Appendix S2. Pre-analysis data processing

General procedure

All downloaded data were, whenever necessary, first converted to a long format with columns: plot, date/period, year, taxon, Number (biomass or abundance), using the 'melt' function from the reshape2 package (Wickham 2007) for the R statistical environment (R Core Team 2019). All information at the plot and data set level was transferred to these respective tables. The abundance or biomass data were then appended to the database.

All plots with fewer than 9 years of data were removed from the analysis. Plots with 9 years of data were only included if: other plots within the same data set had at least 10 years, and for two data sets from underrepresented regions: the Israeli butterfly monitoring scheme (DataSource_ID 1481), and beetle trapping in Ukraine (DataSource_ID 1471).

Ideally, the samples in space and/or time were aggregated to provide one yearly sample, from which also changes in taxon richness or species composition could be calculated. This was often not possible due to differences in spatial or temporal sampling effort among years.

For all plots in all data sets that reported multiple taxa or functional groups at each sampling time, we first excluded all worms, mollusks, crustaceans and myriapods wherever possible, and then summed all species per time point to obtain one estimate of assemblage abundance or biomass using the 'dcast' function from the reshape2 package (Wickham 2007).

For all data sets for which standardization of trapping effort was necessary, the standardization procedure is described below:

DataSource_ID 79: GPDD – Butterfly Monitoring Scheme 1976 – 1985 (UK)

Taxonomic focus: Lepidoptera: Rhopalocera
Number of plots: 22.
Values in database: abundance per transect per year.
Data processing: Some species were not consistently reported in the source data, so we removed all years with missing species. We also excluded years in which 0 butterflies recorded in the entire year. After this exclusion process, we excluded 15 plots of too short duration.

Datasource_ID 249: BioTIME – Light trap on the roof of the NHM in Copenhagen

Taxonomic focus: Lepidoptera and Coleoptera

Number of plots: 1

Values in Database: abundance per year

Data processing: We retained only the months June, July, August, because the other months were not sampled consistently. Years in which no summer samples were taken, including the first year of sampling (1992), were removed. The summer months were summed to provide one value per year.

DataSource_ID 294: BioTIME – Butterflies in Tam Dao (Vietnam)

Taxonomic focus: Lepidoptera: Rhopalocera

Number of plots: 6

Values in database: summed abundance per plot per date

Data processing: One plot (T7) was removed because the sampling duration did not match our inclusion criteria.

DataSource_ID 300: Ladybird beetles at Kellogg station LTER site (USA)

Taxonomic focus: Coleoptera: Coccinellidae

Number of plots: 51.

Values in database: abundance per plot per month.

Data processing: Sampling of insects using yellow sticky traps (5 traps per plot, 51 experimental plots with different land-use types). We use here only the lady beetle data (Coleoptera: Coccinellidae), because other taxa were not identified consistently throughout the sampling period. The months April, May, September and October were only sampled in a few years and were therefore removed. We then summed across species per month.

DataSource_ID 375: BioTIME – ground-dwelling beetles in Japan

Taxonomic focus: Coleoptera

Number of plots: 22.

Values in database: abundance per plot per date.

Data processing: This data set included both abundance and biomass for all individuals. Each site consisted of 5 plots (in a few cases more than 5), with each 4 traps. Seven sites with less than 9 years data were removed, as were all plots that were not sampled consistently (4 plots in 4 sites in total). This left 26

plots at 19 locations (one site contained 7 plots, and one site contained 2 plots). The sum of all plots per site per sampling date was included in the database.

DataSource_ID 465: GPDD – Light trapping near Prague (Czech Republic)

Taxonomic focus: Lepidoptera.

Number of plots: 1.

Values in database: abundance per plot per year.

Data processing: Three years (1977-1979) were removed, as not all species were identified during these years. In all other years we summed across all species.

DataSource_ID 478: Stream insects in the Breitenbach (Germany)

Taxonomic focus: Plecoptera, Ephemeroptera, Trichoptera

Number of plots: 5.

Values in database: abundance per plot per year.

Data processing: The data were downloaded as 6 species × date matrices, which we transformed to the long format. After removing one plot because its sampling duration was less than 10 years, we took the sum of all dates and all species per year for the remaining plots.

DataSource_ID 313: BioTIME – Cedar creek grasshoppers (USA)

Taxonomic focus: Orthoptera: Acrididae

Number of plots: 21

Values in database: abundance per plot per month.

Data processing: We added 17 missing zeroes for sampling dates when no grasshoppers were observed but a survey was known to have been conducted. We then summed all species sampled per month per plot.

DataSource_ID 502: GPDD – Aphid monitoring (UK)

Taxonomic focus: Hemiptera: Aphidoidea

Number of plots: 16

Values in database: abundance per plot per month.

Data processing: One plot was excluded because of its too short duration and we removed four data

points with erroneous dates. We took the sum of all species for each monthly sample.

DataSource_ID 1006: Environmental Change Network – moths (UK)

Taxonomic focus: Nocturnal Lepidoptera

Number of plots: 13.

Values in database: monthly abundance per plot, per 14 nights.

Data processing: The original data set contains 12 sites across the UK, with each one or 2 light traps sampling moths (nocturnal Lepidoptera). The raw data mostly presented all moth individuals caught per trap per night, but at other times the data are summed over more than one night (i.e. the trap was emptied after several nights running). For calculating the abundance, we first removed error code 101 from the 'counts' column. We only retained data from the months June, July and August, because only in these months sampling effort was consistent over time. All other months showed artifactual increases in moth abundance over time due to increasing selectivity in the nights at which sampling took place. We calculated the abundance of moths per 14 trapping nights for the summer months. For months with more than 14 trapping nights, we subsampled 14 nights from the available data and summed all moths over these nights with 100 iterations. We took the mean abundance of these 100 iterations as model input.

DataSource_ID 1261: Ants in Harvard Forest LTER site (USA)

Taxonomic focus: Hymenoptera: Formicide

Number of plots: 8.

Values in database: minimum abundance of ant nests per plot per date.

Data processing: Ants were sampled using pitfall traps in 8 plots, with each a manipulation treatment of the Hemlock trees (girdling of Hemlock, Logging of Hemlock, Hemlock control and hardwood control). Other sampling methods were used, but only the pitfall traps were used consistently over time. The number of sampling dates varied among years. We first added zero's for dates on which sampling took place, but no ants were present, after consulting the data owner. We also corrected the sampling date for two sampling events. To avoid overcounting ant abundance due to nest proximity, we used number of ant species per trap as a proxy for the number of nests. We summed the number of nests over all traps per plot. This is thus the lower bound of the number of ant nests per plot, since ants of the same species could have come from different nests.

DataSource_ID 1263: Environmental Change Network - Butterflies (UK)

Taxonomic focus: Lepidoptera: Rhopalocera

Number of plots: 13.

Values in database: abundance per transect per date.

Data processing: Twelve sites across the UK, with each one or 2 butterfly transects (Pollard walks), sampled at more or less biweekly intervals. Site Cairngorms (T12) was excluded, because it was not sampled for 10 years. Section 4 in Alice Holt was excluded because it was only added to the other 3 sections after 10 years of sampling. We added up all sections per transect.

DataSource_ID 1266: Environmental Change Network- spittlebugs (UK)

Taxonomic focus: Hemiptera: Aphrophoridae

Number of plots: 16.

Values in database: abundance per plot per year.

Data processing: Twelve sites across the UK, with at each site one or 2 grassland plots sampled for spittle bug nymphs using randomly placed quadrants. We used only the Data set of spittlebug nymphs, because the data of adult spittlebugs was not collected in a way that allows the calculation of density changes over time. The subplots were summed to derive one value per plot per year.

DataSource_ID 1267: Environmental Change Network- ground beetles (UK)

Taxonomic focus: Coleoptera: Carabidae

Number of plots: 36.

Values in database: mean abundance per plot per year.

Data processing: Twelve sites across the UK, with each 3 plots, each consisting of 10 pitfall traps to catch ground beetles (Coleoptera: Carabidae). Samples were taken at approximately 14-day intervals, but this was somewhat variable. We first removed all missing dates (Qcodes). Sampling effort, considered as number of traps and number of days per year, was not equal among sites or years. To standardize trapping effort, we calculated the number of trap-days per plot per year. Based on this, we chose for all but one plots a minimum of 154 days of trapping within the growing season (May – October) per year necessary for a reliable estimate of total abundance. The site T12 in the Scottish uplands (Cairngorms) had a shorter sampling season (June – October) therefore we took 126 trap-days as minimum trapping effort per year. This still removed 5 years with lower trapping effort from 3 plots. For plots and years with more than 154 days of trapping, we removed the last 2 weeks of October to homogenize trapping effort. Because the

traps were emptied biweekly, trapping effort could not be 100% equal among plots and years. The following table shows the percentage of plot-years with varying sampling effort:

Sampling days	151	153	154	155	156	157	158	160
% yearly samples	0.53	1.77	87.79	5.84	1.06	0.35	1.062	1.59

To account for randomly missing traps within plots (e.g. due to trampling or leakage, etc.), we averaged beetle abundance over the functioning traps for each plot per date, and summed these over the year.

DataSource_ID 1319: Grasshoppers at Sevilleta LTER site, New Mexico (USA)

Taxonomic focus: Orthoptera: Acrididae

Number of plots: 4.

Values in database: abundance per plot per date.

Data processing: Grasshopper transects at 4 sites, with each 5 'webs', each consisting of 6 50-m transects (in total 30 transects per site per sampling date). Two samples were taken per year, one in spring and one in autumn. Because sampling effort was consistent across years, we summed all transects per site per date.

DataSource_ID 1339: Grasshoppers at Bison Range, Montana (USA)

Data: https://belovskylab.nd.edu/national-bison-range-ltreb-database/survey-data/grasshopper-data/

Taxonomic focus: Orthoptera: Acrididae

Number of plots: 4.

Values in database: mean density per plot per date (half month timesteps)

Data processing: The data were downloaded from the Belovsky lab website (https://belovskylab.nd.edu/national-bison-range-ltreb-database/survey-data/grasshopper-data/), file: Survey Grasshopper Density Data. The data were downloaded as a PDF file, created from an Excel date × species matrix. After converting this back to an excel sheet using the website pdftoexcel.com, we extracted the density of grasshoppers data for all dates (column 'DENSITY (#/m^2)') for the 4 plots with data for over 10 years. These values were used as input for all models.

DataSource_ID 1345: Ground dwelling arthropods at Sevilleta LTER site, New Mexico (USA)

Taxonomic focus: all arthropods. Excluded here: taxa not belonging to the classes Insecta, Arachnida or Entognatha.

Number of plots: 3.

Values in database: abundance per plot per year

Data processing: Three plots, with each 15 pitfall traps over the period 1992-2004. In early years, 30 traps were placed, but this was later reduced to 15 traps (traps with numbers 2,4 and 6 were removed). We thus removed these traps throughout the data to retain equal sampling effort. Trapping occurred year-round, and samples were collected at variable intervals, usually every 2 or 3 months. Zero counts for a given trap at a given date were not recorded, and are thus absent from the data. Only at 3 dates over 13 years was one out of 15 traps reported as broken or missing. As these 3 missing traps constitute a minimal number of individuals compared to the total catch at each plot, we ignored these missing traps and we took the sum across all sampling months and traps per plot per year

Mistakes found in data set: Year and month had partly information about insects in that column = column "comments'. Non-numeric values in Year column were removed. This did not remove any insect counts. Plot B had a too short sampling duration and was removed.

Variable clean up:

```
data$Genus[data$Genus == -888]<-""

data$Species[data$Species == -888]<-""

data$Family[data$Family == -888]<-""

levels(data$Order)[levels(data$Order) == -888]<-"NONE"

data$Order[is.na(data$Order )]<-"NONE"

data$Count[data$Count == -888]<-0

unique(data$Line) # stng weird in Line

data$Line[data$Line == "A "]<-"A"

data$Order[data$Order == "CO "]<-"CO"
```

Exclude some data:

```
data1<-subset(data, Taxon != "NS NS NS NS") # no arthropods collected

data1<-subset(data1, Comments != "sample in bad condition. All bugs
were broken so they weren't saved.")
```

```
data1<-subset(data1, Comments != "lots of individuals brokenand not identifiable")

data1<-subset(data1, Comments != "cup missing.")

data1<-subset(data1, Comments != "no sample taken")
```

DataSource_ID 1346: Ants at Sevilleta LTER site, New Mexico (USA)

Taxonomic focus: Hymenoptera: Formicidae

Number of plots: 6.

Values in database: number of ant nests per treatment per plot per date

Data processing: Ant nests (Hymenoptera: Formicidae) were counted at 2 sites. Each site had 4 blocks with each 3 plots. These plots had one of 3 treatments: Control, Lagomorph exclosure, or rodent exclosure. We changed the erroneous sampling date 10/6/2002 to 10/6/2003 for several plots, matching the sampling dates for the other plots. We summed ant nests across all 4 plots per treatment per site for analysis.

DataSource_ID 1347: Benthic macroinvertebrates at North Temperate Lakes LTER site, Wisconsin (USA)

Taxonomic focus: all freshwater invertebrates, excluded here: taxa not belonging to the classes Insecta, Arachnida or Entognatha.

Number of plots: 14.

Values in database: abundance per plot per year.

Data processing: Three lakes with a variable number of plots within each lake, were sampled for aquatic macroinvertebrates once per year from 1981-2015. At each plot 3 samples should have been taken at each sampling date, but this was evidently not the case, and in the data no distinction was made between true absences of any invertebrates and no sample taken. We solved this by adding 0's for all true absences of species at sampling dates when other species were collected. After this, we averaged the samples per date, and summed the value per date over all species.

DataSource_ID 1349: Ground-dwelling arthropods at Phoenix, Arizona LTER site (USA)

Taxonomic focus: all arthropods. Excluded here: taxa not belonging to the classes Insecta, Arachnida or Entognatha.

Number of plots: 29.

Values in database: abundance per plot per sampling date

Data processing: 29 plots with each 10 or 12 pitfall traps were sampled from 1998-2016. The number of sampling dates was not fully consistent over years. We removed plots with <10 yrs data from the data set (leaving 300 traps). We corrected misspelled taxon names, and removed 20 duplicates. Dates were with fewer than the full complement of traps (10 or 12) were removed. The sum of all traps per date was taken for all taxa.

DataSource_ID 1351: Freshwater invertebrates in Sycamore creek, Arizona LTER site (USA)

Taxonomic focus: all freshwater invertebrates. Excluded here: taxa not belonging to the classes Insecta, Arachnida or Entognatha.

Number of plots: 1.

Values in database: abundance per sampling date.

Data processing: Freshwater invertebrates were sampled in Sycamore Creek at irregular intervals in spring and summer from 1985-1999.

DataSource_ID 1357: Canopy arthropods in Luquillo LTER site Canopy Trimming Experiment (Puerto Rico)

Taxonomic focus: all arthropods. Excluded here: taxa not belonging to the classes Insecta, Arachnida or Entognatha.

Number of plots: 34.

Values in database: abundance / foliage weight per tree species per plot per date.

Data processing: Arthropods were collected from branches of 8 tree species in 12 plots where hurricane damage was simulated (2004-2016). Four plots in each of 3 blocks, sampled once or twice per year. Tree species and blocks with less than 10 years data were excluded. Because each tree species harbors a unique invertebrate fauna, we maintained the separation of the data per tree species and plot.

DataSource_ID 1361: Grasshoppers at Georgia coast LTER site (USA)

Taxonomic focus: Orthoptera: Acrididae

Number of plots: 7.

Values in database: abundance per plot per date.

Data processing: Grasshoppers were sampled at 8 plots, once per year in July or August. We

standardized trapping effort by only using the first 9 transects per plot. We summed all observations across the 9 transects per plot to obtain one value per date per plot. One plot has only one grasshopper observation over the full sampling period, and one plot has 0.

DataSource_ID 1364: Grassland arthropods in Cedar Creek LTER site BEF experiment (USA)

Taxonomic focus: all arthropods, excluded here: taxa not belonging to the classes Insecta, Arachnida or Entognatha.

Number of plots: 172.

Values in database: Abundance per plot per date.

Data processing: Invertebrates were sampled in the Cedar Creek plant biodiversity manipulation experiment from 1996 from 2006 using a sweep net. Each of 172 plots was sampled by 25 net sweeps, between 1 and 3 times per year (June, July and/or August). Any mollusks were removed from the data before aggregation. The original values per date were retained for this analysis.

DataSource_ID 1398: Beetles on spoil heaps in Ostrava (Czech Republic)

Taxonomic focus: Coleoptera

Number of plots: 3.

Values in database: abundance per plot per year

Data processing: Because of unequal sample sizes, all species counts in 1975 and 1976 were multiplied by 5/7. In these years there were 7 traps installed, whereas there were 5 in other years.

DataSource_ID 1404: Arthropod sampling in Greenland

Data: https://data.g-e-m.dk/

Taxonomic focus: all arthropods. Excluded here: taxa not belonging to the classes Insecta, Arachnida or Entognatha.

Number of plots: 6.

Values in database: Pitfalls: mean daily abundance over all active traps per plot per sampling date. Window trap: total abundance per sampling date.

Data processing: One window trap and 6 pitfall stations sampled from 1996-2017. We removed the data from plot "Art6" because sampling period was shorter than 10 years. We also removed all data from 2010 because samples were lost, and from 2011 because in an older version of the database these data

contained errors, and our code was not updated to reflect the newer version of the data. For the pitfall traps, sampling effort varied in sampling interval, the number of traps used, and the sampling period per year, so we calculated mean values per sampling date: we first divided the abundance per trap by the number of days the trap was active for each sampling period. The active traps were then averaged to produce one value per plot. For the window trap, we summed the 4 wind directions per sampling date.
Processing code: https://zenodo.org/record/3691682

DataSource_ID 1416: Caterpillars in Hubbard Brook LTER site, New Hampshire (USA)

Data: http://data.hubbardbrook.org/data/dataset.php?id=82
Taxonomic focus: Lepidoptera
Number of plots: 4.
Values in database: abundance and biomass per plot per year.
Data processing: We only included data from 2 of the 4 tree species sampled (beech and sugar maple), because the others were not sampled long enough. Sampling of both tree species took place at 4 sites. We replaced the tree species codes with species names, the plot numbers with site names, and the taxon codes with family names. There was no consistent sampling in sampling period 6, and therefore we excluded this period. Because the 2 tree species were sampled equally and consistently in all years and all sites, we summed the caterpillars collected on both tree species.
Processing code: https://zenodo.org/record/3691682

DataSource_ID 1429: Freshwater invertebrates in Lake Oneida, New York (USA)

Taxonomic focus: all freshwater invertebrates, excluded here: taxa not belonging to the classes Insecta, Arachnida or Entognatha.
Number of plots: 2 (depth groups).
Values in database: number per depth per season
Data processing: The data were provided as season (spring, summer, fall) × depth-group ('shallow' or 'deep') × species matrix. We first replaced the common names with scientific names to ease later exclusion of non-insect invertebrates. We then converted the matrix to the long format (one taxon observation per line, separated by depth group and season), and replaced all counts marked as '-999' with 'NA'.

DataSource_ID 1430: Freshwater invertebrates in the Green river, Utah (USA)

Taxonomic focus: all freshwater invertebrates, excluded here: taxa not belonging to the classes Insecta, Arachnida or Entognatha.

Number of plots: 3.

Values in database: Abundance per plot per sampling date.

Data processing: We selected only those sites that had repeated samplings at the exact same location, and were sampled using consistent methods over time. This left 3 sites for analysis.

DataSource_ID 1444: Freshwater invertebrates in New Zealand

Taxonomic focus: all freshwater invertebrates, excluded here: taxa not belonging to the classes Insecta, Arachnida or Entognatha.

Number of plots: 66.

Values in database: Abundance per plot per sampling date.

Data processing: On GBIF, 2 data packages were available: the period 1989-2008 and 2009-2015. Because it was unclear why these periods were separated, and no guarantee of methodological consistency was available, we only used the data for the period 1989-2008. We excluded all plots with less than 10 years data, leaving 66 plots, located at 44 rivers. One sample was taken per year.

DataSource_ID 1488: Freshwater invertebrates in Sweden

Taxonomic focus: all freshwater invertebrates, excluded here: taxa not belonging to the classes Insecta, Arachnida or Entognatha.

Number of plots: 264.

Values in database: Ekman and Van Veen grabs: abundance/ m^2 per plot per date and biomass/ m^2 per plot per date; Kick samples: mean abundance per plot per sampling date.

Data processing: We manually downloaded all data from plots that showed 10 or more years of monitoring on the website API of the monitoring program (https://miljodata.slu.se/MVM/Search), totaling 384 plots, and appended these to form one large file. It is possible that some plots were missed. From the middle 1990's onwards, a standard taxon list was used, but there was no guarantee that all taxa had been counted at all plots and in all years before this time. Therefore, we discarded all data before 1996. We selected data with reliable, quantitative methods: Kick sampling (Swedish: 'sparkhåv' and Kicknet) and Ekman / Van Veen grabs (standardized to catch per m^2). Surber samples were removed, because they were only taken in a few years. We first split the data by sampling method (Kick sample or Ekman/Van

Veen grab) and the Ekman data again by metric of assemblage size (biomass or abundance). The kick samples only reported mean abundance per sample per date. We removed all plots with shorter than 10 years sampling periods within each of these groups. This left 264 plots. Three plots reported both kick samples and Ekman grab data, so we assigned separate plot ID for these methods in order not to mix them up.

DataSource_ID 1487: Canopy arthropods in Luquillo LTER site (Puerto Rico)

Taxonomic focus: all arthropods, excluded here: taxa not belonging to the classes Insecta, Arachnida or Entognatha.

Number of plots: 65.

Values in database: abundance ÷ foliage weight, per tree species per plot per date.

Taxonomic focus: Canopy arthropods

Data processing: Canopy arthropods were sampled by collecting foliage from 16 tree species in canopy gaps and in undisturbed canopy from 1991-2016. We found that all data from 2002-2004 had been duplicated and we removed all duplicate values. We excluded the following data because of data inconsistencies: Block 0, Foliage weight = 0, Year = 2000 (no insect data reported). We only included the consistently sampled treatments U (undisturbed) and G (gap), and only 6 tree species that were consistently sampled per plot. Tree species and blocks with less than 10 years data were excluded. Because each tree species harbors a unique invertebrate fauna, we maintained the separation of data per tree species and plot.

DataSource_ID 1491: Freshwater invertebrates in the middle Doce River Basin, Brazil.

Taxonomic focus: all freshwater invertebrates, excluded here: taxa not belonging to the classes Insecta, Arachnida or Entognatha.

Number of plots: 13

Values in database: mean abundance per plot per date.

Data processing: Three plots with less than 10 years of data were removed from the data set. In the remaining plots, sampling effort was not equal over time, so the mean of all samples was taken per sampling event per plot.

DataSource_ID 1501: Butterfly monitoring scheme Ireland

Taxonomic focus: Lepidoptera: Rhopalocera

Number of plots: 2

Values in database: abundance per plot per sampling date.

Data processing: Data obtained from GBIF. Only 2 sites had consistent data for over 10 years. Because the number of sampling events was not consistent among years, we retained the data per sampling date. The separate sections reported for the transect at WX01 - Raven Nature Reserve in 2013 were summed to maintain consistency with the other years.

DataSource_ID 1520: Mosquito monitoring in Iowa (USA)

Taxonomic focus: Diptera: Culicidae

Number of plots: 27.

Values in database: mean daily number of mosquitoes per plot per month.

Data processing: Using a custom python script, we downloaded the yearly data for the 27 sites with 10 or more years of data from the website https://mosquito.ent.iastate.edu/ in September 2019. We appended all 664 date × plot × species matrices and created columns for the month and year of sampling. Since the data for 2019 were not yet complete, and for none of the other data sets data from 2019 were available, we removed this year from the data from the data. We then transformed the matrix into the long format (7.443.220 records), and then took the sum over all species for each date. Because sampling effort was not equal in each month (occasional missing samples or periods), we divided the monthly number of mosquitoes by the number of sampling dates (i.e. for each month the daily mean abundance was used as model input).

DataSource_ID 1525: Mosquito monitoring in Indiana (USA)

Taxonomic focus: Diptera: Culicidae

Number of plots: 13.

Values in database: daily mean per plot per month.

Data processing: All data for the state of Indiana were downloaded from the population database of www.vectorbite.org. All plots with less than 9 years of data were excluded. For the remaining 13 plots, we summed over all species for each date, and then averaged these daily counts per month.

DataSource_ID 1526: Mosquito monitoring in Florida (USA)

Taxonomic focus: Diptera: Culicidae

Number of plots: 32.

Values in database: daily mean per plot per month.

Data processing: We followed the same procedure as for the Indiana mosquitos (DataSource_ID 1526): All data for the state of Florida were downloaded from the population database of www.vectorbite.org. All plots with less than 9 years of data were excluded, as were duplicate values. For the remaining 32 plots, we summed over all species for each date, and then averaged these daily counts per month.

Ecoregions of the world

We intersected the coordinates of all plots with the WWF Terrestrial Ecoregions of the World (TEOF) database (Olson et al. 2001). For plots marked as lakes, and for coastal plots not available in the database, we manually filled out the terrestrial ecoregion and climatic zone based on the online map.

World Database on Protected Areas

We intersected the coordinates of all plots with the World Database on Protected Areas (IUCN and UNEP-WCMC 2018). Protection status was determined by whether a plot was included in this database, regardless of the level of protection. This provided a binary explanatory variable. For 207 plots (12% of the total), a protection status was awarded or its level upgraded during or after the sampling period, but because protected status is typically only awarded to high quality ecosystems, we assumed that before the year of protection, these plots were already of high quality. For analysis, we thus designated 'protected' status to all plots that received protected status at any time before, during or after sampling.

LUH2 (Land Use Harmonization)

We obtained estimates for landscape-scale land-use surrounding the sampling plots from the land-use harmonization (LUH2) database (Hurtt et al. 2018) at 0.25°×0.25° resolution for all years. We extracted the cover of urban land and crop land (of any crop type). To do this, we reprojected the LUH2 to the Eckert IV equal-area map projection and extracted the raster values intersecting with the coordinates of the plots. Because this data set provided data for all years, we calculated both the in-situ

land use at the end of the sampling period, and the changes in land use between the first and last year of sampling.

ESA CCI

We obtained data on local scale land-use at the sampling plots from the ESA CCI database (ESA 2017) (300 × 300m resolution per cell) for the period 1992–2015. This database provides one land-use category for each cell rather than % cover. To account for uncertainty of the provided coordinates, we extracted the land-use information over the 9 cells surrounding each plot, providing data at 900 × 900 m resolution. Over these 9 cells, we calculated the percent urban cover at the plots by dividing the number of cells with an urban classification (category 190) by 9. We calculated crop cover (categories 10, 11, 12, 30 and 40) over the 9 cells in a similar way, but because categories 30 and 40 are mixed cropland and natural vegetation, we down-weighted these to 0.75*number of cells of category 30, and 0.25* number of cells of category 40. The full calculation for crop cover was thus: (# cells in categories 10, 11 and 12 + #cells cat 30× 0.75 + #cells cat 40× 0.25) / 9.

The time series of very few plots fell fully within the timeframe for which these data are available (1992 – 2015), thus we only used the values of the last year of sampling as an indication of in situ land use, assuming that little to no change occurred over the sampling period. This was likely to be reasonable, since over the full period of 1992-2015 only 5% of the plots changed their classification. Plots where sampling ended before 1987 were marked with 'NA'.

Landscape scale climate change - CRU

We calculated changes in mean monthly temperature and precipitation at landscape scales from the weather data of the CRU database (Harris et al. 2014), at 0.5° × 0.5° resolution for the full period. The weather came averaged over 1 to 3 cells, depending on whether the plot was located at the center or edge of the main grid cell. For each plot, we calculated the yearly change in monthly mean temperature and precipitation over the sampled period using generalized additive models, accounting for seasonal patterns with a spline term on month (T or P ~ Year + spline(Month)). From these models, we obtained the slope of temperature and precipitation change, and standardized these to decadal change for ease of interpretation (slope × 10).

We obtained relative climate change values by dividing these decadal change values by the mean absolute temperature (K) and precipitation (mm) for each plot. Because of a lack of water temperature data, and

because for most sampling evens we did not know the exact date or depth of sampling, we used air temperature changes as proxy for water temperature changes for the freshwater data sets, as air and freshwater temperatures are typically correlated (Livingstone and Lotter 1998).

Local scale climate change - CHELSA

We calculated changes in mean monthly temperature and precipitation at local scales from the weather data of the CHELSA database (Karger et al. 2017) at 1 km^2 resolution for 1979-2013). To account for uncertainty of the provided coordinates, we extracted the temperature and precipitation data as the mean of the 9 pixels surrounding each plot. As we did for the CRU data, for each plot we calculated the yearly change in monthly mean temperature and precipitation over the sampled period using generalized additive models, accounting for seasonal patterns with a spline term on month (T or P ~ Year + spline(month)). From these models, we obtained the slope of temperature and precipitation change, and standardized these to decadal change for ease of interpretation (slope × 10).

We obtained relative climate change values by dividing these decadal change values by the mean absolute temperature (K) and precipitation (mm) for each plot. Similar to the CRU database, we did not have water temperature data, but used air temperature as a proxy.