TITLE: Species synonyms depict changing but taxon-independent taxonomic praxis

Vikram Iyer[1,*], Avanthika Prasad[2]

[1]Department of Biology, Trivedi School of Biosciences, Ashoka University, Ashoka University, Sonipat, Haryana, India

[2]Department of Biology, Indian Institute of Science Education and Research (IISER), Pune, Pune, Maharashtra, India

*Corresponding author: Vikram Iyer

Email address: vikram.iyer@ashoka.edu.in

Abstract

What constitutes a single species is ultimately arbitrary, resulting in idiosyncrasies in taxonomic praxis determining species counts. One consequence of this arbitrariness is taxonomic synonymization, wherein variation determined not unique enough to constitute distinct species gets merged into one. Synonymization is non-random and subject to taxon and trait-specific biases. We explore these trends using a comprehensive dataset of the world's marine species for all large phyla and Animalia. We match all accepted species-names from the present day with their synonyms and calculate metrics of synonymization including the time lag between the years of descriptions of a species and its newest synonym. We find that although species description trends across phyla show no consistent, discernible trends, the fraction of names that gets synonymised remains initially constant across phyla, with a subsequent gradual decrease. Most accepted species described around the early 20th century possess synonyms created in the same year, indicating synonyms arising due to new name combinations, suggesting a distinct period of taxonomic praxis. This period was followed by a period of most species descriptions lacking synonyms on moving towards the present. Across taxa, we find the mean lag to be strongly negatively correlated to time, with time not being a limiting factor determining this trend. We also find the number of synonyms per accepted name to also be strongly negatively correlated to time in a taxon-independent manner. Our study considers when the diversity characterising species in the form of once-unique species-names becomes known to science. Our results suggest that across taxa, species described more recently encompass less biological variability, indicating a gradual change in how morphologically delineated species which comprise the vast majority of species described across time are described.

Introduction

The delineation of species boundaries, that is, the range of variation that characterises a single species, is ultimately arbitrary (De Queiroz, 2007). A consequence of this is that idiosyncrasies in taxonomic praxis can determine species counts. For example, using phylogenetic species concepts results in higher species counts relative to morphological species delimitation (Agapow et al., 2004), and persistent taxon-specific practices of lumping together morphological variation into infraspecific taxa, or splitting them into distinct species determine species counts (Isaac et al., 2004). Taxonomic synonymization is another consequence of the arbitrariness of species delimitation. Describing a species generates the testable hypothesis that the holotype individuals formally attached to the species name represent a distinct group, which, if falsified, results in their synonymization with a pre-existing species name (Gaston and Mound, 1997). These pending synonymizations can result in inflated species-counts until rectified by taxonomic revision (Alroy, 2002).

When two formerly unique species are synonymised into one, one of the species-names becomes a synonym of the other following the Principle of Priority (Article 23, "The Code Online | International Commission on Zoological Nomenclature"). Synonymization occurs non-randomly, with species attributes like body size and range size correlating with synonym number (Guedes et al., 2025; Jones et al., 2012). However, synonym count also correlates with indices of taxonomic praxis like how intensively a taxon is studied (Appeltans et al., 2012). These suggest that synonymization displays taxon-specific trends, driven by taxon-specific differences in taxonomic praxis and traits like motility. We explore these trends using the World Register of Marine Species dataset—a comprehensive dataset of the world's marine species-names (Appeltans et al., 2012; WoRMS Editorial Board, 2024). We

choose to use a marine dataset because most large and biologically diverse animal phyla are mostly or completely marine, and the taxonomists describing them likely represent a large and diverse group as well.

Methods

We used the World Register of Marine Species (WoRMS, https://www.marinespecies.org/; WoRMS Editorial Board, 2024; accessed 25-07-2023, doi:10.14284/170) which is a comprehensive and regularly updated species-name database of all marine organisms. From this, we identified large phyla containing more than 10,000 described species-names. This includes both currently accepted species names as well as synonyms. Next, we isolated the year each species-name was described in from the complete binomial name which includes describing authors' names. Describing authors are listed in the field "scientificNameAuthorship" in the WoRMS dataset. If two sets of authors and years were listed, the earlier description was considered. These constitute only 1.90% of the complete Animalia dataset of accepted and unaccepted species names.

For each accepted species, we queried all synonyms listing it as its accepted name and matched each species to its synonyms. For each accepted name described in a given year, we also calculated the time lag in years between the year it was described and the year its newest synonym was described. The first described name of a species remains its accepted name and takes precedence over synonym-names as per the Principle of Priority (Article 23; "The Code Online | International Commission on Zoological Nomenclature") which applies in the overwhelming majority of cases. As an example, *Chama broderipi* Reeve, 1846 and *Chama segmentina* Clessin, 1889 are today synonyms of *Chama pacifica* Broderip, 1835, with *C. pacifica* being the accepted species name since it was described first. This lag acts as a

conservative estimate for the time taken for all the biological diversity—in terms of once-unique species—characterising a species in the present day (2023 CE) to become known to science. For accepted names whose newest synonym bore the same year of publication, we also identified synonyms with similar specific epithets. A specific epithet was regarded as "similar" if its first four letters were the same for the accepted name and its newest synonym. A major source of such similar specific epithets is when taxonomic revisions occur at the genus level and new combinations are created (Article 34.2; Article 48, "The Code Online | International Commission on Zoological Nomenclature").

For each year starting from 1757 CE, we calculated the proportion of accepted names to the total names described in that year, and the mean number of synonyms per each accepted name per year. For these, we also calculated Spearman correlation coefficients for each 30-year period starting from 1757 CE. This acts as an index of the monotonicity of the relationship observed. All data were analysed and plotted in Python ver. 3.12.5. Spearman correlations were calculated using the SciPy function "scipy.stats.spearmanr" (Virtanen et al., 2020) and straight lines and confidence intervals were fit and plotted to the data using the Seaborn function "seaborn.regplot" (Waskom, 2021) after square-root transformation.

Results

Species-description trends vary widely across taxa

Moving averages of accepted species-names revealed that there is no consistent trend across phyla (Figure 1), with larger sliding window lengths being less sensitive to outliers. Linnaeus' continuing legacy could be seen in high starting values at 1757 CE for phyla which are charismatic such as Chordata and for phyla with persistent skeletons such as Mollusca. On the other hand, for soft-bodied phyla such as Platyhelminthes, Nematoda, and Annelida,

species descriptions before around 1850 CE were very low. The largely microscopic Foraminifera also presented low description rates in this period.

Large and prominent phyla like Arthropoda, Mollusca, Chordata, and Cnidaria displayed a continuously increasing trend in species described from the 1950s onward, with Echinodermata as a notable exception. Outlier years where a large number of species are described in a single year became common from around 1850 to 1950 but were rare thereafter except in chordates.

Synonymisation trends reveal changing taxonomic praxis consistently across taxa

For kingdom Animalia, the proportion of all names published in a given year which are valid today increased with time (Figure 2a). This trend held for all large phyla as well (Supplementary figure S1). This is reflected in increasing Spearman correlation coefficients for non-overlapping time bins starting from 1757 in a taxon-independent manner (Figure 2b). Accepted species-names with at least one synonym decreased in proportion from the early 20th century onwards, and the vast increase in named species in the second half of the 20th century was driven by species lacking any synonym (Figure 2c). Accepted species-names for which their newest synonym was published after a time-lag decreased in abundance from the late 19th century onwards. Amongst accepted names with no time-lag in publication, the subset of accepted names for which the newest synonym had a specific epithet similar to theirs displayed similar trends to the overall pattern of zero-lag synonyms with time (Figure 2d).

The mean lag times between the accepted names published in a year and their newest synonyms were strongly negatively correlated to time after square-root transformation

(Spearman R = – 0.88, $p < 0.0001$) and displayed a linear trend (Figure 3a). These mean lag values were smaller than the time remaining till present day until after approximately 2000 CE. Overall, the same trends held across large phyla (Supplementary figure S2). The mean number of synonyms per accepted name was also strongly negatively correlated to time after square-root transformation (Spearman R = – 0.94, $p < 0.0001$) and displayed a linear trend as well (Figure 3b) across all large phyla (Supplementary figure S3).

Discussion

A once-distinct species name which later gets synonymised encompasses biological variability that is not distinct enough to warrant its classification as a unique species. We do not consider when two species-names become synonymised. Instead, we consider when all the diversity characterising a species—represented as different once-unique species which eventually get synonymised into a single species name—becomes known to science. Species described earlier have had more time for the accumulation of synonyms. However, the difference between lag values and years remaining till present day indicates that for most of taxonomic history, this diversity characterising a species becomes known to science well before time becomes limiting (Figure 3a). Exceptions around present day violate the Principle of Priority and represent a very small fraction of the total species described around present day (Figure 2c). That species described earlier are wide-ranging (Collen et al., 2004; Gaston et al., 1995; Guedes et al., 2025; Higgs and Attrill, 2015), are more morphologically variable (Ficetola et al., 2016; Sides et al., 2014), and have more synonyms (Figure 3b) suggests a gradually changing trend in taxonomy of describing species encompassing smaller amounts of biological variability. In agreement, most species described in the latter half of the 20th century lack any synonym (Figure 2c).

Even though there are no cross-taxon trends of when the species accepted today were described (Appeltans et al., 2012; Figure 1), the proportion of accepted-names published each year remains initially constant across taxa (Figure 2a–b, Supplementary figure S1), suggesting that for this period in taxonomic history, the falsification ratio of a described species-name remains similar across taxa. While there is no discernable trend for synonymization for some fossil taxa like dinosaurs (Benton, 2008), our results show that given large sample sizes, there exists a simple and strong relationship for the temporal trends of synonymization across several extant taxa (Figure 3a–b).

The number of species with a lag time of zero peaks around 1900 (Figure 2d). These represent synonyms with a similar specific epithet as the accepted species-name. One major manner by which such synonyms with similar specific epithets arise is by new name combinations (Article 34, ICZN; "The Code Online | International Commission on Zoological Nomenclature"). That is, after the description of widespread, morphologically highly variable species, species described during this time were falsely allied to existing species or were falsely placed into distinct higher-level taxa (like the genus), with later rectification by taxonomic reassignment at higher levels of classification like the genus. For example, *Lacuna uchidai* (Habe, 1953) is a currently accepted mollusc species-name which was originally described as *Stenotis uchidai* (Habe, 1953) and arose as a new combination. This period is therefore characterised by species described for which accurate higher-level taxonomic relationships were not accurately determined during the time of description.

Next, because many recently-described species continue to be morphologically defined (Agapow et al., 2004) and formally described as allied to pre-existing species (Nerlekar et al., 2022), and have no synonyms (Figure 2c), they might represent a decrease in the amount of

variability circumscribed within a species. Our results therefore depict gradually changing taxonomic praxis along with species concepts becoming narrower with time. While there are few studies directly quantifying the quality of species descriptions (Poulin and Presswell, 2016), our results instead suggest them via the novel analyses of trends of synonym-accumulation.

Notably, when a single species is split into multiple species upon taxonomic revision, synonyms of the former parent species may not be transferred to the newly erected species. For example, former deep-water populations of *Ophiacantha bidentata* Bruzelius, 1805 were split off into *Ophiacantha fraterna* Verrill, 1885 by most taxonomists based on morphological evidence. However, existing heterotypic synonyms of *O. bidentata* (*O. spinulosa* Müller & Troschel, 1842 and *O. arctica* Müller & Troschel, 1842) are not synonyms of *O. fraterna* (Martynov and and Litvinova, 2008). However, in spite of this non-transference of parental synonyms to newly split-off taxa, description date correlates with synonym number across taxa, datasets, timespans, and methods of quantification (Appeltans et al., 2012; Eggleton, 1999; Guedes et al., 2025), suggesting that the loss of synonym names in the aforementioned manner is unlikely to majorly impact our results.

Lastly, we note that while trends of synonymization have been used to estimate how many currently-accepted species shall later be falsified (Alroy, 2002; Costello and Wilson, 2011; Wortley and Scotland, 2004), our results instead depict the species concept itself changing with time.

Conclusion

We use a comprehensive dataset of marine species and their synonyms and show that across taxa, while species description lacks discernible trends, the trends of synonymization are remarkably similar across taxa. These similarities suggest overall changes in the manner taxonomy was practised through time, with a reduction in the amount of variability characterising species on moving towards the present day.

Acknowledgments

We gratefully acknowledge the comments and feedback of Akshay Bharadwaj, Arun Ravi, Balaji Chattopadhyay, and Dhrubojyoti Patra. We also acknowledge Devapriya Chattopadhyay for introductions from which this study emerged.

Data and code availability

All data used are available from WoRMS (https://doi.org/10.14284/170). All codes used are available at the Zenodo URL https://doi.org/10.5281/zenodo.15285361.

Author Contributions

VI: Conceptualisation (lead), dataset curation (supporting), formal analysis (equal), visualisation (equal), writing – original draft preparation (lead), writing – review and editing (equal).

AP: Conceptualisation (supporting), dataset curation (lead), formal analysis (equal), visualisation (equal), writing – original draft preparation (supporting), writing – review and editing (equal).

References

Agapow, P.-M., R. P. Bininda-Emonds*, O., A. Crandall, K., L. Gittleman, J., M. Mace, G., C. Marshall, J., Purvis, A., 2004. The Impact of Species Concept on Biodiversity Studies. Q. Rev. Biol. https://doi.org/10.1086/383542

Alroy, J., 2002. How many named species are valid? Proc. Natl. Acad. Sci. 99, 3706–3711. https://doi.org/10.1073/pnas.062691099

Appeltans, W., Ahyong, S.T., Anderson, G., Angel, M.V., Artois, T., Bailly, N., Bamber, R., Barber, A., Bartsch, I., Berta, A., Błażewicz-Paszkowycz, M., Bock, P., Boxshall, G., Boyko, C.B., Brandão, S.N., Bray, R.A., Bruce, N.L., Cairns, S.D., Chan, T.-Y., Cheng, L., Collins, A.G., Cribb, T., Curini-Galletti, M., Dahdouh-Guebas, F., Davie, P.J.F., Dawson, M.N., De Clerck, O., Decock, W., De Grave, S., de Voogd, N.J., Domning, D.P., Emig, C.C., Erséus, C., Eschmeyer, W., Fauchald, K., Fautin, D.G., Feist, S.W., Fransen, C.H.J.M., Furuya, H., Garcia-Alvarez, O., Gerken, S., Gibson, D., Gittenberger, A., Gofas, S., Gómez-Daglio, L., Gordon, D.P., Guiry, M.D., Hernandez, F., Hoeksema, B.W., Hopcroft, R.R., Jaume, D., Kirk, P., Koedam, N., Koenemann, S., Kolb, J.B., Kristensen, R.M., Kroh, A., Lambert, G., Lazarus, D.B., Lemaitre, R., Longshaw, M., Lowry, J., Macpherson, E., Madin, L.P., Mah, C., Mapstone, G., McLaughlin, P.A., Mees, J., Meland, K., Messing, C.G., Mills, C.E., Molodtsova, T.N., Mooi, R., Neuhaus, B., Ng, P.K.L., Nielsen, C., Norenburg, J., Opresko, D.M., Osawa, M., Paulay, G., Perrin, W., Pilger, J.F., Poore, G.C.B., Pugh, P., Read, G.B., Reimer, J.D., Rius, M., Rocha, R.M., Saiz-Salinas, J.I., Scarabino, V., Schierwater, B., Schmidt-Rhaesa, A., Schnabel, K.E., Schotte, M., Schuchert, P., Schwabe, E., Segers, H., Self-Sullivan, C., Shenkar, N., Siegel, V., Sterrer, W., Stöhr, S., Swalla, B., Tasker, M.L., Thuesen, E.V., Timm, T., Todaro, M.A., Turon, X., Tyler, S., Uetz, P., van der Land, J., Vanhoorne, B., van Ofwegen, L.P., van Soest, R.W.M., Vanaverbeke, J., Walker-Smith, G., Walter, T.C., Warren, A., Williams, G.C., Wilson, S.P., Costello, M.J., 2012. The Magnitude of Global Marine Species Diversity. Curr. Biol. 22, 2189–2202. https://doi.org/10.1016/j.cub.2012.09.036

Benton, M.J., 2008. How to find a dinosaur, and the role of synonymy in biodiversity studies. Paleobiology 34, 516–533. https://doi.org/10.1666/06077.1

Collen, B., Purvis, A., Gittleman, J.L., 2004. Biological correlates of description date in carnivores and primates. Glob. Ecol. Biogeogr. 13, 459–467. https://doi.org/10.1111/j.1466-822X.2004.00121.x

Costello, M.J., Wilson, S.P., 2011. Predicting the number of known and unknown species in European seas using rates of description. Glob. Ecol. Biogeogr. 20, 319–330.

De Queiroz, K., 2007. Species Concepts and Species Delimitation. Syst. Biol. 56, 879–886. https://doi.org/10.1080/10635150701701083

Eggleton, P., 1999. Termite species description rates and the state of termite taxonomy. Insectes Sociaux 46, 1–5. https://doi.org/10.1007/s000400050105

Ficetola, G.F., Colleoni, E., Renaud, J., Scali, S., Padoa-Schioppa, E., Thuiller, W., 2016. Morphological variation in salamanders and their potential response to climate change. Glob. Change Biol. 22, 2013–2024. https://doi.org/10.1111/gcb.13255

Gaston, K.J., Blackburn, T.M., Loder, N., 1995. Which species are described first?: the case of North American butterflies. Biodivers. Conserv. 4, 119–127. https://doi.org/10.1007/BF00137780

Gaston, K.J., Mound, L.A., 1997. Taxonomy, hypothesis testing and the biodiversity crisis. Proc. R. Soc. Lond. B Biol. Sci. 251, 139–142. https://doi.org/10.1098/rspb.1993.0020

Guedes, J.J.M., Moura, M.R., Jardim, L., Diniz-Filho, J.A.F., 2025. Global Patterns of Taxonomic Uncertainty and its Impacts on Biodiversity Research. Syst. Biol. syaf010. https://doi.org/10.1093/sysbio/syaf010

Higgs, N.D., Attrill, M., 2015. Biases in biodiversity: wide-ranging species are discovered first in the deep sea. Front. Mar. Sci. 2, 152190. https://doi.org/10.3389/fmars.2015.00061

Isaac, N.J.B., Mallet, J., Mace, G.M., 2004. Taxonomic inflation: its influence on macroecology and conservation. Trends Ecol. Evol. 19, 464–469. https://doi.org/10.1016/j.tree.2004.06.004

Jones, O.R., Purvis, A., Quicke, D.L.J., 2012. Latitudinal gradients in taxonomic overdescription rate affect macroecological inferences using species list data. Ecography 35, 333–340. https://doi.org/10.1111/j.1600-0587.2011.06956.x

Martynov, A.V., and Litvinova, N.M., 2008. Deep-water Ophiuroidea of the northern Atlantic with descriptions of three new species and taxonomic remarks on certain genera and species. Mar. Biol. Res. 4, 76–111. https://doi.org/10.1080/17451000701840066

Nerlekar, A.N., Chorghe, A.R., Dalavi, J.V., Kusom, R.K., Karuppusamy, S., Kamath, V., Pokar, R., Rengaian, G., Sardesai, M.M., Kambale, S.S., 2022. Exponential rise in the discovery of endemic plants underscores the need to conserve the Indian savannas. Biotropica 54, 405–417. https://doi.org/10.1111/btp.13062

Poulin, R., Presswell, B., 2016. Taxonomic Quality of Species Descriptions Varies over Time and with the Number of Authors, but Unevenly among Parasitic Taxa. Syst. Biol. 65, 1107–1116. https://doi.org/10.1093/sysbio/syw053

Sides, C.B., Enquist, B.J., Ebersole, J.J., Smith, M.N., Henderson, A.N., Sloat, L.L., 2014. Revisiting Darwin's hypothesis: Does greater intraspecific variability increase species' ecological breadth? Am. J. Bot. 101, 56–62. https://doi.org/10.3732/ajb.1300284

The Code Online | International Commission on Zoological Nomenclature [WWW Document], n.d. URL https://www.iczn.org/the-code/the-code-online/ (accessed 10.5.24).

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Waskom, M.L., 2021. seaborn: statistical data visualization. J. Open Source Softw. 6, 3021. https://doi.org/10.21105/joss.03021

WoRMS Editorial Board, 2024. World Register of Marine Species. Available from https://www.marinespecies.org at VLIZ. Accessed 25-07-2023. https://doi.org/10.14284/170

Wortley, A.H., Scotland, R.W., 2004. Synonymy, Sampling and Seed Plant Numbers. Taxon 53, 478–480. https://doi.org/10.2307/4135625

Figures

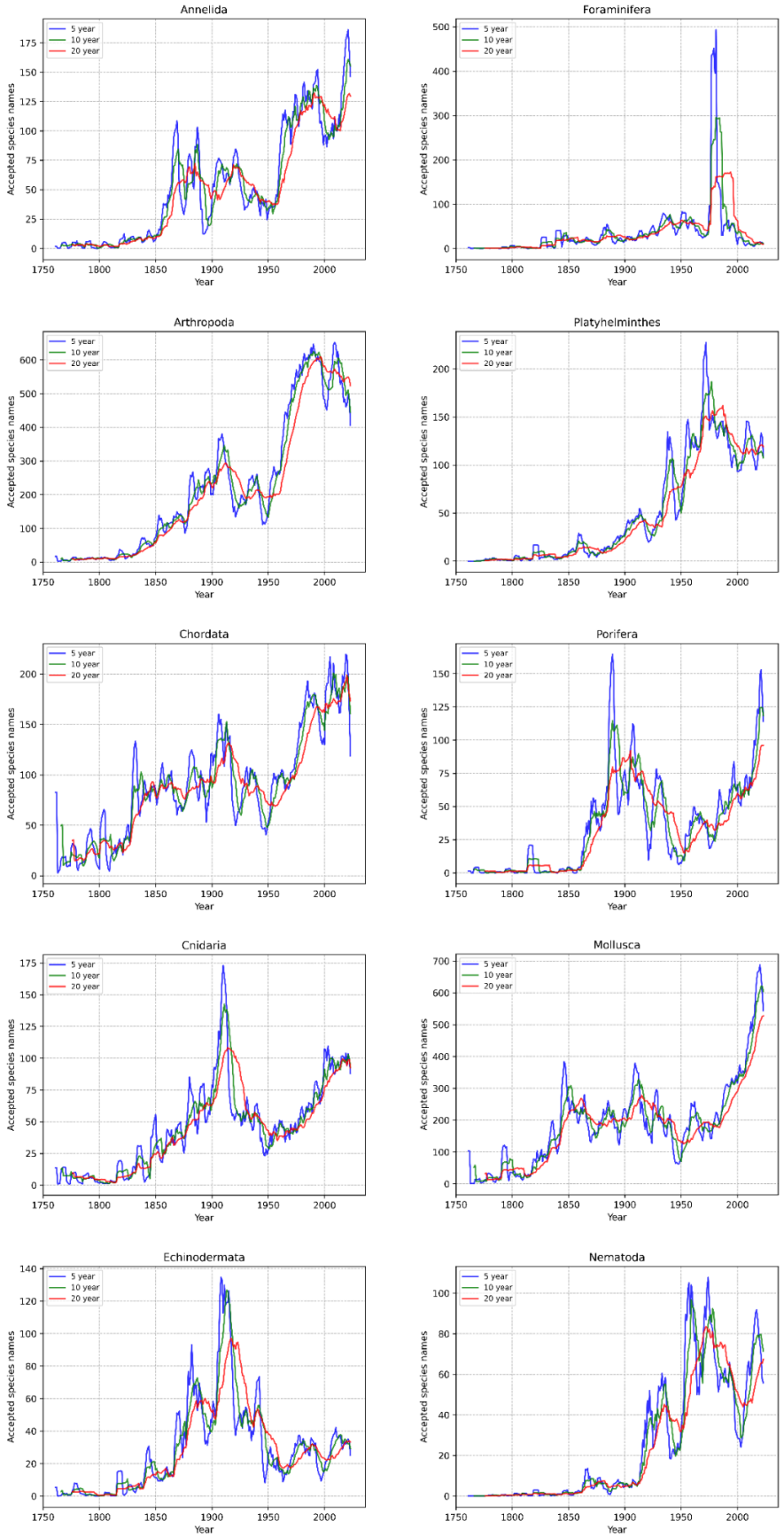

Annelida
Foraminifera
Arthropoda
Platyhelminthes
Chordata
Porifera
Cnidaria
Mollusca
Echinodermata
Nematoda
5 year
10 year
20 year
Accepted species names
Year

Figure 1:

5, 10, and 20 year moving averages of the number of accepted species names with their year of publication for each phylum. For each year, the graph indicates how many species names which are accepted in 2023 were published in that year.

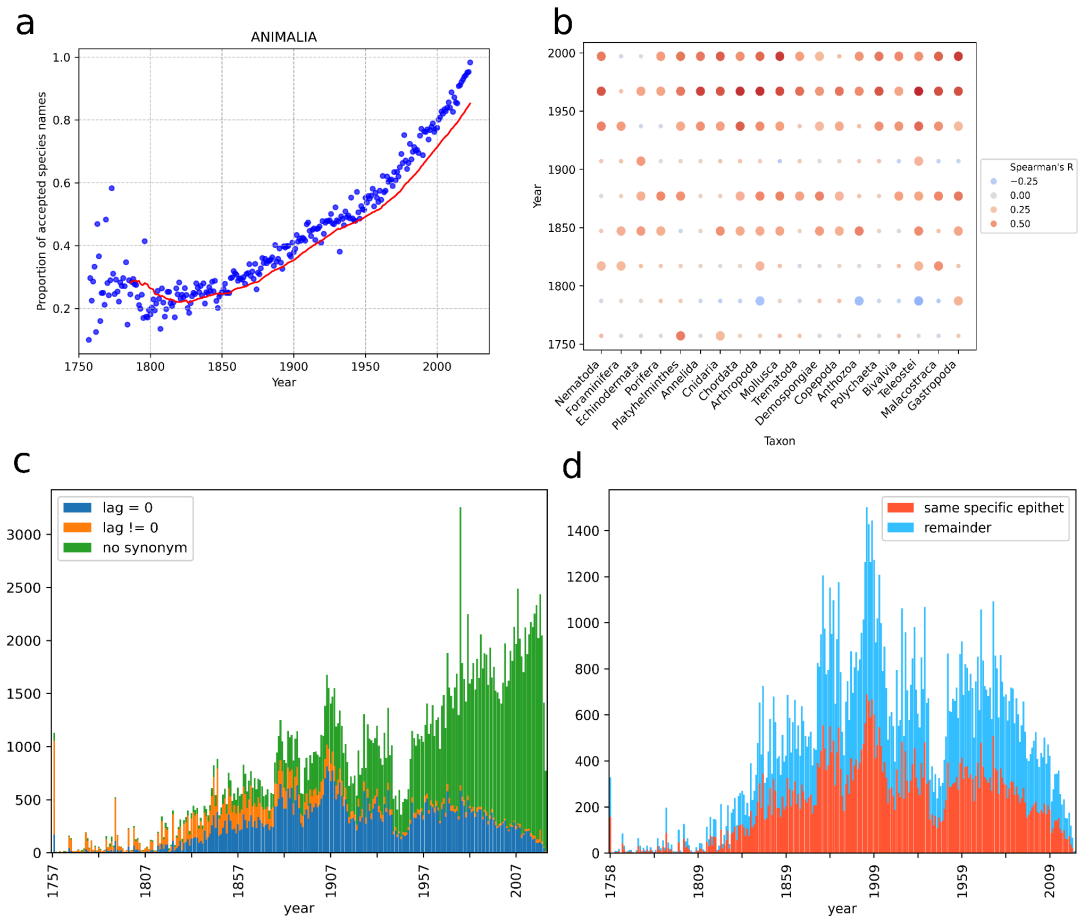
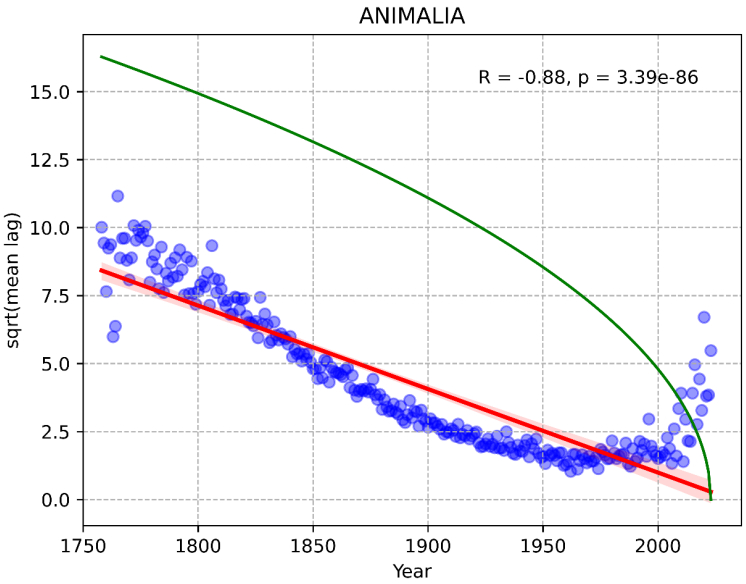


Figure 2:

Synonymization with time for Animalia. **A:** Scatterplot depicting the fraction of all species-names published per year which are accepted today. Points represent values for a year and the red line displays the 20 year moving average. **B:** Spearman correlation coefficients between year and the proportion of accepted species names for all large taxa for each 50 year period starting from 1757. Large circles represent significant ($p < 0.1$) correlations. **C–D:** Stacked histograms depicting the number of accepted species-names (c) in present day grouped by the lag between the accepted species-name and its newest synonym. "Remainder" depicts accepted species-names without a synonym, (d) with a lag time of zero between the name and its newest synonym, grouped by whether or not the specific epithets are similar between them. "Remainder" depicts accepted species-names with their newest synonym bearing a dissimilar specific epithet.



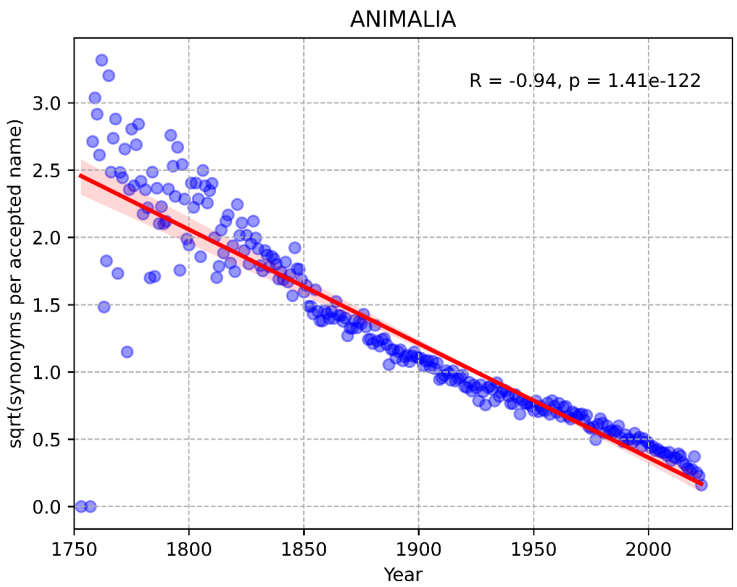


Figure 3:

Metrics of synonymization with time for Animalia. Above: Scatterplot depicting the mean lag value between the years of description of the newest synonym of the accepted species names published in a year. Points represent lag values. The red line represents a straight line fit to the data and red bands indicate 95% confidence intervals. R = Spearman correlation coefficient. Lag values are square-root transformed before plotting. The green line represents the number of years till present day. Below: Scatterplot depicting the mean number of

synonyms for each accepted species-name published with time. Values are square-root transformed before plotting. Points represent the mean number of synonyms. The red line represents a straight line fit to the data and red bands indicate 95% confidence intervals. R = Spearman correlation coefficient.