Article:

Gonzalez-del-Pliego, P., Freckleton, R. orcid.org/0000-0002-8338-864X, Edwards, D.P. et al. (4 more authors) (2019) Phylogenetic and Trait-Based Prediction of Extinction Risk for Data-Deficient Amphibians. Current Biology, 29. pp. 1557-1563. ISSN 0960-9822

https://doi.org/10.1016/j.cub.2019.04.005

Phylogenetic and trait-based prediction of extinction risk for data-deficient amphibians

Pamela González-del-Pliego[1,2]*, Robert P. Freckleton[1], David P. Edwards[1], Michelle S. Koo[3], Brett R. Scheffers[4], R. Alexander Pyron[5], Walter Jetz[2]

Affiliations:

[1]Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield, South Yorkshire S102TN, UK.

[2]Department of Ecology and Evolutionary Biology, Yale University, Prospect Street, New Haven, Connecticut 06520, USA

[3]Museum of Vertebrate Zoology, University of California, Valley Life Sciences Building, Berkeley, California 94720, USA

[4]Department of Wildlife Ecology & Conservation, Newins-Ziegler Hall, University of Florida/IFAS, Gainesville, FL 32611, USA

[5]Department of Biological Sciences, The George Washington University, 22nd St. NW, Washington, DC 20052, USA.

*Lead contact: pgonzalezdelpliego@gmail.com

Summary

Amphibians are among the most highly threatened lineages, with at least 2,000 species estimated to be in danger of extinction [1, 2]. Alarmingly, another ~2,200 species (~25% of all ~7,900 known species) are d~~D~~ata-~~-~~d~~D~~eficient~~ (DD)~~ or Not Evaluated (hereinafter termed data-deficient~~DD~~) by the IUCN [1]. Without an estimate of their status, data-deficient ~~DD~~ species are usually overlooked in conservation planning and resource allocation [3]. Amphibians have the highest proportion of data-deficient ~~DD~~ species of any vertebrate group [1, 4], which highlights the need to estimate their threat status considering potentially imminent extinctions. We apply a trait-based spatio-phylogenetic statistical framework [5] to predict threat status for data-deficient~~DD~~ species. Since ecological, geographical, and evolutionary attributes increase extinction risk [6, 7], we used geographic distribution data [1, 8], phylogenetically imputed ecological traits, and an amphibian phylogeny [9] to provide initial baseline predictions. We estimate that half of the ~2,200 ~~DD~~data-deficient species are threatened with extinction (Vulnerable, Endangered, or Critically Endangered), primarily in the Neotropics and Southeast Asia. This increases the number of amphibian species estimated to be threatened with extinction by ~50%. Of these, we predict that ~500 species are Endangered or Critically Endangered, and three may be extinct already. We highlight families that are most at risk and suggest where urgent conservation is needed to avert their loss. We show that some of the most vulnerable species may also be the most poorly known and offer an analytical framework for preliminary analysis of their threat status in the face of deficient empirical data.

Keywords

Amphibians, Anura, Caudata, data-deficient ~~data deficient~~ species, Gymnophiona, imputation, IUCN threat status, phylogeny

Results and Discussion

We used phylogenetic and geographic data, a human encroachment index as a measure of potential anthropogenic pressure, and a global trait database containing phylogenetically imputed ecological traits for 7,143 species to predict the extinction risk of global data-deficient ~~data deficient (DD)~~ amphibian species (see STAR Methods). These traits include body size, forest association, and reproductive mode. Predictions from our model were generated by evaluating traits together with the degree of spatial and phylogenetic similarity among species (following [5]). We provide imputed threat statuses for ~2,200 data-deficient ~~DD~~ species and note the localization of high predicted threat-status in several lineages and regions. Our validation models, combining trait-based, spatial, and phylogenetic predictors of threat status, accurately predicted the observed threat status of the 998 (20%) assessed species (see STAR Methods). Our model correctly predicted the binary classification of extinction risk (threatened vs non-threatened) for 76%, 70% and 100% of Anura, Caudata and Gymnophiona species already assessed by the IUCN in the validation analysis (Table S1).

As expected, body size ($F_{1, 4981} = 25.34$, $P < 0.001$) and range size ($F_{1, 4981} = 370.8$, $P < 0.001$) proved to be strongly correlated to threat status, as noted by numerous previous authors [10, 11](see Appendix 1 for predictor performance). The latter is true despite the fact that the expert range maps overestimate occupied ranges in a very inconsistent way [12, 13]. Interestingly, similar to reptiles [14, 15], we found that the human-encroachment index was a strong predictor of extinction risk across all orders ($P < 0.05$; Appendix 1)—owing to most identified priority research areas for ~~DD~~ data-deficient amphibian species overlap with regions under high human pressure [16]. This suggests that environmental measures, such as human-encroachment, may also offer strong predictive power in future studies of threat predictions, in addition to range and body size.

Our models predict an additional 1,012 amphibian species (47% of the ~~DD~~ data-deficient ~~-~~species) are threatened with extinction. Of these, nearly half (469 species) are likely to be Endangered or Critically Endangered (‘imperilled’), and three may be extinct already (Appendix 1). Compared with assessed amphibian species, the percentage predicted to be under threat is higher for the data-~~-~~deficient species (47% in data-deficient ~~DD~~ versus 41% in non-~~-~~data-deficient~~DD~~; $\chi^2 = 17.24$, df = 1, $P < 0.001$). Our results, however, found a smaller number of species (47%) predicted to be threatened versus the 63% found by [17]. Similar to [17, 18], we found that the percentage of data-deficient ~~DD~~ species predicted to be threatened with extinction is higher than those assessed by the IUCN. Our model was more accurate in predicting threatened species than non-threatened species indicated by the higher percentage of non-threatened species predicted to be threatened versus the percentage of threatened species predicted to be non-threatened (Table S1). However, the predicted and observed threat scores of species were highly correlated ($R^2 = 0.70$, df = 52, $P < 0.001$; Fig. 1). Our findings highlight the importance for both phylogenetic and geographical distribution data to be incorporated in modelling assessment statuses, which is particularly necessary in amphibians since their population declines can be taxonomically and regionally specific [2].

The differences between average threat score predicted for the data-deficient ~~DD~~ species and the observed for assessed species varied within taxonomic orders. Threat scores were similar in Caudata (F = 1.06, df = 654, P = 0.30) and Gymnophiona (F = 1.83, df = 190, P = 0.17), yet in Anura, assessed species had a slightly higher threat score 2.26 ±0.02 than the predicted threat score 2.14 ±0.02 for data-deficient ~~DD~~ species (F = 10.31, df = 6293, P = 0.001). At the family level, we found no significant differences in the threat score between assessed species and data-deficient ~~DD~~ species for Caudata (F = 0.96, df = 642, P = 0.41) and Gymnophiona (F = 1.65, df = 8, P = 0.11). However, we did find differences in threat scores among families of Anura (F = 1.41, df = 40, P = 0.05; Fig. 2).

Only four Anura families showed significant differences in threat score between assessed species and data-deficient ~~DD~~ species. Specifically, we found that the predicted threat score for data-deficient ~~DD~~ species in Ptychadenidae and Dicroglossidae was, on average, higher than the assessed species (Fig. 2). Conversely, the assessed species from Microhylidae and Eleutherodactylidae had a higher threat status than their data-deficient ~~DD~~ counterparts (Fig. 2). Microhylids have a global distribution, but both of these families overlap in distribution in the Americas, and both are in the top ten families with the highest species richness (see Fig. 2 for number of species per family). The same differences on threat scores between orders and families were found when comparing the average threat score of predicted assessed species (instead of observed assessed species) and predicted data-deficient ~~DD~~ species (see Supplementary information and Fig. S1).

The phylogenetic distribution of threat in data-deficient ~~DD~~ species is non-random with respect to phylogeny, especially at the family level (Fig. 3). For Anura, ~~D~~data-deficient ~~D~~ threatened species belonged primarily to Neotropical families, such as: Dendrobatidae (poison frogs), Centrolenidae (glass frogs), Strabomantidae (*Pristimantis*; South American rain frogs), Hylidae (*Hypsiboas*; gladiator frogs), and Bufonidae (*Atelopus*; harlequin toads). We also found the Asian bush frogs in Rhacophoridae are particularly threatened (Fig. 3). Within Caudata, data-deficient ~~DD~~ species predicted to be threatened were found mostly in

Plethodontidae (76 species, mostly the Neotropical mushroom-tongued salamanders *Bolitoglossa*). Most Gymnophiona species are data-deficient ~~DD~~ (67%); however, Caeciliidae (common caecilians from the Neotropics) and Herpelidae (African caecilians from tropical Africa) showed the highest number of data-deficient ~~DD~~ species predicted to be threatened (2 species each; see Table S2 for geographical distributions). The non-random distribution of threatened status with respect to phylogenetical relatedness of these genera and families increase the risk of losing whole branches of the tree of life [11]. This risk is higher for *Bolitoglossa* and *Pristimantis* clades as they are strongly susceptible to habitat loss [19]. Also, evolutionarily distinct amphibians, such as Microhylidae and Caeciliidae, can be particularly at risk as they tend to disappear when forested areas are lost [20]. Of concern, the only three species that our model predicted as Extinct occur in the Neotropics (*Atelopus carauta* and *Craugastor cuaquero*) and India (*Nyctibatrachus poocha*). This suggest they are highly threatened, and we strongly suggest immediate assessment of these species.

Data-deficient ~~DD~~ species are found across all continents, except Antarctica (Fig. S2), but most threatened species, both assessed and predicted, are in the tropics (Fig. 4). Overall, the distribution of data-deficient species predicted to be threatened is similar to the distribution of assessed threatened (Figs. 4B, 4D) and small-ranged species [21]. This suggests that the geographical similarity between data-deficient species predicted to be threatened and assessed threatened species is partially a result of data-deficient species predicted to be threatened having small ranges and a geographically overlap with known threatened species [21]. Yet, there are also areas, for example the Atlantic forest in Brazil, where the distribution of assessed threatened and small-ranged species is similar to both data-deficient species predicted to be threatened and not-threatened [22](Fig. 4). Indeed, a ~~A~~cross the tropics, the

Atlantic forest harbours most of the data-deficient ~~DD~~ species predicted to be non-threatened (Fig. 4C). ~~Moreover~~

T~~, t~~here were two tropical regions that held most of the data-deficient ~~DD~~ species predicted to be threatened: the Neotropics and Southeast Asia (Fig. 4D). The geographical distribution of data-deficient ~~DD~~ species predicted to be threatened is similar to that of assessed threatened species (Figs. 4 and S3). Particularly in the Neotropics, assessed and predicted threatened species have a high geographical overlap (Figs. 4 and S3A). This may have important implications for amphibian conservation strategies. For example, conservation initiatives in the Neotropics aimed to protect current threatened species would also have a positive effect on predicted threatened species. The Neotropics also house the largest number of threatened amphibians, both assessed and predicted likely due to extensive habitat loss in this region, since land-use change is currently the largest threat for amphibians [23]. The Tropical Andes, for example, has lost about 75% of its original forests, while a large part of its remaining forests are severely fragmented [24].

In Southeast Asia, contrary to the Neotropics, we found countries like Sri Lanka, Malaysia, and Myanmar where there is little or no geographic overlap between data-deficient ~~DD~~ species predicted to be threatened and assessed threatened species (Figs. 4 and S3B). We also find countries like the Philippines and Sri Lanka that have 49% and 70% of their amphibians listed as Threatened, respectively [1, 8]. Moreover, Southeast Asia harbours a high proportion of rhacophorid species, which have high numbers of related data-deficient ~~DD~~ species predicted to be threatened. This, in turn, increases the risk of losing this phylogenetically clustered group [17], particularly because they are strongly susceptible to habitats loss [19]. Countries with a low number of threatened species such as Papua New

Guinea and Thailand may be artificially low due to an extreme lack of data, a pattern also seen in squamates [25]. In support of this point, interestingly, Papua New Guinea and Sri Lanka are among the top four countries for new species discovery, along with Brazil and Peru [8].

Finally, in the Central Africa region, the geographical distributions of assessed and predicted threatened amphibians do not geographically overlap nearly as much as in the Neotropics (Figs. 4 and S3C). The presence of data-deficient ~~DD~~ species in this region might be partially caused by high inaccessibility due to human conflict and political isolation [8]. There are also several countries, such as Australia (mainly in the east), Bolivia, Jamaica, and Cuba that have a large number of assessed threatened species, and a low number of data-deficient ~~DD~~ species predicted to be threatened (Fig. 4). Again, these situations are likely related to knowledge gaps including Linnaean, Darwinian, and Wallacean shortfalls.

Amphibians continue to have a high rate of newly discovered species [26], suggesting that true amphibian richness is significantly greater than we currently know [27, 28]. Thus, the number of threatened species could also be much greater than we estimate. Overall, our results agree with [16], suggesting that priority research areas for data-deficient ~~DD~~ amphibians are in the South American Tropical Andes and Atlantic forest, and in Southeast Asia, and research funds should be allocated there. Relative to other taxa, amphibians still remain grossly underfunded [29]. Our results allow for informed decisions as to which species and regions to target with this limited budget (i.e. those that we identified as threatened or imperilled).

Our predictions represent a large increase of amphibian species at risk compared to the IUCN data, specifically an additional 1,012 amphibian species (47% of the data-deficient ~~DD~~ species). Of these, we predict that half (469 species) are likely to be Endangered or

Critically Endangered, and three may be extinct already. This suggests that many new species should be added to the threatened categories of the Red List, with potentially great consequences for geographical conservation prioritisation, especially in Southeast Asia and the Central Africa region. Unfortunately, the regions that harbour the highest amphibian richness are also experiencing the greatest rates of forest loss and increases in human population size, as well as greater fertiliser use, agricultural production, and irrigation [30]. Our study adds to a building consensus that the fate of data-deficient ~~DD~~ species is uncertain due to their intrinsic traits linked to extinction risk combined with their exposure to external disturbances and thus, they require urgent conservation attention [17, 18, 31]. As amphibian declines continue worldwide [32], we need to move quickly to consider amphibians as a high conservation priority and integrate data deficient species into conservation strategies.

Acknowledgments

Massive thanks to Ajay Ranipeta. We thank Julie Hart, Diego Ellis-Soto, Nathan Upham, Dan Greenberg, Michelle Duong and Felix Lim for methods support. Many thanks to Mario Ribeiro de Moura for his comments. Thanks to Gavin Thomas for the original tree colour code. A big thank to David Bickford for sharing the trait database [10] that served as a base for ours. P.G. was supported by CONACyT, Scholarship 359063. Support from the VertLife Project (NSF DEB-1441652 to MSK and DEB-1441719 to RAP) contributed to range, trait, and phylogenetic data. We thank students at San Francisco State University and University of California Berkeley for their contributions.

Authors contribution

Conceptualization: PG, RF, DE, MSK, WJ; Methodology: PG, RF, WJ; Validation: PG; Formal Analysis: PG, RF; Investigation: PG, MSK; Writing – original draft: PG; Writing – review & editing: PG, RF, DE, MSK, BRS, RAP, WJ; Visualization: PG, RAP: Supervision: RF, DE, BRS, WJ.

Declaration of interests

The authors declare no competing interests.

References

1. IUCN (2017). The IUCN Red List of threatened species. Version 2016-3. Volume 2017.
2. Stuart, S.N., Chanson, J.S., Cox, N.A., Young, B.E., Rodrigues, A.S.L., Fischman, D.L., and Waller, R.W. (2004). Status and trends of amphibian declines and extinctions worldwide. Science *306*, 1783-1786.
3. Sousa-Baena, M.S., Garcia, L.C., and Townsend Peterson, A. (2014). Knowledge behind conservation status decisions: Data basis for “Data Deficient” Brazilian plant species. Biological Conservation *173*, 80-89.
4. Böhm, M., Collen, B., Baillie, J.E.M., Bowles, P., Chanson, J., Cox, N., Hammerson, G., Hoffmann, M., Livingstone, S.R., Ram, M., et al. (2013). The conservation status of the world's reptiles. Biological Conservation *157*, 372-385.
5. Jetz, W., and Freckleton, R.P. (2015). Towards a general framework for predicting threat status of data-deficient species from phylogenetic, spatial and environmental information. Philosophical Transactions of the Royal Society B-Biological Sciences *370*.
6. Bielby, J., Cunningham, A.A., and Purvis, A. (2006). Taxonomic selectivity in amphibians: ignorance, geography or biology? Animal Conservation *9*, 135-143.
7. Lee, T.M., and Jetz, W. (2011). Unravelling the structure of species extinction risk for predictive conservation science. Proceedings of the Royal Society B: Biological Sciences *278*, 1329-1338.
8. AmphibiaWeb (2016). AmphibiaWeb: information on amphibian biology and conservation. (Berkeley, California:).
9. Jetz, W., and Pyron, R.A. (2018). The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. Nature Ecology & Evolution.
10. Sodhi, N.S., Bickford, D., Diesmos, A.C., Lee, T.M., Koh, L.P., Brook, B.W., Sekercioglu, C.H., and Bradshaw, C.J.A. (2008). Measuring the meltdown: drivers of global amphibian extinction and decline. PLoS One *3*.
11. Veron, S., Penone, C., Clergeau, P., Costa, G.C., Oliveira, B.F., Sao-Pedro, V.A., and Pavoine, S. (2016). Integrating data-deficient species in analyses of evolutionary history loss. Ecology and Evolution *6*, 8502-8514.
12. Ocampo-Peñuela, N., Jenkins, C.N., Vijay, V., Li, B.V., and Pimm, S.L. (2016). Incorporating explicit geospatial data shows more species at risk of extinction than the current Red List. Science Advances *2*, e1601367.
13. Jetz, W., Sekercioglu, C.H., and Watson, J.E. (2008). Ecological correlates and conservation implications of overestimating species geographic ranges. Conserv Biol *22*, 110-119.
14. Tingley, R., Hitchmough, R.A., and Chapple, D.G. (2013). Life-history traits and extrinsic threats determine extinction risk in New Zealand lizards. Biological Conservation *165*, 62-68.
15. Böhm, M., Williams, R., Bramhall, H.R., McMillan, K.M., Davidson, A.D., Garcia, A., Bland, L.M., Bielby, J., and Collen, B. (2016). Correlates of extinction risk in squamate reptiles: the relative importance of biology, geography, threat and range size. Global Ecology and Biogeography *25*, 391-405.
16. Nori, J., Villalobos, F., and Loyola, R. (2018). Global priority areas for amphibian research. Journal of Biogeography *45*, 2588-2594.
17. Howard, S.D., and Bickford, D.P. (2014). Amphibians over the edge: silent extinction risk of Data Deficient species. Diversity and Distributions *20*, 837-846.
18. Morais, A.R., Siqueira, M.N., Lemes, P., Maciel, N.M., De Marco, P., and Brito, D. (2013). Unraveling the conservation status of Data Deficient species. Biological Conservation *166*, 98-102.
19. Nowakowski, A.J., Frishkoff, L.O., Thompson, M.E., Smith, T.M., and Todd, B.D. (2018). Phylogenetic homogenization of amphibian assemblages in human-altered habitats across the globe. Proceedings of the National Academy of Sciences.

20. Greenberg, D.A., Palen, W.J., Chan, K.C., Jetz, W., and Mooers, A.Ø. (2018). Evolutionarily distinct amphibians are disproportionately lost from human-modified ecosystems. Ecology Letters *21*, 1530-1540.

21. Jenkins, C.N., Pimm, S.L., and Joppa, L.N. (2013). Global patterns of terrestrial vertebrate diversity and conservation. Proceedings of the National Academy of Sciences *110*, E2602-E2610.

22. Jenkins, C.N., Alves, M.A.S., Uezu, A., and Vale, M.M. (2015). Patterns of Vertebrate Diversity and Protection in Brazil. PLOS ONE *10*, e0145064.

23. CBD, S.o.t.C.o.B.D. (2014). Global Biodiversity Outlook 4. (Montréal), p. 155.

24. Amézquita, A., Kahn, T., Kraus, K., Marca, E., Medina-Rengifo, R., and Chaves Portilla, G. (2016). Amphibian Conservation in the Tropical Andes and Amazon Basin.

25. Tonini, J.F.R., Beard, K.H., Ferreira, R.B., Jetz, W., and Pyron, R.A. (2016). Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. Biological Conservation *204*, 23-31.

26. Parra, G., Brown, R., Hanken, J., Hedges, B., Heyer, R., Kuzmin, S., Lavilla, E., Lötters, S., Pimenta, B., Richards, S., et al. (2007). Chapter 10. Systematics and Conservation *in* Gascon C., Collins J.P., Moore R.D., Church D.R., McKay J.E., Mendelson J.R. III (eds) Amphibian Conservation Action Plan. IUCN/SSC Amphibian Specialist Group. . In Gland, Switzerland and Cambridge, UK. (The World Conservation Union (IUCN), Gland, Switzerland), p. 64.

27. Köhler, J., Vieites, D.R., Bonett, R.M., García, F.H., Glaw, F., Steinke, D., and Vences, M. (2005). New amphibians and global conservation: a boost in species discoveries in a highly endangered vertebrate group. BioScience *55*, 693-696.

28. Glaw, F., and Köhler, J. (1998). Amphibian species diversity exceeds that of mammals. Herpetological Review *29*, 11-12.

29. ASG and ASA, A.S.G.a.A.S.A. (2014). Amphibian Survival Alliance and Amphibian Specialist Group. Volume 2017. (London, UK).

30. Collins, J.P., and Halliday, T. (2005). Forecasting changes in amphibian biodiversity: aiming at a moving target. Philos Trans R Soc Lond B Biol Sci *360*, 309-314.

31. Nori, J., and Loyola, R. (2015). On the Worrying Fate of Data Deficient Amphibians. PLoS One *10*, e0125055.

32. Pounds, J.A., Bustamante, M.R., Coloma, L.A., Consuegra, J.A., Fogden, M.P.L., Foster, P.N., La Marca, E., Masters, K.L., Merino-Viteri, A., Puschendorf, R., et al. (2006). Widespread amphibian extinctions from epidemic disease driven by global warming. Nature *439*, 161-167.

33. Freckleton, R.P., and Jetz, W. (2009). Space versus phylogeny: disentangling phylogenetic and spatial signals in comparative data. Proceedings of the Royal Society B: Biological Sciences *276*, 21-30.

34. Tapley, B., Michaels, C.J., Gumbs, R., Böhm, M., Luedtke, J., Pearce-Kelly, P., and Rowley, J.J.L. (2018). The disparity between species description and conservation assessment: A case study in taxa with high rates of species discovery. Biological Conservation *220*, 209-214.

35. Goolsby, E.W., Bruggeman, J., and Ané, C. (2017). Rphylopars: fast multivariate phylogenetic comparative methods for missing data and within-species variation. Methods in Ecology and Evolution *8*, 22-27.

36. Theodore Garland, J., and Ives, A.R. (2000). Using the Past to Predict the Present: Confidence Intervals for Regression Equations in Phylogenetic Comparative Methods. The American Naturalist *155*, 346-364.

37. Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution *3*, 217-223.

38. ESA (2017). Climate Change Initiative Land Cover - https://www.esa-landcover-cci.org/ European Space Agency.

39. Hurlbert, A.H., and Jetz, W. (2007). Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. PNAS *104*, 13384-13389.

40. Jetz, W., Sekercioglu, C.H., and Watson, J.E.M. (2008). Ecological Correlates and Conservation Implications of Overestimating Species Geographic Ranges. Conserv Biol *22*, 110-119.
41. Therneau, T.M. (2015). Package ‘coxme’. coxme: Mixed Effects Cox Models. CRAN.
42. R (2016). R Core Team: a language and environment for statistical computing. R.F.f.S. Computing, ed. (Vienna, Austria).

Figure titles and legends

Figure 1. Similar threat status across assessed and data deficient amphibian species. Relationship between the predicted threat status of data-deficient ~~DD~~ and the assessed by the IUCN threat status of amphibians (solid line). Each point represents the mean (±se) threat status per family from Anura, Caudata and Gymnophiona. Diagonal dashed line is equivalence line with a slope of 1. See also Fig. S1.

Figure 2. Threat status across amphibian families. Mean threat status (±se) per family of fully assessed species (blue points) and predicted threat status for data deficient species (red points). Threat status ranges from 1 (Least Concern) to 6 (Extinct). Green asterisks represent significant differences ($p < 0.05$). Number on the right are the total number of species per family. Number in brackets is the percentage of data-deficient ~~DD~~ species per family. Dashed line is the threshold between threatened (right) or non-threatened (left) species.

Figure 3. **Phylogenetic distribution of threat.** Phylogenetic position and threatened status of assessed species (dark colours) and predicted threatened status of data deficient species (light colours and black dots). Grey and black outlines represent different families for the orders Gymnophiona, Caudata and Anura.

Figure 4. Regions most at risk. Global richness of non-threatened (A, C) and threatened (B, D) assessed (A, B) and data deficient species (C, D). Different colours represent number of overlapping species ranges. Non-threatened species: near threatened or least concerned. Threatened species: critically endangered, endangered, and vulnerable. See also Figs. S2 and S3.

STAR Methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Pamela González del Pliego (pgonzalezdelpliego@gmail.com).

METHOD DETAILS

Data

We collected data on 7,675 amphibian species from three orders and 74 families. This represents essentially all known, extant amphibian species as of 2018 (~98% of current species diversity as of 1st March 2018 [8]). Of these, 4,983 had some category of threat by the IUCN; 1,426 species were recognized but not assessed (category ‘d~~D~~ata-~~-~~deficient’~~ – DD~~); and 1,266 species had not been recognized nor evaluated by the IUCN. As we found similar amount of ecological and spatial data between the data-deficient ~~DD~~ and the not recognized species, the latter were also considered as data-deficient ~~DD~~ in our dataset.

Amphibian phylogeny

We used 100 amphibian phylogenetic trees sampled from the posterior distribution generated by [9] sampling 7,238 species. For our analyses, we excluded 62 species that have been taxonomically split or lumped in the interim and could no longer be matched to existing trait data with certainty. Therefore, from the total trait dataset, 7,176 species (from 3 orders and 74 families) could be placed in the phylogeny [9] representing ~92% of current species diversity (as of 1st March 2018 [8]).

Natural history traits

Natural history data representing 30 traits were initially collected for all amphibian species – using the database of [10] as a starting point. For the final analyses, we excluded traits that had more than 40% missing values and traits that had high collinearity ($R^2 > 0.80$). This resulted in a highly complete database of nine natural-history traits (traits definition and values in Table S3), whose strong importance for predicting threat status has been previously illustrated [10]. These traits include habit, fertilization type, reproductive mode, egg deposition site, parental care, forest association and body size, which is regarded as the second most important factor to determine threat status [10].

Spatial data

We included range size, because it is the largest contributor to extinction risk [10, 11, 17] and is used by the IUCN as a primary indicator to assess threat status. We obtained the geographical range of 6,489 species from the IUCN and AmphibiaWeb [1, 8].

For 684 of the data-deficient ~~DD~~ species, there were no spatial data. We thus collected all the coordinates available from the literature and from AmphibiaWeb (654 species; for list of species see Appendix 2, for spatial data see https://amphibiaweb.org, and https://mol.org). We then estimated the range size for these species by calculating a 10 km radius buffer around the available coordinates. For 405 of these species, we had only a pair of coordinates, thus the 10 km radius buffer was considered as the range size. If the species had two or more pairs of coordinates, we then draw a convex polygon across the points and then include a 10 km buffer around the polygon. However, if two points were more than 40 km apart, then two polygons were drawn. We decided to use a 10 km radius buffer as it represents the lower quartile of the range size distribution. This means that for most species

(75%) using a 10 km radius is a conservative approach, yet it might overestimate the range size of 25% of species (see supplementary information).

For the 7,143 species, we projected the geographical ranges to Mollweide equal-area projection and estimated the total area per species (km^2). We also calculated a matrix of geographical distances to reduce the configuration of the data to a series of pairwise comparisons that measures the distance between the centroids of each species geographic range, following [33]. Both the total area per species and the matrix of geographical distances were used to predict threat status (see 'Predicting threat status' section below).

The resulting database contained 7,143 amphibian species. From these, 2,069 species are deemed threatened (i.e. categories Vulnerable: 665 species; Endangered: 841 species; Critically Endangered: 541 species; and Extinct and Extinct in the wild: 22 species), 2,914 are considered non-threatened (i.e. Least Concern: 2,521 species; and Near-threatened: 393 species), and 2,160 are data-deficient~~DD~~. Of these, 23 families had no data-deficient ~~DD~~ species, and one family had only data-deficient ~~DD~~ species (Chikilidae: Gymnophiona). ~~The~~ Data-deficient ~~DD~~ species were comprised of the orders: Anura, 1,883 species; Caudata 148; and Gymnophiona 129. It is worth noting that a high proportion of the IUCN assessments for amphibians are now quite old [34], but they remain the only global assessments available to date.

QUANTIFICATION AND STATISTICAL ANALYSIS

Imputation of amphibian traits

The trait database, although highly comprehensive, still contained missing values, a common problem in comparative studies [17, 35]. To address this, we used the R package *Rphylopars* to impute the missing values of the trait database [35, 36]. *Rphylopars* contains tools for phylogenetic imputation of missing data, using the phylogenetic position of each species to impute new values under the assumption of strong phylogenetic signal [35]. However, if the traits do not show a strong phylogenetic signal, then the imputed values will have high variance. Thus, we first calculated Pagel's lambda (λ) for all traits using the *phylosig* function from the R package *phytools* [37]. Lambda values tend to range from 1 (strong phylogenetic signal) to 0 (no phylogenetic signal). We imputed values for all selected traits with a strong phylogenetic signal (>0.60; Table S3).

Human encroachment index

Following [7], we calculated an index of human encroachment across the range of each species as a measure of potential anthropogenic pressure. Specifically, we overlaid each species range map with a global land cover map providing data on anthropogenic transformations. We used the land cover map provided for 2016 by the European Space Agency [38] at 300 m resolution, aggregated at 1 km resolution, with land cover 'cropland' and 'urban areas' considered as human-dominated. We used a 1 km resolution, because the available expert range maps, when analysed at a high resolution, can show false presences and overestimate range sizes in a potentially biased way [39, 40]. This might constrain our results, yet much of the range overestimation will be captured by our Human encroachment index. This index ranges from 100 (range overlaps entirely with human modified land cover) to 0 (no overlap). In other words, this index is the inverse of the 'natural cover remaining' in each pixel.

Predicting threat status

Traits are frequently influenced by both spatial and evolutionary factors [33]. Therefore, predictions from our model were generated by using the trait values together with the degree of phylogenetic similarity among species (following [5]). The phylogenetic component allows for similar species to share similar levels of threat because they are evolutionarily related [5] and was defined by a variance-covariance matrix generated from the full phylogeny [5]. Traits can also be affected by spatial processes, because species that live in the same place tend to be similar to each other and experience similar environmental (particularly anthropogenic) factors [5]. The possible spatial component of trait variation was measured by a matrix of geographic distances, comparable to the phylogenetic matrix.

To model spatial and phylogenetic effects simultaneously, we obtained the proportion of variation of the threat status in assessed species attributed to the spatial (spatial fraction) and phylogenetic (phylogenetic fraction) component for each order (Anura, Caudata, and Gymnophiona) using the function *lmekin* (R package *'coxme'* [41]; see supplementary information) and using 100 phylogenetic trees from [9]. We used the spatial and phylogenetic variation fractions and obtained 100 spatial-phylogenetic matrices. For each order, we selected the traits with strongest explanatory power for species threat status using a generalized linear model approach (see Appendix 1). We then predicted the threat status of data-deficient ~~the DD~~ species using the generalized least-squares (GLS) approach described in [33] and implemented in [5]. We ran the models for each of the 100 spatial-phylogenetic matrices for each order.

The response variable in our model (threat status) is ordinal, yet to date no satisfactory method is available to address the phylogenetic and spatial covariance for ordinal responses [5]. The alternative would be an ordinal logistic model; however, this is rather complex and requires the estimation of a large number of parameters for both the fixed and random (phylogeny and spatial) components. We therefore modelled the response as a continuous variable ranging from ‘Least Concern’, 1, to ‘Extinct’, 6, retaining the ordinal nature of the IUCN scale. To validate our model, we randomly deleted the threat status for 20% (998 species) of fully assessed species per order and we predicted their threatened status using the GLS approach described previously. We then examined the prediction performance of the binary classification: threatened and non-threatened by comparing the real values versus the predicted values. To categorize a species as threatened or non-threatened we used ‘2’ as a threshold (i.e. if a species’ mean predicted threatened status is >2 we consider them as threatened, otherwise it is considered as non-threatened). We used a threshold of ‘2’ because it decreased by ~50% the number of species that were predicted non-threatened but were indeed threatened compared to a ‘2.5’ threshold. Also, it had similar predictive performance compared to the ‘2.5’ threshold (82% vs 83%). We also determined if threat scores were different between assessed and data-deficient ~~DD~~ species across families as well as orders, using an analysis of variance (‘*aov*’). All analyses were carried out in R version 3.3.1 [42].

DATA AND SOFTWARE AVAILABILITY

Datasets are available online as Appendix 1:

Data S1: Threat status of assessed by the IUCN species and predicted threat status for data deficient species.

Data S2. Spatial data collected for the 654 species.

Data S3: Predictor performance for Anura, Caudata and Gymnophiona. Related to Results: Model performance.

Supplemental information titles and legends

Supplementary information. Details on: a) estimating range size for species; b) lmekin function; c) taxonomy; d) predicted threat score for assessed species versus predicted threat score for data deficient species.

Figure S1. Predicted threat status of data-deficient ~~DD~~ versus predicted threat status from the validation analysis for assessed by the IUCN threat status of amphibians.

Figure S2. Global distribution of data deficient species richness.

Figure S3. Global distribution of assessed and data deficient threatened species in three tropical regions.

Table S1. Prediction percentage of binary classification of extinction risk for amphibians.

Table S2. Geographical distribution of the families with the higher abundance of data deficient species.

Table S3. Phylogenetic signal for each trait used in the analyses.

Appendix 1. Data S1: Threat status of assessed by the IUCN species and predicted threat status for data deficient species; Data S2: Spatial data collected for the 654 species; Data S3: Predictor performance showing the selected traits used for the final model to predict the threat score.