

ECOGRAPHY

Research

The impact of empirically unverified taxonomic concepts on ecological assemblage patterns across multiple spatial scales

Jeffrey C. Nekola and Michal Horsák

J. C. Nekola (<https://orcid.org/0000-0001-6073-0222>) ✉ (nekola@sci.muni.cz) and M. Horsák, Dept of Botany and Zoology, Faculty of Science, Masaryk Univ., Brno, Czech Republic.

Ecography

2022: e06063

doi: 10.1111/ecog.06063

Subject Editor: Miguel Nakamura

Editor-in-Chief:

Jens-Christian C Svenning

Accepted 29 January 2022



Accurate ecological assemblage analysis requires that underlying taxonomic divisions reflect biological reality. However, the validity of many taxonomic hypotheses have never been rigorously confronted with replicable data. As a result, these categories might say more about human psychology than biology. We consider here the ways that statistically unchallenged taxonomic concepts can bias observed ecological pattern across multiple spatial scales. We compared assemblage lists for a group of Holarctic land snail species/subspecies hypotheses both before and after empirical confrontation. Of the 124 taxa initially believed to exist, only 105 were validated, with 34 being oversplit, another 15 being incorrectly lumped within 11 taxa and a final 33 having incorrect identification features. This led to considerable scale-dependent alteration in assemblage patterns across 2528 sites within 42 regions and 9 biogeographic zones: depending upon location up to 90% of sites had revised composition with 10% showing a reduction in richness. Additionally, 6–60% of site pairs had altered turnover. These impacts became larger and more frequent with increasing sample grain and extent: 80–90% of regions and 100% of global biogeographic zones demonstrated altered composition, with 70–80% of regions and 100% of zones having changed richness. Reductions in richness also became more prevalent with increasing scale. Likewise, 80–94% of regional and zone pairs were found to have altered turnover, with increased similarity predominating at small and decreased similarity at large extents. Because these issues appear general and are not limited to land snail taxonomy, it is crucial for macroecologists to consider how unchallenged taxonomic concepts might bias their results. The use of objective and replicable data-driven validation of all recognized taxonomic concepts is thus an essential foundation for accurate ecological hypothesis testing.

Keywords: bias, ecological error, lumping, richness, splitting, taxonomic revision, turnover



www.ecography.org

© 2022 The Authors. Ecography published by John Wiley & Sons Ltd on behalf of Nordic Society Oikos

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Introduction

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; <<https://web.archive.org/web/20101002130705/https://royalsociety.org/about-us/history/>>)

It goes without saying that ecological analyses can only be as accurate as the data upon which they are based. For instance, problems with museum accession information – based on labeling errors, specimen misidentification and the differential goals of ecologists versus collectors – can lead to inaccurate documentation of composition, abundance and body size patterns (Nekola et al. 2019). And, reliance on 'expert opinion', encumbered with its many inherent biases, leads to significant error in conservation prioritization (Perez et al. 2020). But an even more profound issue may exist: what if taxonomic classifications themselves are incorrect? What if a given species actually represents multiple biologically distinct entities? Or if supposedly distinct species are actually the same entity? Or if they are being identified using the wrong traits?

While potential taxonomic acts should be seen as hypotheses requiring rigorous data-driven confrontation before being accepted (Hey et al. 2003, Dayrat 2005, Agnarsson and Kunter 2007, deCarvalho et al. 2008), outside of a few charismatic groups (Burgin et al. 2018, Upham et al. 2019) many taxonomists still rest their actions on personal expertise (Evenhuis 2007, Páll-Gergely 2017). Although physical features almost always underlie their conclusions, because statistical confrontation has never taken place the reader must accept on faith the word of the author that their statements are true. This is likely one of the root causes for the seemingly endless and irresolvable taxonomic debates regarding splitting and lumping (Vaidya et al. 2018). And without the mediation of data, any disagreements are often perceived as personal attacks. Fontaine et al. (2012) additionally point out that an increasing number of species are being described by amateur taxonomists and 'citizen scientists' who lack the analytical training and access (both physical and financial) required to accomplish high-standard systematics.

At least for some groups – like mollusks – less than 5% of new species descriptions even attempt to document morphological variability (Rosenberg 2008) let alone to confront the null hypothesis that the new form falls within some previously named entity. In combination with the fact that replicable hypothesis testing using morphometrics or genetics has only been feasible for the last 75 years, the reality is that the great majority of nomens have never been statistically vetted. While there has been substantial consideration on the impact of taxonomic effort (Clark and May 2002), misidentification (Sheppard 1998, Nekola et al. 2019) and differing species concepts (Genner et al. 2003, Agapow et al. 2004, Dillon and Fjeldså 2005, Jones et al. 2012, Vinarski

and Kramarenko 2015) in biasing community, macroecology and biodiversity analyses, as far as we can determine none have considered this more fundamental issue. We accomplish this here by comparing composition, richness and turnover estimates between opinion-based versus statistically-vetted taxonomy for a Holarctic-wide suite of land snails from site-to-biogeographic zone grains and local-to-global extents.

Methods

Study group and species concepts

Terrestrial gastropods represent an ideal system to address these issues: with over 35 000 species believed to exist globally (Barker 2001) they represent the second most diverse terrestrial invertebrate group (Zhang 2011). At the same time the description of new species has slowed in North America and Europe to less than 1% of the total fauna per decade (Welter-Schultes 2012, Nekola 2014). Both continental (Pilsbry 1948, Sysoev and Schileyko 2009, Welter-Schultes 2012) and regional (Baker 1939, Cheatum et al. 1974, Kerney and Cameron 1979, Horsák et al. 2013a) identification resources extend back over 70 years and arguably represent the earliest whole-fauna treatments for any non-lepidopteran invertebrate group.

However, the great majority of currently recognized taxonomic entities have never faced potential falsification using replicable data. Doing this requires an explicit enunciation of underlying decision-rules (Laamanen et al. 2003, Denise et al. 2008). Ours is to consider a species distinct once it has become a quasi-independent evolutionary unit. While close to the evolutionary (Simpson 1961) and biological (Mayr 1942) species concepts, ours is more statistically-focused with the rare appearance of fertile hybrids or genetic introgression not necessarily requiring the lumping of two entities, especially when consistent divergence is seen across a suite of traits. It also differs from the phylogenetic species concept (Cracraft 1983) by acknowledging that not all reciprocally monophyletic groups deserve species-level status (Agapow et al. 2004).

When evolutionary independence is achieved, divergence across multiple traits is highly likely because of differential selection pressures and/or random walks (Bickford et al. 2007, Schlick-Steiner et al. 2007). As a result we employ an integrative taxonomic approach (Will et al. 2005, Schlick-Steiner et al. 2010) and consider a given species hypothesis to have survived empirical challenge when its distinctness can be demonstrated across a consensus of data streams including (for terrestrial gastropods) some reasonable subset of mtDNA sequence, nDNA sequence, conchology, genitalic anatomy, behavior, ecological preference and/or biogeography. We reserve subspecies-level assignments to those taxa strongly supported in mtDNA but unsupported in the more slowly evolving nDNA amplicons which we employ. These taxa tend to also possess weakly differentiated shell characters and unique ecology/biogeography (for full methods see Nekola et al. 2015, 2018, Horsáková et al. 2020). Thus, while mtDNA

and nDNA sequence data represent two critical channels of independent information, they must be used in conjunction with other more traditional signals (Dayrat 2005).

Our analyses are restricted to three abundant and diverse Holarctic land snail genera which possessed a stable and uncontroversial taxonomy (Pilsbry 1948, Kerney and Cameron 1979, Pokryszko 1990): the *Euconulus fulvus* group, the bulk of Holarctic *Pupilla* and all global *Vertigo* with ranges extending north of 24°N. In total 10 *Euconulus*, 24 *Pupilla* and 105 *Vertigo* taxa were revised (Supporting information). More than 50% of these nomens originated before 1900, with only a little over 10% having been erected in the last half-century. We vetted each species- or subspecies-level hypothesis considered valid by current taxonomic workers (Hubricht 1985, Sysoev and Schileyko 2009, Welter-Schultes 2012). We did not include behavioral signals due to lack of information and anatomical signals due to the simplified genitalia of these genera, which are of little taxonomic use (Pilsbry 1948, Pokryszko et al. 2009). Following this statistical vetting, the fate of each taxon was noted: 1) both taxonomic status and identification criteria were valid; 2) taxa that were biologically valid but possessing incorrect identification features; 3) taxa requiring lumping with prior entities; and 4) taxa needing to be split from prior entities.

Composition data

Site scale

We considered 2847 100–1000 m² sites observed prior to 1 June 2015 (see Horsák et al. 2010 and Nekola 2010 for sampling methodology). This cutoff was used as all identifications made prior were based solely on statistically-unchallenged taxonomic concepts. We then excluded 319 sites which contained no target taxa. For the remaining 2528, we generated species lists both before and after revision for target species only. Site locations ranged across the Holarctic and were recorded in decimal degrees (Fig. 1a).

Regional scale

To generate intermediate-scale data we demarcated 42 50 000–500 000 km² regions (25 North America, 17 Eurasia) whose boundaries reflected sampling patterns and regional ecology (Fig. 1a). The centroid for each region was determined as a geographical center of each polygon. Using site scale data in combination with other published, verified reports (Pilsbry 1948, Kerney and Cameron 1979, Hubricht 1985, Nekola and Coles 2010, Welter-Schultes 2012, Horsák et al. 2013b), regional occupancy was coded for all taxa both before and after revision (Supporting information).

Global scale

Finally, the entire Holarctic was divided into 9 roughly equivalent-sized biogeographic zones (Fig. 1b). The centroid for each was calculated as a geographical center of a polygon defined based on the area supporting at least one taxon and limited by 30°N from the south. As before, we used our site-scale data in conjunction with other published sources to

determine zonal occupancy for all taxa both before and after revision (Supporting information).

Analyses

Taxonomic revision

Significance of unchanged, misdiagnosed, oversplit and overlumped taxa frequency from a uniform distribution was estimated using Fisher's exact test for the entire dataset; between Europe, Beringia and North America; and across all nine biogeographic zones. Fisher's exact test was also used to estimate significance of unchanged versus oversplit names originating from 1758–1914 versus 1915–2015.

Site scale

Taxon richness was determined both before and after revision. The number of sites with increased, decreased or unchanged scores was recorded with significance being estimated via the sign test. The compositional change within individual sites due to taxonomic revision was estimated using Jaccard dissimilarity.

Pairwise surface straight-line geographic distance was calculated between all sites within a continent using the Haversine equation. Jaccard similarity was calculated before and after revision for each pair with score change being recorded. The frequency of increased or decreased values was calculated across all pairs from a given focal site across five distance classes. This process was repeated across all sites. Boxplots were then used to illustrate the frequency of increasing or decreasing similarity versus distance class. Significance was estimated using both Kruskal–Wallis and ANOVA tests. All calculations were made in R 3.5.2; code is available on request.

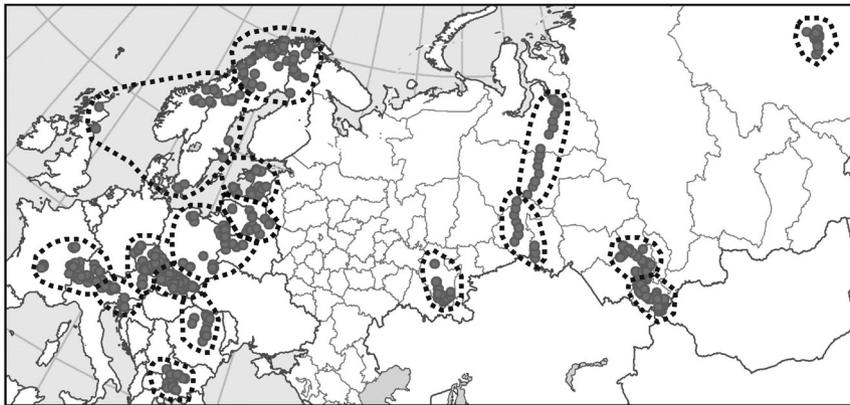
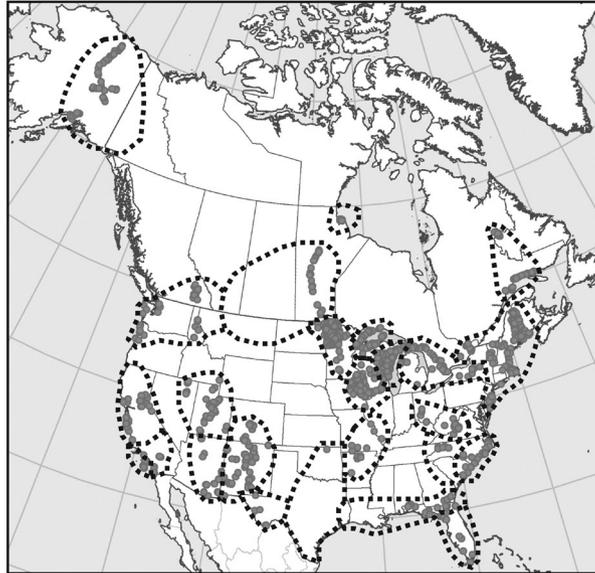
Regional scale

Richness impacts were determined by comparing the total taxa present in each region before and after revision. The number with increased, decreased and unchanged values were recorded, with significance of differences between increased and decreased richness faunas being estimated using the sign test. To illustrate how much change in individual regional lists was generated, Jaccard dissimilarity was calculated for regional species lists before versus after empirical revision. Jaccard similarity was used to calculate turnover between each pair of regions within a continent both before and after revision. Change in similarity was then calculated for each pair and compared to surface geographic distance as calculated between centroids using the Haversine equation. Because this relationship appeared linear, least-squares linear regression was used to determine significance and to estimate explained variance. To ensure that apparent heteroscedasticity did not bias outcomes, analysis was repeated using generalized least squares (GLS) regression which relaxes this assumption (Kantar 2015).

Global scale

Richness impacts were determined for each biogeographic zone by comparing total taxa present before and

(a)



(b)

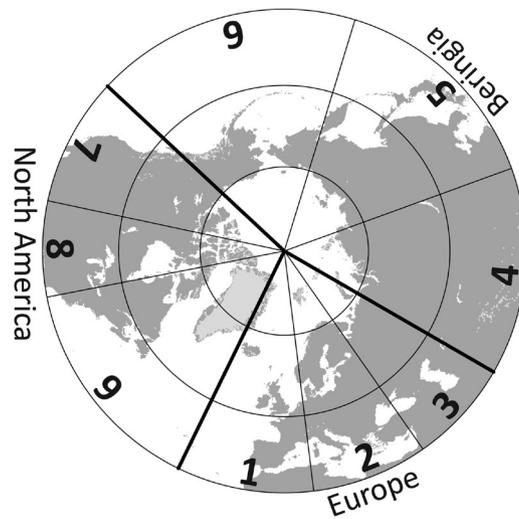


Figure 1. (a) Location of study sites observed before 31 May 2015 which supported analyzed *Euconulus*, *Pupilla* and *Vertigo* taxa. Boundaries for the 25 North American and 17 Eurasian regions are shown. (b) Locations of the nine biogeographic zones.

after revision. The number with increased, decreased and unchanged richness was recorded, with significance of differences between increases and decreases being estimated using the sign test. In addition, the total number of zones occupied by each taxon before and after revision was determined, with significance of differences being estimated using Fisher's exact test. Compositional change within a given zone was documented using Jaccard dissimilarity. Jaccard similarity was also calculated before and after revision for each zone pair, with score change being recoded and plotted against surface geographic distance based on straight-line routes between centroids using the Haversine equation. Routes were not allowed to cross 70°N (e.g. no transpolar routes) or the Atlantic Ocean. Because AICc demonstrated that the similarity change versus distance relationship was better explained by linear (−110) versus exponential/power law (−99/−98) forms, least-squares linear regression was used to estimate significance and explained variance. GLS regression was also performed to ensure that heteroscedasticity did not influence observed results.

Results

Taxonomic revision outcomes

Originally 112 species and 12 subspecies were reported (Table 1). Following consensus-based, replicable, data-driven confrontation, 57 (46%) were confirmed both in terms of their taxonomic status and identification features; 33 (27%) were biologically valid but had incorrect diagnostic features; and 34 (27%) were invalidly split into distinct entities (Supporting information). All three classes were represented across the entire 1758–2015 description year timeline with the ratio of valid to oversplit names changing from 4:1 (62:15) for names originating before 1915 to 3:2 (28:19) after. Fisher's exact test showed this difference to be marginally significant ($p=0.05$). Additionally, 1/3 of misdiagnosed taxa represented multiple distinct entities, with 15 new taxa being split from them. Thus, following revision 97 species and 8 subspecies were recognized. There was no significant variation in the four revision outcomes between European, Beringian and North American taxa ($p=0.7452$) or across the nine biogeographic zones ($p=0.41$).

Site scale

The distribution of site richness scores remained unchanged, ranging from 1 to 11 in North America and 1 to 8 in Eurasia with a median of 3 and approximately 75% of sites harboring 2 or more taxa (Supporting information). In Eurasia, 99.6% of sites had unchanged richness, with the number of increasing and decreasing sites being identical (Table 2). However, in North America 10% of sites had revised richness levels, with 96% of these representing reductions. This difference was highly significant ($p \ll 0.0000001$). Almost a half of North American and a quarter of Eurasian sites had altered compositions following revision. Considerable variation existed across Eurasia, however, with 3% of European and almost 90% of Beringian site lists changing. These differences were highly significant ($p \ll 0.0000001$). Median Jaccard compositional change within altered sites was 0.57 in North America and 0.4 in Eurasia (Fig. 2).

In North America, approximately 10% of site pairs at < 1000 km extents had altered similarity, with increasing scores being approximately four times more likely (Fig. 3). At distances exceeding 4000 km over 50% of pairs changed similarity, with increasing values being more than 20 times as frequent. This scale dependence was highly significant ($p \ll 0.0000001$), with distance classes explaining 17% of observed variation in increased-similarity and 4% in reduced-similarity pairs. Likewise in Eurasia the proportion of changed pairwise similarity increased from roughly 6% at < 1000 km to > 30% at distances > 4000 km. This scale-dependence was highly significant ($p < 0.00001$), with distance classes explaining 3% of variation in increased-similarity and 20% in reduced-similarity pairs. Unlike North America, reduced similarity was more frequent across most distance classes, becoming an order-of-magnitude more common at extents > 4000 km. The Kruskal–Wallis test demonstrated that these continental differences were highly significant ($p \ll 0.0000001$).

Regional scale

Regional taxa richness in North America ranged from 4–22 (before; median = 11) to 5–19 (after; median = 12), and in Eurasia from 5–20 to 6–17 (median = 14 both before and after; Supporting information). In North America 70% of regions demonstrated altered richness, while in Eurasia it was 80%, with scores changing by 1–6 per zone in North

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

	Total	Europe	Beringia	North America
Correct	57	10	23	36
Oversplit	34	4	14	22
Overlumped	15	1	6	10
Changed Features	33	9	23	22
Fisher's exact test				
Total vs uniform			$p=0.003732$ (all taxa)	$p=0.0916$ (only changed taxa)
Europe vs Beringia vs North America:			$p=0.7452$	

Table 2. Impact of taxonomic revision on site richness and composition (U = unchanged; C = changed; – = decreased; + = increased).

Continent	Number of sites				
	Richness			Composition	
	U	–	+	U	C
North America	1322	139	6	782	685
Eurasia	1057	2	2	776	285
Europe				745	20
Beringia				31	265

Sign test between number of reduced vs increased richness sites: North America $p < 2.2 \times 10^{-16}$
Eurasia $p = 1.0$
Fisher's exact test for altered composition frequency: Eurasia vs North America: $p < 2.2 \times 10^{-16}$
Europe vs Beringia: $p < 2.2 \times 10^{-16}$

America and 1–3 in Eurasia (Table 3). In both areas richness reductions were 80–90% more likely. While richness reduction was highly favored versus increase in North America ($p = 0.0024$), this only approached significance in Eurasia ($p = 0.0574$). Over 90% of regions in both continents experienced altered composition, with median change being 0.38 in North America and 0.22 in Eurasia (Fig. 4).

In both continents most regional pairs demonstrated altered similarity (80% in North America, 94% in Eurasia), with 75% of North America pairs demonstrating an increase in scores ($p \ll 0.0000001$). In both continents similarity decrease was more common at large extents (Fig. 5): 80% of regional pairs at < 1000 km extents in both continents had similarity increase while at > 4000 km extents this ratio fell to 60% in North America and 0% in Eurasia. Linear least-squares regression showed this relationship to be highly significant in both North America ($p = 0.0004$) and Eurasia ($p \ll 0.0000001$) explaining 4% and 54% of observed variation, respectively. GLS regression replicated these results ($p < 0.001$ in both continents) indicating that heteroscedasticity was not responsible for observed results.

Global scale

The number of biogeographic zones occupied by taxa (Table 4A) demonstrated a clear power law – lognormal distribution being modal for single zone endemics (50% before and 43% after revision) and with few taxa being found across five or more zones (4.1% before; 3.8% after). These distributions were statistically similar before and after revision ($p = 0.7386$). No differences were seen in zone occupancy between the four revision categories (Fisher's exact test, $p = 0.94$ – 0.13 ; tables not shown). Taxon richness ranged from 16–53 per zone before to 14–43 following revision, with median values being 22 and 19. Richness fell in all zones ($p = 0.0039$), with the number of reductions ranging from 1 to 10 (Table 4B). All zones had altered composition, with dissimilarity ranging from 0.0625 (Eastern Europe) to 0.6333 (western Beringia). Median change was 0.4 (eastern North America).

All zone pairs demonstrated changed similarity. Similarity increases dominated at smaller and decreases at larger extents, especially when three outliers – representing comparisons of western, central and eastern Europe to western Beringia – were removed (Fig. 6). These had abnormally large similarity reductions over short distances because of four Beringian

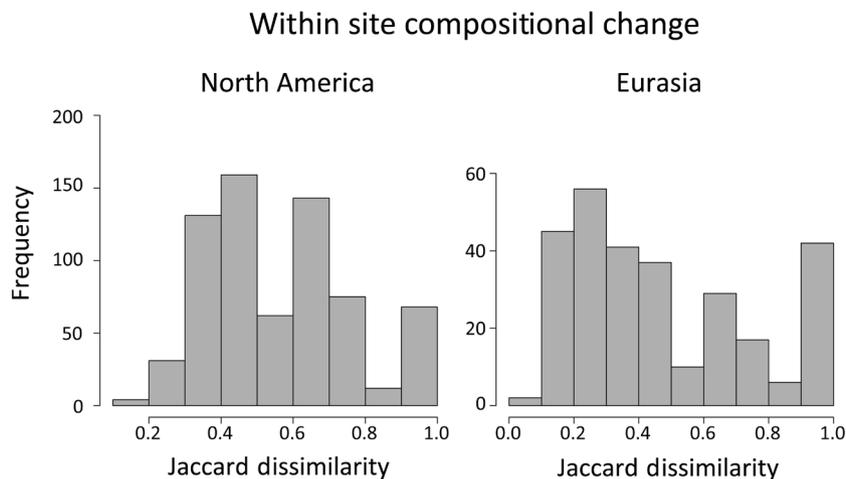


Figure 2. Site-scale compositional change following revision.

Site pairwise comparisons

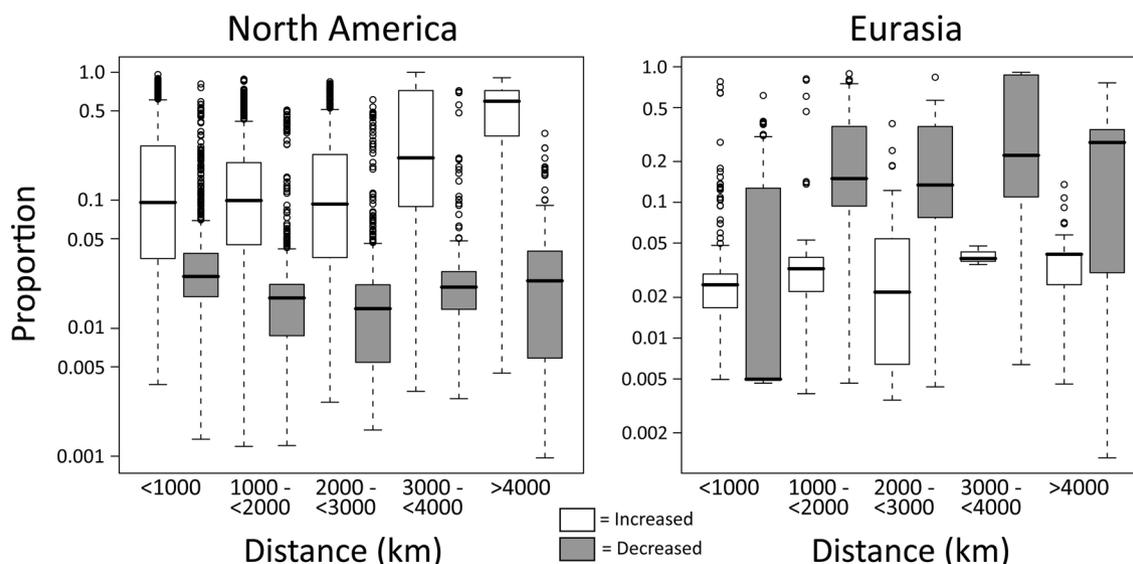


Figure 3. Proportion of site-scale pairs which changed similarity following revision. Significance of frequency change with increasing extent: North America – Increased; Kruskal–Wallis $p=0$, ANOVA $p=0$, $r^2=0.1683$. – Decreased; Kruskal–Wallis $p=0$, ANOVA $p=0$, $r^2=0.043$. Eurasia – Increased; Kruskal–Wallis $p=0.00000993$, ANOVA $p=0.002298$, $r^2=0.02878$. Decreased; Kruskal–Wallis $p=0$, ANOVA $p=0$, $r^2=0.2008$.

taxa (*Euconulus fulvus egenus*, *Pupilla turcmenia*, *Vertigo beringiana*, *V. genesioides*) that had initially been incorrectly lumped with European taxa (*E. fulvus fulvus*, *P. sterrii*, *V. alpestris*, *V. genesii*). Following outlier removal, linear least-squares regression showed the relationship to be highly significant ($p < 0.0009$) explaining 30% of observed variation. Again, observed heteroscedasticity was not responsible for these results with GLS regression on both the full and outlier-removed datasets showing high statistical significance ($p < 0.001$)

Discussion

Our data-driven confrontation of prior taxonomic concepts demonstrated that more than half were in error, with roughly 25% having incorrect identification characters, another 25% being oversplit and 10% being improperly lumped into previously described entities. This led to considerable error in composition lists, richness and turnover with impacts becoming greater at larger observational grains and extents. While

approximately 90% of Beringian, 50% of North American and 3% of European site composition lists were altered, this occurred in over 90% of regions on both continents, and within all biogeographic zones. Site richness was more stable, with only 10% of North American, 1.4% of Beringian and no European sites exhibiting altered scores. In North America, when scores varied there was an almost 25-times greater chance of site richness reduction. Between 70 and 80% of regions had altered richness, with reductions being significantly more likely. All biogeographic zones experienced a reduction in richness, ranging from 6 to 19% of the fauna. Considerable scale-dependent change was also noted in turnover: in both continents 2–10% of pairwise site similarities changed at shorter extents, increasing to 50% or more at longer extents. In North America increased pairwise similarity predominated while in Eurasia decreased similarity prevailed. This difference is largely related to the greater tendency for small-scale oversplitting in North America and large-scale overlumping in Eurasia. At the regional scale, 80–95% of pairs demonstrated altered similarity, with increasing scores being significantly ($> 80\%$) more common at < 1000 km

Table 3. Impact of taxonomic revision on regional richness, composition and pairwise similarity (U=unchanged; C=changed; –=decreased; +=increased).

Continent	Richness			Composition		Similarity		
	U	–	+	U	C	U	+	–
North America	8	15	2	2	23	58	176	66
Eurasia	3	11	3	1	16	8	65	63

Sign test: Decreased vs increased richness; North America $p=0.00235$; Eurasia $p=0.05737$; unchanged vs changed composition; North America $p=0.000019$; Eurasia $p=0.00027$; increased vs decreased similarity; North America $p=3.3 \times 10^{-12}$; Eurasia $p=0.9296$.

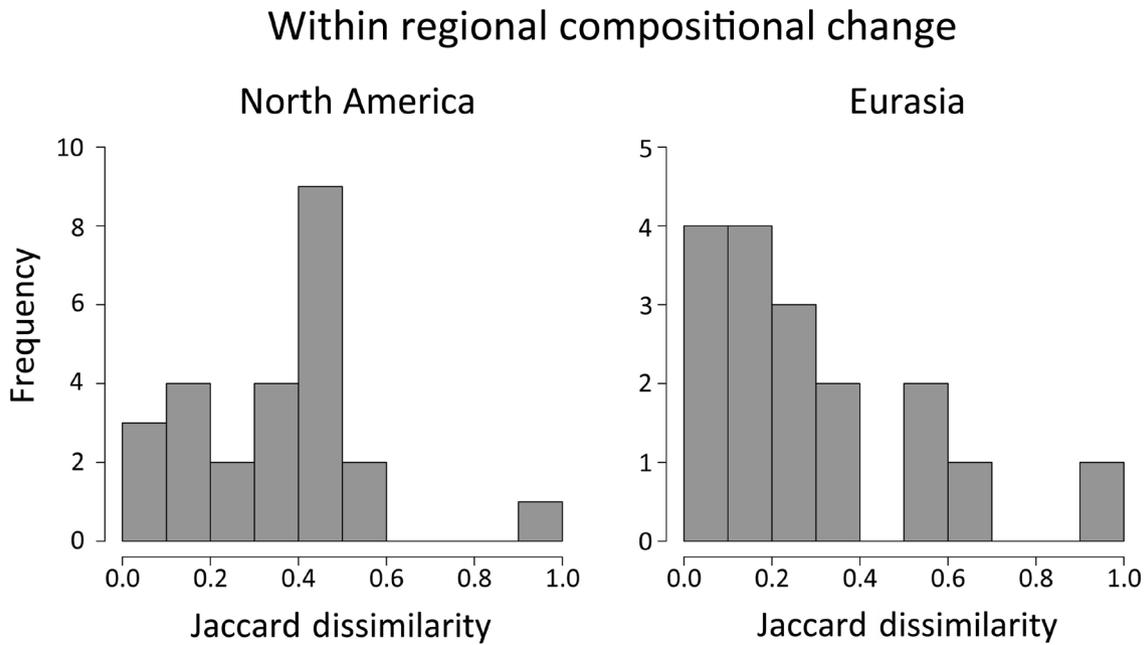


Figure 4. Regional-scale compositional change following revision.

extents. As extent increased, however, this value significantly fell. At the biogeographic zone scale, all pairs demonstrated changed similarity, with values tending to increase at shorter and decrease at larger extents. Unverified taxonomic concepts thus lead to underestimation of turnover rates.

In some ways these results are not surprising. A number of previous works have also demonstrated that altering species concepts will change reported richness (Genner et al.

2003, Agapow et al. 2004, Evangelista et al. 2013). However, unlike those who suggested that the relative shape and placement of richness gradients does not change (Isaac et al. 2004, Dillon and Fjeldså 2005, Jones et al. 2012, Vinarski and Kramarenko 2015), we document that regions prone to unwarranted taxonomic exuberance (such as the upper Midwestern USA) can have significant site-scale over-reporting bias.

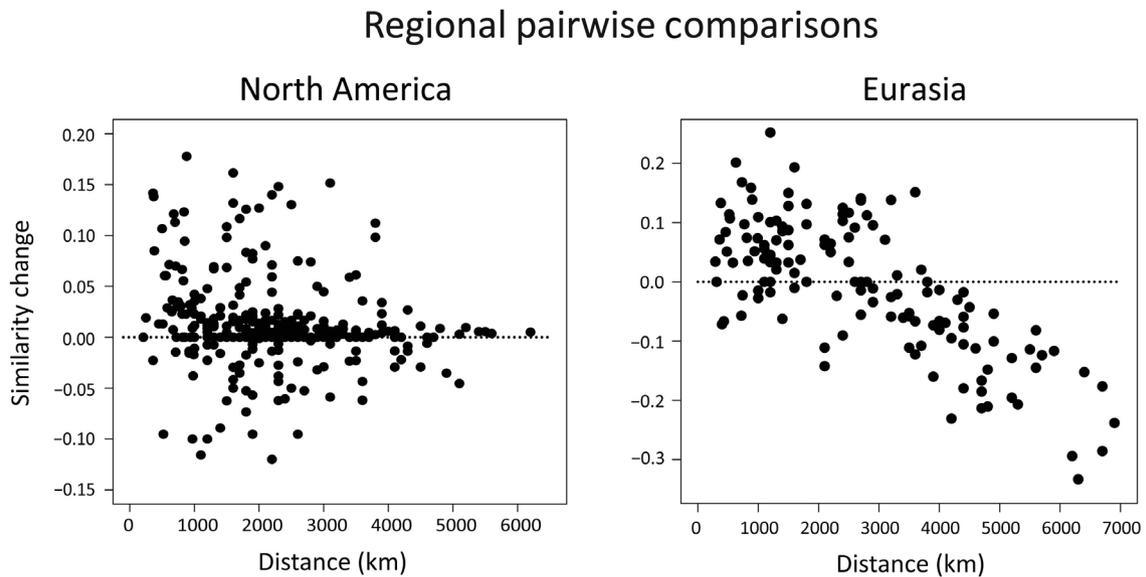


Figure 5. Regional-scale pairwise Jaccard similarity change versus distance. Least-squares linear regression: North America: $p=0.0004$, $r^2=0.04$; Eurasia: $p << 0.0000001$, $r^2=0.54$.

Table 4. Impact of empirically rigorous taxonomic revision on (A) number of occupied zones per taxa and (B) zone richness.

(A) Revision	Number of zones								
	1	2	3	4	5	6	7	8	9
Before	62	30	14	13	1	0	2	1	1
After	45	33	12	11	1	2	1	0	0
Fisher's exact test: $p=0.7386$									

(B) Richness	Zone								
	1	2	3	4	5	6	7	8	9
Before	16	22	16	26	22	20	53	42	33
After	14	19	15	23	19	17	43	34	31
Sign test for universal reduction in richness across all 9 zones: $p=0.003906$.									

Sources of error and bias in opinion-based taxonomy

Identified errors in traditional species assignments appear to be generally related to three non-mutually-exclusive human perception issues: 1) focus on easily observed but taxonomically unreliable plastic features; 2) disregard of less readily observable but robust diagnostic traits; and 3) poor understanding of variability within taxa. An excellent example is found in the genus *Pupilla*: of the 20 taxa initially believed to exist, only 5 were both biologically valid and correctly diagnosed. The remaining $\frac{3}{4}$ had been improperly split into distinct entities (7 taxa) or were biologically valid but inaccurately diagnosed (8 taxa). Additionally, traditional concepts for three of these latter taxa included four distinct additional species requiring description. Errors in this genus alone led to 20% of all North American species lists being incorrect. Almost all diagnostic feature errors were due to unwarranted

focus on easily-observed but highly variable shell aperture calcification traits, with the shell microsculpture traits that actually distinguished valid taxa tending to be subtler (Nekola et al. 2015). In other genera soft-body pigmentation features have also proved useful (Horsáková et al. 2020).

These perceptual errors can be exacerbated by the social/political environment in which taxonomists work. For instance, the most avidly oversplit species was the North American *Vertigo arthuri* which had been divided into eight taxa based on differing aperture calcification levels. Five of these were thought limited to charismatic and endangered algific talus slope habitats in the upper Mississippi River valley which were in need of protection (Frest 1991). Almost 40% of within-site oversplitting in North America is traceable to this one group alone.

While oversplitting is not the only outcome of non-statistical taxonomy – as a whole misdiagnosis and overlumping are jointly more frequent – it is of particular interest given that new taxa description has historically been used to indicate expertise. There thus exists a natural tendency for some workers to generate spurious new taxa, especially in charismatic groups. For instance, the large and colorful *Liguus fasciatus* Müller, 1774 tree snails of Cuba and south Florida have over 120 color morphs which have been afforded official subspecies status (Pilsbry 1948) even though essentially no genetic variability exists between them (Hillis et al. 1991). And, orchids have a 5-fold higher rate of species description in the European flora than other less attractive but equally diverse plant groups (Pillon and Chase 2006). New species descriptions are also more likely to occur in large versus small body size mammals (Zachos et al. 2013). And as we show here the desire to describe becomes increasingly problematic once the majority of valid entities have already been named, leading to increasing oversplitting rates over time.

Workers who ‘describe new species without supportive evidence’ have been referred to as having caught the ‘*mihich*’ (Calvert 1898, Evenhuis 2008) with the term ‘*mihilism*’ being coined to summarize their activities (Bruun 1950). The propagation of unwarranted new names has been variously referred to as ‘synonymy load’ or ‘taxonomic inflation’ with estimates ranging from 20% in insects (Stork 1997) to 60% in amphibians (Dubois 1977). In this regard the ~25% load we document falls well within the lower bound. Synonymy

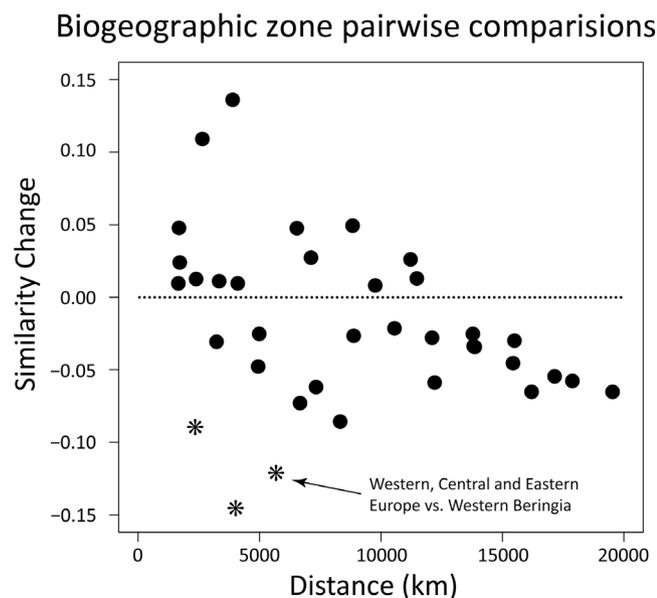


Figure 6. Biogeographic zone-scale pairwise Jaccard similarity change vs distance. Least squares fit of a linear model following deletion of the three Europe vs western Beringia outliers: $p < 0.0009$; $r^2 = 0.30$.

load appears to strongly correlate to the number of taxonomists who have worked on a group (Dubois 2008).

Also of concern is the fact that taxonomic activity follows a power law/Pareto distribution – with the majority of names being traceable to a few individuals (Nekola 2014, Isaac and Pocock 2015). For instance 109 clausiliid land snails in Europe (almost 30% of the group total) have been erected since 1960 by a single worker (as detailed in Welter-Schultes 2012). This should not be construed as a personal condemnation of this individual; for all we know most of these taxa could be valid. Rather, the issue is that any psychological eccentricities, errors or preferences related to those few people will have an outsized impact on overall pattern (Whitaker and Kimmig 2020).

We would be remiss to not also mention errors related to poor scholarship, which often happens with inadequately trained amateurs, citizen scientists or professionals operating outside of their areas of expertise (see examples provided in Bolton et al. 2008, Schleip and O’Shea 2010, Horsák and Cameron 2015). For instance, both oversplit European *Euconulus* taxa should never have been afforded species rank given that type material was never accessed. Had they been, it would have been immediately clear that these nomens were simply junior synonyms of previously described forms (Horsáková et al. 2020). Such errors will generate systematic bias because taxonomist activities tend to be spatially-auto-correlated and largely limited to their home regions.

Scale-dependence in taxonomy error

Previous considerations of differing species concepts on ecological pattern have been limited to single observational grains, usually site scales. This is one reason why some prior studies (Isaac et al. 2004, Dillon and Fjeldså 2005, Jones et al. 2012, Vinarski and Kramarenko 2015) have concluded that errors due to changing species criteria are of little concern. Our work not only demonstrates that ecological impacts occur at the site scale, but that they also exhibit considerable idiosyncratic variability. The lack of changed richness in European sites, for instance, is simply related to the happenstance that three regional oversplit taxa (*Euconulus callopisticus*, *E. trochiformis*, *Pupilla bigranata*) were not encountered in our field sampling, while the remaining taxon (*P. pratensis*) has never been reported to coexist at site scales with its senior synonym (*P. alpicola*). Idiosyncratic differences in history and taxonomist temperament also led to higher site-scale oversplitting in North America at small and overlumping in Eurasia at large scales, leading to site turnover overestimates in North America and underestimates in Eurasia.

Our crucial finding is that the amount of composition, richness and turnover error is positively related to observational scale: over 90% of regions on both continents, and all biogeographic zones, had changed taxon lists. As a result, systemic biases are most likely to occur in macroecological analyses, where 70–80% of regions had altered richness with reductions being significantly more likely. All biogeographic zones also experienced a reduction in richness,

ranging from 6 to 19% of the fauna. These results are likely due to the positive correlation between observational scale and the number of taxonomic workers. Thus, at larger scales there will be an increased likelihood that someone will have had poor skills, questionable motives, and/or a case of the mihi itch.

Observed changes in turnover above the site level – with initial pairwise similarity estimates being too low over short distances and too high over longer – is also almost certainly an artifact of human psychology. It seems likely that workers 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 must relate to underlying biology. However, across larger distances workers are likely less confident in their knowledge base and more likely to assume that unobserved intermediate forms exist. Nomenclatural rules also help determine where overlumping is more frequent: because initial taxonomic descriptions were generally based in Europe or North America, Beringian identifications were reasonably first assigned to these already existing entities (Horsák et al. 2013b). As a result, fully 2/3 of the Yakutian fauna was initially (and incorrectly) lumped with European forms. This also occurred at a slightly lower rate in the Altai Mountains (Horsák et al. 2010).

Concluding remarks

We appreciate that not all may agree with our revised taxonomy. However, because our concepts are rooted in replicable data and analyses we strongly feel they will prove more stable than if they had simply been based on our personal opinions. But the main take-home message here is that taxonomic concepts employed by different workers will generate different ecological patterns. It is thus crucial for ecologists to carefully consider the taxonomic classifications being used, and when a choice exists, to preferentially select those which have been validated using rigorous empirical confrontation. Otherwise, patterns may reflect the perceptual and societal biases of the specialists who last worked on a group.

Are the issues observed here simply an indictment of land snail taxonomists? While this would limit the scope of our findings, sadly we do not think this is the case. In fact, synonymy load in our group is in lower range of reported values, and is fully three times less than for groups like amphibians. It also seems likely that taxonomic inaccuracy may be more pronounced in larger or more charismatic groups which have attracted the attention of more workers, not all of them careful and well-trained. And the lack of statistical validation of newly described taxa is not unique to land snails – for instance the species-level hypothesis for the Beringian spider recently named after the senior author has never been statistically challenged (*Porrhomma nekolai* Růžička 2018). For these reasons, neither of us in our editorial responsibilities will send out for review any manuscript that describes a new taxon that lacks statistical confrontation of the associated taxonomic null hypotheses.

The analyses presented here are contingent upon seeing species not only as testable hypotheses, but also being biologically real. If, as some contend, taxonomic categories simply represent inherently untestable philosophic constructs (Isaac et al. 2004), then there will be no way to disentangle biology from the human experience. While this is a sobering thought, we do not believe it to be true. Our experience with multi-channel data-based taxonomic confrontation is that valid species are almost always clearly separated with considerably more genetic distance existing between them in both mtDNA and nDNA than are contained within. And when ambiguous results exist in one data stream they are not replicated in others. As a result, we tend to assess taxon validity by first consulting multiple DNA sequence datastreams, not the least because they provide replicable insights regarding genetic relatedness. We then use these groupings to search for additional corresponding morphological/ecological/behavioral/biogeographic differences. Once identified, these become essential in helping determine phylogenetically informative signal in cases of DNA disagreement (e.g. incomplete sorting, mitochondrial introgression). We would be remiss to not acknowledge, however, that in some systems DNA sequence data cannot serve as the initial sorting criteria. For instance, when horizontal transfer of genetic information is common (e.g. the land snail *Cerion*; Woodruff and Gould 1987), DNA signals will not be generated by simple divergence from a common ancestor and may provide conflicting or uninterpretable signals. In such situations, other datastreams, such as ecology or shell features, may need to take precedence. Therefore, a single set of rules determining the order of processing of multiple empirical channels does not exist.

What is vital to remember, however, is that as long as taxonomic categories remain largely based on statistically unvetted statements they have the potential of saying as much about human psychology as they do biology. Our most important wish for the taxonomic community (sensu Gotelli 2004) is thus for it to fully embrace '*Nullius in verba*' and confront all taxonomic hypotheses, no matter their age, with a wide range of replicable data. By doing this at least some human bias will be filtered out. And, while it is unreasonable to expect all ecologists to take up the mantle of revision, none-the-less they should avoid treating taxonomy as a black box. Rather, they need to determine the approaches being used by the taxonomic worker whose concepts they are using and then consider what biases and errors those decision rules could generate.

Acknowledgements – We thank Jan Divíšek for generation of the maps illustrated in Figure 1 and assistance in other geographical analyses.

Funding – This project was supported by the Czech Science Foundation (no. 20-18827S).

Author contributions

Jeffrey C. Nekola: Conceptualization (lead); Data curation (lead); Formal analysis (lead); Funding acquisition (equal); Investigation (lead); Methodology (lead); Writing – original

draft (lead); Writing – review and editing (lead). **Michal Horsák:** Conceptualization (equal); Data curation (supporting); Formal analysis (supporting); Funding acquisition (equal); Investigation (supporting); Methodology (supporting).

Transparent Peer Review

The peer review history for this article is available at <<https://publons.com/publon/10.1111/ecog.06063>>.

Data availability statement

Because analyzed data include specific locality information for many protected species, and because access to private land was often granted only after agreeing to not publically broadcast findings, we are unable to make the full datasets upon which these analyses are based available to the general public. However, the corresponding author can share this data with individual researchers upon request.

Supporting information

The supporting information associated with this article is available from the online version.

References

- Agapow, P.-M. et al. 2004. The impact of species concept on biodiversity studies. – *Quart. Rev. Biol.* 79: 161–179.
- Agnarsson, I. and Kunter, N. 2007. Taxonomy in a changing world: seeking solutions for a science in crisis. – *Syst. Biol.* 56: 531–539.
- Baker, F. C. 1939. Fieldbook of Illinois land snails. – Illinois Natural History Survey Division, Illinois Dept of Registration and Education.
- Barker, G. M. 2001. Gastropods on land: phylogeny, diversity and adaptive morphology. – In: Barker, G. M. (ed.), *The biology of terrestrial molluscs*. CABI Publishing, pp. 1–139.
- Bickford, D. et al. 2007. Cryptic species as a window on diversity and conservation. – *Trends Ecol. Evol.* 22: 148–155.
- Bolton, B. et al. 2008. New synonyms in Myrmicine ants (Hymenoptera: Formicidae). – *Zootaxa* 1732: 61–64.
- Bruun, A. F. 1950. The systema naturae of the twentieth century. – *Science* 112: 342–343.
- Burgin, C. J. et al. 2018. How many species of mammals are there? – *J. Mammol.* 99: 1–14.
- Calvert, P. P. 1898. A biographical notice of George Henry Horn. – *Trans. Am. Entomol. Soc.* 25: i–xxiv.
- Cheatum, E. P. et al. 1974. The aquatic and land mollusca of Texas. – Dallas Museum of Natural History.
- Clark, J. A. and May, R. M. 2002. Taxonomic bias in conservation research. – *Science* 297: 191–192.
- Cracraft, J. 1983. Species concepts and speciation analysis. – *Curr. Ornithol.* 1: 159–187.
- Dayrat, B. 2005. Towards integrative taxonomy. – *Biol. J. Linn. Soc.* 85: 407–417.
- deCarvalho, M. R. et al. 2008. Systematics must embrace comparative biology and evolution, not speed and automation. – *Evol. Biol.* 35: 150–157.
- Denise, T. S. H. et al. 2008. The need for specifying species concepts: how many species of silvered langurs (*Trachypithecus cris-*

- tatus* group) should be recognized? – *Mol. Phylogenet. Evol.* 49: 688–689.
- Dillon, S. and Fjeldså, J. 2005. The implications of different species concepts for describing biodiversity patterns and assessing conservation needs for African birds. – *Ecography* 28: 682–692.
- Dubois, A. 1977. Les problèmes de l'espèce chez les amphibiens anoures. – *Mém. Soc. Zool. France* 39: 161–284.
- Dubois, A. 2008. A partial but radical solution to the problem of nomenclatural taxonomic inflation and synonymy load. – *Biol. J. Linn. Soc.* 93: 857–863.
- Evangelista, D. A. et al. 2013. Species richness estimates of *Blattodea* s.s. (Insecta: Dictyoptera) from northern Guyana vary depending upon methods of species delimitation. – *Syst. Entomol.* 39: 150–158.
- Evenhuis, N. L. 2007. Helping solve the 'other' taxonomic impediment: completing the eight steps to total enlightenment and taxonomic nirvana. – *Zootaxa* 1407: 3–12.
- Evenhuis, N. L. 2008. The 'Mihi itch' – a brief history. – *Zootaxa* 1890: 59–68.
- Fontaine, B. et al. 2012. New species in the Old World: Europe as a frontier in biodiversity exploration, a test bed for 21st century taxonomy. – *PLoS One* 7: e36881.
- Frest, T. J. 1991. Summary status reports on eight species of candidate land snails from the driftless area (Paleozoic Plateau), Upper Midwest. Final Report, Contract no. 30181-01366. – U.S. Fish and Wildlife Service Region 3.
- Genner, M. J. et al. 2003. How does the taxonomic status of allopatric populations influence species richness within African cichlid fish assemblages. – *J. Biogeogr.* 31: 93–102.
- Gotelli, N. J. 2004. A taxonomic wish-list for community ecology. – *Phil. Trans. R. Soc. B* 359: 585–597.
- Hey, J. et al. 2003. Understanding and confronting species uncertainty in biology and conservation. – *Trends Ecol. Evol.* 18: 597–603.
- Hillis, D. M. et al. 1991. Minimal genetic variation in a morphologically diverse species (Florida tree snail, *Liguus fasciatus*). – *J. Hered.* 82: 282–286.
- Horsák, M. and Cameron, R. A. D. 2015. COMMENT on Altaba (2015): a case of species misidentification? – *Biol. J. Linn. Soc.* 118: 1103–1106.
- Horsák, M. et al. 2010. Habitats of relict terrestrial snails in southern Siberia: lessons for the reconstruction of palaeoenvironments of full-glacial Europe. – *J. Biogeogr.* 37: 1450–1462.
- Horsák, M. et al. 2013a. Molluscs of the Czech and Slovak Republics. – Kabourek.
- Horsák, M. et al. 2013b. Exceptionally poor land snail fauna of central Yakutia (NE Russia): climatic and habitat determinants of species richness. – *Polar Biol.* 36: 185–191.
- Horsáková V. et al. 2020. Integrative taxonomic consideration of the Holarctic *Euconulus fulvus* group of land snails (Gastropoda, Stylommatophora). – *Syst. Biodivers.* 18: 142–160.
- Hubricht, L. 1985. The distributions of the native land mollusks of the eastern United States. – *Fieldiana Zool.* 24: 1–191.
- Isaac, N. J. B. and Pocock, M. J. 2015. Bias and information in biological records. – *Biol. J. Linn. Soc.* 115: 522–531.
- Isaac, N. J. B. et al. 2004. Taxonomic inflation: its influence on macroecology and conservation. – *Trends Ecol. Evol.* 19: 464–469.
- Jones, O. R. et al. 2012. Latitudinal gradients in taxonomic overdescription rate affect macroecological inferences using species list data. – *Ecography* 35: 333–340.
- Kantar, Y. M. 2015. Generalized least squares and weighted least squares estimation methods for distributional parameters. – *REVSTAT-Stat. J.* 13: 263–282.
- Kerney, M. and Cameron, R. A. D. 1979. Field guide to the land snails of Britain and northwestern Europe. – Collins Press.
- Laamanen, T. R. et al. 2003. Kelp flies and species concepts – the case of *Coelopa frigida* (Fabricius, 1805) and *C. nebularium* Aldrich, 1929 (Diptera: Coelopidae). – *J. Zool. Syst. Evol. Res.* 41: 127–136.
- Mayr, E. 1942. Systematics and the origin of species from the viewpoint of a zoologist. – Columbia Univ. Press.
- Nekola, J. C. 2010. Acidophilic terrestrial gastropod communities of North America. – *J. Mollusc. Stud.* 76:144–156.
- Nekola, J. C. 2014. Overview of the North American terrestrial gastropod fauna. – *Am. Malacol. Bull.* 32: 225–235.
- Nekola, J. C. and Coles, B. F. 2010. Pupillid land snails of eastern North America. – *Am. Malacol. Bull.* 28: 29–57.
- Nekola, J. C. et al. 2015. Species assignment in *Pupilla* (Gastropoda: Pulmonata: Pupillidae): integration of DNA-sequence data and conchology. – *J. Mollusc. Stud.* 81: 196–216.
- Nekola, J. C. et al. 2018. A phylogenetic overview of the genus *Vertigo* O. F. Müller, 1773 (Gastropoda: Pulmonata: Pupillidae: Vertigininae). – *Malacologia* 62: 21–161.
- Nekola, J. C. et al. 2019. Caveat consumptor notitia museo: let the museum data user beware. – *Global Ecol. Biogeogr.* 28: 1722–1734.
- Páll-Gergely, B. 2017. Should we describe genera without molecular phylogenies? – *Zootaxa* 4232: 593–596.
- Perez, K. E. et al. 2020. Poorly vetted conservation ranks can be more wrong than right: lessons from Texas land snails. – *Nat. Areas J.* 40: 309–317.
- Pillon, Y. and Chase, M. W. 2006. Taxonomic exaggeration and its effects on orchid conservation. – *Conserv. Biol.* 21: 263–265.
- Pilsbry, H. A. 1948. Land Mollusca of North America (North of Mexico). Monograph no. 3. – Academy of Natural Sciences of Philadelphia.
- Pokryszko, B. M. 1990. The Vertiginidae of Poland (Gastropoda: Pulmonata: Pupilloidea) – a systematic monograph. – *Ann. Zool.* 43: 133–257.
- Pokryszko, B. M. et al. 2009. Pupilloidea of Pakistan (Gastropoda: pulmonata): Truncatellinae, Vertigininae, Gastrocoptinae, Pupillinae (in part). – *Ann. Zool.* 59: 423–458.
- Rosenberg, G. 2008. What gets left out of descriptions of new species of mollusks? – Abstracts, Annual Meeting of the American Malacological Society, Carbondale, IL, USA, p. 51.
- Růžička, V. 2018. A review of the spider genus *Porrhomma* (Araneae, Linyphiidae). – *Zootaxa* 4481: 1–75.
- Schleip, W. D. and O'Shea, M. 2010. Annotated checklist of the recent and extinct pythons (Serpentes, Pythonidae), with notes on nomenclature, taxonomy, and distribution. – *ZooKeys* 66: 29–79.
- Schlick-Steiner, B. C. et al. 2007. Without morphology, cryptic species stay in taxonomic crypsis following discovery. – *Trends Ecol. Evol.* 22: 391–392.
- Schlick-Steiner, B. C. et al. 2010. Integrative taxonomy: a multi-source approach to exploring biodiversity. – *Annu. Rev. Entomol.* 55: 421–438.
- Sheppard, C. R. C. 1998. Biodiversity patterns in Indian Ocean coals and effects of taxonomic error in data. – *Biodivers. Conserv.* 7: 847–868.
- Simpson, G. G. 1961. Principles of animal taxonomy. – Columbia Univ. Press.
- Stork, N. E. 1997. Measuring global biodiversity and its decline. – In: Reaka-Kudla, M. and Wilson, E. O. (eds), *Biodiversity II*. Joseph Henry Press, pp. 41–68.
- Sysoev, A. and Schileyko, A. 2009. Land snails of Russia and adjacent countries. – Pensoft.

- Upham, N. S. et al. 2019. Inferring the mammal tree: species-level sets of phylogenies for questions in ecology, evolution and conservation. – *PLoS Biol.* 17: e3000494.
- Vaidya, G. et al. 2018. The tempo and mode of taxonomic correction process: how taxonomists have corrected and recorrected North American bird species over the last 127 years. – *PLoS One* 13: e0195736.
- Vinarski, M. V. and Kramarenko, S. S. 2015. How does the discrepancies among taxonomists affect macroecological patterns? A case study of freshwater snails of western Siberia. – *Biodivers. Conserv.* 24: 2079–2091.
- Welter-Schultes, F. W. 2012. European non-marine molluscs, a guide for species identification. – Planet Poster Editions.
- Whitaker, A. F. and Kimmig, J. 2020. Anthropologically introduced biases in natural history collections, with a case study on the invertebrate paleontology collection from the middle Cambrian. – *Palaentol. Electron.* 23: a58.
- Will, K. W. et al. 2005. The perils of DNA barcoding and the need for integrative taxonomy. – *Syst. Biol.* 54: 844–851.
- Woodruff, D. S. and Gould, S. J. 1987. Fifty years of interspecific hybridization: genetics and morphometrics of a controlled experiment on the land snail *Cerion* in the Florida Keys. – *Evolution* 41: 1022–1045.
- Zachos, F. E. et al. 2013. Species inflation and taxonomic artifacts – a critical comment on recent trends in mammalian classification. – *Mammal. Biol.* 78: 1–6.
- Zhang, Z. Q. 2011. Animal biodiversity: an outline of higher-level classification and survey of taxonomic richness. – *Zootaxa* 3148: 1–237.