

Supplementary Materials for
**Major shifts in biogeographic regions of freshwater fishes as evidence of the
Anthropocene epoch**

Boris Leroy *et al.*

Corresponding author: Boris Leroy, leroy.boris@gmail.com

Sci. Adv. **9**, eadi5502 (2023)
DOI: 10.1126/sciadv.adi5502

This PDF file includes:

Supplementary Text
Figs. S1 to S8
Tables S1 to S3

Supplementary Text

Changes in biogeographical structure

To further understand the changes in biogeographical structure between natural and Anthropocene regions, we explored the distribution of area per cluster for each clustering level (Figure S2). Natural regions showed a large gap in cluster area between minor and major regions (10 to 15-fold increase) at clustering levels 1 and 2 (Figure S2a,b, mind the log₁₀ axis for cluster area), whereas they showed a continuous increase at level 3 (Figure S2c). We also observed that the number of major clusters increased with hierarchical level. We used this biogeographical structure as a reference for comparison with the other pools.

For Anthropocene regions, we found two notable differences with natural regions. First, we found that level 1 and level 2 of Anthropocene regions (Figure S2d,e) respectively resembled levels 2 and 3 of natural regions (Figure S2b,c), suggesting a loss of biogeographical structure at the coarsest level. Second, the number of clusters decreased between level 2 and 3 of Anthropocene regions (Figure S2e-f), contrary to natural regions (Figure S2b-c). This observation suggests that there were clusters detected at a fine scale in the natural regions which are no longer detected in the Anthropocene regions.

When accounting for introductions only, we found a biogeographical structure very similar to Anthropocene regions (Figure S2g-i), with variations (e.g., Figure S2e vs S2h, Figure S2f vs S2i). Last, when we accounted for extinctions only, we found a biogeographical structure resembling the natural regions (Figure S2j-l), with noticeable variations (Figure S2l vs. S2c). These variations suggest that extinctions did impact the biogeographical structure at levels 2 and 3.

Changes in biogeographical networks

The biogeographical network of the natural pool of species displayed very distinct groups of nodes for each cluster, with only a limited number of links between clusters (Figure S3). For the Anthropocene pool, the network structure changed dramatically (Figure S4), with a very large number of links between clusters, resulting in a cluttered network with important overlap among clusters. We noted that the Pan-Anthropocenic Global North and East Asia region (PAGNEA) and Neo-Oriental clusters are very close to each other on the graph, suggesting a very large number of links (i.e., shared species co-occurrences) between these two clusters. The introduction-only network (Figure S5) was very similar to the Anthropocene network (Figure S4), and the extinction-only network (Figure S6) was very similar to the natural network (Figure S3).

Patterns of natural distribution range of introduced species

To further understand the mechanisms explaining changes in bioregion delineations, and the importance of introduced species, we investigated the natural occurrences of introduced species. We specifically sought to understand which types of species were introduced in terms of distribution range, i.e., rather restricted-range or widespread species.

First, we compared the observed distribution of frequencies of occurrences for the 453 introduced species to an expected distribution if introduced species had been selected at random. We simulated the expected distribution by randomly sampling 453 species in the total pool of freshwater fish species 100 times. For all intervals of 10 occurrences between 1 and the

maximum, we calculated the expected number of species as the average across all simulations. For the same intervals, we calculated the observed number of introduced species. We compared the expected and observed distributions with a χ^2 test. Because multiple intervals had low counts, we simulated the p-values with a Monte-Carlo simulation with 50 000 replicates. To emphasise which intervals had the most significant differences, we extracted the standardised Pearson residuals, which we deemed as significant when their absolute values were superior to 2. We used the `chisq.test()` functions from the R package `stats` for the statistical analyses.

We found that introduced species were not randomly selected in the pool of freshwater fishes ($\chi^2=199.36$, $p=0.00002$, Figure S3). Specifically, introduced species were less-often-than random selected in the pool of rarest species (natural occurrence ≤ 10); conversely, introduced species were disproportionately sampled in the pool of high-occurrence species (natural occurrence > 10 ; Figure S3). In fact, we found a strong increase in the percentage of introduced species with the increase in interval values (Figure S4, all values supplied in Table S2). These results suggest that although introduced species represent only 4% of the total number of freshwater fishes (a proportional identical to plants (37)), they represent a substantial proportion of the pool of widespread species (Figure S4). Introduced species represent 20% of species with a native occurrence over 10, 48.7% of species with an occurrence over 50, and 100% of species with an occurrence over 200 (Table S3).

Description of introduced species

The most introduced species were the common carp *Cyprinus carpio* (731 introductions), the goldfish *Carassius auratus* (339) and two mosquito fish species (*Gambusia holbrooki* 262; *G. affinis* 254). Despite fewer species originated from the Ethiopian and Neotropical regions, three of them were among the most introduced species (the Mozambique and the Nile tilapia *Oreochromis mossambicus* and *O. niloticus*, and the guppy *Poecilia reticulata*).

We observed a large diversity of locations of introductions, from species introduced only in their native regions to species only introduced outside their native regions (Figure 5). Intriguingly, we found that the majority of the introduced species were introduced in other basins within their native region, hence retaining their endemism (286 species) with respect to major biogeographic regions. Still, a large number of species were introduced outside their native region, and, as such, were no longer endemic to their native region (131 species). The top introduced species (over 100 introductions) were mostly introduced outside their native regions, except for *Trichopodus pectoralis* which was introduced in 115 basins in its native region out of its 120 introductions, and which, consequently, had a low dilution value.

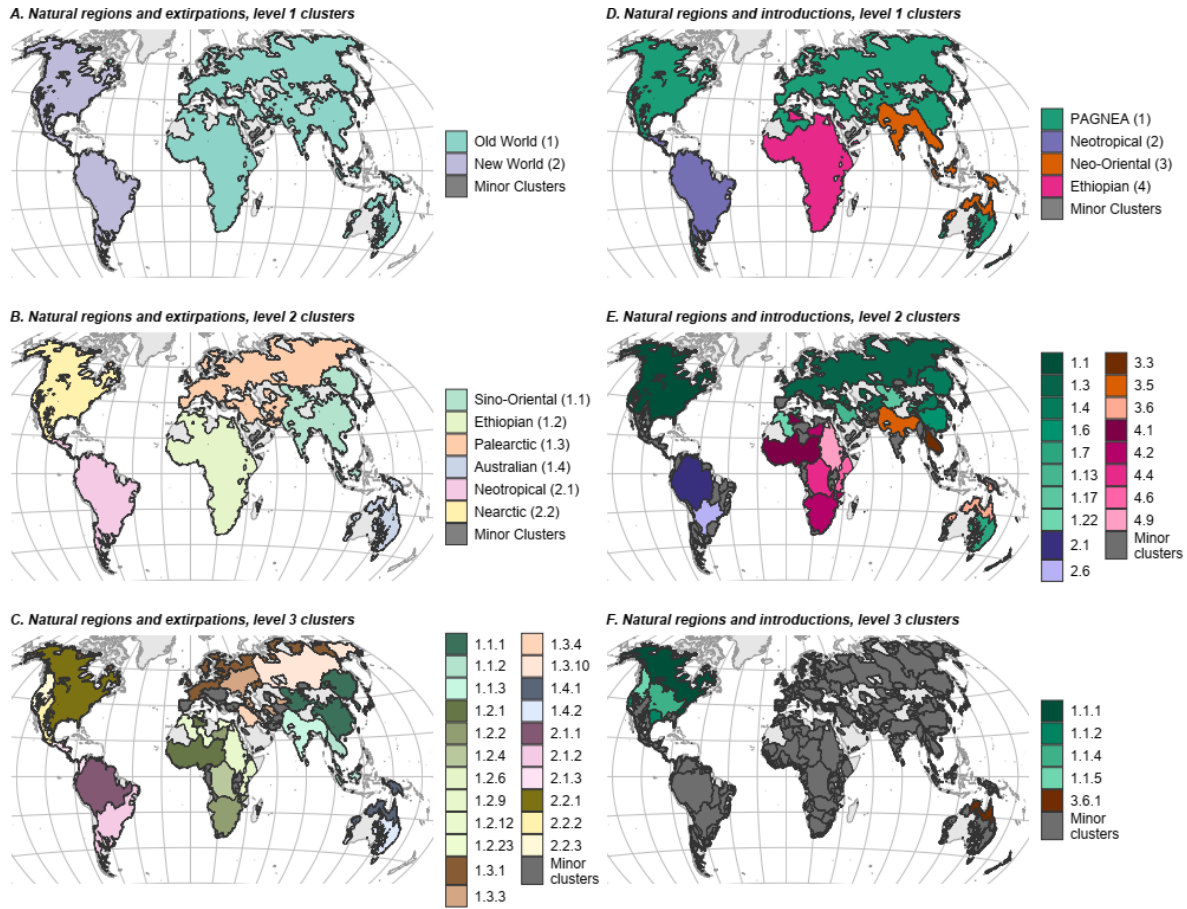


Figure S1. Biogeographical regions of freshwater fishes based on native minus extirpated occurrences (A-C) and native plus introduced occurrences (D-F) at the first three hierarchical levels of the bioregionalisation. Numbers in parentheses indicate (panels A, B, D) indicate cluster codes. PAGNEA in panel D. is the acronym for “Pan-Anthropocenic Global North and East Asia” region.

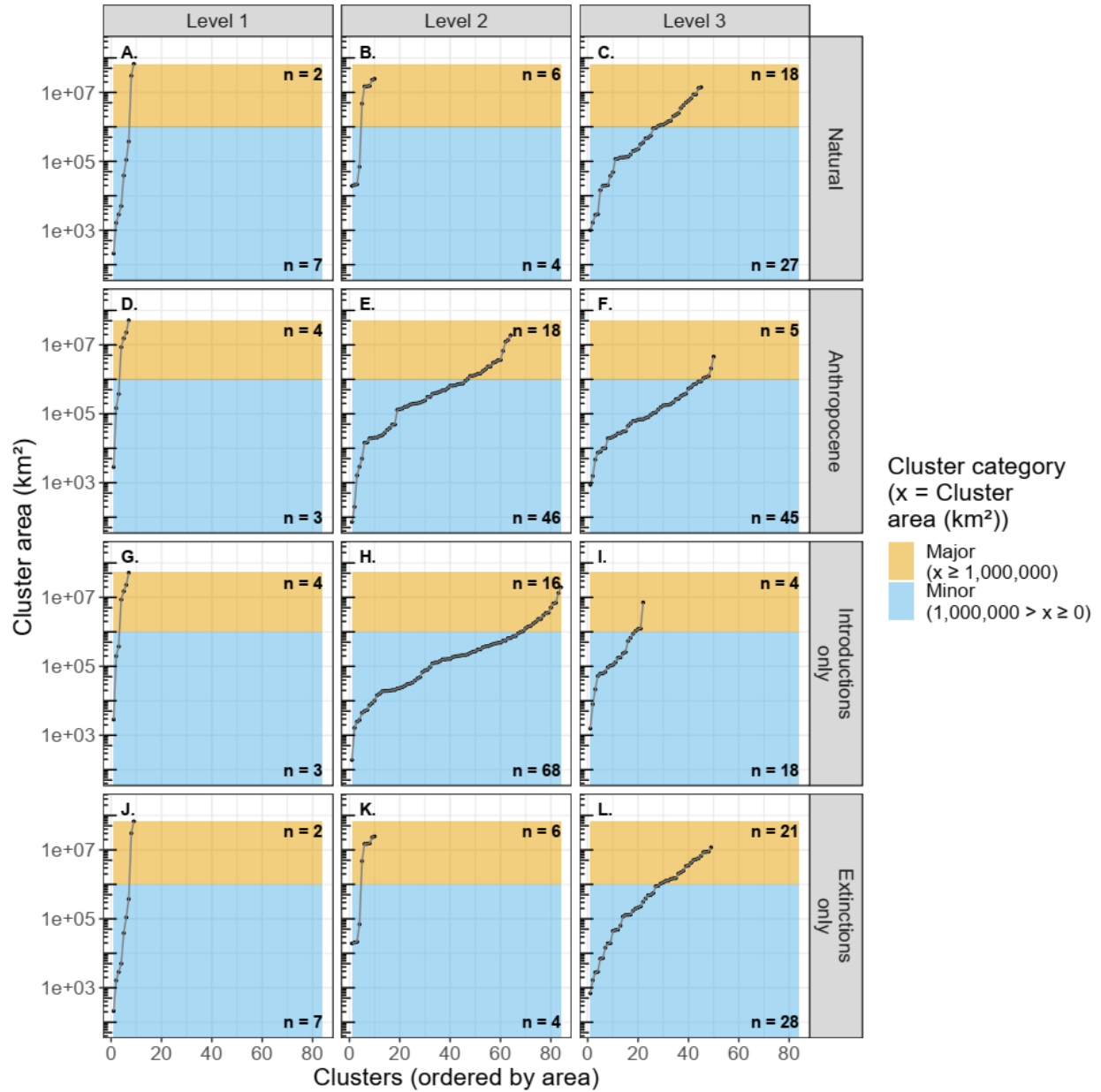


Figure S2. Ordered distribution of cluster area per cluster at each hierarchical level of bioregions, for each pool of species (Natural (A-C), Anthropocene (D-E), Introductions only (G-I) and Extinctions only (J-L)). The two colours on the graph illustrate the difference between major and minor clusters with a cut-off at 1,000,000 km^2 . The numbers on the right of each panel indicate the number of major clusters (yellow part) and minor clusters (blue part).

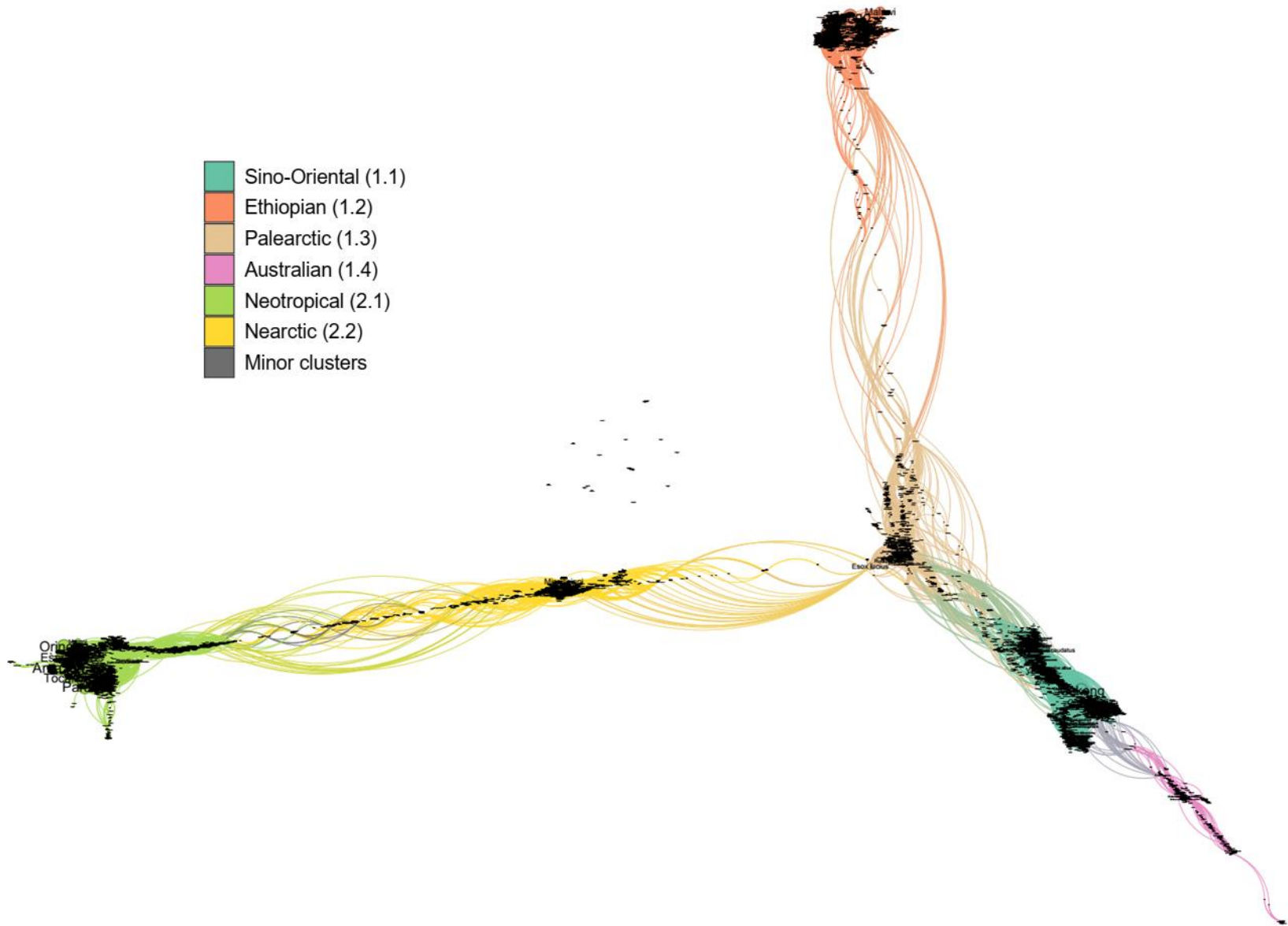
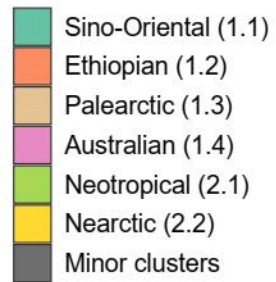


Figure S3 (previous page). Global biogeographical network of freshwater fishes based on the natural pool of species. In such a biogeographical network, nodes depict both species and drainage basins. A link is established between a species and a basin when a species is documented to occur in that basin. This is a very complex network because there is a high number of nodes (13 919 in total, including 11 331 species and 2 588 basins) and links (60 488 occurrences). We spatialised the network – i.e. structured it spatially by aggregating groups of nodes with numerous links in common (biogeographical regions) – with the same parameters as in the study that established the natural biogeographical regions of freshwater fishes (12). We coloured species and basin nodes according to the regions identified with the Map Equation algorithm (Figure 2B).

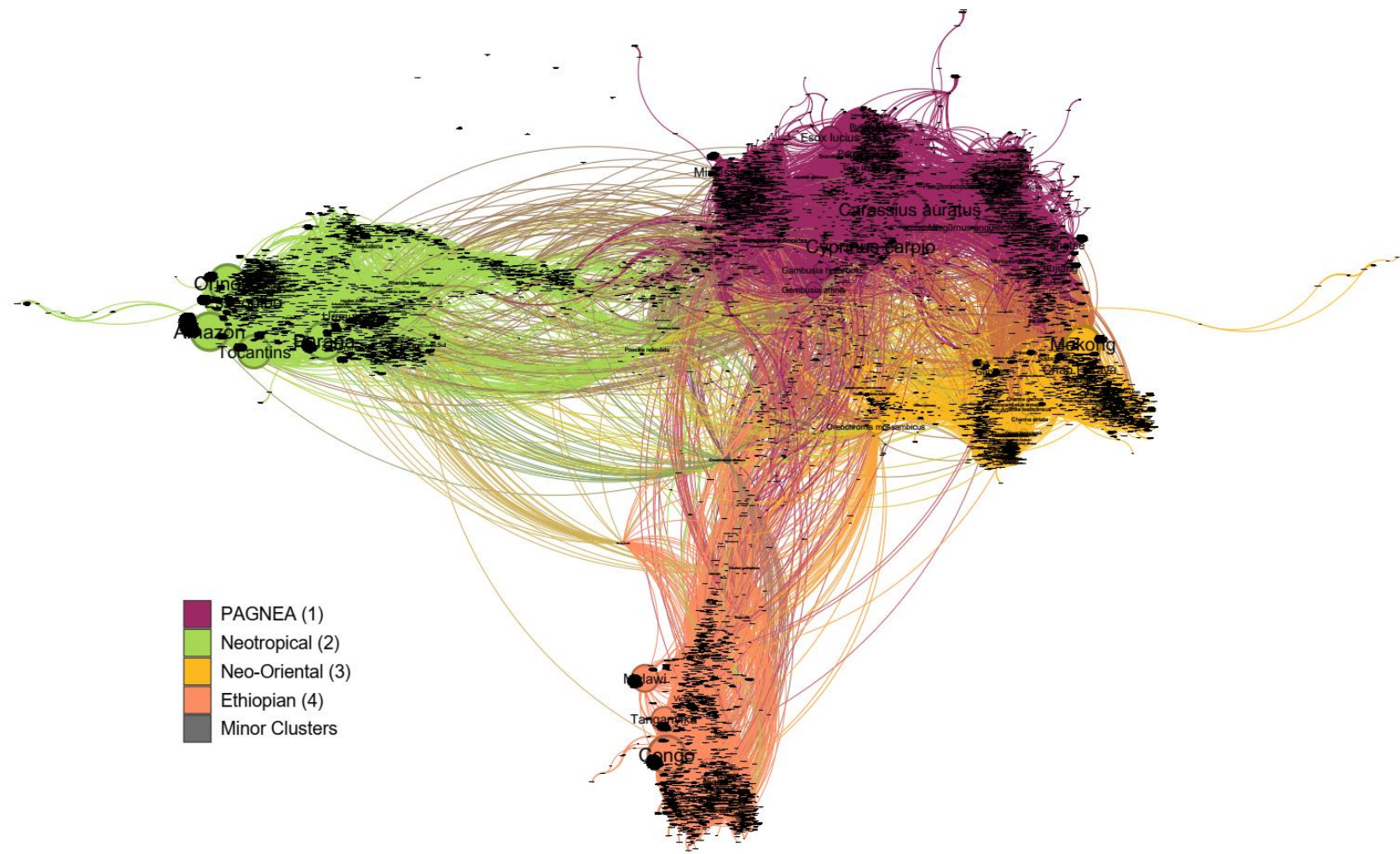


Figure S4. Global biogeographical network of freshwater fishes based on the Anthropocene pool of species. See caption S3 for more details on how this figure was generated.

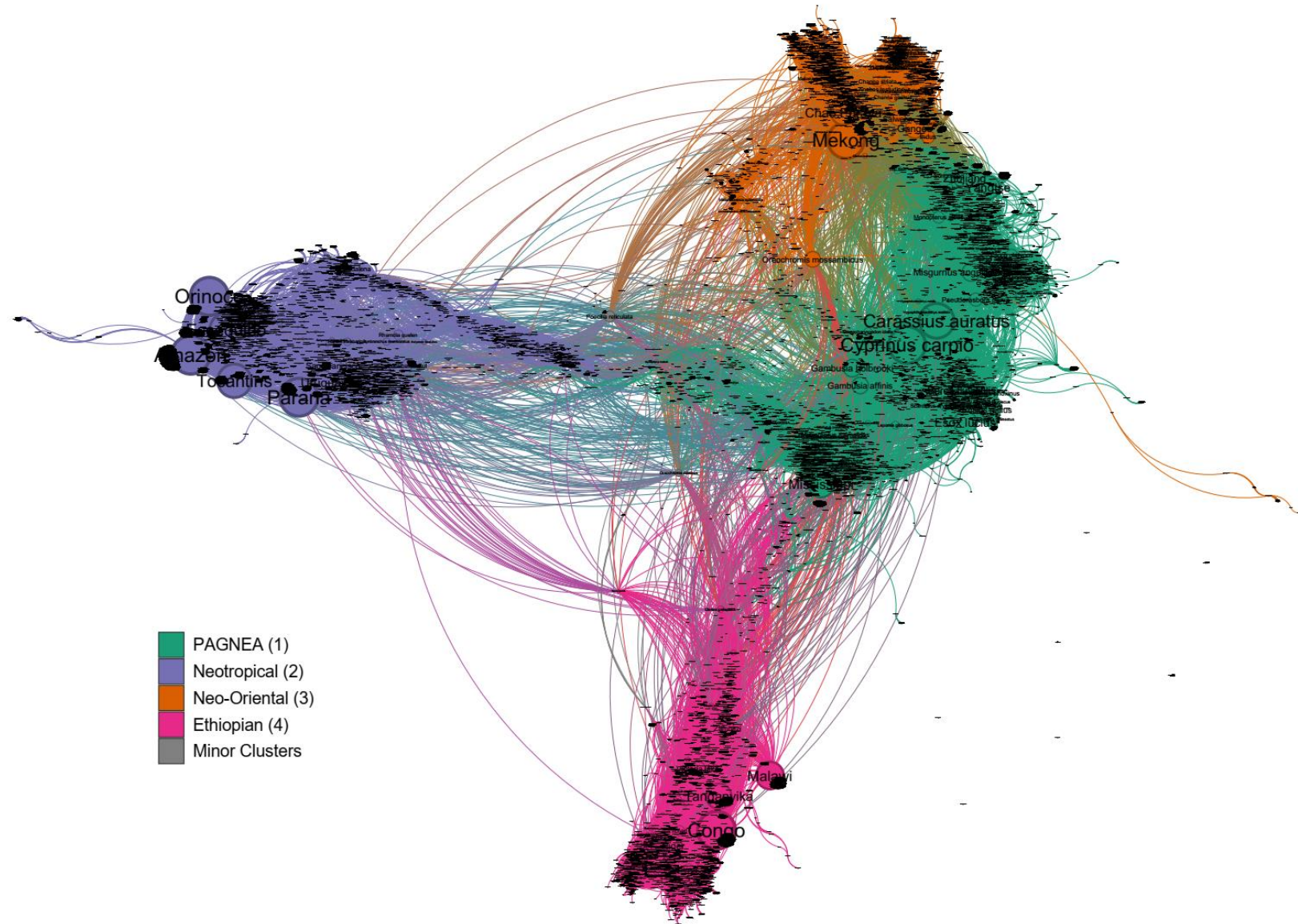


Figure S5. Global biogeographical network of freshwater fishes based on the introduced-only pool of species. See caption S3 for more details on how this figure was generated.

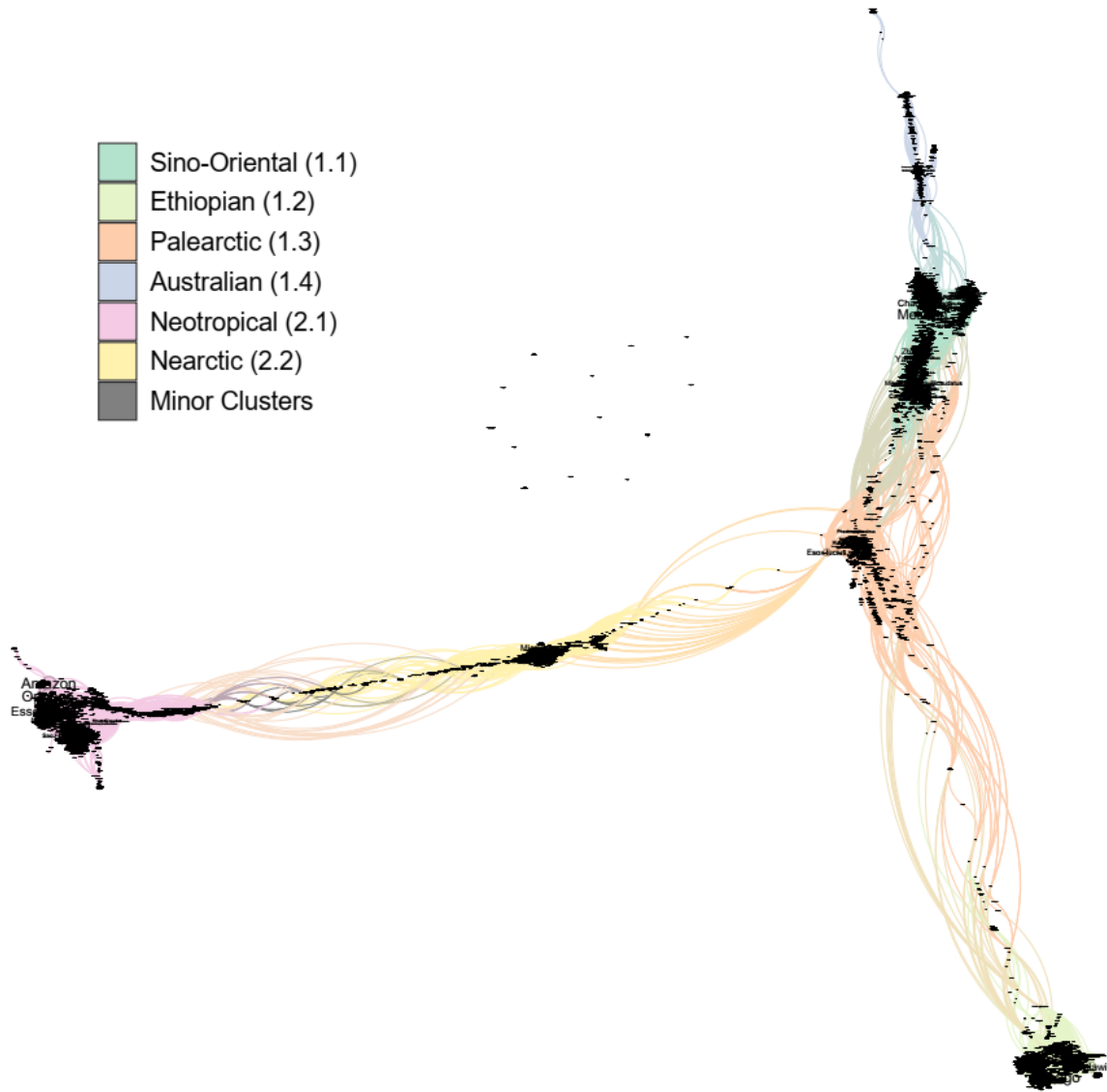


Figure S6. Global biogeographical network of freshwater fishes based on the extinction-only pool of species. See caption S3 for more details on how this figure was generated.

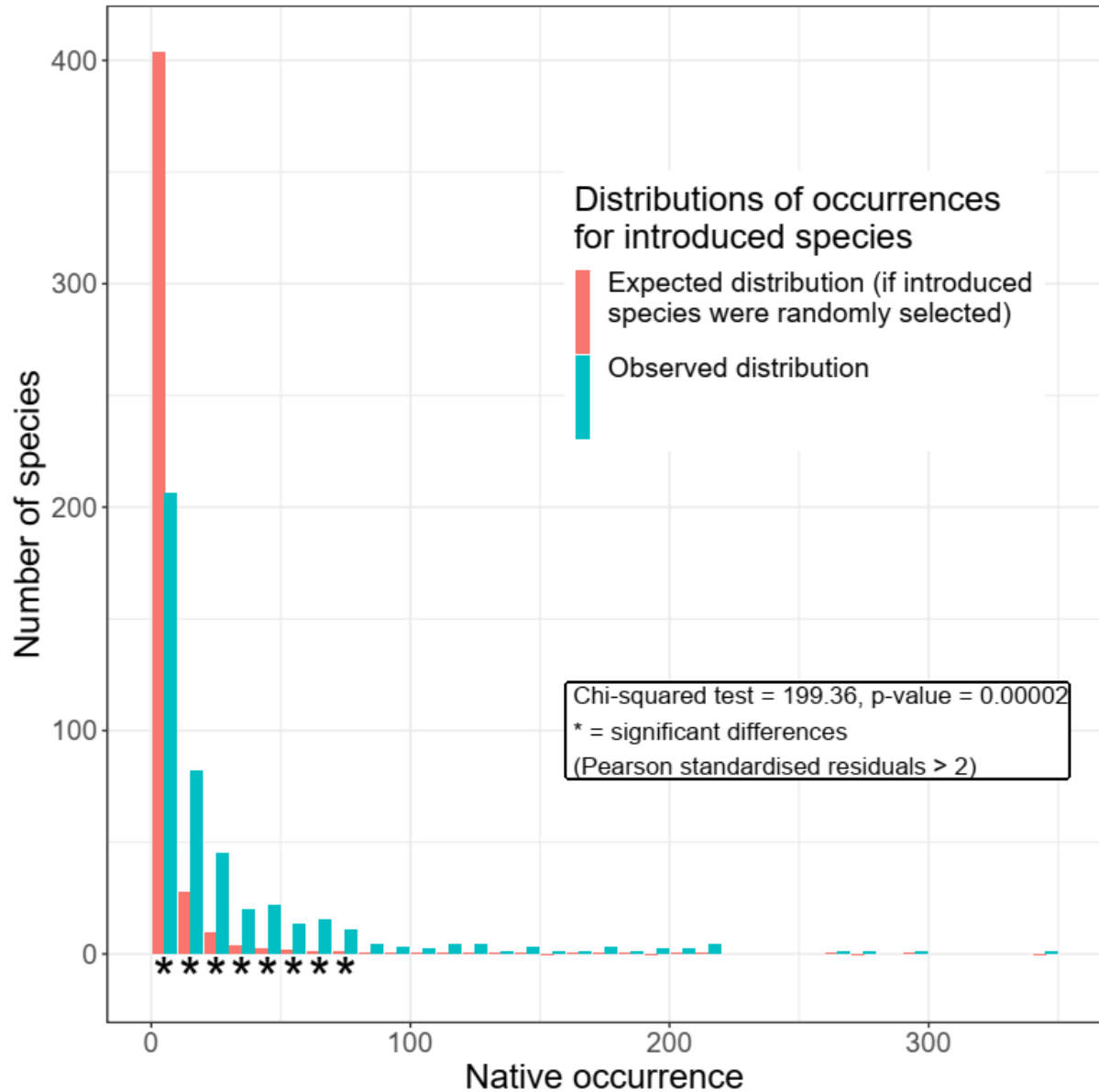


Figure S7 Frequency distribution of native occurrences for introduced species (blue bars) compared with an expected frequency distribution if introduced species were randomly sampled in the total pool of species (orange bars). We compared both distributions with a χ^2 test, we simulated the p-values with a Monte-Carlo simulation with 50 000 replicates. Asterisks indicate significant differences, i.e., standardised Pearson residuals superior to 2. See Supporting Information Text S2 for the methods used to generate the expected number of species.

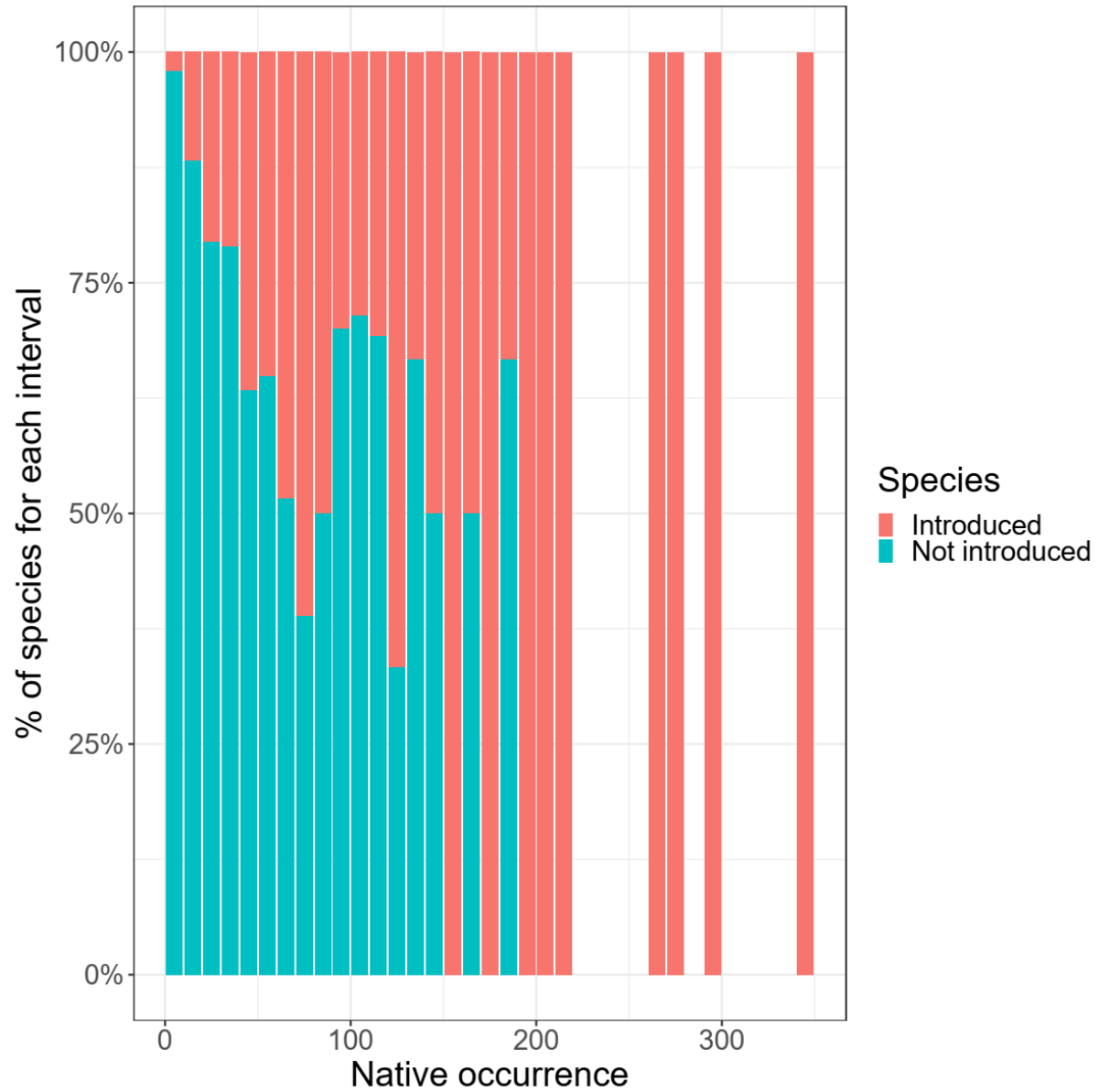


Figure S8 Relative proportion of species that were never introduced versus species that were introduced. Proportions were analysed for each interval of 10 occurrences between 0 and the maximum (values provided in Table S2).

Table S1. Fixed effect estimates from the generalized linear mixed model on the percentage of endemism per drainage basin. The Sino-Oriental was used as the baseline category by the model, hence the interaction Anthropocene:Sino-Oriental is equal to the Anthropocene effect. The model was fitted with function glmer() from the R package lme4, and marginal effects and 95% confidence intervals (CI) were estimated with the function emmeans() from the R package emmeans.

Effects	Estimate	Std. Error	z value	p-value	Marginal effects	Upper 95%CI	Lower 95% CI
Natural: Sino-Oriental	2.48	0.05	51.11	<0.001	0.923	0.916	0.930
Natural: Ethiopian	3.51	0.10	36.37	<0.001	0.971	0.965	0.976
Natural: Palearctic	0.63	0.06	11.39	<0.001	0.653	0.628	0.677
Natural: Australian	2.63	0.12	21.51	<0.001	0.932	0.916	0.946
Natural: Neotropical	3.16	0.07	42.79	<0.001	0.959	0.953	0.965
Natural: Nearctic	3.12	0.09	33.35	<0.001	0.958	0.950	0.965
Anthropocene	-1.60	0.03	-46.07	<0.001	0.707	0.689	0.724
Anthropocene: Ethiopian	0.38	0.09	4.34	<0.001	0.907	0.893	0.919
Anthropocene: Palearctic	0.21	0.06	3.72	<0.001	0.317	0.295	0.340
Anthropocene: Australian	0.21	0.13	1.63	0.103	0.774	0.743	0.802
Anthropocene: Neotropical	0.02	0.07	0.25	0.800	0.829	0.811	0.845
Anthropocene: Nearctic	-0.53	0.08	-6.54	<0.001	0.730	0.701	0.756

Table S2. Statistics for introduced species for all intervals of 10 native occurrences between the minimum and the maximum in our database. Occurrence is here calculated using only the native occurrences (excluding introduced occurrences). See Supplementary Text section “Patterns of natural distribution range of introduced species” for the methods used to generate the expected number of species.

Occurrence intervals	Expected number of species	Observed number of introduced species	Standardised Pearson residuals	Total number of species (native and introduced)	Percentage of introduced species
]0,10]	403.66	206	-14.00	10100	2.04%
]10,20]	27.6	82	5.54	697	11.76%
]20,30]	9.32	45	4.99	219	20.55%
]30,40]	3.41	20	3.47	95	21.05%
]40,50]	2.53	22	3.99	60	36.67%
]50,60]	1.55	13	3.03	37	35.14%
]60,70]	1.11	15	3.49	31	48.39%
]70,80]	0.87	11	2.96	18	61.11%
]80,90]	0.33	4	1.77	8	50.00%
]90,100]	0.39	3	1.42	10	30.00%
]100,110]	0.26	2	1.16	7	28.57%
]110,120]	0.57	4	1.61	13	30.77%
]120,130]	0.23	4	1.84	6	66.67%
]130,140]	0.13	1	0.82	3	33.33%
]140,150]	0.27	3	1.51	6	50.00%
]150,160]	0.02	1	0.97	1	100.00%
]160,170]	0.05	1	0.93	2	50.00%
]170,180]	0.09	3	1.66	3	100.00%
]180,190]	0.11	1	0.85	3	33.33%
]190,200]	0.04	2	1.37	2	100.00%
]200,210]	0.09	2	1.32	2	100.00%
]210,220]	0.19	4	1.87	4	100.00%
]260,270]	0.05	1	0.93	1	100.00%
]270,280]	0.04	1	0.94	1	100.00%
]290,300]	0.05	1	0.93	1	100.00%
]340,350]	0.04	1	0.94	1	100.00%

Table S3. Number and percentage of introduced species above different occurrence cutoffs (0, 10, 50, 100, 200). Occurrence is here calculated using only the native occurrences (excluding introduced occurrences).

Occurrence intervals	Total number of species	Number of introduced species	Percentage of introduced species
]0,350]	11331	453	4.00%
]10,350]	1231	247	20.06%
]50,350]	160	78	47.75%
]100,350]	56	32	57.14%
]200,350]	10	10	100.00%