frest 1991) the presence of eight other endangered, endemic and undescribed land snail taxa, and helped lead to protection recommendations for over additional 70 sites.

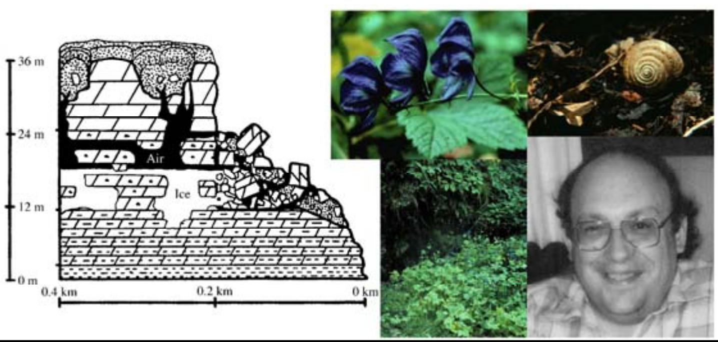

Figure 2: Clockwise from left – Geological diagram of an algific talus slope; *Aconitum noveboracense*; *Discus macclintocki*; Terry Frest in the summer of 1989; Elk Creek East site in Delaware County, Iowa, with cold fog emanating from a fissure cave;.

Frest (1991) provided highly detailed descriptions and/or biological illustrations to what he termed *Vertigo briarensis, V. hubrichti variabilis,* and *V. iowaensis* (Figure 3). The problems began when others attempted to replicate his findings: not only did resampling document populations at five-times more sites than he reported, but out of the 1000s of shells observed not a single one looked like those from his illustrations. And, his claimed diagnostic traits ignored the fact that continual gradation existed between all forms. Could it be that these taxa only existed within his mind? Clearly some type of independent assessment was required. And when Jeff did this using DNA sequence data following Terry's death (Nekola *et al*. 2009), it became clear that these endemic taxa were simply weakly-defined shell forms of *Vertigo arthuri*, a species demonstrating considerable shell variability across its range extending from Newfoundland to Alaska and down the Rocky Mountains to northern New Mexico (Figure 3).

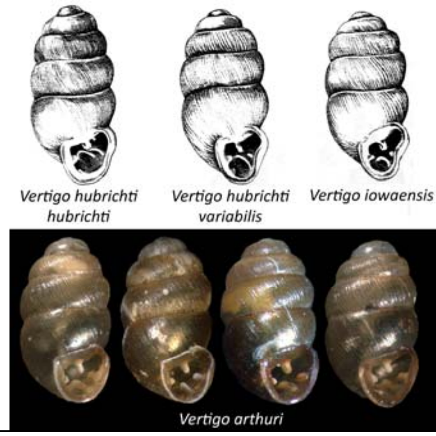

Figure 3: Top row – biological illustrations drawn by Frest of three claimed algific talus slope endemics. Bottom row – visible light photomicrographs for the four main shell forms of *Vertigo arthuri*, including (left to right): *arthuri, basidens, paradoxa,* and *hubrichti*.

How could Terry have gotten this so wrong? How could he describe – and illustrate! – shells that no one else could see? Like Percival Lowell, the answers seem to lie within the realm of psychology: His personal academic training was grounded in paleontology where "species" don't necessarily represent biologically valid entities, but convenient physical categories. And while he had discovered one of the most charismatic, fragile, and important habitats for central North America biodiversity, the large majority of sites could only be conserved by increasing the number of endangered species known from the system. By subconscious application of Gestalt reconfiguration, he was able to "see" multiple discrete taxa, all of which were endangered. Sociocognitive factors then came into play, as these new "species" afforded protection for more of these amazing habitats. To facilitate this he subconsciously applied cognitive dissonance to filter out the presence of the intermediates which tied all these forms together. Because he was the sole expert for this group in the region, it took over 15 years for his ideas to be challenged.

To withstand the domination of authority and verify all statements by an appeal to facts...

"The Royal Society's motto '*Nullius in verba'* [dates to 1663 and] roughly translates as 'take nobody's word for it.' It is an expression of the determination of Fellows to withstand the domination of authority and to verify all statements by an appeal to facts determined by experiment." – Royal Society, 2010.

Clearly one human mind can see things that others do not. How can we then find common ground to discuss ideas and advance knowledge? How can we know that patterns really exist rather than being individual manifestations of personal psychology? This was the dilemma facing natural philosophers at the dawn of the Renaissance. One solution was the use of empirical data, which was emotionally neutral and could be challenged without personally attacking the proponent of a given viewpoint. And as it also allowed for replicate observations across multiple individuals, its use mitigated some of the most egregious examples of psychological bias. The confrontation of statements with empirical data from multiple observers thus became one of the foundational concepts underlying modern science. As indicated by the motto of the *Royal Society of London* – the world's first scientific society

– no matter the authority of the person making a claim, it should only be accepted if it could be verified by others through use of data.

While science has in general embraced this approach, some fields have lagged behind. Given the ongoing biodiversity crisis, it is discouraging to note that taxonomy can represent a holdout. A large part of this is due to rules governing naming: the Codes of Nomenclature only require that a statement be given describing how a taxonomic concept is distinguished. Empirical challenge is not required – it is good enough for the author to simply say it. While this allows taxonomy to be among the most pluralistic of scientific disciplines – giving any single person the ability to name life (provided they follow the rules) – it also provides safe haven for authority-driven, non-empirical, pre-Renaissance thinking. When this happens, readers – in the absence of statistical tests – are required to accept on faith the veracity of an author's statements.

Because replicable hypothesis testing of taxonomic concepts has only been feasible for the last 75 years, the reality is that the majority of names in many groups (especially for invertebrates which make up so much of the planet's biodiversity) have never been empirically vetted. This means that biologists, ecologists, conservation biologists, policy makers and general public who use them cannot "withstand the domination of authority" by "an appeal to facts." Rather we are left to trust that the authority championing a given name has not been overly impacted by psychological observational biases and is not advocating for the biological equivalent of Martian canals.

Correcting Errors through Integrative Taxonomy

> "When you follow two separate chains of thought, Watson, you will find some point of intersection which should approximate to the truth." – Sherlock Holmes in *The Disappearance of Lady Frances Carfax*. Arthur Conan Doyle, 1911

How can taxonomic units be empirically vetted? The answer lies in the fact that biological species represent largely independent evolutionary units. As a result they will be exposed to different selection pressures and random walks. Legitimate taxa should thus demonstrate uniqueness across multiple empirical signals. Astronomers who search for consensus across gravitational wave, neutrino, photon, and cosmic ray signals term their investigative process 'multi-messenger astronomy' (Branshesi 2016). An analog exists with 'integrative taxonomy' in which a species-level hypothesis is vetted across multiple empirical data streams and is accepted as valid only after it is shown to be distinct across a consensus of these data (Box I). We have used this approach to empirically vet and firmly ground the taxonomy of a number of land snail groups.

Pupilla hebes (Ancey, 1881) provides an excellent example of this process. This species was thought to range from the SW USA through Alaska and down the western Pacific coast to Hokkaido (Pilsbry 1948). Nuclear and mitochondrial DNA sequence both illustrated that Alaskan and Japanese material were distinct species (*P. alaskensis* Nekola *et al*. 2015 and *P. hokkaidoensis* Nekola *et al*. 2015), with true *P. hebes* being limited to the western USA (Figure 4). Genetically-verified *P. hebes* populations demonstrated an astonishing range of shell sizes and aperture calcification levels (Figure 5A-E). Because these were historically used to identify *Pupilla*, most *P. hebes* populations had been previously identified as either *P.*

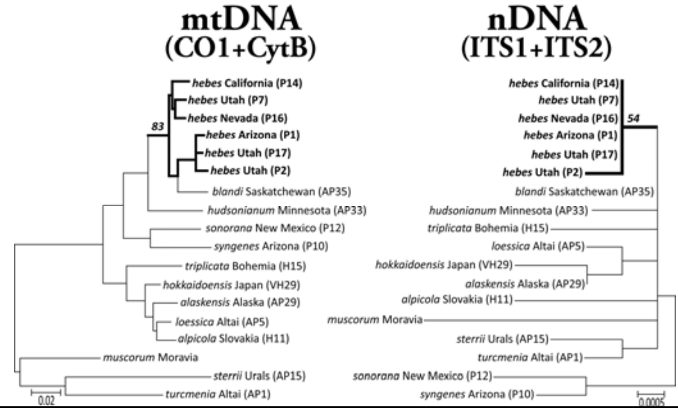

Figure 4: Highly supported Maximum Likelihood mtDNA and nDNA clades, corresponding to *Pupilla hebes*, with representatives of other Holarctic *Pupilla* species. Note that *P. alaskensis* and *P. hokkaidoensis* were both previously thought to represent *P. hebes*. The nDNA support value is low because of low statistical power from there being only four variable base pairs out of ~1575 which distinguish it from *P. blandi*, its nearest cousin. However, though few in number they are completely diagnostic.

blandi Morse, 1865 or *P. muscorum* (Linnaeus, 1758). However, a previously unnoticed shell feature did allow for accurate identification: shell microscuplture, which consisted of numerous, sharp, small, wavy ribs in *P. hebes* as opposed to almost smooth shells in *P. blandi* and larger, straighter, and more rounded ribs in *P. muscorum* (Figure 5F-G). *Pupilla hebes* was determined to be the correct name for this clade because all material near the Type Locality at White Pine, Nevada, USA, possessed numerous sharp, small wavy ribs. Microsculpture also distinguished *P. hebes* from *P. alaskansis* and *P. hokkaidoensis* which both possess stronger, more widely-spaced, divaricating ribs.

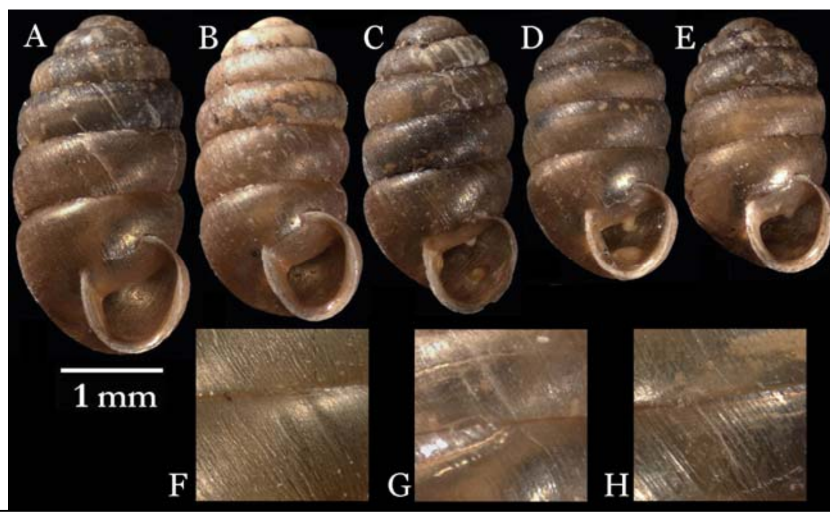

Figure 5: Profound shell plasticity within genetically-confirmed *Pupilla hebes* from the western USA, including shell size, number of apertural lamellae (0-3), and calcification of aperture margin (none to great). A. Lake Tahoe, California (identified as *P. hebes* using traditional characters); B. Utah mountains (identified as *P. muscorum* using traditional characters); C. Nevada mountains (identified as *P. blandi* using traditional characters); D. Utah Juniper savanna (identified as *P. blandi charlestonensis* using traditional characters); E. Arizona Juniper savanna (identified as *P. hebes kaibabensis* using traditional characters). Diagnostic shell microsculpture for *P. hebes* (F; sharp, narrow, dense, scalloped ribs; matte luster), *P. blandi* (G, almost smooth, shiny luster), and *P. muscorum* (H, rounded, wide, remote, straight ribs, shiny luster).

From Taxonomic to Biodiversity Error

"To form a new Government, requires infinite care, and unbounded attention; for if the foundation is badly laid the superstructure must be bad."
– George Washington, 31 May 1776, in a letter from Philadelphia to his brother John Augustine

There are three principle ways in which untested taxonomic groups may not reflect biologically reality: (1) incorrect identification features were used; (2) they were not valid and needed to be subsumed into previously-named entities; (3) they were valid but hidden within previously-known entities.

To determine how often this happens, we conducted integrative revisions on 124 Holarctic land snail taxa across three genera (*Euconulus*, *Pupilla*, and *Vertigo*; Nekola & Horsák, *in review*). Even though these groups possessed a stable and presumably well-known taxonomy, after revision over half of the concepts were found to be incorrect, with misdiagnosis and over-splitting being equally frequent. And, roughly half of the misdisgnosed taxa also contained valid entities which had been incorrectly lumped together. Error rates did not statistically vary across the Holarctic, showing that taxonomists from Europe, Asia, and North America were making the same mistakes at similar frequencies (Table 1).

Table 1: Type of changes generated through empirically rigorous taxonomic revision

	Total	Europe	Beringia	North America
Oversplit	34	4	14	22
Overlumped	15	1	6	10
Changed Features	33	9	23	22
Total Correct	57	10	23	36
Total in Error	82	14	44	54

To determine how this impacted accurate documentation of ecological and biodiversity pattern, we next compared taxonomic lists from both before and after integrative revision for over 2500 sites, 42 regions, and 9 biogeographic provinces across the entire Holarctic north of 40° N (Nekola & Horsak, *in review*). The taxa lists from almost 40% of sites were in error, with this ratio increasing to over 90% of regions and 100% in large biogeographic provinces. While less than 1% of Eurasian sites had incorrect numbers of reported taxa, 10% of North American sites had incorrect values, with diversity being over-reported 95% of the time. Reported diversity was wrong in 75% of regions and 100% of biogeographic provinces, with overestimates being by far the most prevalent. 10-15% of site pairs also had altered similarity, increasing to 80% of regions and 100% of biogeographic provinces. Distance between pairs strongly impacted this error, with similarity values based on unvetted taxa being in general too low at short distances and too high at longer ones (Figure 6).

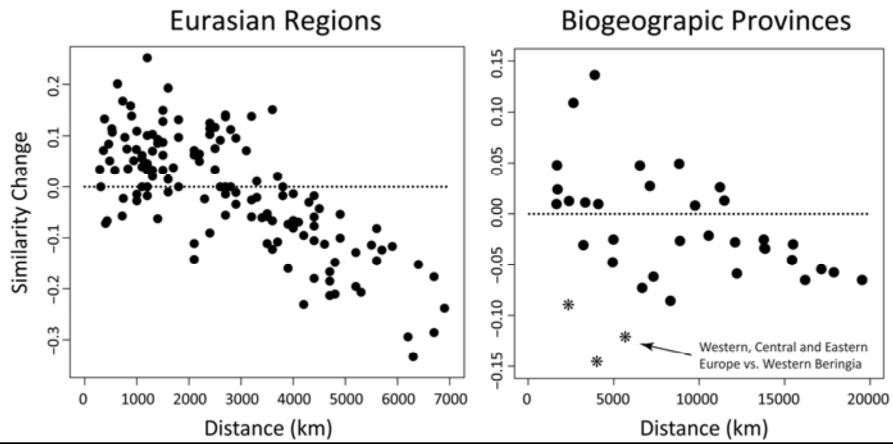

Figure 6: Changes in observed Jaccard similarity vs. intersample distance between 17 Eurasian regions and 9 Holarctic biogeographical provinces based on differences between values calculated from traditional vs. empirically-vetted taxa lists from each. Positive values indicate cases in which traditional taxonomic categories underreported actual similarity; negative values indicate when traditional taxonomy over-reported actual similarity. Least-squares linear regression: **Eurasia** – $p<<0.0000001$, $r^2 = 0.54$; **Biogeographic Provinces** (following deletion of the three Europe vs. western Beringia outliers) – $p<0.0009$; $r^2 = 0.30$

Thus, untested taxonomy frequently leads to considerable error in the data used to measure biodiversity and test ecological hypotheses. Besides making for inaccurate taxa lists, these errors also commonly overestimate site/region diversity while underestimating the actual distinctness of biotas. Psychological bias is certainly responsible for at least some of these errors: it seems likely that taxonomists are more willing to split forms at local scales because they believe they know the fauna and environment well and trust that differences they perceive relate to underlying biology. However, across larger distances they are likely less confident in their knowledge base and more likely to assume that unobserved intermediate forms exist, even when they don't.

Why Epistemology Matters

"How does it happen that a properly endowed natural scientist comes to concern himself with epistemology? .... When I think about the ablest students whom I have encountered in my teaching, that is, those who distinguish themselves by their independence of judgment and not merely their quick-wittedness, I can affirm that they had a vigorous interest in epistemology. They happily began discussions about the goals and methods of science, and they showed unequivocally, through their tenacity in defending their views, that the subject seemed important to them." - Albert Einstein, 1916, *Physikalische Zeitschrift* 17:101-102.

As seen here, epistemology matters greatly in the real world: when fully half of taxonomic categories are not supported by empirical data, the fields using them can have their patterns driven by human psychological biases as much as by actual biology. Obviously this will negatively impact our ability to accurately document and protect the diversity of life on Earth. Such issues can be at least partially resolved by epistemologically refusing to use concepts that have not been independently vetted with empirical data. Until this happens, we can't know if we are measuring and protecting real entities or the psychological projections of taxonomists.

As scientists we must remember that various levels of "proof" will change the "objective" data we analyze, and perhaps the patterns generated by that "data" as well. Because those of us who have achieved a PhD have actually been awarded a doctorate in *philosophy*, we should not be afraid to consider such philosophical limitations of our work, and be able to make recommendations to ourselves, students, and colleagues that make our search for truth more robust.

References:

Branshesi, M. 2016. Multi-messenger astronomy: gravitational waves, neutrinos, photons, and cosmic rays. *Journal of Physics: Conference Series.* 718: 022004.

Festinger, L. 1957. *A Theory of Cognitive Dissonance*. Evanston, IL: Row, Peterson.

Frest, T.J. 1991. *Summary Status Reports on Eight Species of Candidate Land Nnails from the Driftless Area (Paleozoic Plateau), Upper Midwest*. Final Report, Contract #30181-01366, U.S. Fish & Wildlife Service Region 3, Ft. Snelling, Minnesota, 54 pp

Nekola, J.C. 1999. Paleorefugia and neorefugia: the influence of colonization history on community pattern and process. *Ecology*. 80:2459-2473.

Nekola, J.C., B.F. Coles & U. Bergthorsson. 2009. Evolutionary pattern and process in the *Vertigo gouldii* (Mollusca: Pulmonata, Pupillidae) group of minute North American land snails. *Molecular Phylogenetics and Evolution*. 53:1010-1024.

Nekola, J.C., B.F. Coles & M. Horsák. 2015. Species assignment in *Pupilla* (Gastropoda: Pulmonata: Pupillidae): integration of DNA-sequence data and conchology. *Journal of Molluscan Studies*. 81:196-216.

Pilsbry, H.A. 1948. *Land Mollusca of North America (North of Mexico)*. Monograph #3. — Philadelphia: Academy of Natural Sciences of Philadelphia.

Sharps, M.J. 2018. Percival Lowell and the canals of Mars. *Skeptical Inquirer*. 42:41-46.

Box I: A Roadmap for Conducting Integrative Taxonomic Revisions

Specimens:

(1) Select individuals representing the full morphological, geographic and ecological range for each recognized species or subspecies within a group.

(2) Include material originating at or near the type locality for each species, which will allow determination of the valid name for a given clade.

DNA Sequence Analysis:

(1) Obtain mitochondrial DNA (which is only transferred from mothers and experiences no recombination), and nuclear DNA (which is transferred by both parents and does experience recombination) sequence from each specimen.

(2) Separately conduct phylogenetic reconstructions on mitochondrial and nuclear DNA data.

(3) Identify likely species-level units by noting highly supported clades containing the same individuals in both datasets.

Additional Data:

(1) Sort material into these genetically-identified groups

(2) Look separately for additional traits that distinguish groups in physical / ecological / behavioral / biogeographic data. When possible, use physical measurements in combination with statistics to provide p-values for the significance of differences between groups.

(3) A species can be considered empirically validated when the two DNA signals are shown to correspond to at least one of these other macroscopic signals.

(4) Use Type Locality DNA to determine the available name for each supported clade. If none exists a new species will need to be described.

(5) Compile results across all signals to provide a list of the diagnostic traits that allow for accurate identification of each supported taxon.