A dire need for better standards of data quality, transparency, and reproducibility in IUCN RedList assessments

Alice C. Hughes[1*], Michael C. Orr[2,3+], Ruben D. Palacio[4], Yan Xuan[5], Huijie Qiao[3*]

[1] School of Biological Sciences, University of Hong Kong *achughes@hku.hk

[2] Entomologie, Staatliches Museum für Naturkunde Stuttgart, Stuttgart, Germany.

[3] Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, China. * qiaohj@ioz.ac.cn

[4] Fundación Ecotonos, Cra 72 No. 13A-56, Santiago de Cali, Colombia.

[5] 7912 Heritage Palms TRL, Mckinney, TX 75070, USA.

+equal contributions.

Abstract

The IUCN RedList is the most extensive source of information on the global extinction risk including over 157000 species. The sheer scale of this initiative presents challenges in data standards and reporting, especially given that legacy issues may reduce accuracy. Here, we assess the bibliographic underpinnings of RedList assessments for five taxa with fairly complete assessments (four terrestrial vertebrate and one invertebrate group, including 41647 species). We assess the number of publications referenced, their age, their specificity, and use of primary data. Body-size and popularity are then explored as potential drivers of bibliographic trends. Disturbingly, many references are old and general (especially in smaller and less popular taxa), with many lacking specific references (e.g., only 1.3% of Odonata species have species-specific references). Public data are virtually never mentioned (GBIF is cited once in Odonata and Reptiles) and private databases are often cited. Furthermore, the use of data for mapping of species remains completely opaque. Better methods and standards are urgently needed for data inclusion, wider participation, mapping, and data citation if the RedList is to fulfil its remit.

Introduction

We face staggering ongoing losses of global biodiversity and our conservation actions can only be as good as the data informing them (Isbell et al., 2023; Jaureguiberry et al., 2022). Since 1964, the International Union for the Conservation of Nature (IUCN) RedList has become the major tool for assessment for extinction risk assessments, and is now the basis for prioritisation worldwide (Palacio et al., 2023). In fact, the IUCN RedList provides the only species-level indicator within the Kunming-Montreal Global Biodiversity framework (Hughes & Grumbine 2023). The IUCN has undoubtedly provided a useful tool for assessments (Betts et al. 2020), and nurtured networks of specialists around certain taxa (eg. State of the World's Amphibians: The Second Global Amphibian Assessment). However, it has also been recognised that the RedList assessments often fall far short of what is needed, and can hamper conservation efforts if used uncritically for conservation prioritizations (Palacio et al., 2023). The methods and consistency of the RedList species maps have previously been assessed (Hughes et al., 2021a), and shortfalls in inclusivity and representativeness (Hughes et al., 2021b), with consequences for the accuracy of species assessments (Palacio et al. 2021). However, there have been no assessments of the reproducibility of approaches, the standardisation of assessments, or the traceability of information used. Not only is this problematic from a scientific standpoint, it can also propagate errors across assessments if overly reliant on prior erroneous information, especially when the origins or age of such information are untraceable.

Whilst it is understandable that information for many species assessments comes from experts within specialist groups, a lack of documentation of such information impedes traceability, precludes assessments of uncertainty and timeliness, and dramatically limits comparability between assessments for different taxa or regions. FAIR data-standards now exist to ensure that data are findable, accessible, interoperable and reusable (Wilkinson et al., 2016), and yet interoperability is dependent on consistently applied standards. This is important, as whilst unpublished expert insights may be invaluable, tracing their contributions is just as important as it is for peer-reviewed publications, enabling greater transparency, facilitating material updates, and providing clear metadata that ultimately enhances the standards and quality of information included within the RedList and downstream research.

Here, we explore the bibliographic data underlying IUCN RedList assessments, exploring how many publications are cited for each species, when these resources were published (to assess how updated they may be), their specificity, and how the number of references varies within and between orders based on taxonomy, body size, and popularity. We expected that larger and more popular taxa would be better studied, while smaller and less popular groups would be more poorly documented in assessments, but that these relationships may vary within different orders where overall "popularity" is low. We also wanted to determine the empirical underpinning of distribution data and other data-sources used within assessments, such as the use of data from GBIF. Following from this assessment, we make recommendations for better standardizing IUCN RedList assessments, to ensure their transparency, reproducibility, and accuracy. Following these new standards would ensure not only the accuracy and reliability of RedList assessments, but also the representativeness of the hundreds of publications leveraging the results of these assessments for global scientific and policy purposes.

Methods

IUCN documentation was first downloaded as JSON format (Supplemental code 1). For counts per time-period we removed references referring to former RedList assessments, and then calculated the number of references before 1990, between 1990-2000, from 2001-2010, and post-2010. Analysis was then conducted to explore, between and within each taxon, how the number of references used varied per species. Overall patterns were determined as well as their relationships to body-mass and popularity within and between orders and families (see below).

As larger species are often assumed to be more charismatic, and therefore have more research (Tensen 2018), we assessed if there was a relationship between body-weight and mean publications within birds and mammals. For body-mass data, various sources were combined to determine the mean mass for each taxonomic grouping examined (Myhrvold et al., 2015; MammalBase, 2023; Soria et al., 2021; Wilman et al 2014; Jones et al., 2009). Genus was the highest level for which an average was calculated as weights were not available for all species, (especially for diverse but small-bodied taxa such as bats, rodents, and insectivores as well as various passerine birds).

To assess popularity, we used Google Trends and assessed the level of online attention each order within each class had received globally throughout 2023 based on averaging weekly data which uses global search history for the common-names (https://trends.google.com/trends/). To do this, we used common-names (excluding pets (cats, dogs, rabbits), livestock, pests), ensuring the search recognised them as an animal rather than a general search-term; all keywords are available in Table S1. First, we determined which were the most popular groups to provide a comparator, as scores are calculated relative to the most popular term within a search. In mammals, elephants were the most popular, so this was retained in all mammal searches, likewise penguins were used in all bird searches. We used the common-names of most families within the groups (see Table S1), downloaded the scores as CSVs, then merged all scores and calculated the mean and standard deviation within each score then assessed the relationship with the mean number of references via regression.

We also assessed the mean number of publications per species, per realm for each order. We downloaded all species assessments in XLSX format from the RedList website, connected them by species name within R (R Core Team, 2024), then calculated the mean and total number of publications per species within each realm (using WWF 2017 realms). We ran analysis both for all species, and then for only species found within a single realm, as interpreting outcomes may be challenging, so only species found within a single realm were included. The average numbers were then tabulated to calculate how the averages may vary within orders between different realms.

In addition, we tabulated the exact references for each species to assess the total number of references included in each class (Supplemental Code 2). The total number of references was calculated for each class once duplicates were cleaned and removed, which was a very challenging process as the IUCN does not standardise reference formatting thus DOIs had to be found to ensure references were not counted multiple times (Supplemental Code 3). We then used the "=ISNUMBER(SEARCH(" with a substring of the referenced genus to determine if the genus was noted within each reference for each genus, then calculated overall coverage levels. For many "charismatic" mammals, common-names were often used rather than Latin names, thus this process was repeated with common-names for large primates (Apes) and carnivores, as a check for this specific phenomenon. From de-duplicated references we collates the use of certain key terms including the number of GBIF references, listings of use of major data repositories (birdlife, ebird, reptileDB), the use of data from various ministries, and the number of references from books. It was also noted on visual inspection that many assessments had very general listings, such as books of regional faunas, or the IUCN assessment guidelines, thus as a further metric of specificity we calculated how many assessments each reference had appeared in, and averaged this for each taxa. Very high values would mean that the average number of times references had been reused was very high, whereas low values indicated little reuse.

Results

Mammals: Within the mammals, the number of publications per species varied from the highest with a mean of 90.3 per species in Proboscidea (elephants, three species) to the lowest in the rodents with 4.2 publications, followed by Didelphimorphia (American opossums) at 4.4, then 4.5 in Dermoptera (flying lemurs, two species; see supplemental data). Only 7/27 orders had more than 20 publications on average per species (generally larger taxa), whereas 16 had ten or less. When total publications per taxa was plotted relative to number of species, the only orders (with >5 species) with disproportionately high numbers of publications were Carnivora, Primates, and Artiodactyla (Figure 2a). At a family level in mammals, smaller-bodied taxa seemed particularly under-studied, with the rat-kangaroos (Hypsiprymnodontidae) and honey-possums (Tarsipedidae) averaging one paper per species, followed by bumblebee bats (Craseonycteridae), feathertail gliders (Acrobatidae) and beavers (Myocastoridae) with two. The most studied groups also frequently had a higher recent proportion, and very few pre-1990 publications (Figure S1). When different time periods were considered, 78% of publications for elephants were post-2000, whereas conversely 50% of publications were pre-2000 for the flying lemurs (Figure 1a, Figure S1). When plotting mean number of publications against body weight for all taxa under 1kg mean weight (Figure S3), there was an average of under 6.9 publications per species, those between 1-5kg had under 9.3 (except primates, which had 21), whereas larger taxa had far more, with a significant relationship between bodysize and mean number of references (family: R^2 0.08666, $P<0.0001$; Order R^2 0.6630, $P<0.0001$). Within mammals, there were more publications than other areas: in carnivores high-income regions (Nearctic, Palearctic, Australia, as well as Antarctica), Palearctic for Perissodactyla. The Nearctic realm had more publications than elsewhere for Lagomorpha, Eulipotyphla, and Rodentia, showing that higher-income economies typically exhibited better taxonomic coverages (Figure S2). Within mammals, the average popularity correlated very strongly with the mean number of publications per species within each order (R^2 0.4414, $P<0.0002$, Table S1). We also assessed the mean number of species assessments each reference had appeared in to provide a metric for reference reuse (and generality). Within mammals at an order level, this varied from a mean of 10.84 assessments in Sirenians and 11.1 in elephants. To further contextualise this, elephants had 207 of their 247 references only used in their assessments, and only 8 appeared in over 10 assessments (Table S2). Similarly, for Sirenians 370 of the 391 references only appeared in a single species assessment, and only 12 references were used in more than 10. Conversely the mean number of assessments each reference appeared in was 447 in rodents and 365.5 in Eulipotyphla. Patterns are stronger at the family level, ranging from 2.4 for the narwals (Monodontidae) and 2.6 in River dolphins (Platanistidae), to 879.8 in beavers (Myocastoridae).

Odonata: Dragon- and damselflies were examined at a family-level as all fall within order Odonata. Nine families had a mean of two publications per species, and only one (Philogangidae) had more than 10 (11), with a further three having more than five of the 41 families considered (Figure S5). In total, 29.4% of references were from before 1990, whereas 22% were from post-2010, but this varied considerably by group, with 67% of data for some families (e.g Rimanellidae) dating from before 1990, and four families having at least 40% of references dated pre-1990 (Figure S4). In terms of regional differences, the Nearctic had far more publications on average (9.6) than other regions (with the most in the Palearctic region at 5.1, and most other regions falling below 5 (Figure S6)). In terms of reference reuse, the monotypic family Neopetaliidae had the least, with references used an average of 12 times, followed by Rimanellidae at 19 times. At the other end of the spectrum are the damselflies of Isostictidae, where references were reused in an average of 202 assessments (Table S2), followed by Lestoideidae damselflies at 190, also notably both these groups are Australasian. Overall, the average level of reuse was 126 assessments per reference.

Birds: The penguins (Sphenisciformes) were dramatically better documented than any other group, averaging 41.3 references per species, with 14 in the next-best Phoenicopteriformes (Flamingos) and Otidiformes (Bustards) (Figure 2b, Figure S7). At the other end of the spectrum, some smaller orders averaged only one publication per species, including the Seriemas (Cariamiformes), Mousebirds (Coliiformes), Cuckoo-rollers (Leptosomiformes), and Hoatzins (Opisthocomiformes). At a family-level, 34 of the 247 families had a mean of only one paper per species, most of which were small (and evolutionarily distinct) families, 11 of which were monotypic. Penguins were also the group with the most recent references, at 51% since 2010 (Figure 1b, Figure S8). Weight was a significant correlate of number of publications at an order level (R^2 0.3631, $P<0.0001$). However, this is largely due to low coverage of very small species, and larger bodied groups did not necessarily have high coverage, for example Struthioniformes (mean of 6.8 publications per species) despite large body size (possibly in part because tinamous as the IUCN uses old taxonomy and includes four orders within Struthioniformes (IOU 2024; Clements et al., 2023)) (Figure 2f). Clearly, other factors played a much greater role in determining taxonomic coverage. There were also geographic patterns in the average numbers of publications within various orders, with developed economies having many more publications for three widespread orders than in the rest of their range. Seven orders had more in the palearctic, with up to 66 publications in the Otidiformes (Bustards) (Figure S9). Popularity was also a significant correlate of the mean number of publications per species (R^2 0.2423, P<0.0023, Figure 2c; Table S1).

Reference reuse in birds varied from a mean of 4130 in the Opisthocomiformes (Hoatzins), this alarmingly high number is because the one species only has two references (Table S2), one of which is the IUCN 2016 guidelines, and the other a generic text on Neotropical birds from 1996; both of which are used across thousands of assessments. Unsurprisingly, once again the penguins were the best served with a mean of 71.5 (highly citied IUCN guidelines appeared in virtually all assessments). For example, the Penguins had 582 references only cited in a single assessment, 75 used in between 2-10 assessments, and only 13 with more than 10. Patterns at a family level showed similar trends, with five families equally poorly documented as Hoatzins, all with identical references. Conversely, birds like the Stitchbird (Notiomystidae) (*Notiomystis cincta*) have the least reuse, averaging 19.2. These families, such as Australian shrubbirds (Atrichornithidae) and the New Zealand Stitchbird were also some of best studied with 25.5 publications and 24 publications each per species, and most references not used in other assessments (35/55 and 20/24 unique).

Reptiles: Crocodylia had exponentially more documented research than any other group, averaging 62 publications per species, relative to only 15 in Testudines, nine in Rhynchocephalia, and only four in Squamata (Figure 1c, Figure S11). At a family level within reptiles, the elapoid snakes (Pseudoxyrhophiidae) were the least well studied, with only 1.8 publications per species, followed by Alopoglossidae lizards at 2.05. At the other end of the spectrum were the Gharials (Gavialidae) with 151 publications, followed by Cheloniidae tortoises at 133. In total five families (2 Testudine and 3 Crocodylia) averaged over 50 publications per species, whereas 78 had ten or less (of the 95 total families). For most groups, around 20% of references were from before 1990, though there were differences in the most recent publications varying from 34% of publications in Crocodylia data coming from post-2010 to only 22% in Rhynchocephalia (Figure 1c, S13). In Squamata, Nearctic species and Palearctic species (8 and 10 publications per species, respectively) had more than the rest of their ranges (4-5 publications). In Testudines, Afrotropical was the best represented (15 publications), followed by the Nearctic (14), with Australasian (9.5) and Palearctic (8.2) the least represented. Crocodylia were well-studied across their ranges, but the highest level was in the Neotropics (75) followed by the IndoMalaysian realm (64) (Figure S12). The impact of mass was not assessed due to the small number of orders and clear phylogenetic differences in level of study. Squamata was the most popular but the least well-studied (but varied between families), whereas Tuataras were the least popular and the second-worst studied, and Testudines were the second-most popular and second-best studied (Table S1). Reference reuse echoed other trends, with Squamata showing the most reuse at 145 assessments per reference to 8.3 in the case of Crocodylia (Table S2). At a family level, the disparity unsurprisingly grows, varying from 335 in the worm lizards (Cadeidae) down to only 4.5 in the leatherback turtles (Dermochelyidae).

Amphibians: Amphibians comprise three orders, with Caudata having more publications on average (10.5 per species; Figure S14), but approximately half of publications for all three groups fell after 2000 (for Caudata, 29% of all references are post-2010, 26.7% for Anura, but only 20.7% for Gymnophiona) (Figure 1d, Figure S14). At a family-level, there were also very different patterns across groups; Odontobatrachidae was the least studied with only two publications on average per species, ranging to a mean of 37 in Pelobatidae followed by 36.5 in Cryptobranchidae (Giant Salamanders). However, of the 77 orders, only seven averaged over 20 publications per species. Whilst most groups have under 50% of publications from pre-1990, Conrauidae had 54.4%, and Sooglossidae has 52. Within each order, high-income economies had more publications on average within most groups (Figure S15). The impact of mass was not assessed due to the small number of orders and clear phylogenetic differences in level of study. No relationship between number of publications and popularity was found in amphibians, though Caecilians (Gymnophiona) was lowest for both mean number of publications and popularity (Table S1), Anura was the most popular group, but Caudata was better studied (possibly due to high levels of research in the Nearctic). For Amphibians reference reuse at an order level varied less than for other groups, ranging from a maximum of 190.6 in Gymnophiona to 112.6 in Caudata (Table S2). Like the worm lizards in the case of reptiles, the least specific references in amphibians were in the Indian caecilians (Chikilidae) at 390.6, to the Seychelles frogs (Sooglossidae) at only 9.2.

Trends between groups

Analysis of unique references highlighted the lack of standardisation between specialist groups, with certain taxa relying more on overarching references (most references were used multiple times, in Odonata unique references represent 15% of total use, followed by 26% in Amphibians, 29% in both reptiles and birds) and the least reuse in mammals at 48% retention (Table S3). The level of older references was similar, with 39% of Odonata references coming from before 1990, whereas in birds only 7.3% were pre-1990 (mammals 18.9% Amphibians 20.2 and reptiles 23.4%). Amphibians, birds, and mammals showed a steady increase in the total number of references for each successive period, whereas for Reptiles and Odonata the numbers fluctuated and showed no overall trend. Looking at the overall number of references, Odonata had the equivalent of under one unique reference per species (0.85), increasing to 1.71 in birds, 2 in reptiles, 2.38 in amphibians, and 5.73 in mammals.

Breaking this down further, the sources of references also varied (Table S3). Mammal assessments cited the greatest number of former IUCN assessments (1043), whereas Odonata had the least at 23. Data from various ministries were also used in some assessments, with over 332 referencing ministries in mammals, 169 in reptiles, 141 in birds and 142 in Amphibians, yet only 25 in Odonata. Specialist websites were also used, like Amphibiaweb (167 for Amphibians), ReptileDB and the Australian Reptile Database (203 for Reptiles), and eBird for birds (221). BirdLife data were also used for all groups, but surprisingly GBIF data were almost never referenced, with some references ironically only citing the GBIF citation guidance rather than sourceable DOIs. In total, Odonata referenced GBIF once, reptiles once, reptiles 8 times, amphibians 10 times, and birds 326 times. In addition, books and reports were frequently used, with upto 1903 used for mammals. However, private and personal databases were frequently used with 342 separate private sources in mammals, and whilst there were less different private sources in Odonata (58), there were 592 citations of them overall; highlighting that these private sources are the main source of information for these species.

Obviously, a specialised reference is likely to contain much more assessment information than a short general description within a generalist review, or encyclopaedia of mammals or birds of a region (Table S4); thus assessing both specificity and reuse helps understand how much detailed information likely went into each review. The level of specificity also varied, based on the number of reference titles directly listing the species being assessed (Supplemental Data 3). Odonata had the lowest level of specificity, with only 0.3% stating the genus, this was followed by birds (15%), reptiles (20.5%), Amphibians (21.5%), and then mammals (24.5%). In addition, the number of genera not explicitly referenced varied between classes. At the highest, 94.3% of genera and 98.7% of species were never explicitly noted in the reference title in in Odonata, and 54% of genera and 75% of species in birds. On the other end of the spectrum, only 16% of genera and 30% of species were never referenced in titles in amphibians, similarly 20% genera and 29% of species for mammals, and in reptiles (31%, and 37%, respectively). However, for mammals, the level of coverage was actually considerably better than the scientific names alone show in terms of the number of references citing the specific species under assessment (Table S4), as common names were used rather than Latin names for many "charismatic" species. For example, for the Pandas and red panda only 24% of its references used the Latin name, but a remarkable 80% referenced the common-name. Similar trends existed across carnivores, and primates, where about 30% of references listed the species name, and 60-85% the common name (Table S4). Mammal referencing was particularly good for large-bodied taxa, but smaller-bodied taxa assessments were dominated by more general references.

Discussion

The IUCN RedList has been developed to facilitate the assessment of extinction risk to enable successful management intervention and policy. Clearly, the accuracy and timeliness of assessments is crucial for their effective use and interpretation, and yet here we find a lack of transparency in data use, and the "popularity" of most groups is the best indicator of the level of detail in species assessments. The importance of reproducibility and traceability within science has been recognised in recent years, and yet the RedList clearly lags, especially when it comes to the approaches used for mapping species (which is a key part of assessments), and ironically assessing threats. For example, whilst GBIF is mentioned in the useful external links on the RedList website, virtually no species assessments cite the DOIs in their repositories (a total of ten links for mammals, reptiles, Odonata, and amphibians combined, with more frequent use in birds). Thus, if species distribution data are used in maps, those data are untraceable, consistent with a study showing that nearly 50% (25–46%) of species points fall outside their expert-designated ranges (Hughes et al., 2021a; Li et al., 2019) and that the overall accuracy of these maps is consistently lower than in data-driven approaches (Palacio et al. 2021). Furthermore, whilst recent publications stated that IUCN maps were consistent with data-driven approaches (Aronsson et al., 2024), this is not in fact the case as such publications failed to reflect landcover (and habitat is a major driver of species distributions), used a coarse resolution, and failed to account for bias in distribution data (Hughes et al., *In Press*). Accounting for biases in point-based data or modelling of ranges is critical to obtain informative and accurate range data, and in any instance, clear citations of data-sources is crucial to assess data uncertainty and provide transparency needed to utilise data outputs, and analysis consistently shows major inaccuracies in IUCN mapped ranges, likely as a consequence of variable use of primary data.

Furthermore, private databases are cited in some cases (58 for Odonata, cited over 300 times), inhibiting replicability or assessment of representativeness. This is problematic, as key data are likely being ignored as there is no mechanism to add and cite such data, and conversely inaccurate data within assessments is hard to flag by those outside the specialist group. Again, although expert opinion can be invaluable, the role of empirical data relative to expert knowledge must be clear. As IUCN RedList data are frequently used for mapping conservation priorities, understanding their limits is critical, as current approaches have led to inconsistencies between specialist groups and the frequent use of political barriers to delimit species ranges (especially between regions which may not collaborate extensively, such as China and Southeast Asia), hindering the added value of these maps for developing spatial priorities and targets. Clearly the underlying data of IUCN RedList assessments should follow the FAIR principles (Findable, Accessible, Interoperable, Reusable) whereas at present much data for assessment, and virtually all data for mapping species ranges falls outside these criteria (Dunning et al., 2017).

Families and orders perceived as more charismatic and popular (large mammals, penguins), typically had more publications, more recent publications, and often higher levels of specificity within publications. Species coverage was higher for larger and more charismatic groups, and reference reuse was far lower. Alarmingly for some groups, the entire bibliography was sometimes limited to small numbers of generic references, and no information on where any species-specific information may have originated from. Levels of research vary several orders of magnitude within groups, with species such as elephants, penguins, and gharials including many more publications than any other families within their classes, and mammals having many more publications per species than any other group. In attempts to justify the greater attention on larger species ascertains are sometimes made that “larger species are evolutionarily distinctive”. Yet many other evolutionarily distinctive and monotypic groups (e.g., Dermoptera) at least as evolutionarily distinctive than the best-studied taxa, where there are upto 23x more publications per species (e.g., elephants). Similar patterns are present in other groups; many of the least studied bird, reptile and amphibian families are monotypic, thus, more attention is clearly needed for overlooked taxa, and distinctiveness is obviously not a good argument for the observed disparities in attention. This underscores the subjectivity of prioritisation approaches, where despite the frequent assertions in defence of the amount of attention some species receive, these are not supported by more comprehensive or objective analysis. As popularity, rather than threat or evolutionary distinctiveness, is most associated with the level of knowledge and effort in assessment, efforts are needed to counteract this. Furthermore, many taxa are much better studied in higher-income parts of their ranges than in developing economies. Evidently, more work is needed to upgrade assessments of neglected taxa and regions. Our results demonstrate an urgent need to counterbalance biases, and highlight that the major NGOs such as WWF that often fund IUCN specialist groups should reallocate resources to ensure that all taxa get an adequate level of attention irrespective of their wider popularity.

Here, we show that not only are major gaps in standardisation and transparency across assessments, but that popularity and region where taxa reside are primary drivers of data availability, and the degree of "specificity" of citations in many taxa. Furthermore, the IUCN RedList purports to detail threats to species, but the resources used are of variable value. Many references are old, general, or focus on species taxonomy rather than evaluating threat, and for the majority of taxa there are not references that would enable population-level assessments. Clearly, efforts are urgently needed to overcome these biases, and to develop more standardised and temporally explicit data across taxa. Furthermore, the overreliance on statements from experts, with no meta-data on what information is sourced from where or when, makes it impossible to gauge the accuracy or uncertainty in assessments. These issues are compounded by the reuse of prior IUCN assessments, as any inaccurate information likely remains in place, and for most species it is impossible to trace and assess if assessments are still (or ever were) accurate reflections of species threats or distribution.

Whilst the IUCN has come a tremendous way since it began to develop and share species assessments, since the initiative started in 1964, data availability has increased exponentially, and it is time for updated processes that reflect new techniques, leverage new data, and clearly share it based on FAIR standards, all enabling more effective and appropriate use.

Recommendations

The IUCN serves a pivotal role in global conservation prioritisation, and recognising elements that are not in line with current approaches is essential for the effective fulfilment of this role. The use of empirical data is crucial to assessing the veracity and timeliness of inputs, and thus ensuring data is traceable and reliable, as well as timely is essential if the IUCN RedList is to genuinely enable it to provide the data needed to monitor biodiversity and develop priorities.

Accurate data are key to accurate mapping and assessment, yet we see virtually no evidence that GBIF data are used for mapping species ranges, with virtually no datasets cited (with the partial exception of birds). This is especially concerning, as in addition to the AOO and EOO maps used within species extinction risk assessments, each assessment must generally include a more detailed map of the species range. Maps created as part of assessments must be based on the best available data and cite data DOIs to be reproducible (possibly via models which could then be checked and annotated by experts to ensure they accurately capture species ranges, without artifacts such as political borders Hughes et al 2021a).

How data are used and recorded should be standardized, including not just pipelines but also even reference formatting. Automating a system of reference input, as most journals do today, could both enable clearer referencing standards and reduce effort. Similarly, some groups used private databases, and these should clearly be made transparent, and data made available to enable clear and accurate mapping of species ranges, while also enabling validations. Increased data availability for poorly-known groups, could enable the automation of assessments (and validation; Orr et al 2022), and identification of gaps, enhancing the ability to assess neglected groups and identifying areas where work is most urgently needed.

Whilst there will inevitably be resistance to updating IUCNs approaches to RedListing, all methods must evolve to reflect new data and technologies. We suggest that with new guidelines, trainings could be developed to streamline the process of standardisation, enhanced approaches of mapping (such as the use of species distribution models), and details of meta-data recording to provide the necessary transparency, and that these approaches could be included as species assessments are updated. As versions of the IUCN guidelines are also dated, noting which assessments reflect the new guidelines would also be simple as assessments are progressively updated, enabling the evolution of IUCN approaches to provide the dynamism needed to accurately assess threats and map species ranges across all scales. By making assessments fully versionable, biases and gaps could be better accounted for.

We also recognise that there is a lack of transparency on how assessments are conducted, or how to ensure expertise is adequately reflected within specialist groups. Thus, some groups may be more academic, and others representing NGOs as well as having different balances of national and international expertise. This is reflected in assessments, with some specialist groups relying much more on assessments from ministries or governmental departments. Groups like the penguin specialist group also have well maintained websites, and likely clear mandates and guidelines that may underlie the good practices underlying their assessments. More explicit best practice guidelines from successful specialist groups would enhance the quality of data for less well-studied groups. Use of data from ministries varies from very high levels in birds (especially USA) and amphibians (especially Japan) to lower levels for others. This clearly relies on the work of small numbers of individuals within specialist groups, but by standardising approaches and bettering information transfer between specialist groups, better data could be accessed from more regions. Better tracking and interoperability could be facilitated by a data submission portal, where experts could submit species-specific information, with both observations and the source (either from literature or personal observation). This would enable expert knowledge, for example species in trade, to be added to RedList assessments, provide access to publications in local languages, and provide a means of connecting people (particularly nationals of countries that lack UN languages) into specialist groups. Furthermore, the potential accuracy and modes of validation could be verified, providing another citable source of information directly feeding into assessments.

The RedList is expected to continue its role in providing a global index of species vulnerability to extinction, so it should follow similar standards and expectations in data quality, transparency and analysis to other forms of science. Outside experts should be readily able to comment on assessments alongside data uploads, to input their updates and caveats, and such information could also be noted through a centralised portal. Not only would this allow identification and flagging of any misleading information, but it would also provide a mechanism for detecting new threats, and likely a more accurate indicator of species threatened by trade (or other specific threats), especially for specialist groups generally composed of international academics (which may not observe such information). To make best use of the RedList, to monitor trends in biodiversity, and appropriately allocate resources, it must reflect the best knowledge in a traceable and open framework. Given that species assessments are conducted periodically to assess changes in species statuses, we have the opportunity to then update approaches to better reflect more diverse data, and to make the use of those data transparently. Whilst it is almost guaranteed that there will be resistance to such initiatives by some who feel it is too much work, the regular updates of assessments means a transition could not only be simple, but is also urgently needed for the RedList to effectively fulfil its mandate. By doing so, the RedList can provide a much more accurate and data-driven approach to effectively monitor, target, and conserve species.

Box 1. Recommendations of updating RedList guidelines and standards.

- Provide better metadata, evidence from experts must be cited and dated to ensure data used can be accurately dated temporally. Versioned assessments should continue to be archived, with the versions of guidelines applied noted to enhance retroactive replicability of prior analyses and facilitate use.
- Species maps should be based on data-driven approaches such as modelling, data should be cited using a GBIF DOI for transparency and reproducibility. The only exceptions should be for species threatened by trade and justified.
- Cautious preliminary assessments should be automated for groups via methods such as machine or deep learning, to provide a basis for experts to build on.
- References should use a standard system and format for easy use, access and updating.
- Updated guidelines to liaise more effectively with government departments should be created to enable representative data to be accessed from all regions.
- A data submissions portal should be created for the submission of relevant data (with evidence) for observations of distribution, trade, etc., alongside flagging mechanisms for erroneous or incomplete assessments in need of updates.
- Better indices of population-level data should be added for assessments of population trends.
- These changes should be finessed by a taskforce then implemented with the reassessments of each species (as most species are assessed at least decadally).

References

Aronsson, H., Zizka, A., Antonelli, A., & Faurby, S. (2024). Expert-based range maps cannot be replicated using data-driven methods but macroecological conclusions arising from them can. Journal of Biogeography. DOI: 10.1111/jbi.14847

Clements, J. F., P. C. Rasmussen, T. S. Schulenberg, M. J. Iliff, T. A. Fredericks, J. A. Gerbracht, D. Lepage, A. Spencer, S. M. Billerman, B. L. Sullivan, and C. L. Wood. 2023. The eBird/Clements checklist of Birds of the World: v2023. Downloaded from https://www.birds.cornell.edu/clementschecklist/download/

Dunning, A., De Smaele, M., & Böhmer, J. (2017). Are the FAIR data principles fair?. International Journal of digital curation, 12(2), 177-195.

Hughes, A. C., Orr, M. C., Yang, Q., & Qiao, H. (2021a). Effectively and accurately mapping global biodiversity patterns for different regions and taxa. Global Ecology and Biogeography, 30(7), 1375-1388.

Hughes, A. C., Orr, M. C., Ma, K., Costello, M. J., Waller, J., Provoost, P., ... & Qiao, H. (2021b). Sampling biases shape our view of the natural world. Ecography, 44(9), 1259-1269.

Hughes, A. C., Qiao, H., & Orr, M. C. (2021c). Extinction targets are not SMART (Specific, measurable, ambitious, realistic, and time Bound). BioScience, 71(2), 115-118.

Hughes, A. C., & Grumbine, R. E. (2023). The Kunming-Montreal global biodiversity framework: what it does and does not do, and how to improve it. Frontiers in Environmental Science, 11.

Hughes, A. C., Dorey, J.D., Bossert, S., Qiao, H., Orr, M.C. (*In Press*) Big data, big problems? How to circumvent problems in biodiversity mapping and ensure meaningful results. Ecography. DOI: 10.1111/ecog.07115

IOU (2024) Working Group Avian Checklists Update (January 2024) https://www.internationalornithology.org/working-group-avian-checklists

Isbell, F., Balvanera, P., Mori, A. S., He, J. S., Bullock, J. M., Regmi, G. R., ... & Palmer, M. S. (2023). Expert perspectives on global biodiversity loss and its drivers and impacts on people. Frontiers in Ecology and the Environment, 21(2), 94-103.

Jaureguiberry, P., Titeux, N., Wiemers, M., Bowler, D. E., Coscieme, L., Golden, A. S., ... & Purvis, A. (2022). The direct drivers of recent global anthropogenic biodiversity loss. Science advances, 8(45), eabm9982.

Jones, K. E., et al. PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals: Ecological Archives E090-184. Ecology 90, 2648–2648 (2009).

MammalBase (2023). MammalBase — Database of traits, measurements and diets of the species in class Mammalia. Licensed under CC BY 4.0. Retrieved [download date] from https://www.mammalbase.net. http://doi.org/10.5281/zenodo.7462864

Myhrvold, N. P. et al. An amniote life-history database to perform comparative analyses with birds, mammals, and reptiles: Ecological Archives E096-269. Ecology, 96, 3109–3109 (2015).

Orr, M. C., Hughes, A. C., Costello, M. J., & Qiao, H. (2022). Biodiversity data synthesis is critical for realizing a functional post-2020 framework. Biological Conservation, 274, 109735.

Palacio, R. D., Abarca, M., Armenteras, D., Balza, U., Dollar, L. J., Froese, G. Z., ... & Hughes, A. C. (2023). The global influence of the IUCN Red List can hinder species conservation efforts. Authorea Preprints. https://www.techrxiv.org/doi/full/10.22541/au.169945445.50394320

R Core Team (2024). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Soria, C. D. et al. COMBINE: a coalesced mammal database of intrinsic and extrinsic traits. Ecology, 102, e03344 (2021).

Tensen, L. (2018). Biases in wildlife and conservation research, using felids and canids as a case study. Global Ecology and Conservation, 15, e00423.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3(1), 1-9.

Wilman, H., et al. (2014) Traits 1.0: Species-level foraging attributes of the world's birds and mammals: Ecological Archives E095-178. Ecology, 95, 2027-2027.

WWF (2017) RESOLVE. https://ecoregions.appspot.com/

Figures

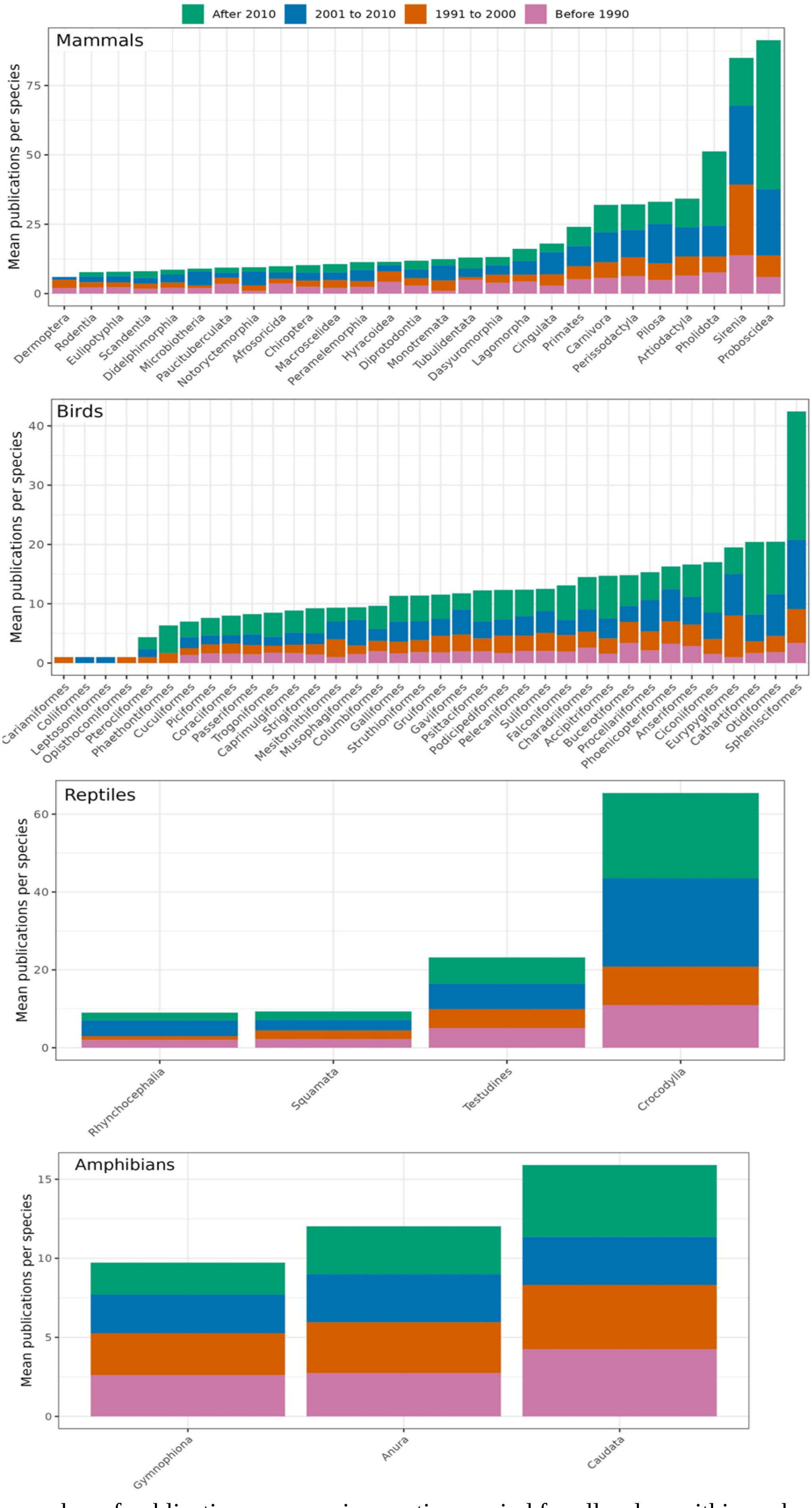

Figure 1. Mean number of publications per species per time period for all orders within each of the four vertebrate classes examined.

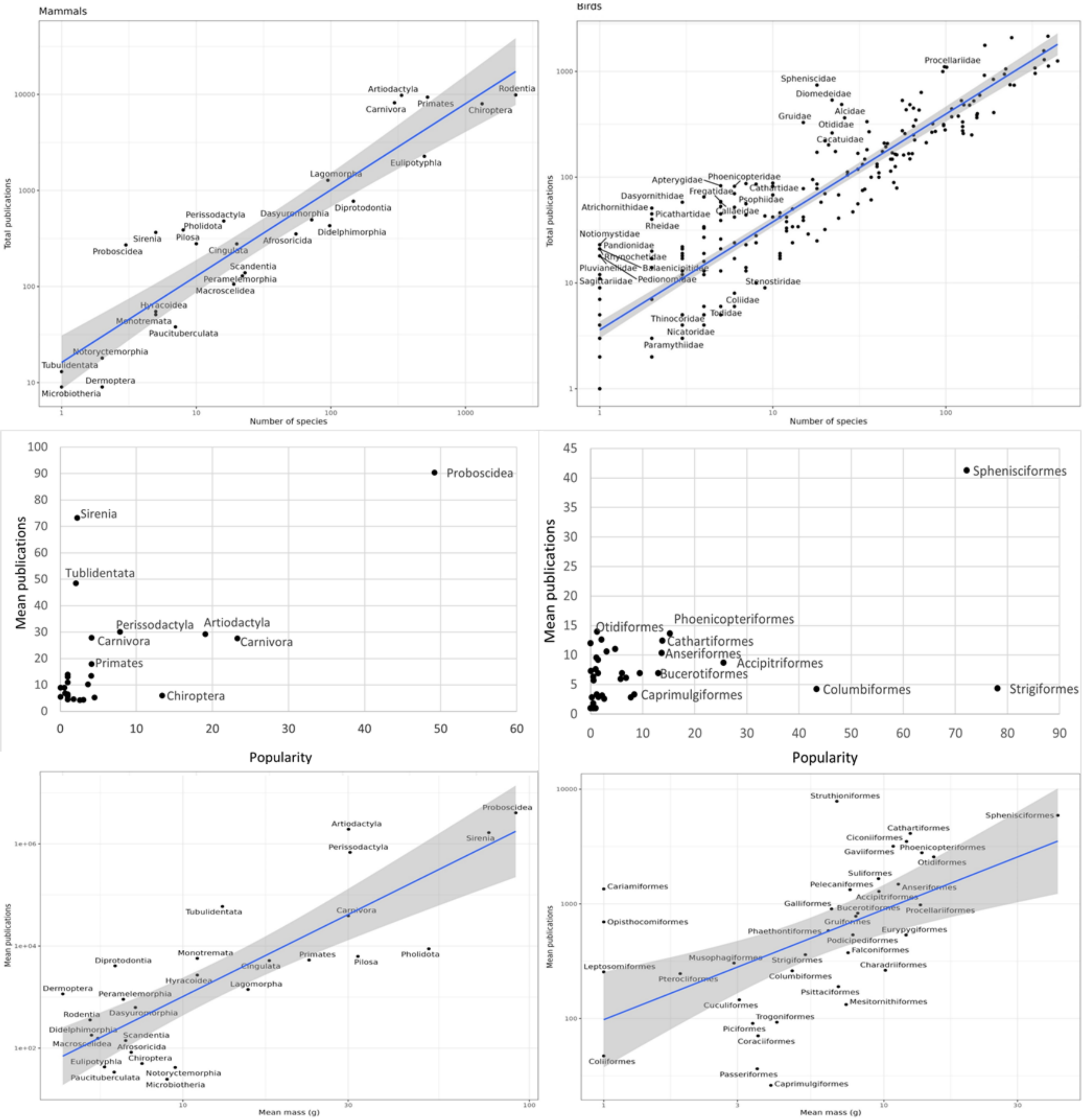

Figure 2. Trends between publications and various correlates for mammals and birds. A. Total publication number vs number of species for mammal orders. B. Total publication number vs number of species for bird families (family was given as trends were clearer in this case). C. Mean publication number vs popularity for mammal orders. D. Mean publication number vs popularity for bird orders. E. Mean publication number vs mean bodyweight for mammal orders. F. Mean publication number vs mean bodyweight for bird orders.

Supplements

Supplemental tables

Table S1. Mean popularity scores of each order based on googletrends analysis, and keywords for each taxa.

Table S2. Reuse of references within each of the five taxa, showing the mean number of assessments each reference appeared in for each order and family.

Table S3. Summary of general trends, and use of key datasets, such as private databases and GBIF data within each of the five groups.

Table S4. Specificity of each assessment, based on the number of species directly referenced within their associated references.

Supplemental Code

Supplemental Code 1:

https://github.com/qiaohj/Literatures_R/blob/master/IUCN_References/Prepare/IUCN_API.r.

Supplemental Code 2:
https://github.com/qiaohj/Literatures_R/blob/master/IUCN_References/Prepare/extract_references_from_json.r

Supplemental Code 3:
https://github.com/qiaohj/Literatures_R/blob/master/IUCN_References/Prepare/reference_similarity.r

Supplemental Data

All analysis associated with figures and supplemental figures.

Supplemental Figures

Figure S1. Mean number of publications per family per year in mammals.

Figure S2a. Mean number of publications per species for each order per region in mammals for species with distributions restricted to a single realm.

Figure S2b. Mean number of publications per species for each order per region in mammals for all species.

Figure S3. Mean number of publications vs mean mass (g) per group in mammals.

Figure S4. Mean number of publications per order per year in Odonata.

Figure S5. Number of species compared to number of publications per family in Odonata.

Figure S6a. Mean publication per species in Odonata per realm, for species with distributions restricted to a single realm.

Figure S6b. Mean publication per species in Odonata per realm, for all species.

Figure S7. Number of bird species relative to number of publications per order.

Figure S8. Mean publications per time period per family within birds.

Figure S9a. Mean publications per order per region for birds with distributions restricted to a single realm.

Figure S9b. Mean publications per order per region for all bird species.

Figure S10a. Total publications vs total number of species per order within reptiles.

Figure S10b. Total publications vs total number of species per family within reptiles.

Figure S11. Mean publications per period per family for reptiles.

Figure S12a. Mean publications per region per order for reptiles with distributions restricted to a single realm.

Figure S12b. Mean publications per region per order for all reptiles.

Figure S13a. Total publications vs total number of species per order for amphibians.

Figure S13b. Total publications vs total number of species per family for amphibians.

Figure S14. Mean publications per period per family for amphibians.

Figure S15a. Mean publications per region per order for amphibians for species with distributions restricted to a single realm.

Figure S15b. Mean publications per region per order for amphibians for all species.

Mammals

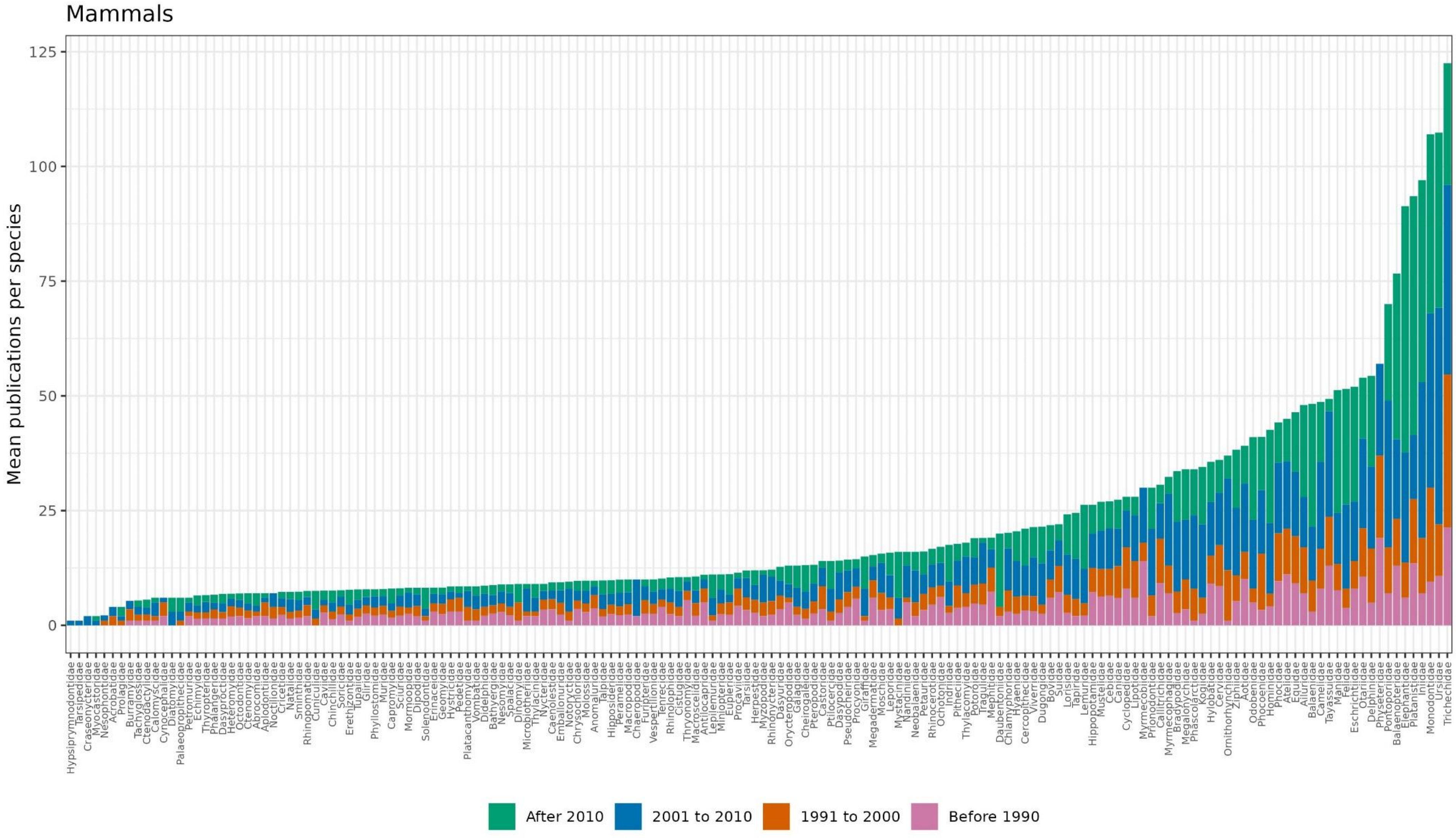

Figure S1. Mean number of publications per family per year in mammals.

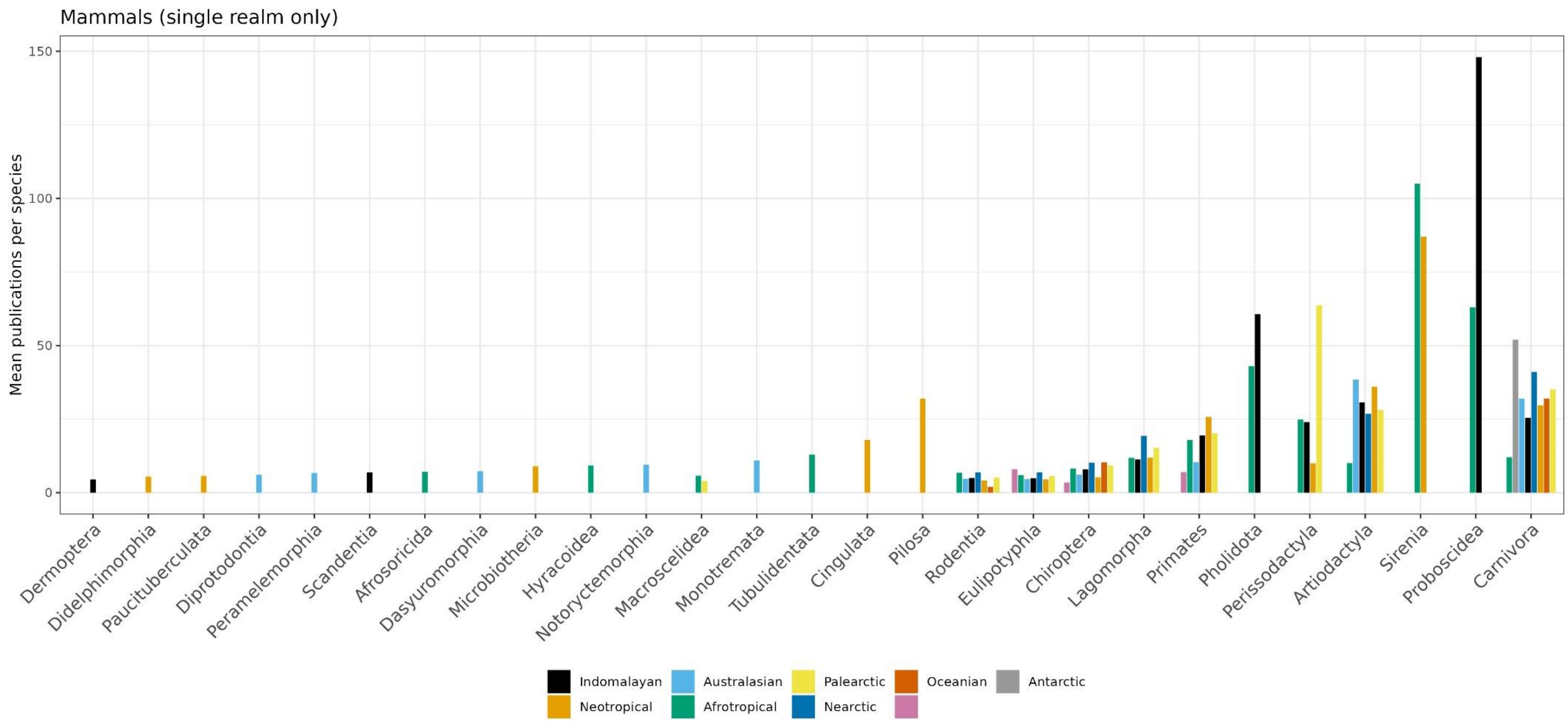

Figure S2a. Mean number of publications per species for each order per region in mammals for species with distributions restricted to a single realm.

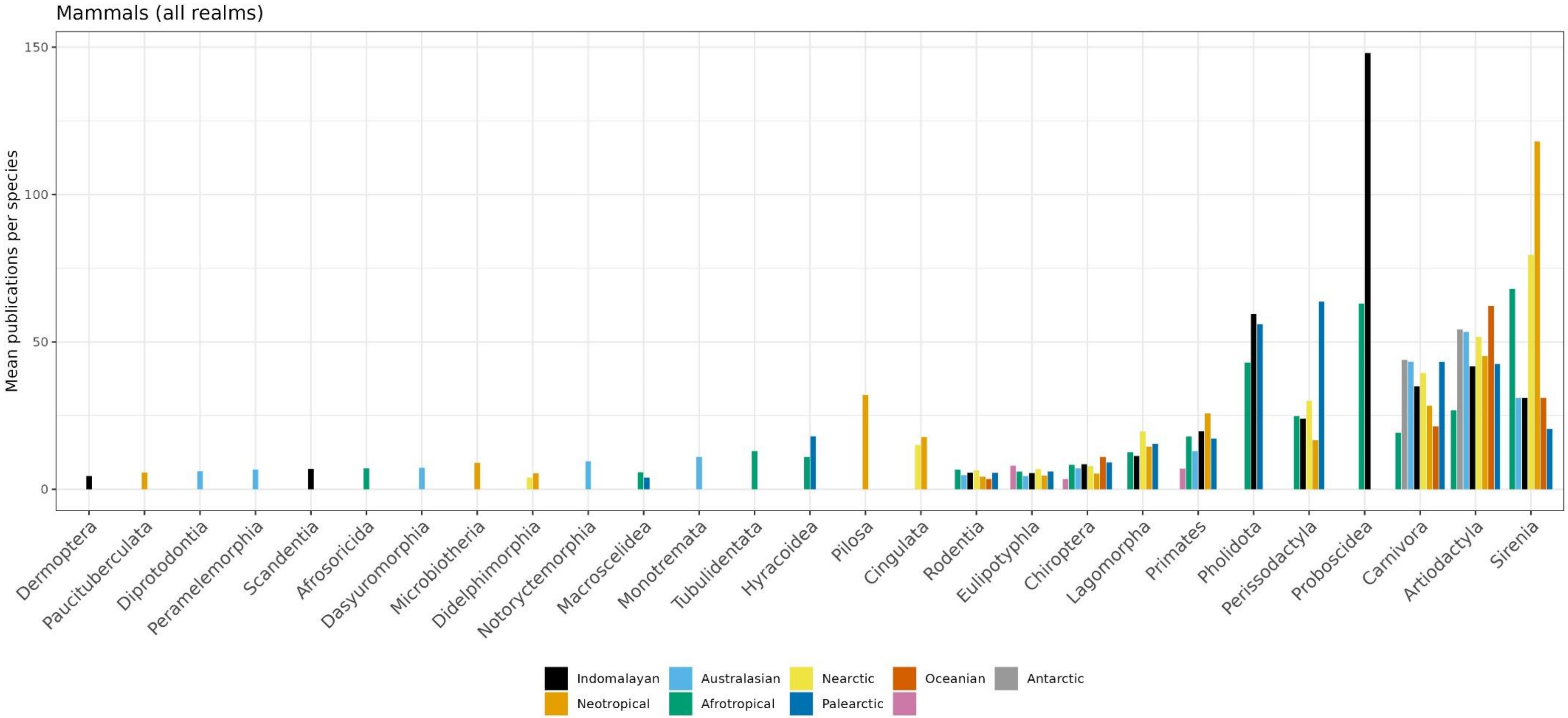

Figure S2b. Mean number of publications per species for each order per region in mammals for all species.

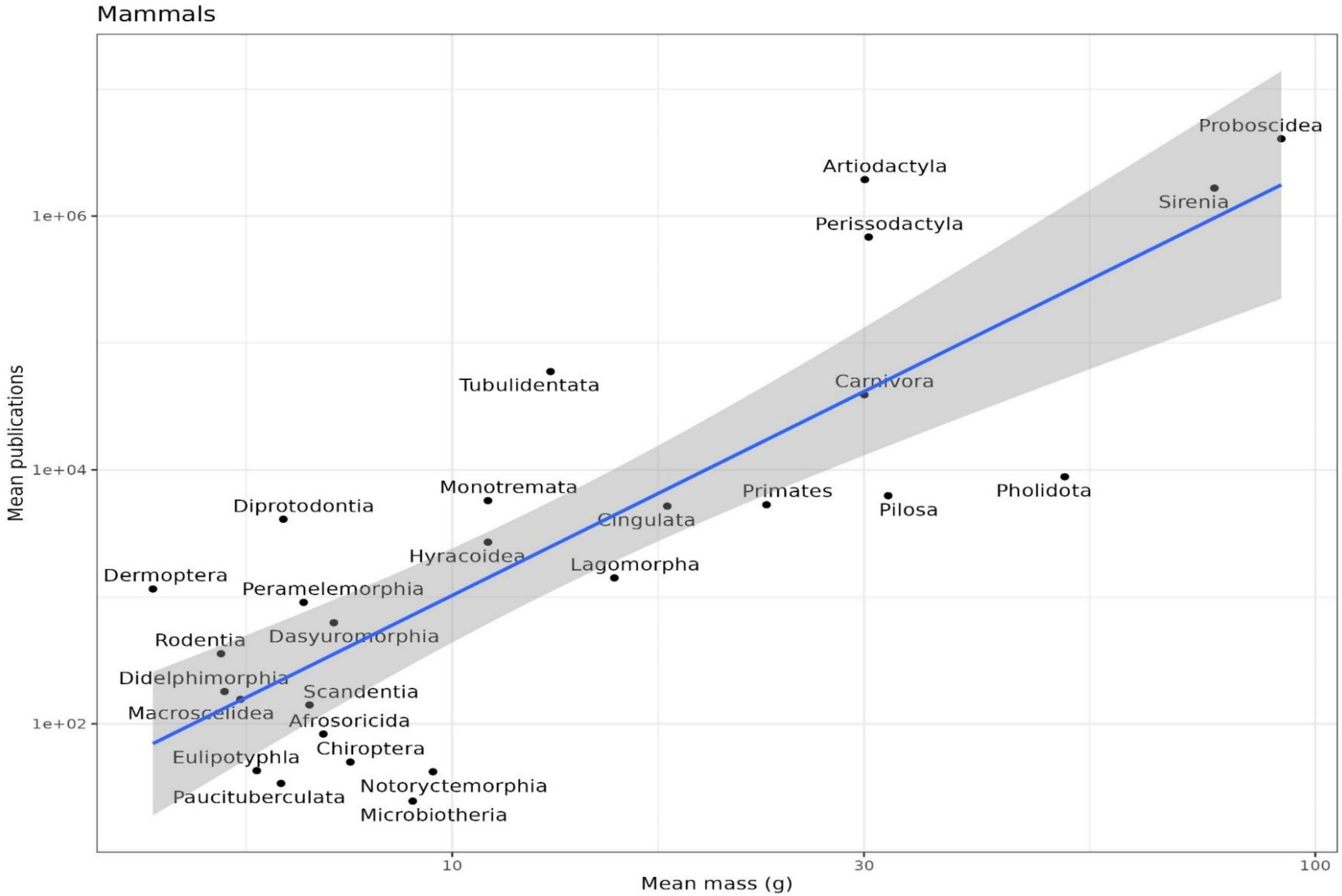

Figure S3. Mean number of publications vs mean mass (g) per group in mammals.

Odonata

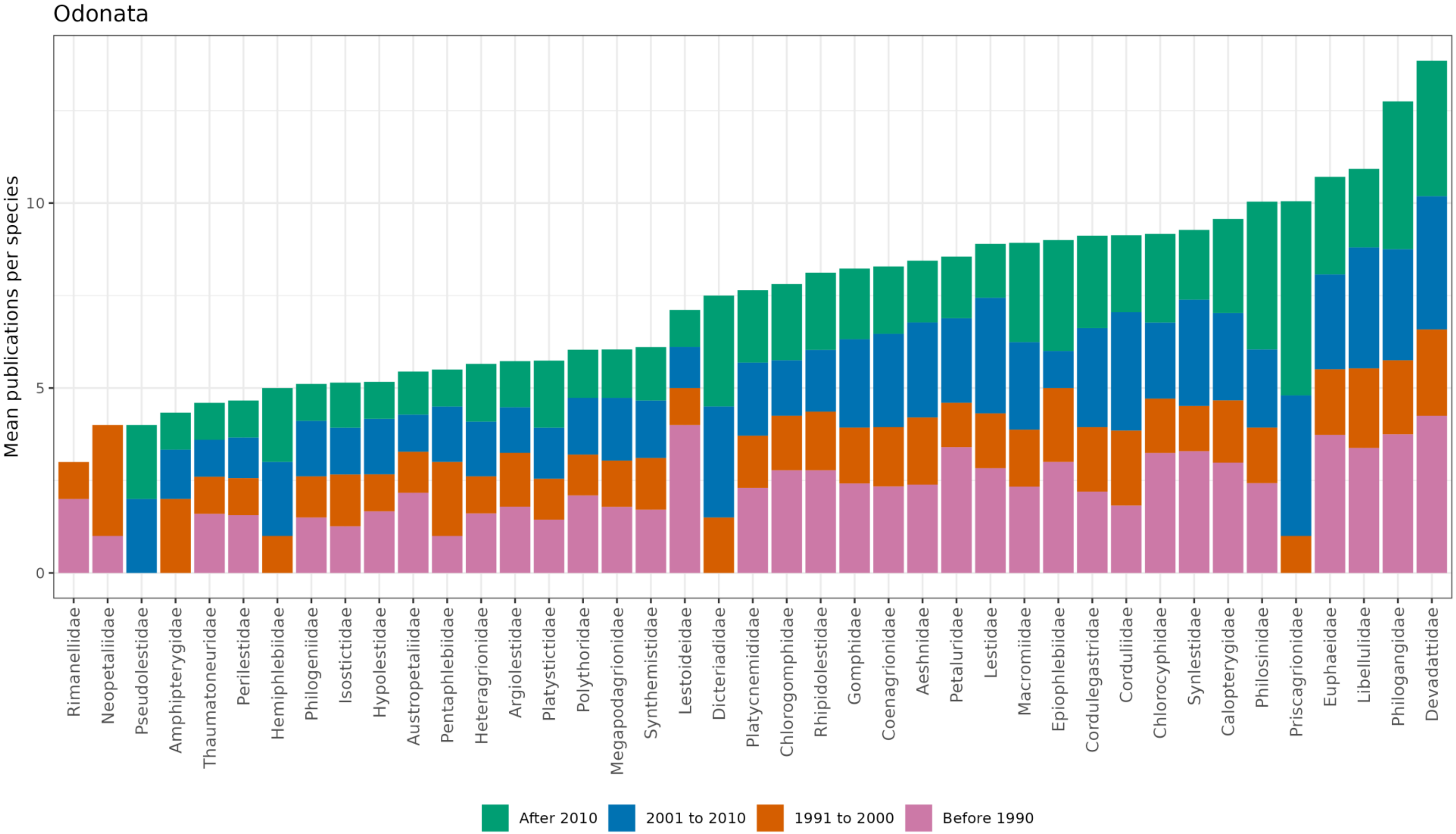

Figure S4. Mean number of publications per order per year in Odonata.

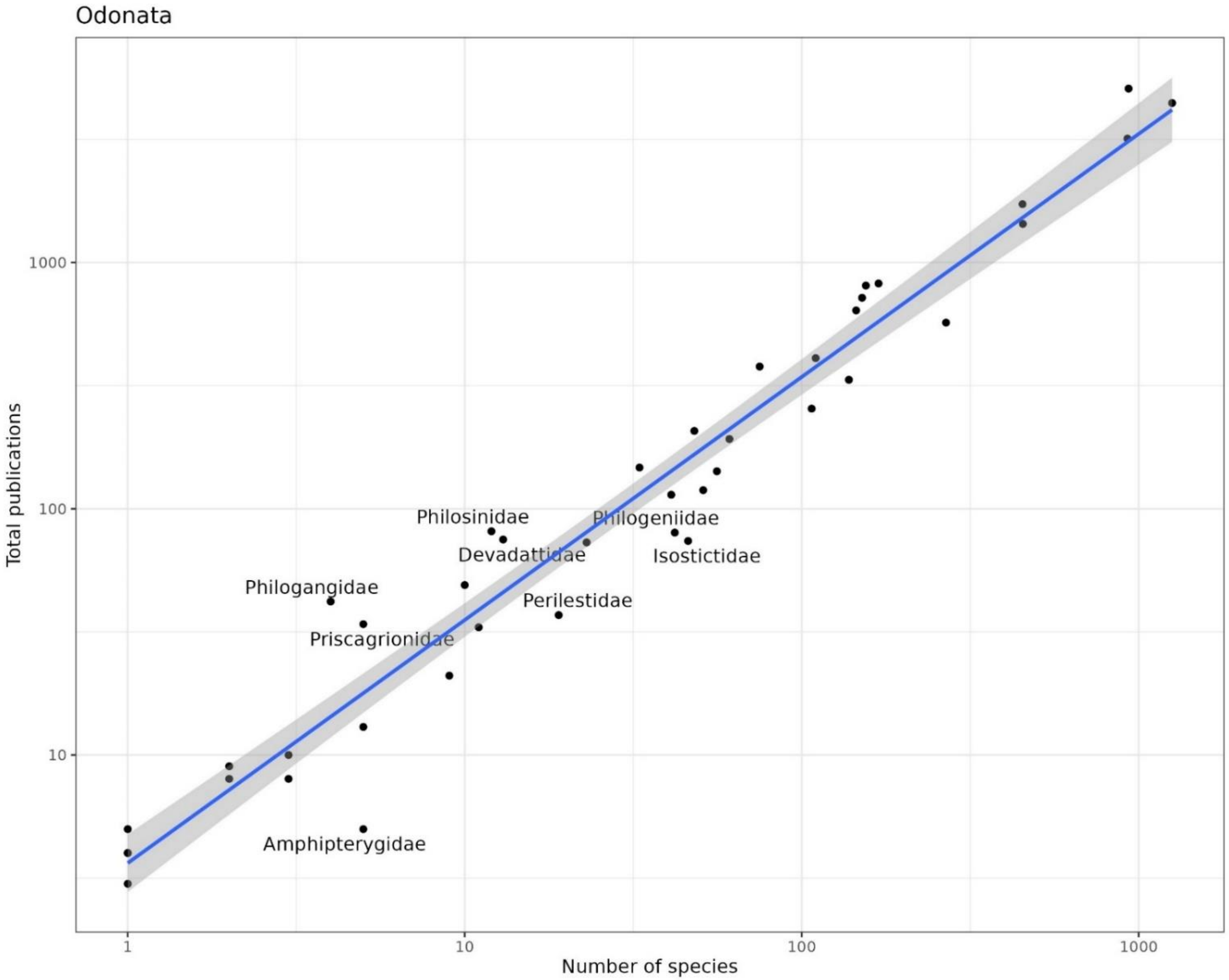

Figure S5. Number of species compared to number of publications per family in Odonata.

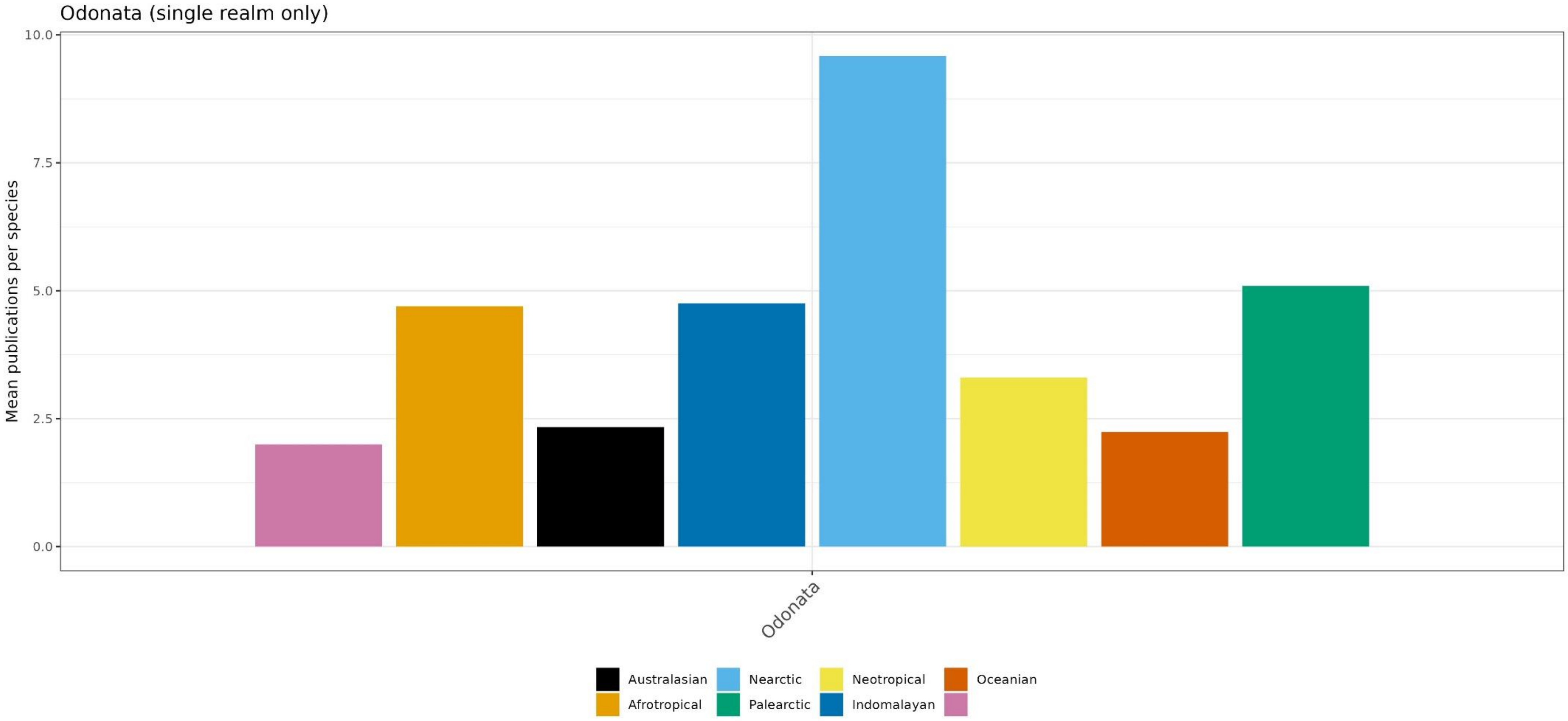

Figure S6a. Mean publication per species in Odonata per realm, for species with distributions restricted to a single realm.

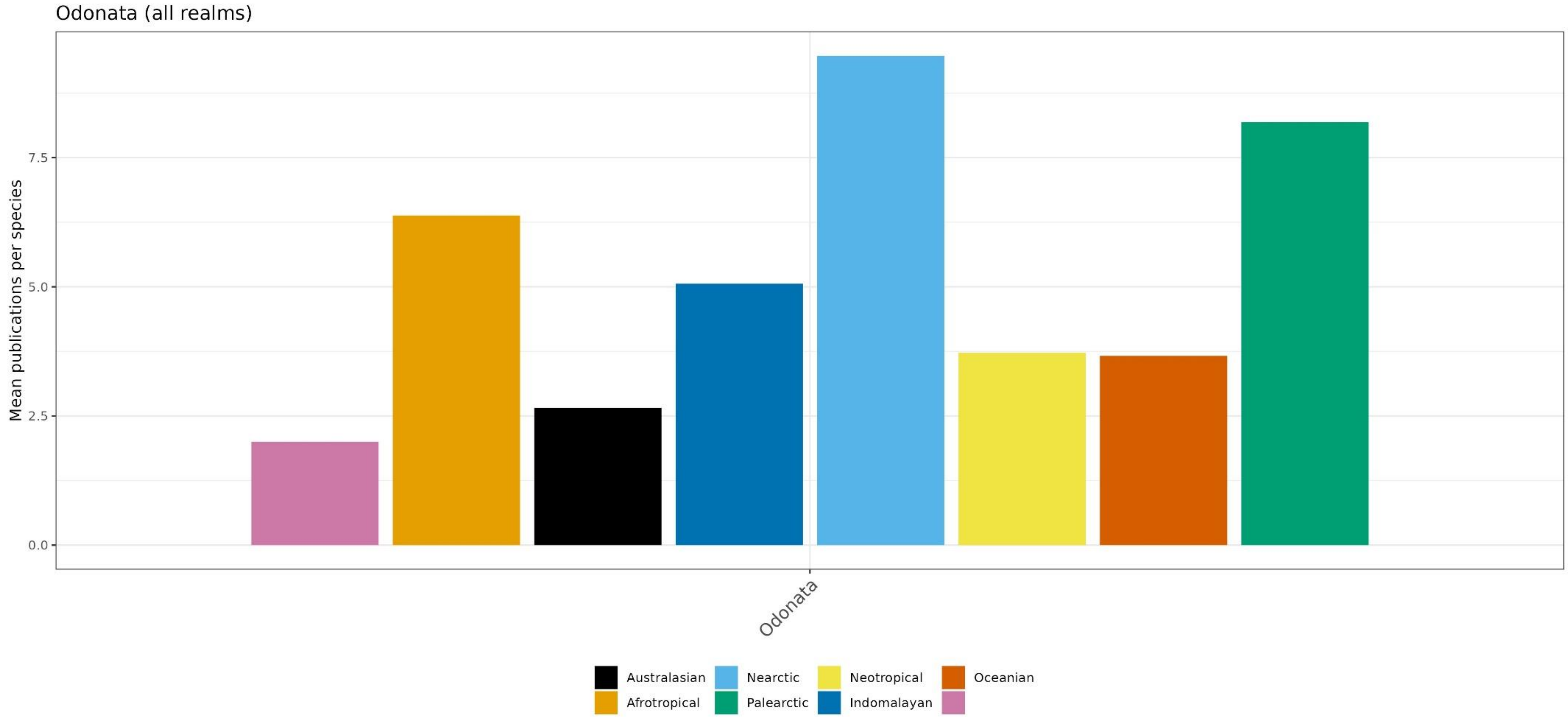

Figure S6b. Mean publication per species in Odonata per realm, for all species.

Birds

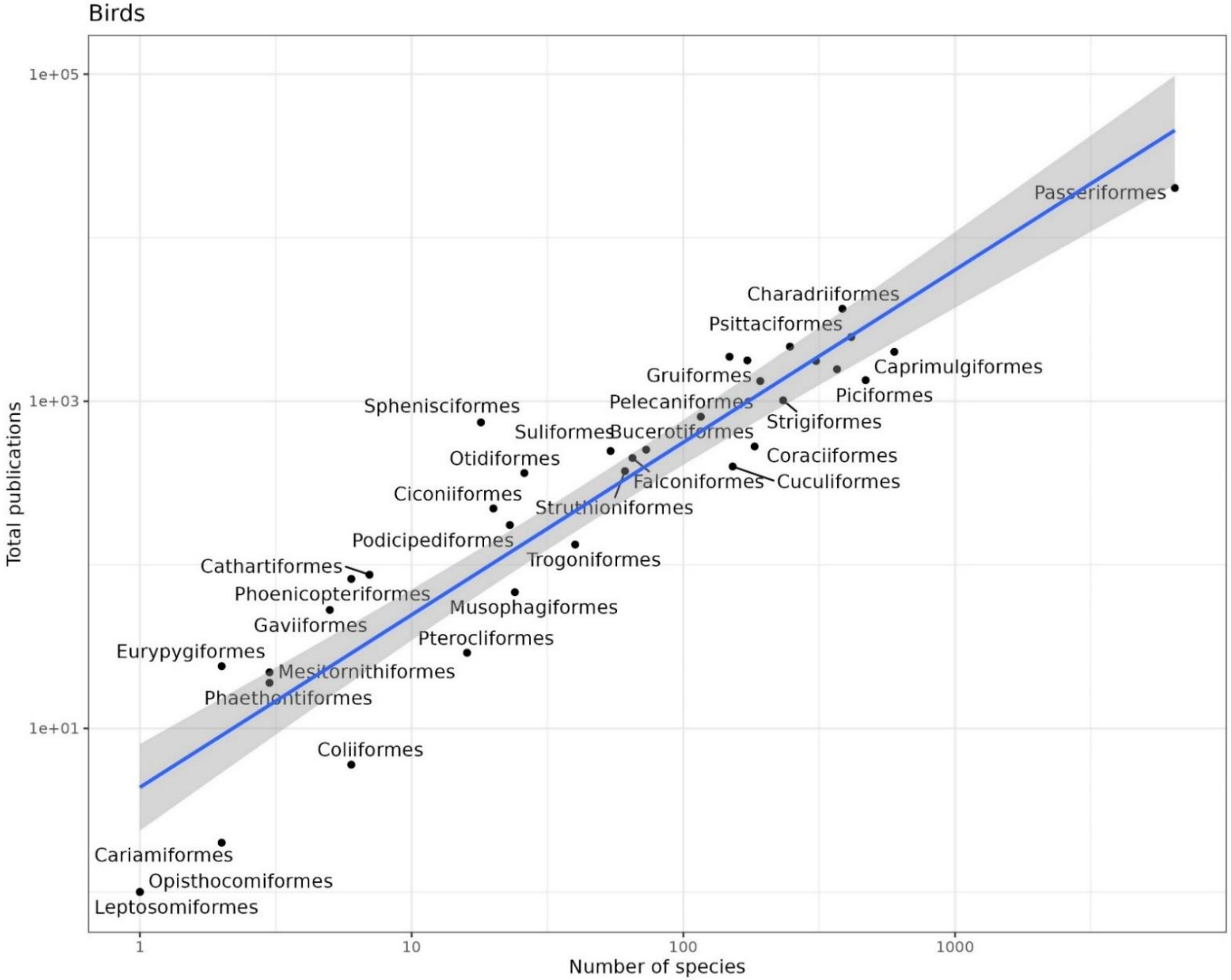

Figure S7. Number of bird species relative to number of publications per order.

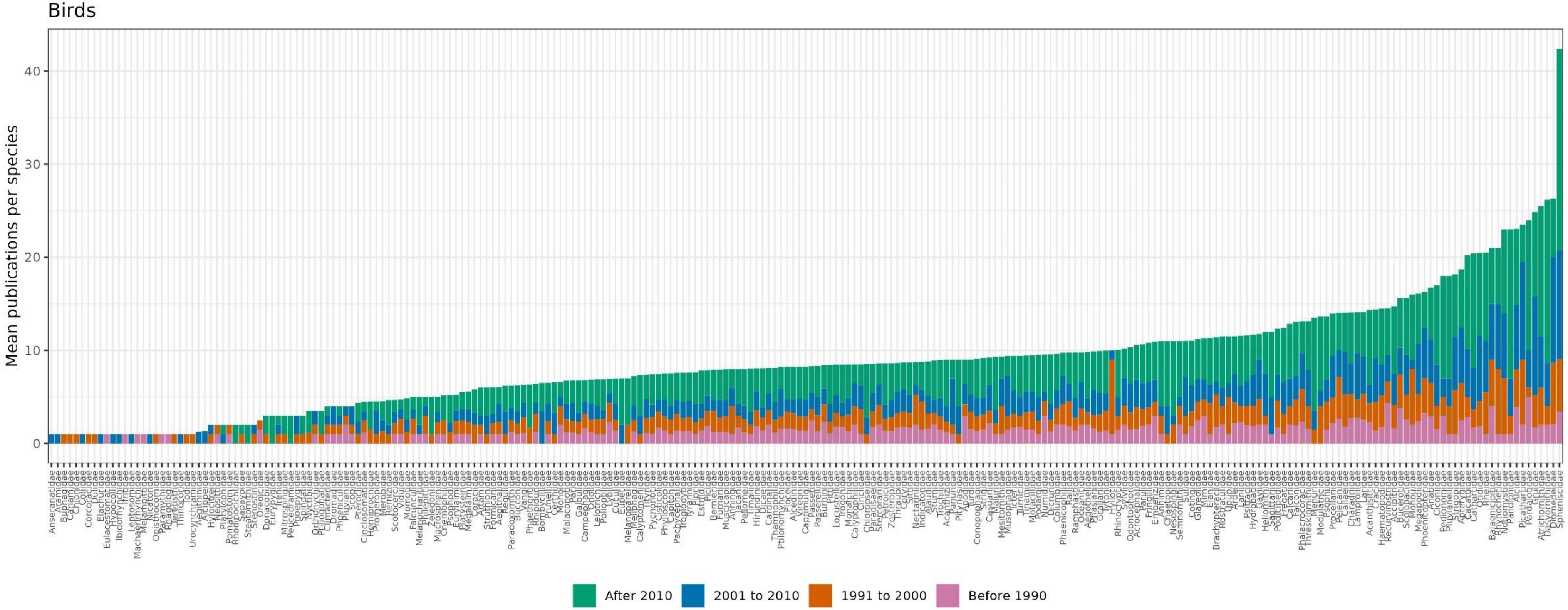

Figure S8. Mean publications per time period per family within birds.

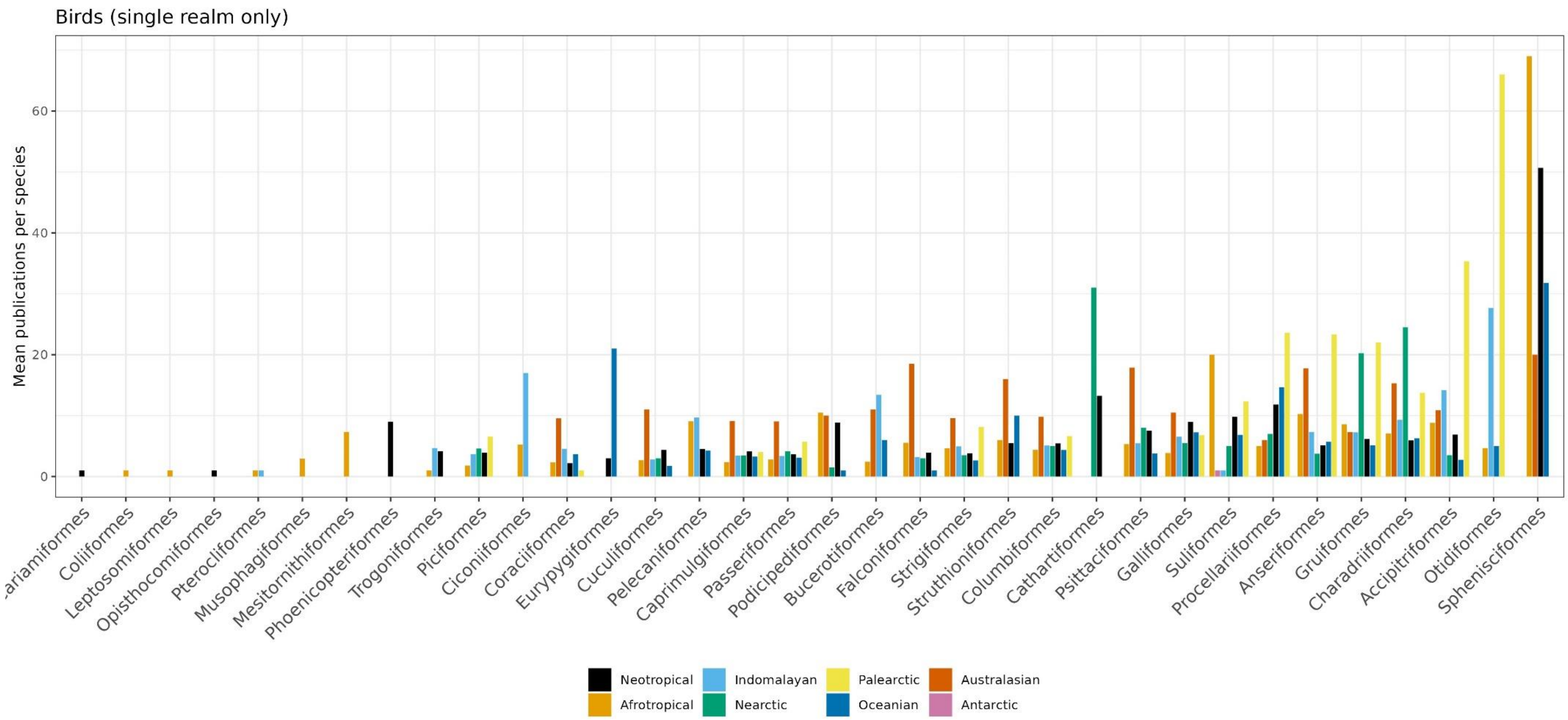

Figure S9a. Mean publications per order per region for birds with distributions restricted to a single realm.

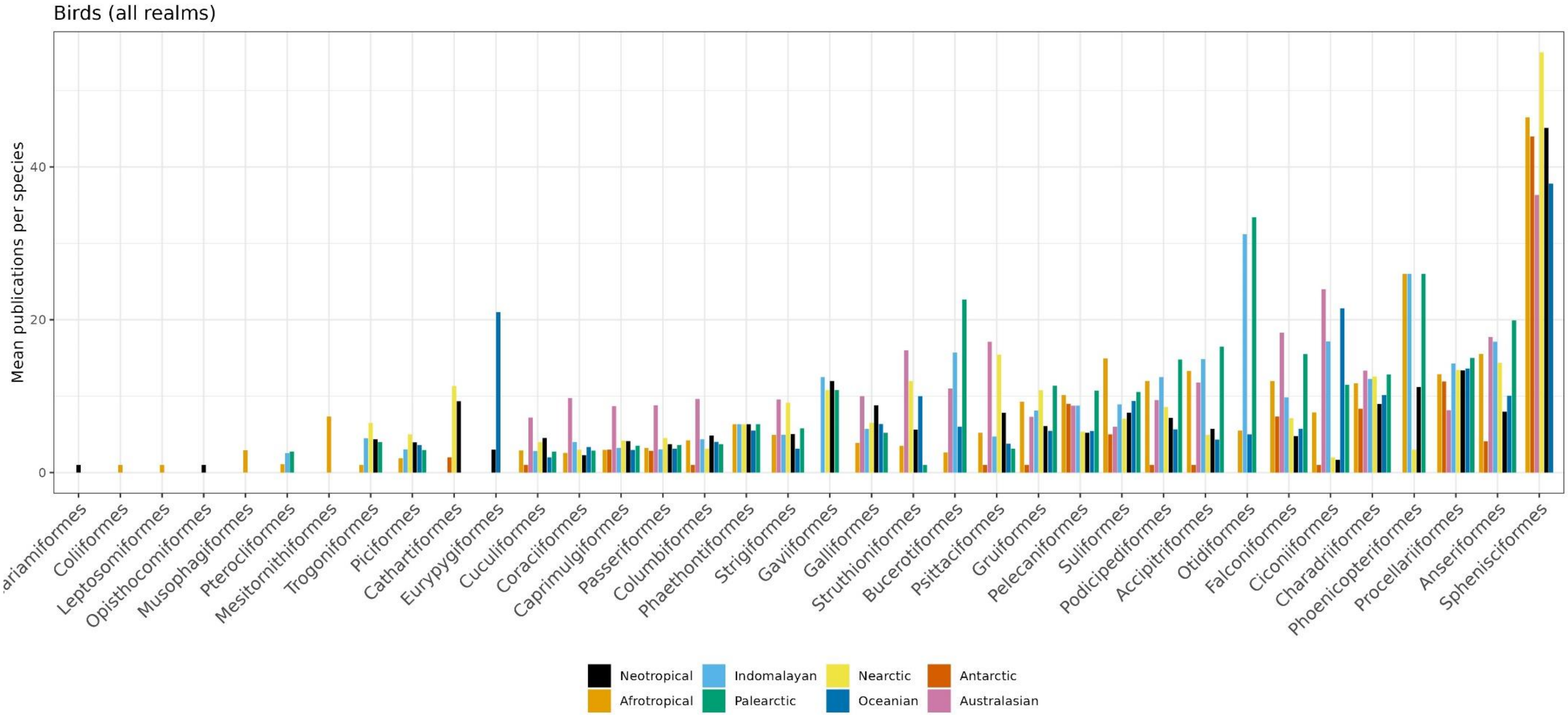

Figure S9b. Mean publications per order per region for all bird species.

Reptiles

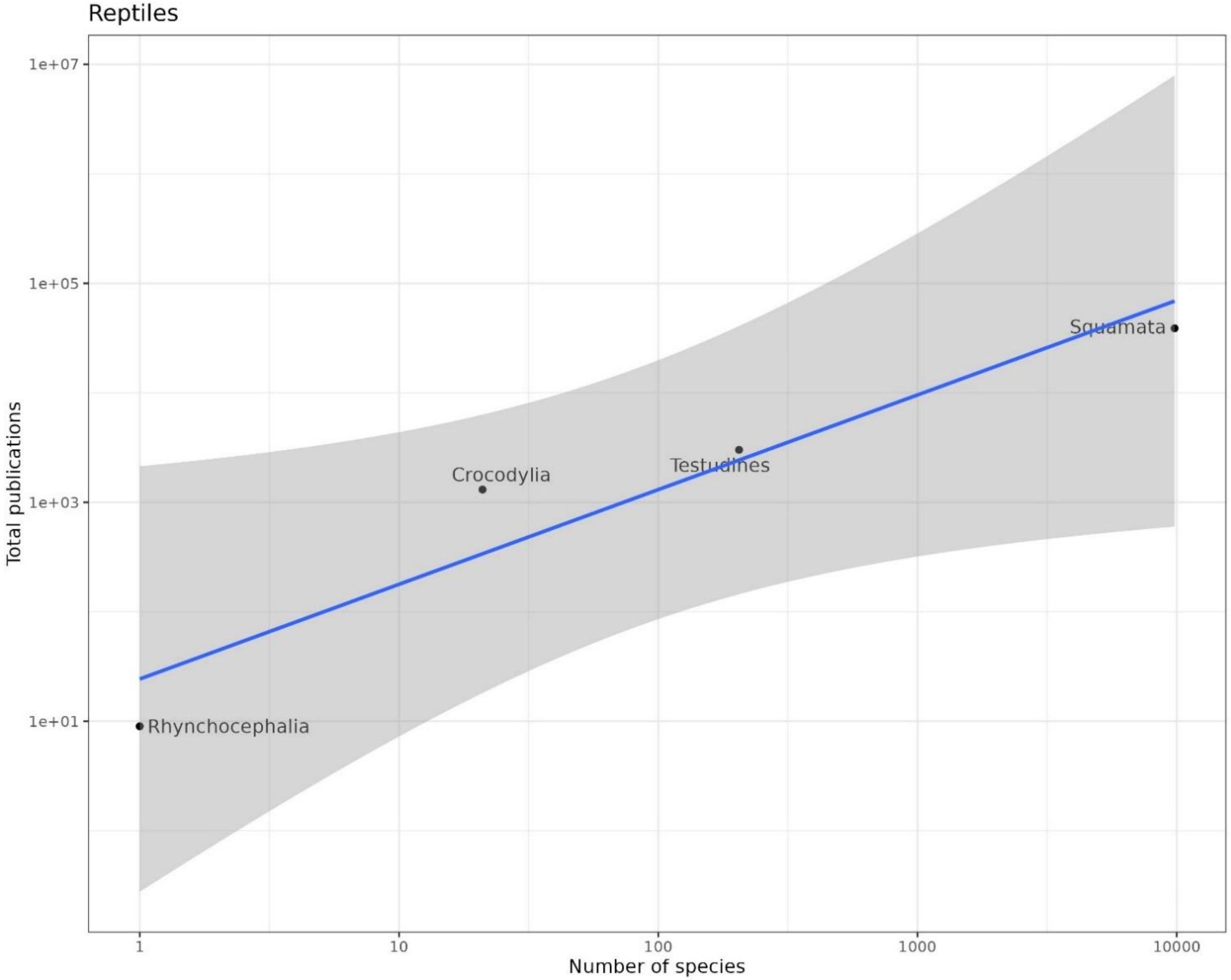

Figure S10a. Total publications vs total number of species per order within reptiles.

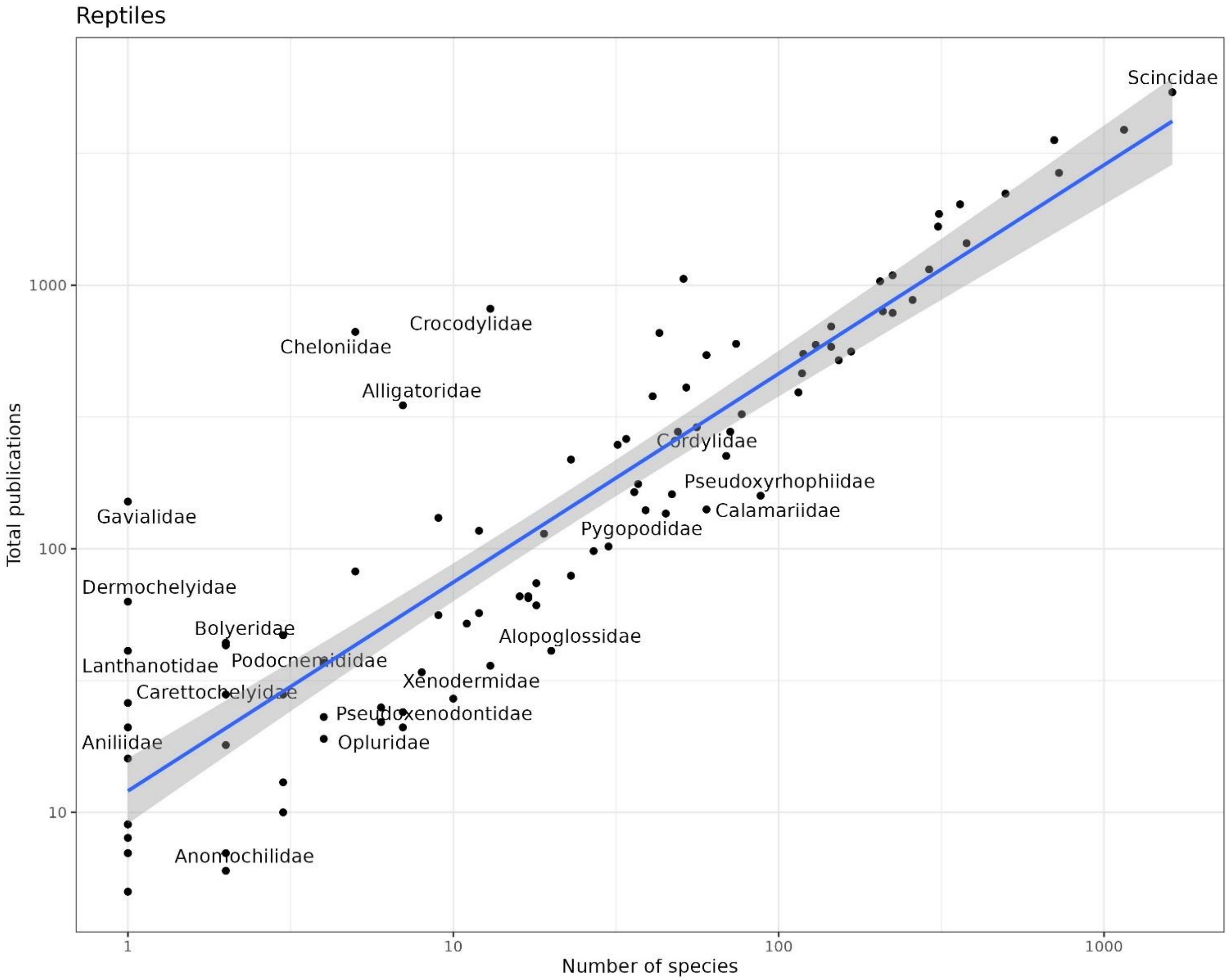

Figure S10b. Total publications vs total number of species per family within reptiles.

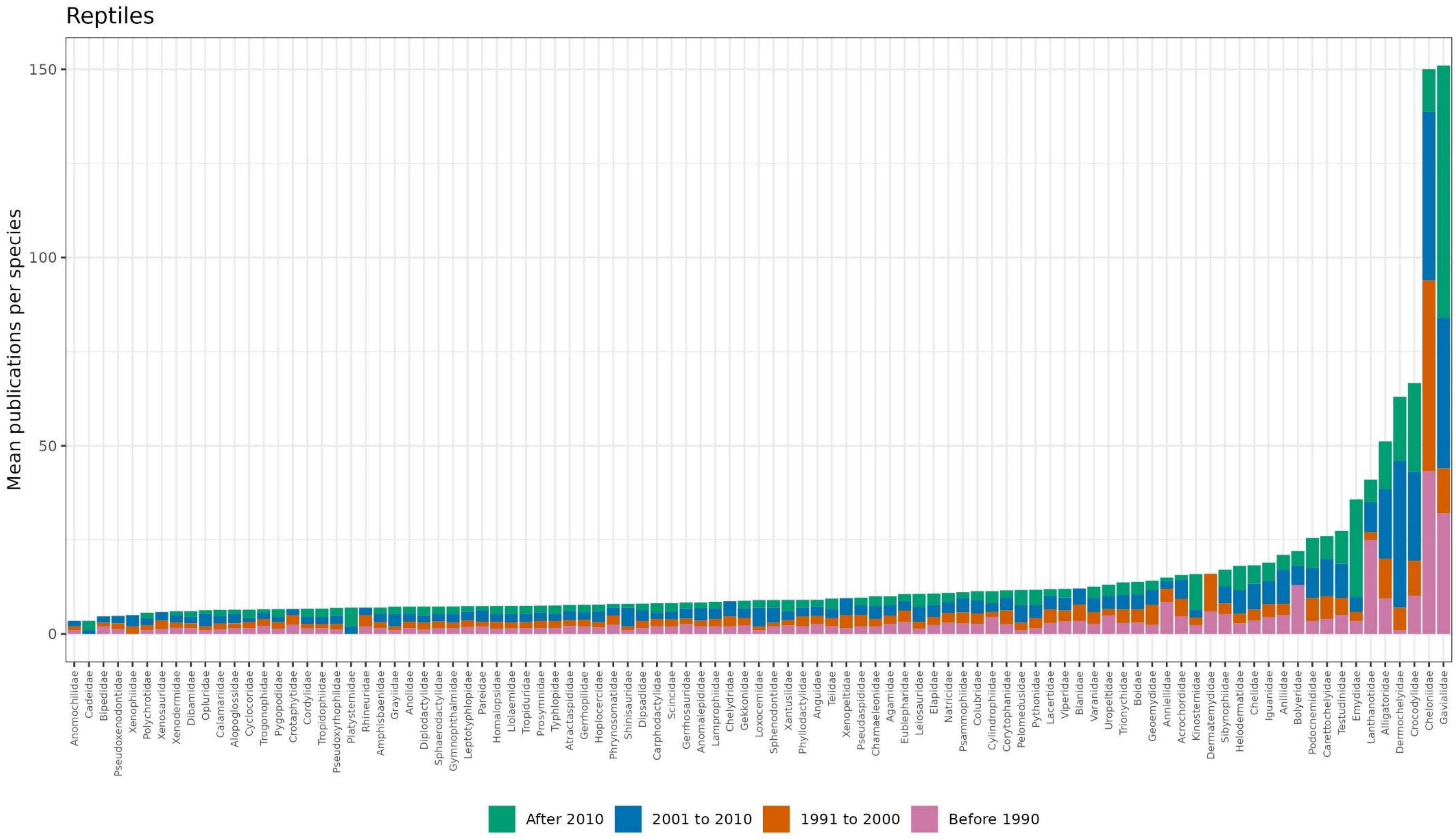

Figure S11. Mean publications per period per family for reptiles.

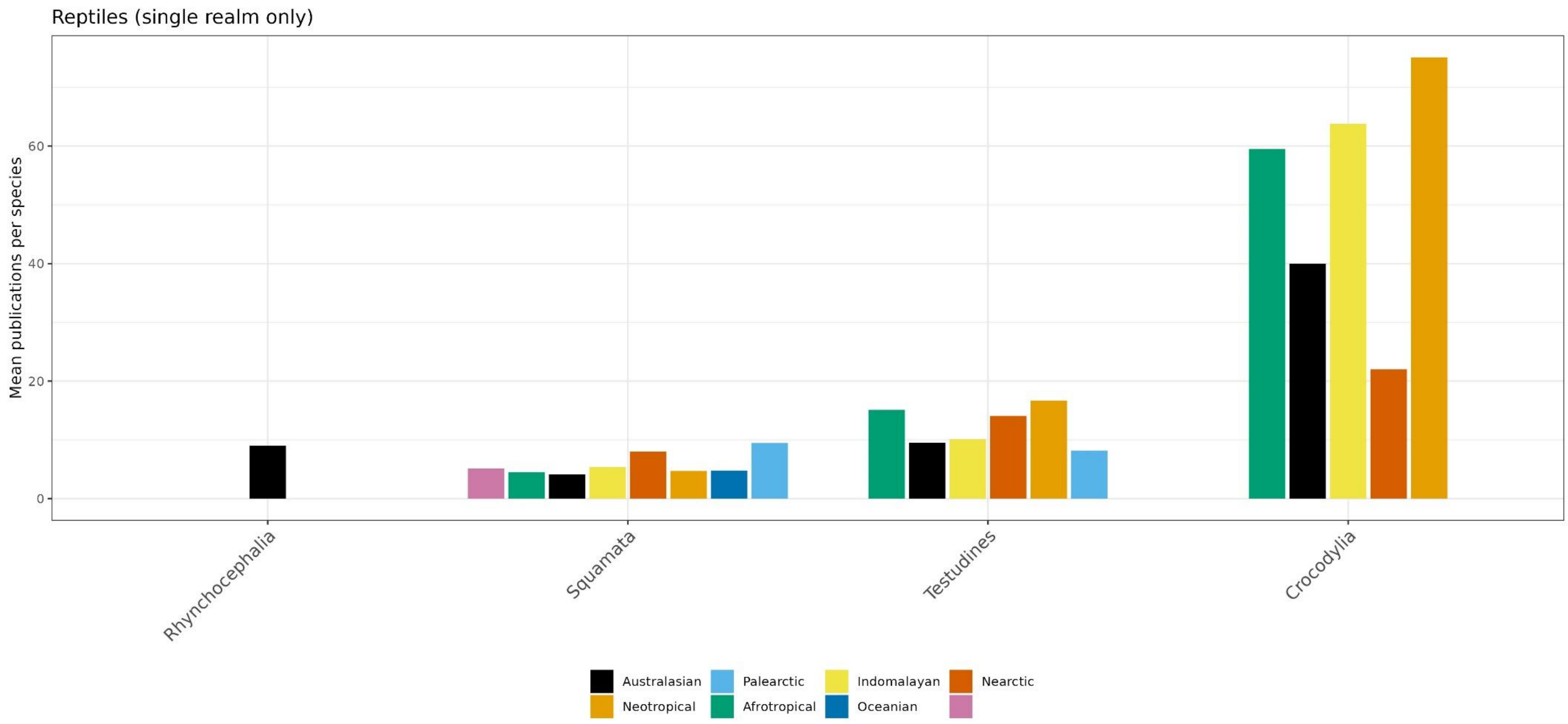

Figure S12a. Mean publications per region per order for reptiles with distributions restricted to a single realm.

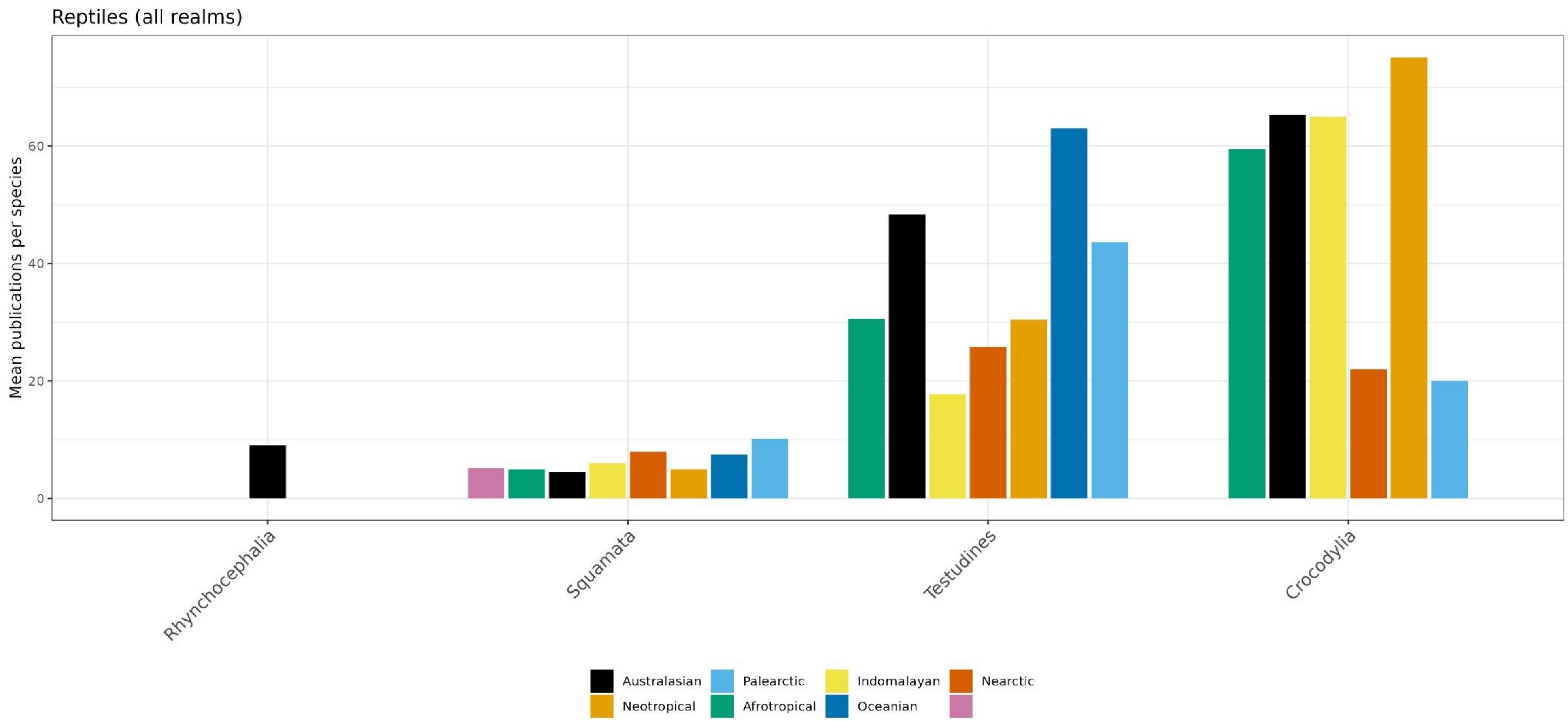

Figure S12b. Mean publications per region per order for all reptiles.

Amphibians

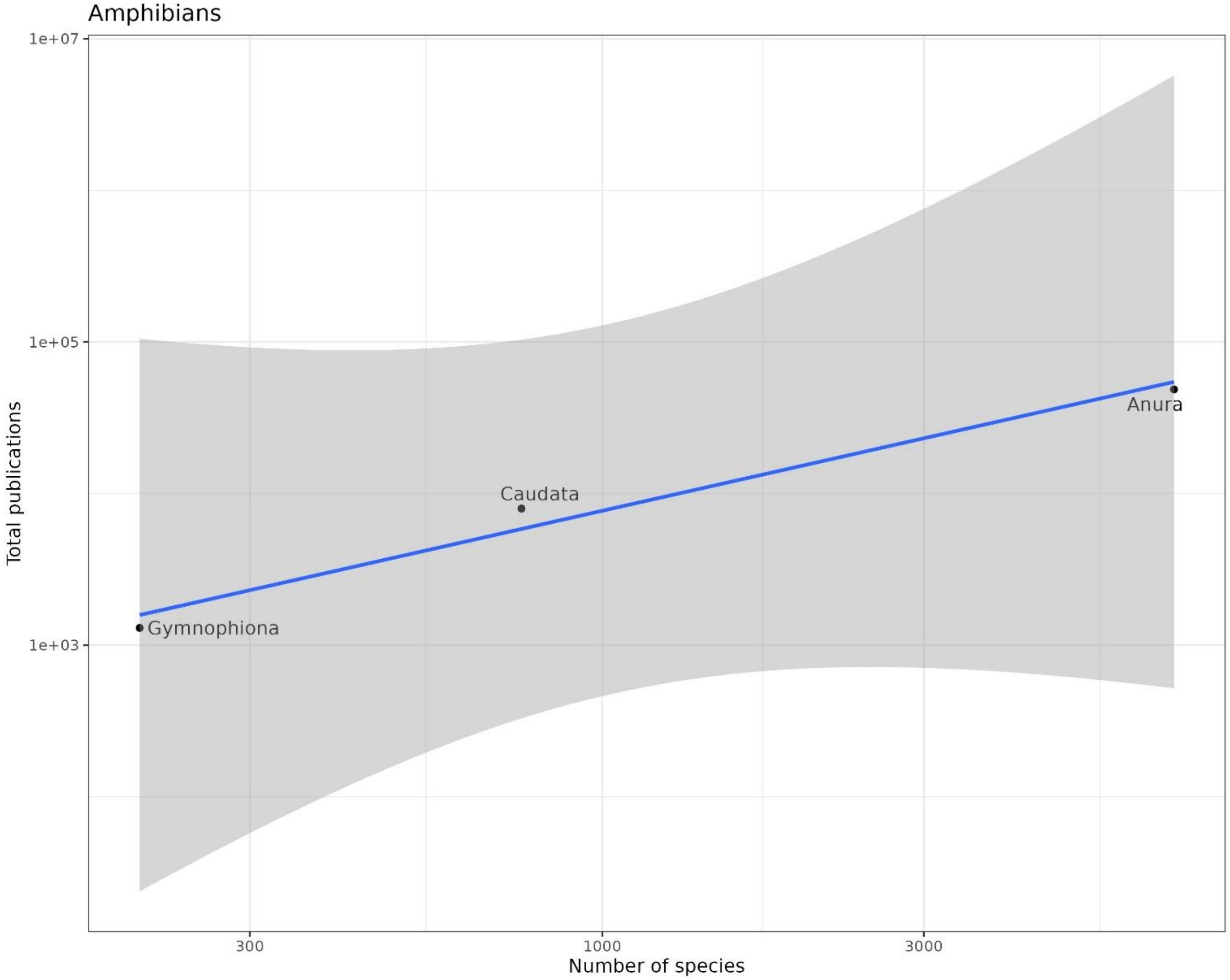

Figure S13a. Total publications vs total number of species per order for amphibians.

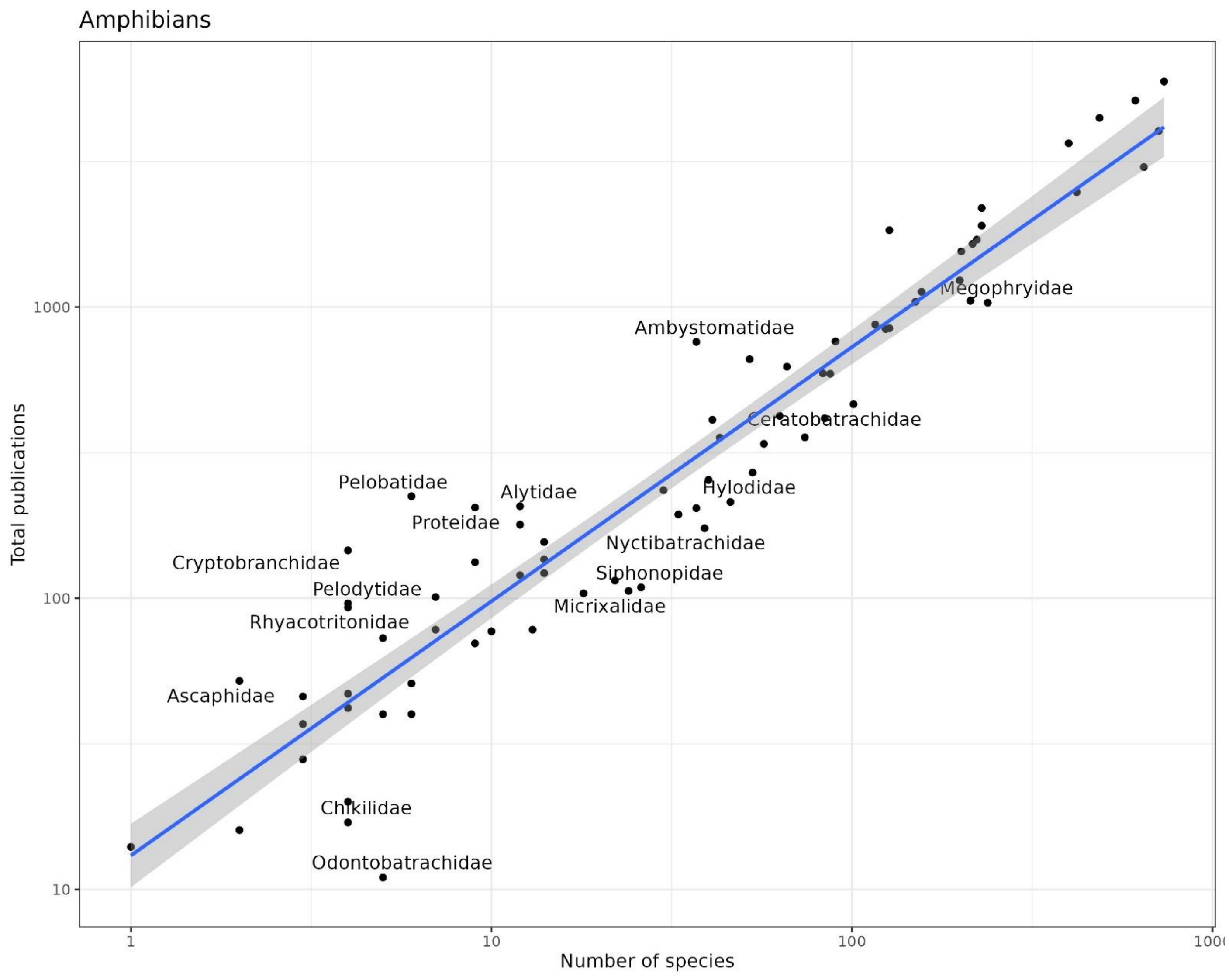

Figure S13b. Total publications vs total number of species per family for amphibians.

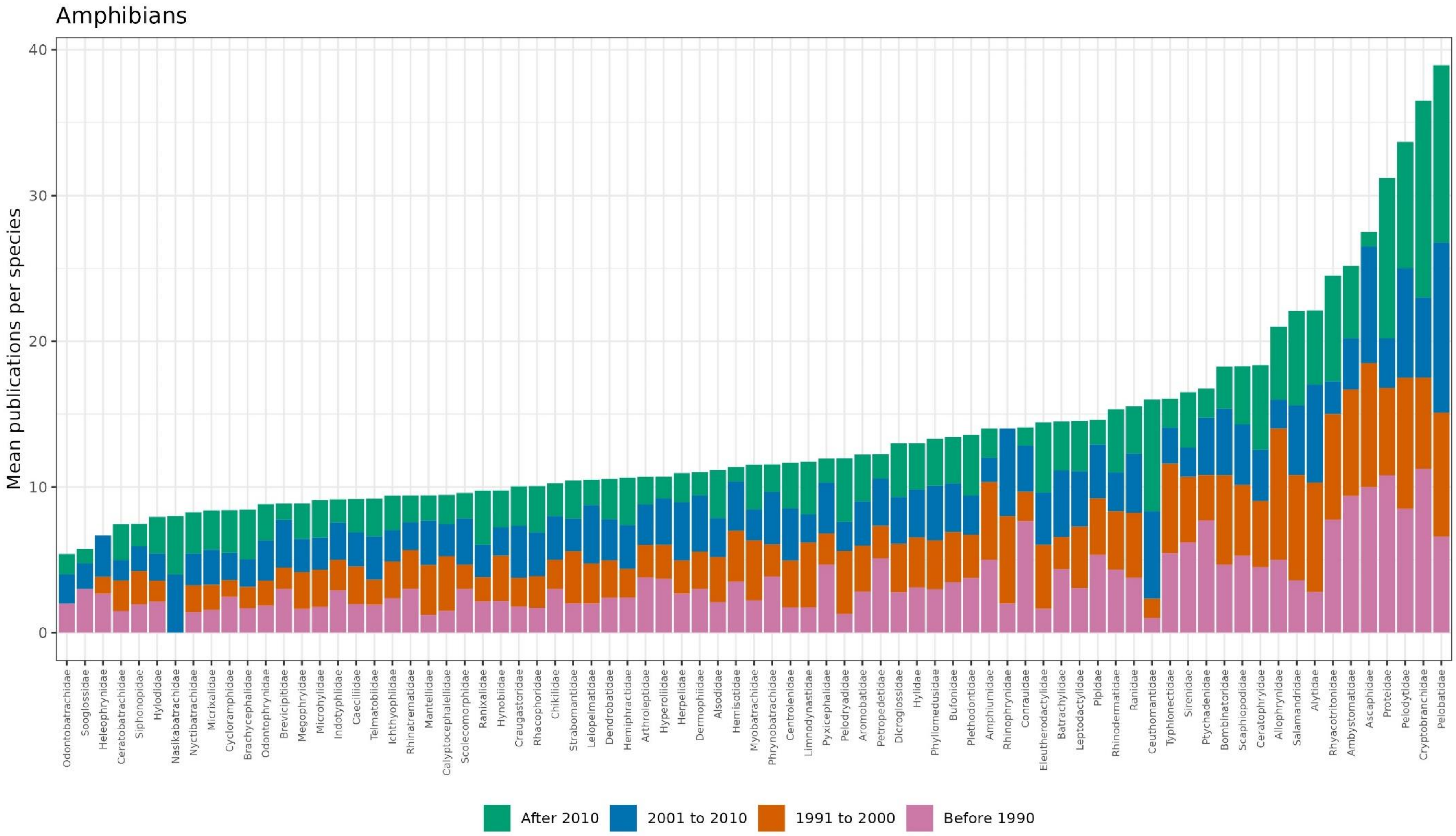

Figure S14. Mean publications per period per family for amphibians.

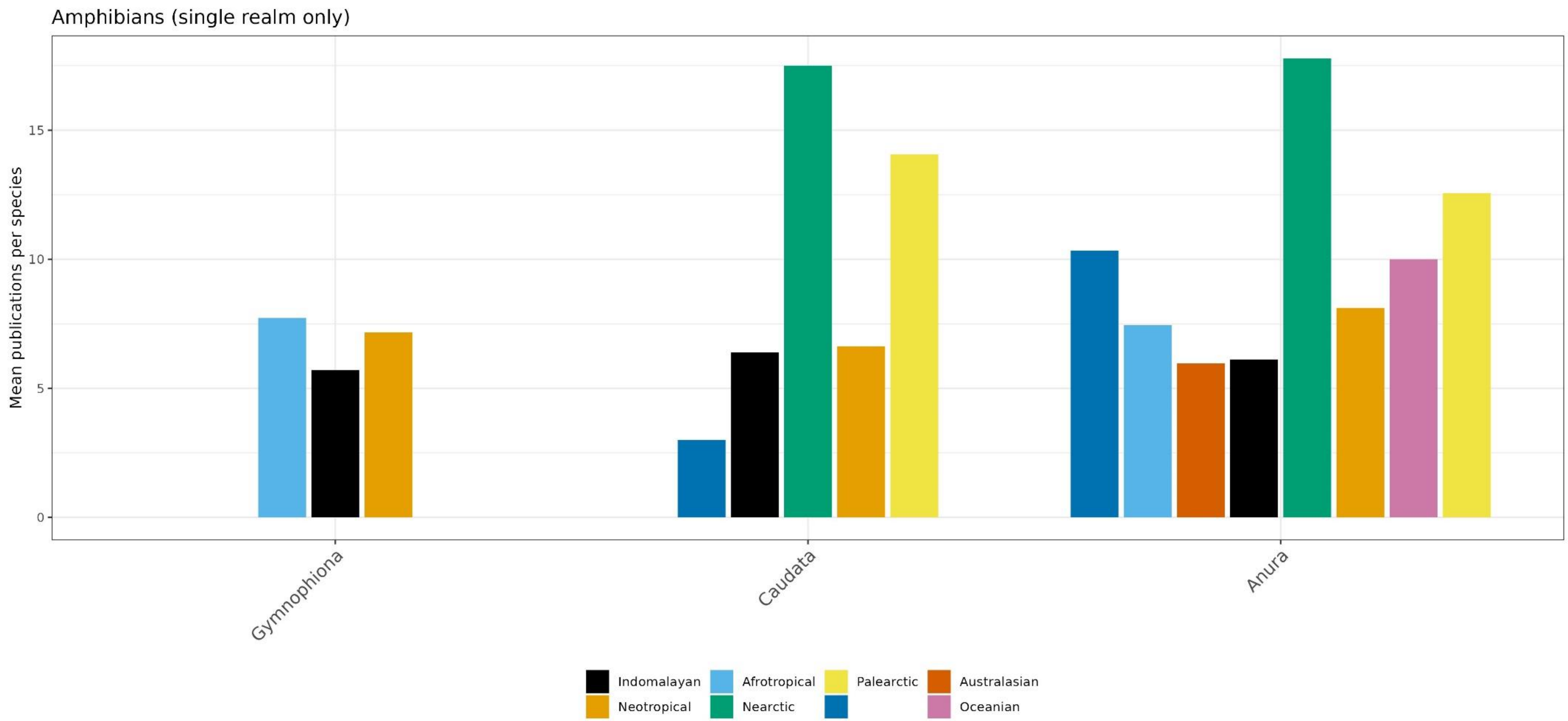

Figure S15a. Mean publications per region per order for amphibians for species with distributions restricted to a single realm.

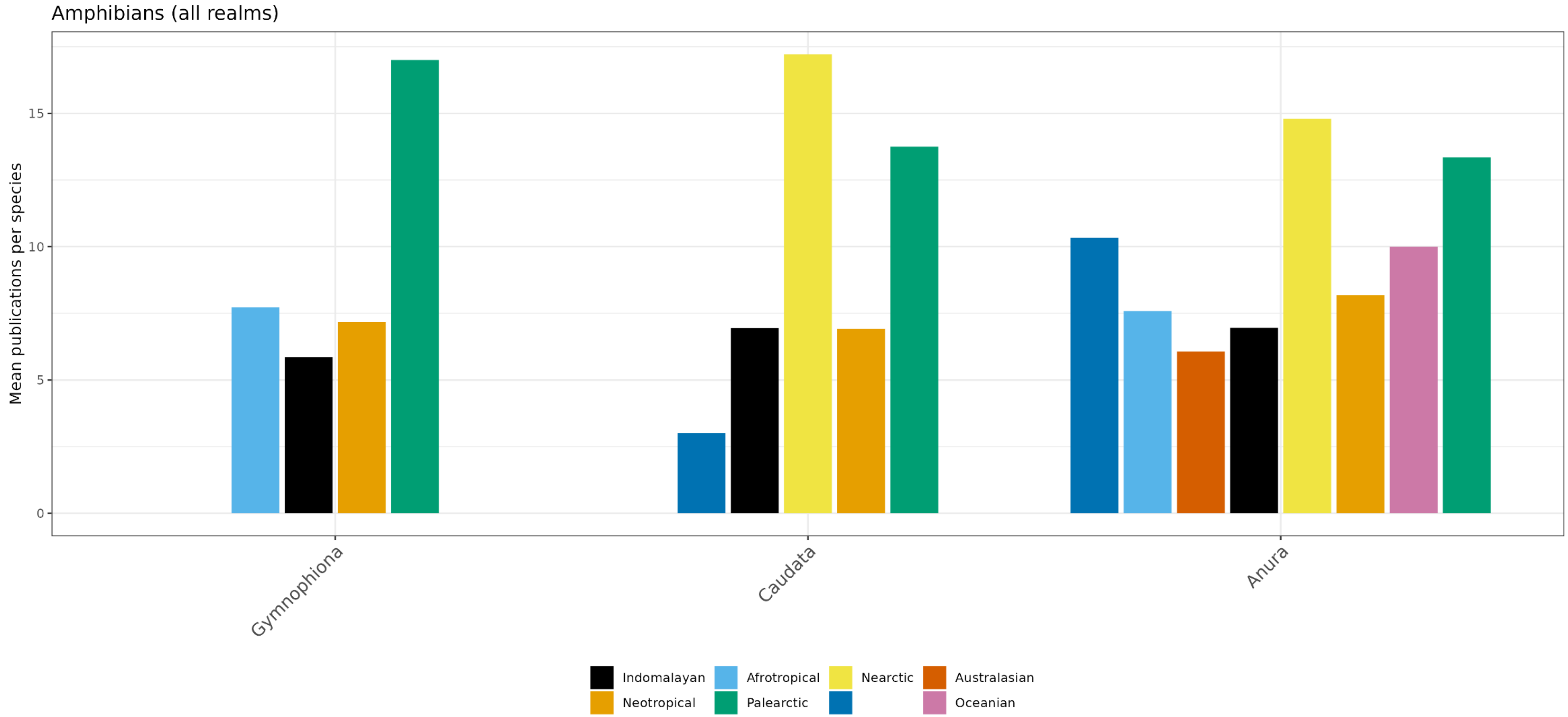

Figure S15b. Mean publications per region per order for amphibians for all species.