Supplemental Methods

May 15, 2018

1 Data overview

This document briefly describes the data preparation prior to analysis for each dataset. Table 1 shows metadata on the datasets included in the meta-analysis. We plot species accumulation curves and spatio-temporal variation in the number of taxa observed, and identify any potential issues with the data that we should address.

For each dataset, we evaluated whether all sites were observed with equal effort in all years of study and we document any years, sites, sampling occasions, or species that were droppped from the dataset prior to analysis. We also document the method used to aggregate the data to an annual value for each species at each site. We visually evaluated the species accumulation curves and temporal patters in species abundances to screen for changes in sampling methods or data recording methods over the course of the study that could affect results. For example, a sudden jump in the number of taxa in a given year could indicate that taxa were recorded to a finer resolution (not that more taxa were actually present at the site). We also checked to be sure an abundance of zero was recorded for each absence (i.e. we made sure zeros were filled in for site-years in which a species was not observed). We removed species codes that obviously represent'unknown' species. **Do we use cutoff criteria to remove rare species prior to analysis? If so, we should probably include that code in formatting script for each site so that the Supplemental figures and L3 data product we publish match the data used in analysis.** We also include a table of datasets that were considered, and that we prepared into the L3 format, but ultimately were not included in the analysis (Table 2).

[Table 1 about here.]

[Table 2 about here.]

2 Marine datasets

2.1 csun.usvi-coral

Data are from Max Castorani. Need more metadata for this dataset. Margaret is looking for data API. Data are in Figure 1.

[Figure 1 about here.]

2.2 sbc-survey-algae

Waiting for updated data to be archived on EDI. Two of the eleven sites were initiated in the third year of study. I think we should remove the two sites. Data are shown in Figure 2

[Figure 2 about here.]

2.3 sbc-survey-sessile-inverts

Waiting for updated data to be archived on EDI. Two of the eleven sites were initiated in the third year of study. I think we should remove the two sites. Data are shown in Figure 3

[Figure 3 about here.]

2.4 sbc-survey-mobile-inverts

Waiting for updated data to be archived on EDI. Two of the eleven sites were initiated in the third year of study. I think we should remove the two sites. Data are shown in Figure 4

[Figure 4 about here.]

2.5 sbc-survey-fish

Waiting for updated data to be archived on EDI. Two of the eleven sites were initiated in the third year of study. I think we should remove the two sites. Data are shown in Figure 5

[Figure 5 about here.]

2.6 mcr-inverts

Need to get updated data. Need to separate by habitat. Needs data package citation. Data were downloaded from EDI (knb-lter-mcr.7.28). Data were aggregated across habitats and transects. Thus each of the six sites contains data from very different habitats lumped together (back, fringing, outer reefs) and there is likely very little overlap in the species found in each habitat. This is different than the other datasets, and it would be good to discuss whether or not to separate by habitat. Non-relevant taxa and taxa observed outside the quadrat were removed from the dataset. Abundance was averaged across subplots, transects, and habitats for each species at each site in each year. Data are in Figure 6. It's unclear whether these are mobile inverts, sessile inverts, or both (need to know for comparison with SBC).

[Figure 6 about here.]

2.7 mcr-fish

Updated data in ecocomdp format on EDI. Need to prepare it for this analysis. Need to separate by habitat. Needs data package citation. Data were aggregated across habitats and transects. Thus each of the six sites contains data from very different habitats lumped together (backreef, forereef, fringing) and there is likely very little overlap in the species found in each habitat. This is different than the other datasets, and it would be good to discuss whether or not to separate by habitat. Non-relevant taxa codes (e.g. "No fish present") were removed from the dataset. Abundance was recorded as dry biomass per 250 m^2, averaged across subplots, transects, and habitats for each species at each site in each year. Data are in Figure 7. Six extra locations in forereef habitats were sampled in 2015. These appear to be in addition to the four transects per

habitat per site performed as part of the long term data collections, and thus they were removed from the dataset prior to analysis.

[Figure 7 about here.]

2.8 mcr-coral

Need to get updated data. Need to separate by habitat. Needs data package citation. Data were downloaded from EDI (Edmunds, 2018). The corals are identified to the genus level. Data were aggregated across habitats and transects. Thus each of the six sites contains data from very different habitats lumped together (back, fringing, outer reefs) and there is likely very little overlap in the species found in each habitat. This is different than the other datasets, and it would be good to discuss whether or not to separate by habitat. Non-relevant taxa were removed from the dataset. Abundance was averaged across subplots, transects, and habitats for each species at each site in each year. Data are in Figure 8.

[Figure 8 about here.]

2.9 mcr-algae

Need to get updated data. Need to separate by habitat. Needs data package citation. Data were downloaded from EDI (Carpenter, 2015). Data were aggregated across habitats and transects. Thus each of the six sites contains data from very different habitats lumped together (back, fringing, outer reefs) and there is likely very little overlap in the species found in each habitat. This is different than the other datasets, and it would be good to discuss whether or not to separate by habitat. Non-relevant taxa were removed from the dataset. Abundance was averaged across subplots, transects, and habitats for each species at each site in each year. Data are in Figure 9. The cumulative number of taxa was still increasing at the end of the time series.

[Figure 9 about here.]

3 Freshwater datasets

3.1 fce-diatoms

Need to document where the data came from and clean the data. There are over 171 sites in this dataset, but only a fraction of those were sampled in all seven years. This could contribute to variation in total number of taxa observed - the earlier years have the most missing sites and also the lowest total number of taxa. Since the time series is so short to begin with, I suggest only including sites will all seven years in the analysis.

3.2 fce-algae

Need to document where the data came from and clean the data. There are duplicated records in the dataset - NOT ready for analysis. The duplicates cause the errors in the metadata table and the missing figures. There are over 150 sites in this dataset, but only a fraction of those were sampled in all seven years. Also, about 1/3 of all sited were not sampled in 2009, and it seems systematic (the sites with higher numbers were not sampled). Notice the dip in number of taxa observed in 2009. Since the time series is so short to begin with, I suggest only including sites will all seven years in the analysis.

3.3 fce-fish

Rolando obtained the data from the PI and is working on cleaning (Rehage, 2017).

3.4 ntl-zooplankton

The data were downloaded from the EDI Data Portal (Center for Limnology, 1983). Samples were taken via vertical tows at fortnightly intervals on a minimum of five occasions per year (range = 5 - 18 occasions per year). Density was recorded as number of individuals per liter for each taxa, integrated volumetrically over the water column. Many taxa are identified to species level, but some are identified to genus level. Lake 'Tr' was only sampled in one year and was assumed to be the same as lake 'TR'; thus we changed the lake identification code. The initial year (1981) was

removed from analysis because only five of the seven lakes were sampled. We additionally removed 165 records with missing or unknown taxa designations. Data were aggregated annually for each taxa in each lake by taking the maximum density observed in a tow sampling occasion. Data are shown in Figure 10.

[Figure 10 about here.]

3.5 ntl-fish

Data on the abundance of fish species were downloaded from the EDI Data Portal (Magnuson et al., 2010). Two types of gill nets (VGN and VGN127) were rarely used and removed from the dataset as per a conversation with Noah Lottig at NTL. The gear method "ESHOCK" was also rarely used and a follow up call is needed to determine how best to handle those data. They are currently included. Catch per unit effort (CPUE) was calculated for each species in each lake per year across all gear types (electrofishing, gill nets, baited traps) as the total catch divided by the total effort. We should double check with dataset contacts to be sure we did this correctly. 88 rows containing unidentified species were removed. Also, the two bog lakes (CB, TB) had 1-3 species each and were removed. Three of the 11 sites (FI, MO, WI) were initiated in 1995, and the time series was subsetted to include only years after 1994. Data are shown in Figure 11.

[Figure 11 about here.]

4 Terrestrial datasets

4.1 nwt-plants

no provenance. The data were obtained from Eric Sokol. Not all years were sampled, but all plots were sampled in each year of observation. Number of taxa increased rapidly toward the end of the time series. Data in Figure 12. It would be good to collect more metadata and provenance for this dataset. The plots comprise at least four distinct habitats. Some plots have very few species.

[Figure 12 about here.]

4.2 and-plants-mtStHelens

Needs cleaning still (Figure 13).

[Figure 13 about here.]

4.3 jrn-lizards

The data were downloaded from the EDI Data Portal (Whitford et al., 1991). This was a mark-recapture study. Pitfall traps were opened for two weeks four times per year (quarterly). The monthly samples from 1990 and 1991 were removed. Individual lizards were identified and the number of unique individuals per site per year were summed. Two sites that were established five years after the start of the study (SUMM and NORT) were excluded. We don't have a key to the species codes. Data are shown in Figure 14. The cumulative number of taxa was still increasing at the end of the time series, although with only 20 species total one or two species introductions could cause this pattern.

[Figure 14 about here.]

4.4 and-birds

Needs work. The data were downloaded from the EDI Data Portal (knb-lter-and.4781.2). These are point count data. Need to re-think how the 'count' is calculated. Should not be the simple sum of all six rounds per summer; perhaps max would be better. Also, the same bird could be observed multiple times during the 10-minute observation interval if he sang more than once. The NEW_RECORD column indicates this status with binary 0/1, and our final count should only include NEW_RECORD = 1. Note that all six rounds per growing season were only done in the initial five years of study. Only one or two rounds per growing season were done in 2014-2016, likely causing the dip in richness in those years. Fix this. Data not propagated. Data are shown in Figure 15.

[Figure 15 about here.]

4.5 hbr-birds

Need to subset by date so only 1-2 sampling occasions per year are included. These point count data were downloaded from the EDI Data Portal (Rodenhouse and Sillett, 2016). Years with highly incomplete sampling were removed (1999, 2000, 2008) and only plots that were sampled in all of the remaining years were retained for analysis. The number of individuals of each species within 50 m of the 'point' was summed for each 10-minute plot visit, using the 'NewRecord' column to identify unique individuals for data collected after 2005. The maximum number of individuals of each species for each plot-year (across 1-3 sampling occasions in June/early July per year) was calculated. Data are shown in Figure 16.

[Figure 16 about here.]

4.6 cap-birds

Riley obtained the data from EDI (Bateman et al., 2017) and is cleaning them. Data are shown in Figure 17.

[Figure 17 about here.]

References

Bateman, H., Childers, D., Katti, M., Shochat, E., and Warren, P. (2017). Point-count bird censusing: long-term monitoring of bird abundance and diversity in central arizona-phoenix, ongoing since 2000. Environmental Data Initiative. http://dx.doi.org/10.6073/pasta/201add557165740926aab6e056db6988. Dataset accessed 5/15/2018.

Carpenter, R. (2015). Mcr lter: Coral reef: Long-term population and community dynamics: Benthic algae and other community components, ongoing since 2005. Environmental Data Ini-

tiative. http://dx.doi.org/10.6073/pasta/79a6edbcf3aa2380d43deed778856416. Dataset accessed 5/07/2018.

Center for Limnology, N. L. (1983). North Temperate Lakes LTER: Zooplankton - Trout Lake Area 1982 - current. Environmental Data Intiative. http://dx.doi.org/10.6073/pasta/c866e3663bae76388f63233a5fdfb3d4.

Edmunds, P. (2018). Mcr lter: Coral reef: Long-term population and community dynamics: Corals, ongoing since 2005. Environmental Data Initiative. http://dx.doi.org/10.6073/pasta/263faa48b520b7b2c964f158c184ef96. Dataset accessed 5/07/2018.

Magnuson, J., Carpenter, S., and Stanley, E. (2010). North Temperate Lakes LTER: Fish Abundance 1981 - current. Environmental Data Intiative. http://dx.doi.org/10.6073/pasta/7ed3313d08fbfc92656262b977508340.

Rehage, J. (2017). Seasonal electrofishing data from rookery branch and tarpon bay, everglades national park (fce) from november 2004 to present. Environmental Data Initiative. http://dx.doi.org/10.6073/pasta/ed3febe89ff59f68ae2aedf6c87b7eff. Dataset obtained from PI 5/15/2018.

Rodenhouse, N. L. and Sillett, T. S. (2016). Valleywide Bird Survey, Hubbard Brook Experimental Forest, 1999-present. Environmental Data Initative. http://dx.doi.org/10.6073/pasta/0a2354315d4cc76b3695f13cae31e60e.

Whitford, W., Lightfoot, D., and Anderson, J. (1991). Lizard pitfall trap data (LTER-II, LTER-III). Environmental Data Intiative.http://dx.doi.org/10.6073/pasta/411d2828c578c4777218fce541cc4291.

Table 1: Metadata on the data sets included in the meta-analysis. This table is automatically generated directly from the datasets in the L3 folder. I can make it look prettier once it is final.

"dataset"	"initial.year"	"study.length"	"n.years"	"n.plots"	"n.taxa"	"organism"
"and-birds-wisnoski"	2009	8	8	184	88	"birds"
"cap-birds-banville"	2001	16	16	35	104	"birds"
"fce-algae-marazzi"	Inf	-Inf	0	0	249	"algae"
"fce-diatoms-marazzi"	2005	7	7	171	367	"diatoms"
"hbr-birds-sillett"	2001	13	8	64	61	"birds"
"jrn-lizards-hope"	1990	16	16	9	20	"lizards"
"mcr-algae-castorani"	2006	10	10	6	73	"algae"
"mcr-coral-castorani"	2005	11	11	6	31	"coral"
"mcr-fish-castorani"	2006	10	10	6	376	"fish"
"mcr-inverts-castorani"	2005	11	11	6	13	"inverts"
"ntl-fish-stanleyLottig"	1995	23	23	9	81	"fish"
"ntl-zooplankton-stanleyLottig"	1982	34	34	7	143	"zooplankton"
"nwt-plants-hallett"	1989	26	11	88	109	"plants"
"sbc-algae-castorani"	2001	16	16	11	62	"algae"
"sbc-fish-castorani"	2001	16	16	11	64	"fish"
"sbc-mobileInverts-castorani"	2001	16	16	11	36	"mobileInverts"
"sbc-sessileInverts-castorani"	2001	16	16	11	70	"sessileInverts"
"usvi-coral-castorani"	1992	24	24	6	34	"coral"

Table 2: This is a record of the datasets we considered using and created an L3 dataset for, but did not make the final cut for one reason or another.

Dataset	Reason excluded
mcm-diatoms	highly irregular sampling
ntl-macroinvertebrates	only four taxa

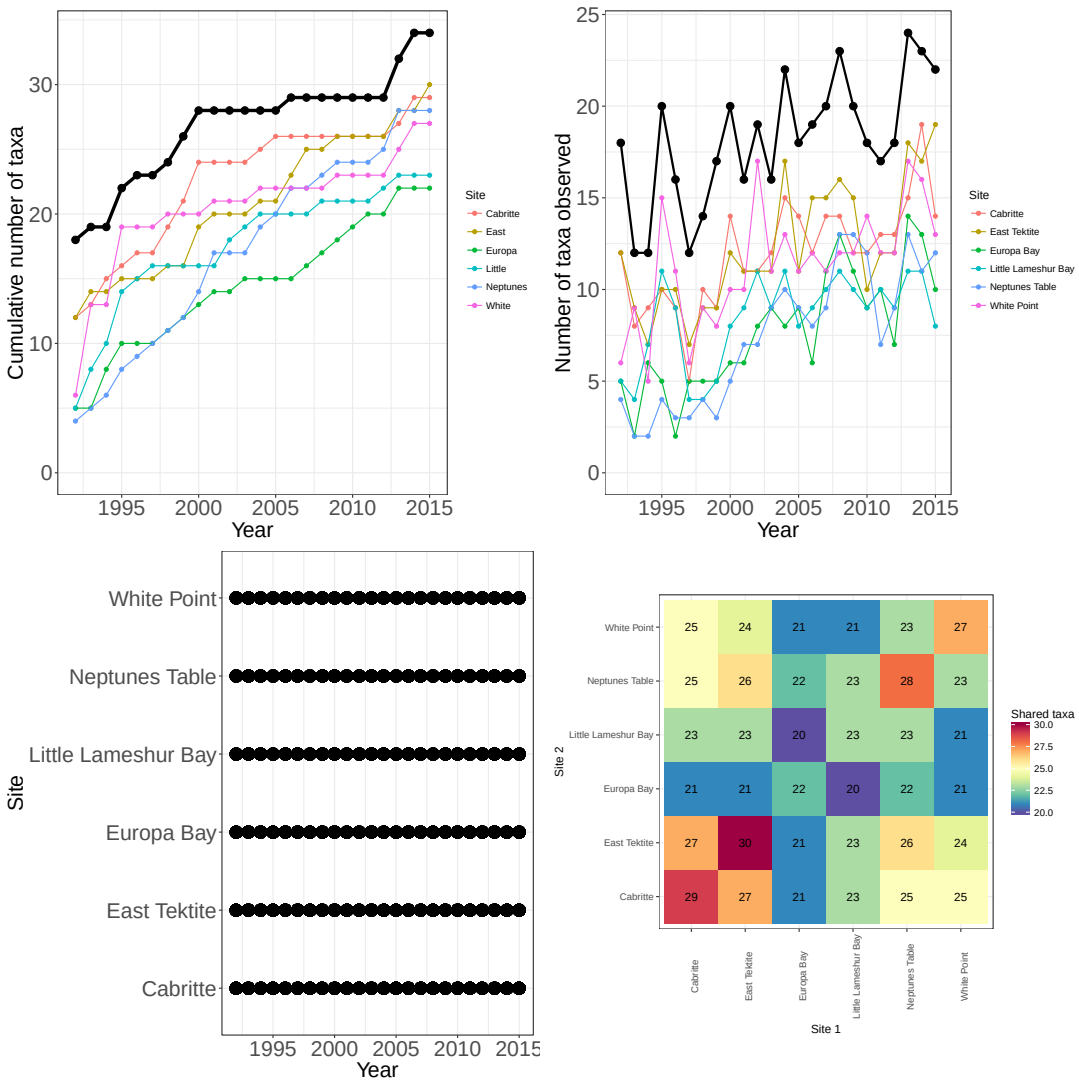

Figure 1: **USVI-corals:** Species accumulation curves (top left), annual richness (top right), and sampling effort (bottom) for 34 coral taxa observed at 6 plots on St. John, US Virgin Islands (1990-2005). The black lines represent total site-level values across all plots.

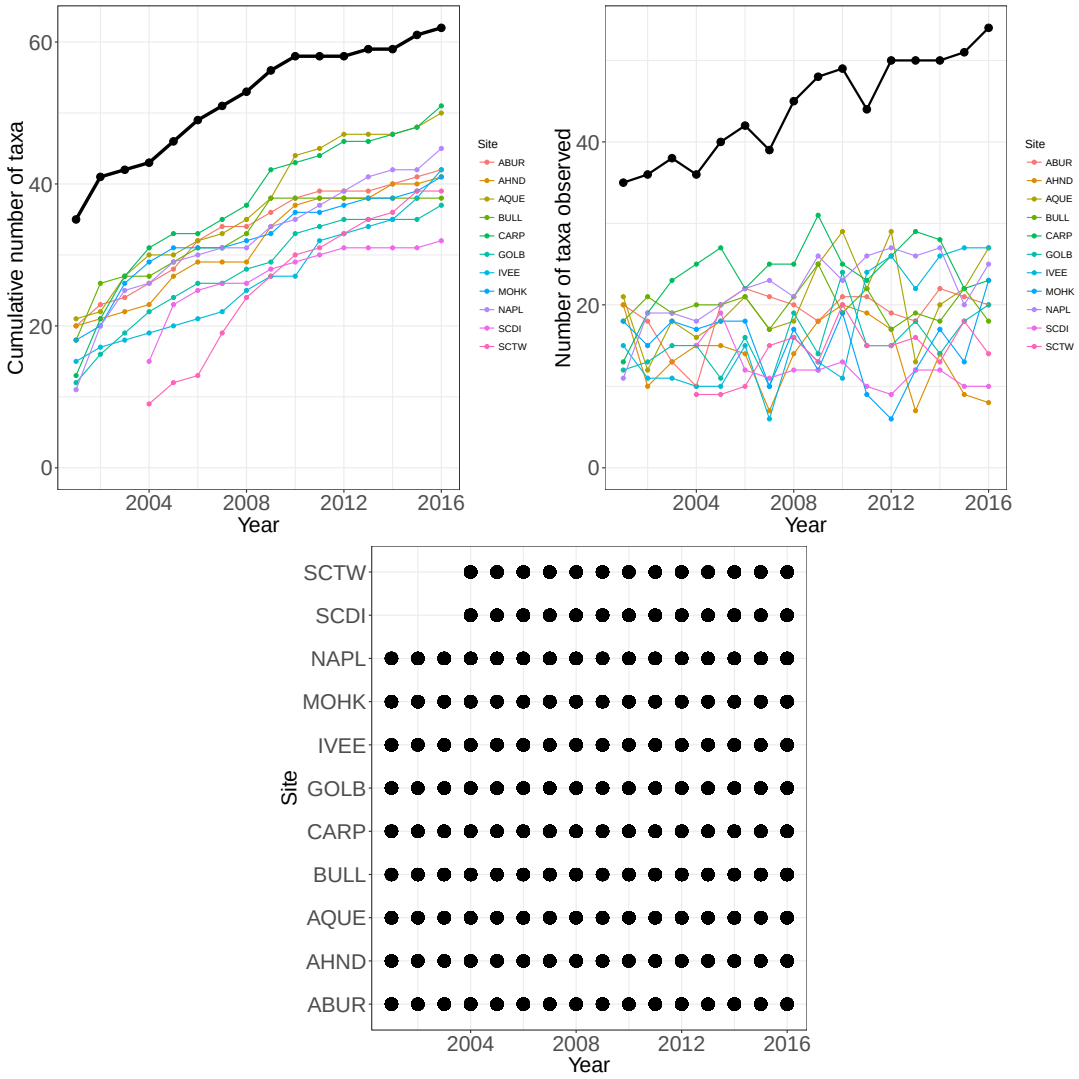

Figure 2: **SBC-algae:** Species accumulation curves (top left), annual richness (top right), and sampling effort (bottom) for 62 algae taxa observed at 11 plots in the Santa Barbara Channel LTER (2001-2016). The black lines represent total site-level values across all plots.

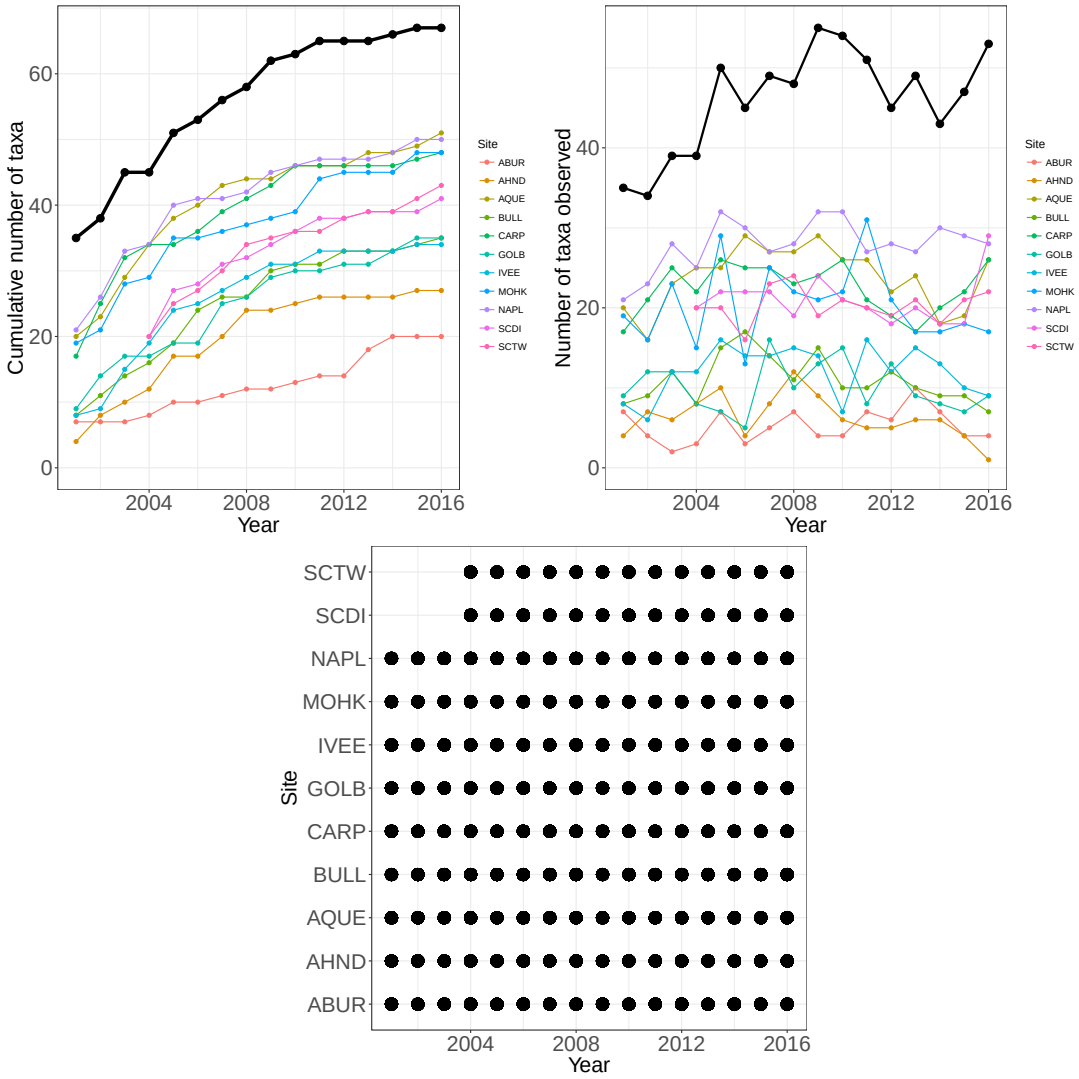

Figure 3: **SBC-sessile invertebrates:** Species accumulation curves (top left), annual richness (top right), and sampling effort (bottom) for 70 sessile invertebrate taxa observed at 11 plots in the Santa Barbara Channel LTER (2001-2016). The black lines represent total site-level values across all plots.

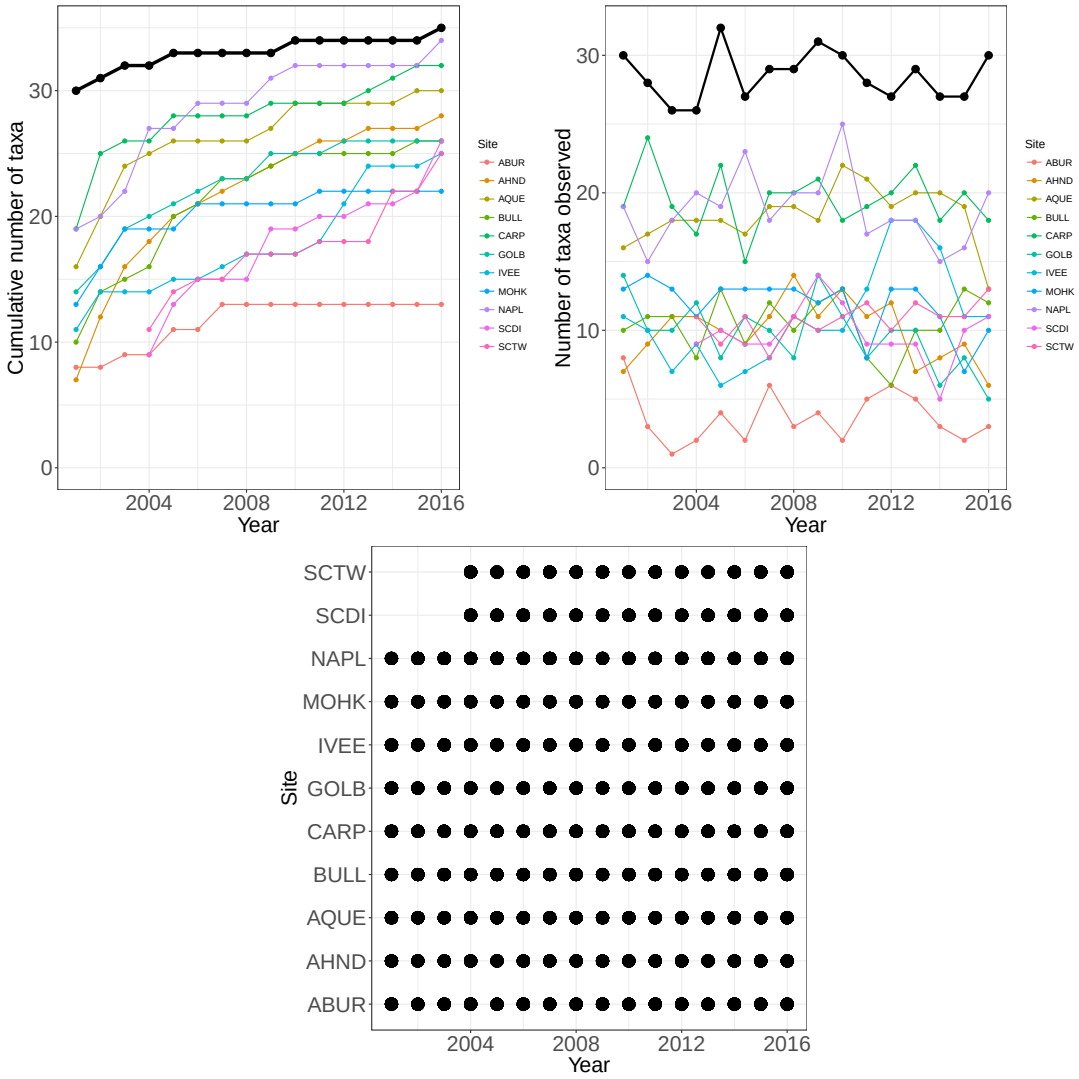

Figure 4: **SBC-mobile invertebrates:** Species accumulation curves (top left), annual richness (top right), and sampling effort (bottom) for 36 mobile invertebrate taxa observed at 11 plots in the Santa Barbara Channel LTER (2001-2016). The black lines represent total site-level values across all plots.

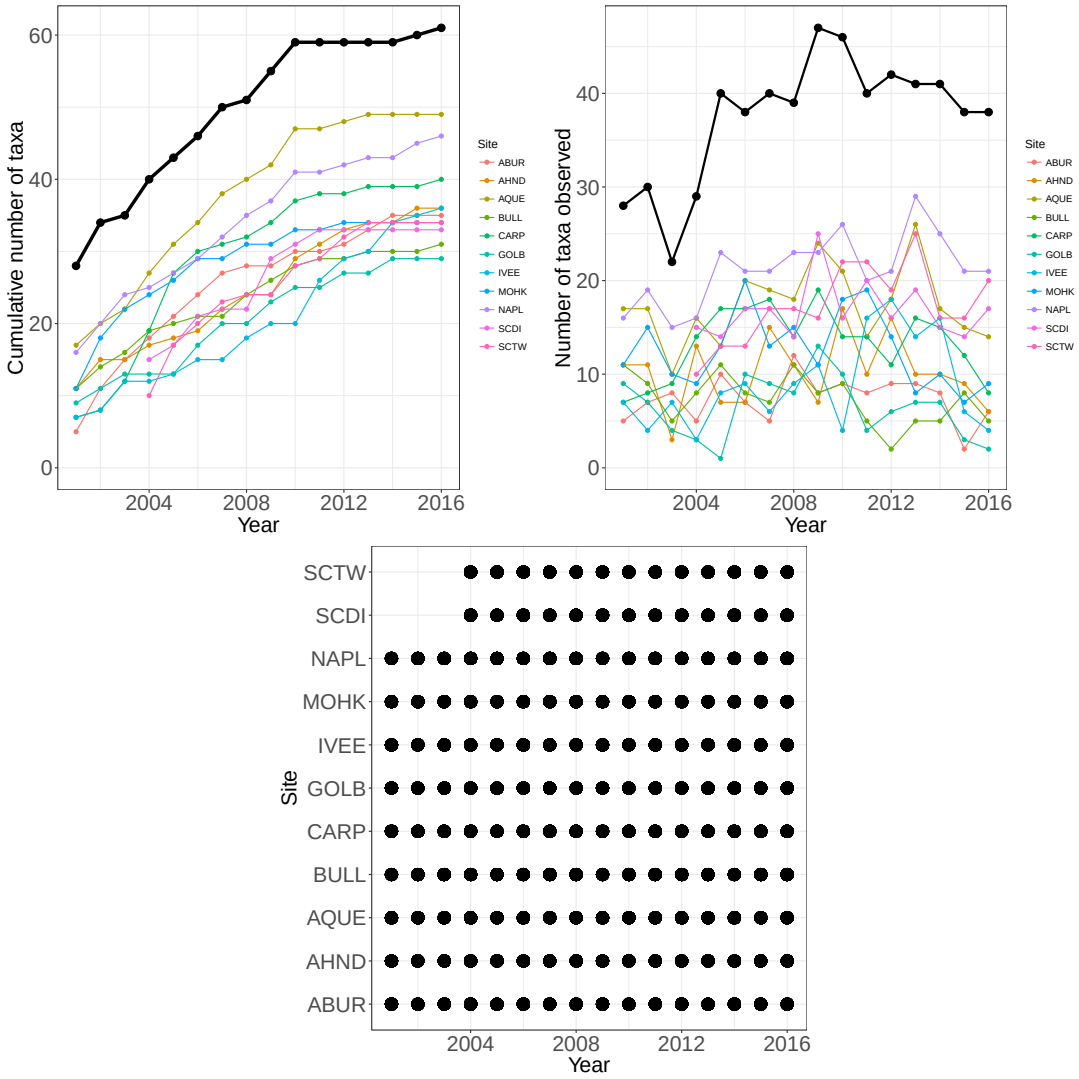

Figure 5: **SBC-fish:** Species accumulation curves (top left), annual richness (top right), and sampling effort (bottom) for 64 fish species observed at 11 plots in the Santa Barbara Channel LTER (2001-2016). The black lines represent total site-level values across all plots.

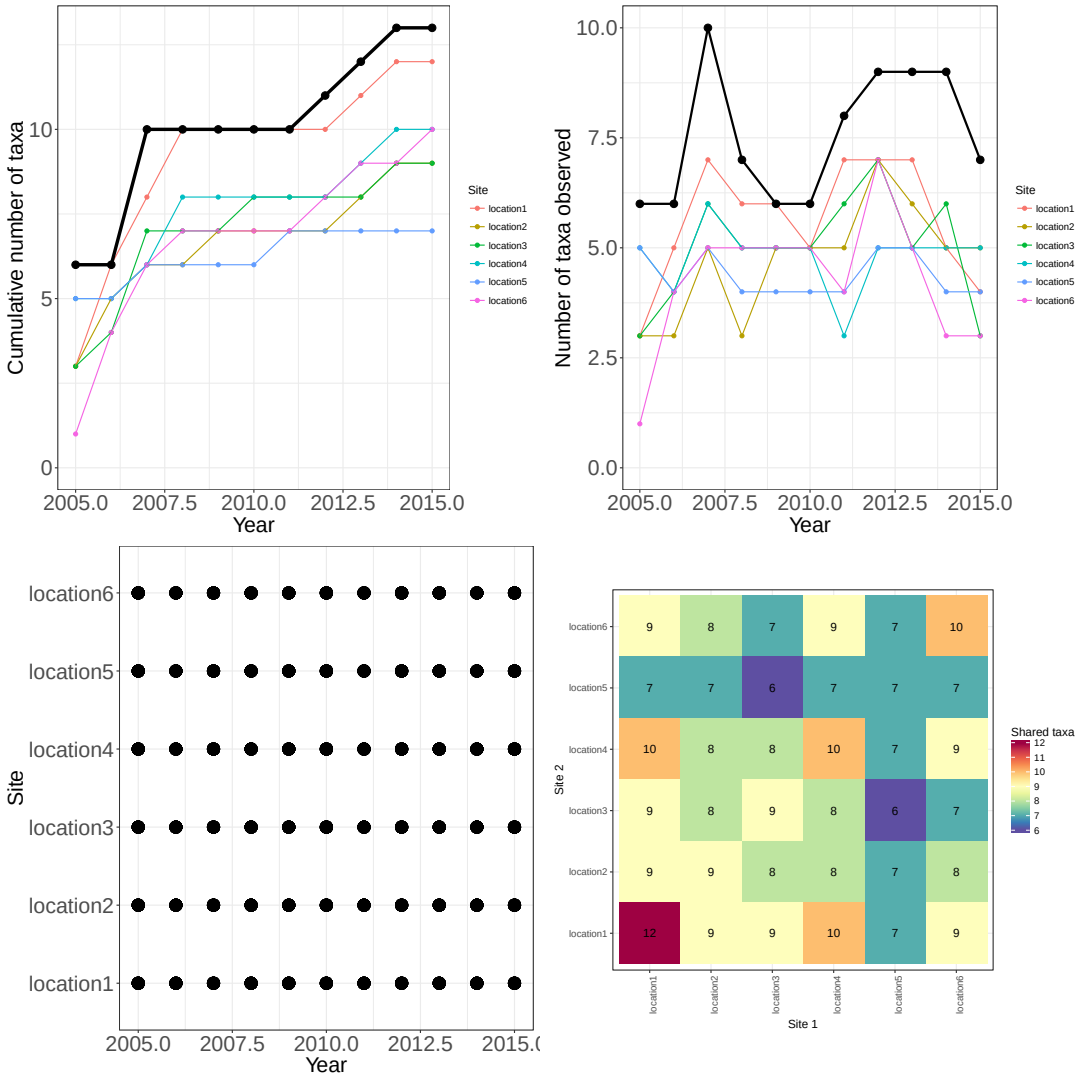

Figure 6: **MCR-inverts:** Species accumulation curves (top left), annual richness (top right), and sampling effort (bottom) for 13 invertebrate taxa observed at six sites on Moorea coral reef LTER (2006-2015). The black lines represent total site-level values across all plots.

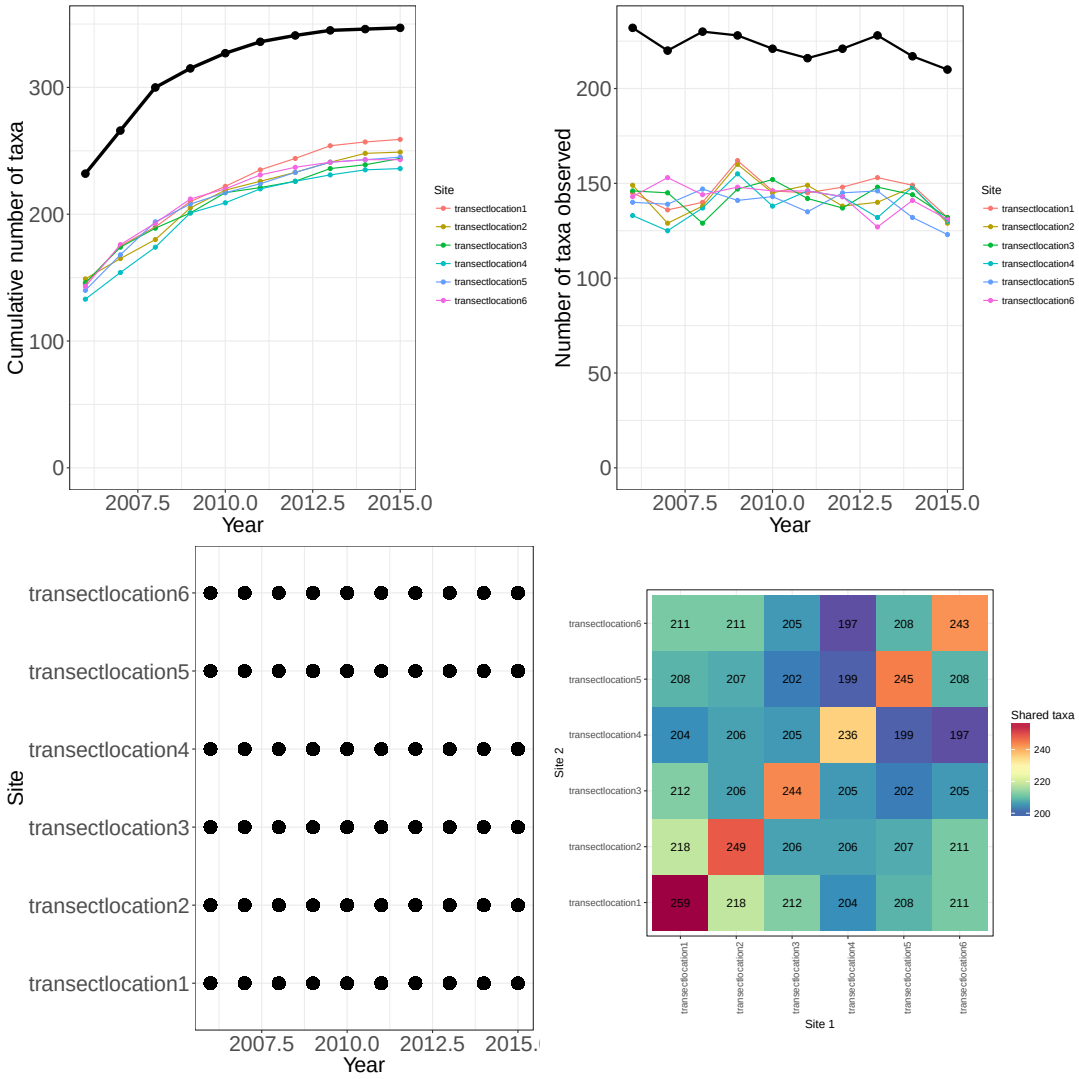

Figure 7: **MCR-fish:** Species accumulation curves (top left), annual richness (top right), and sampling effort (bottom) for 377 fish taxa observed at six sites on Moorea coral reef LTER (2006-2015). The black lines represent total site-level values across all plots.

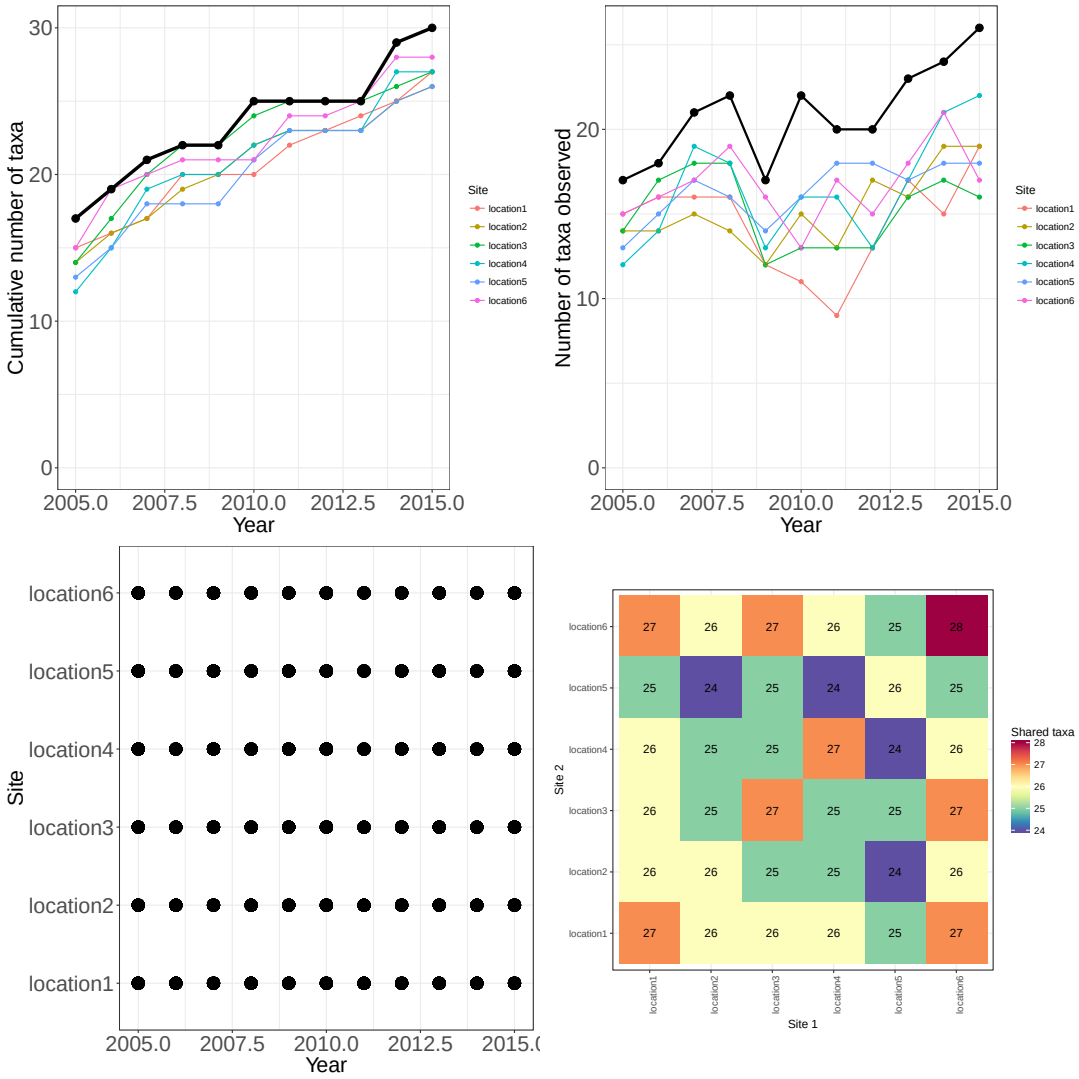

Figure 8: **MCR-coral:** Species accumulation curves (top left), annual richness (top right), and sampling effort (bottom) for 31 coral taxa observed at six sites on Moorea coral reef LTER (2006-2015). The black lines represent total site-level values across all plots.

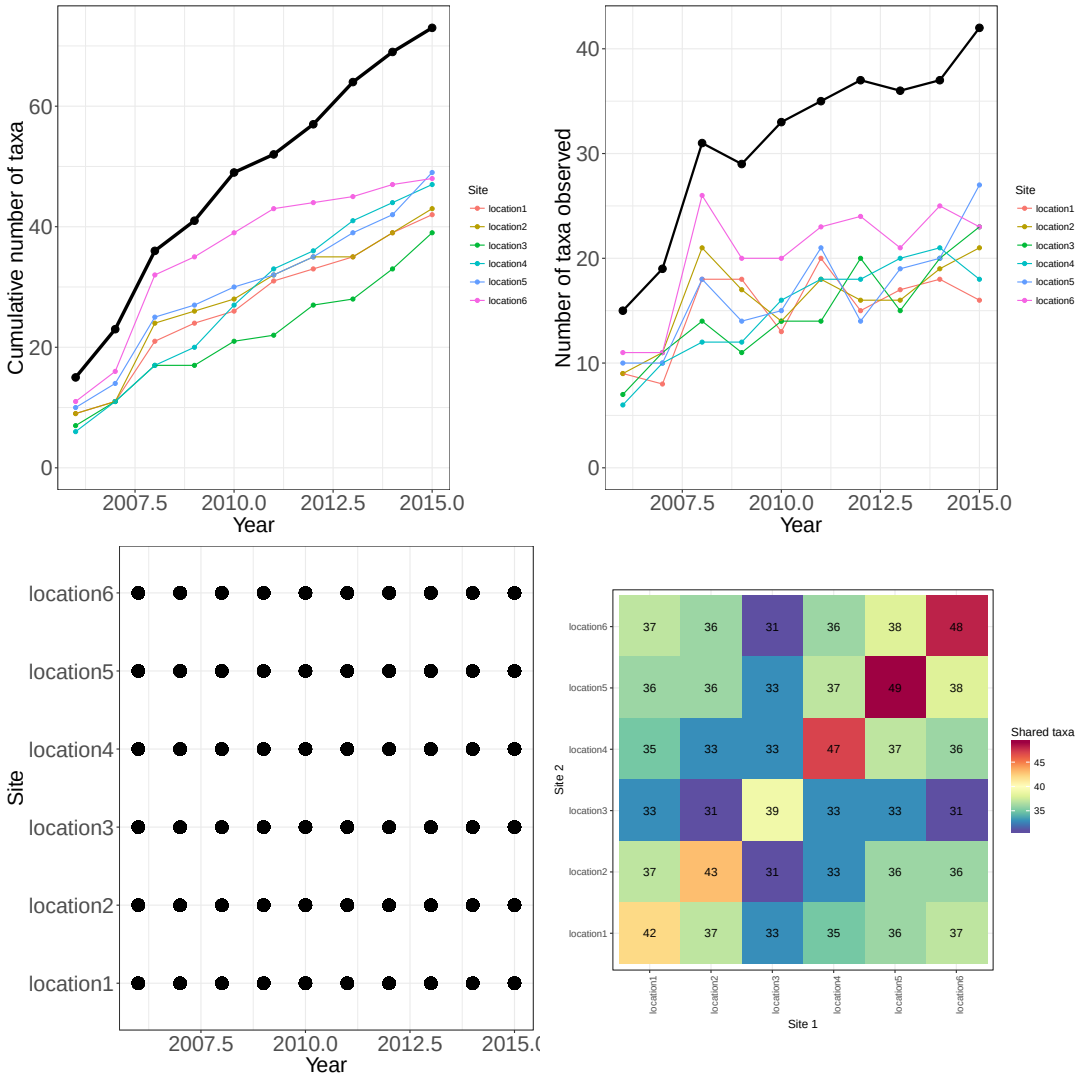

Figure 9: **MCR-algae:** Species accumulation curves (top left), annual richness (top right), and sampling effort (bottom) for 73 algae taxa observed at six sites on Moorea coral reef LTER (2006-2015). The black lines represent total site-level values across all plots.

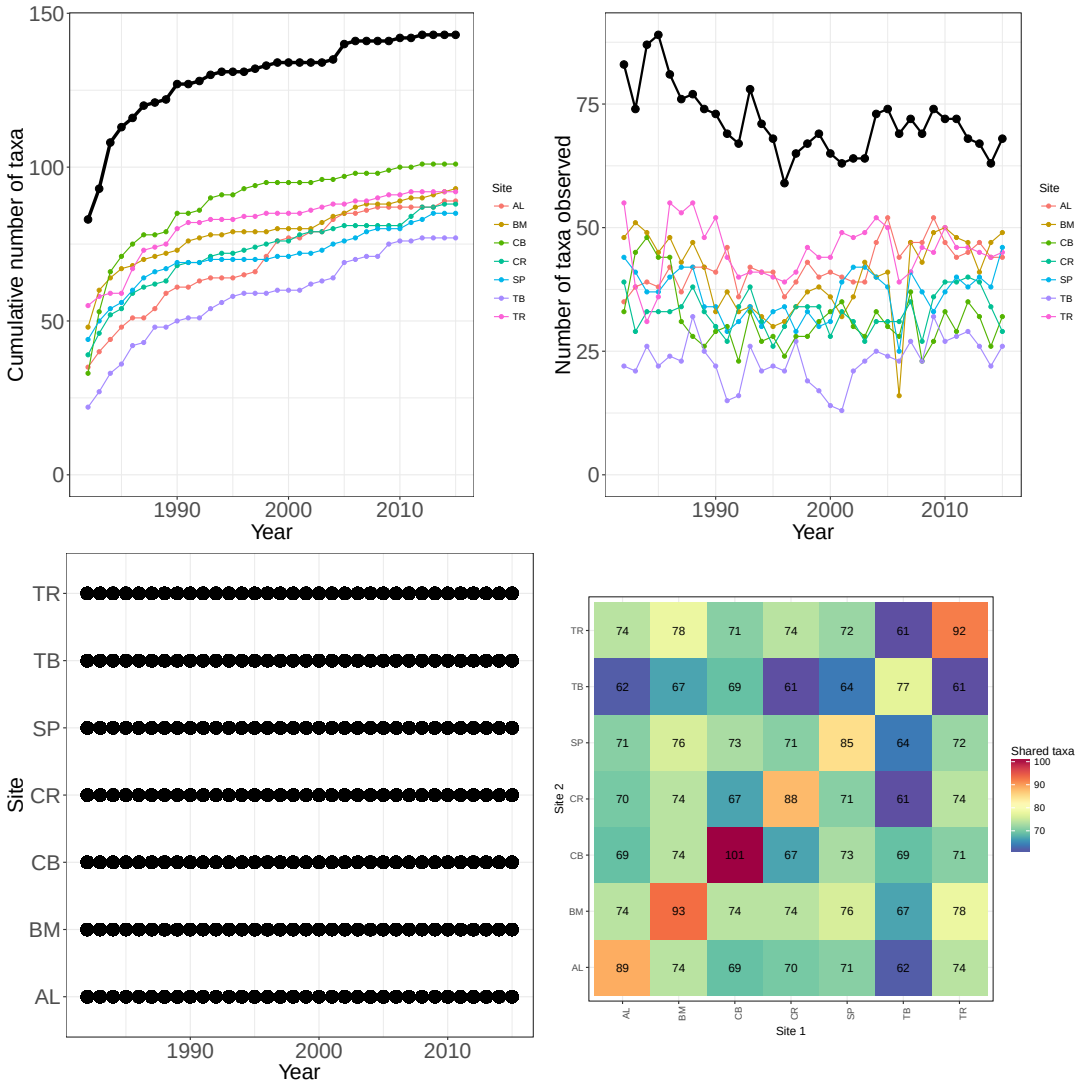

Figure 10: **NTL-zooplankton:** Species accumulation curves (top left), annual richness (top right), and sampling effort (bottom) for 143 zooplankton taxa observed at 7 sites North Temperate Lakes LTER . The black lines represent total site-level values across all plots. The plot of the species accumulation curve failed because one site was sampled in only one year, and will be fixed once that site is removed.

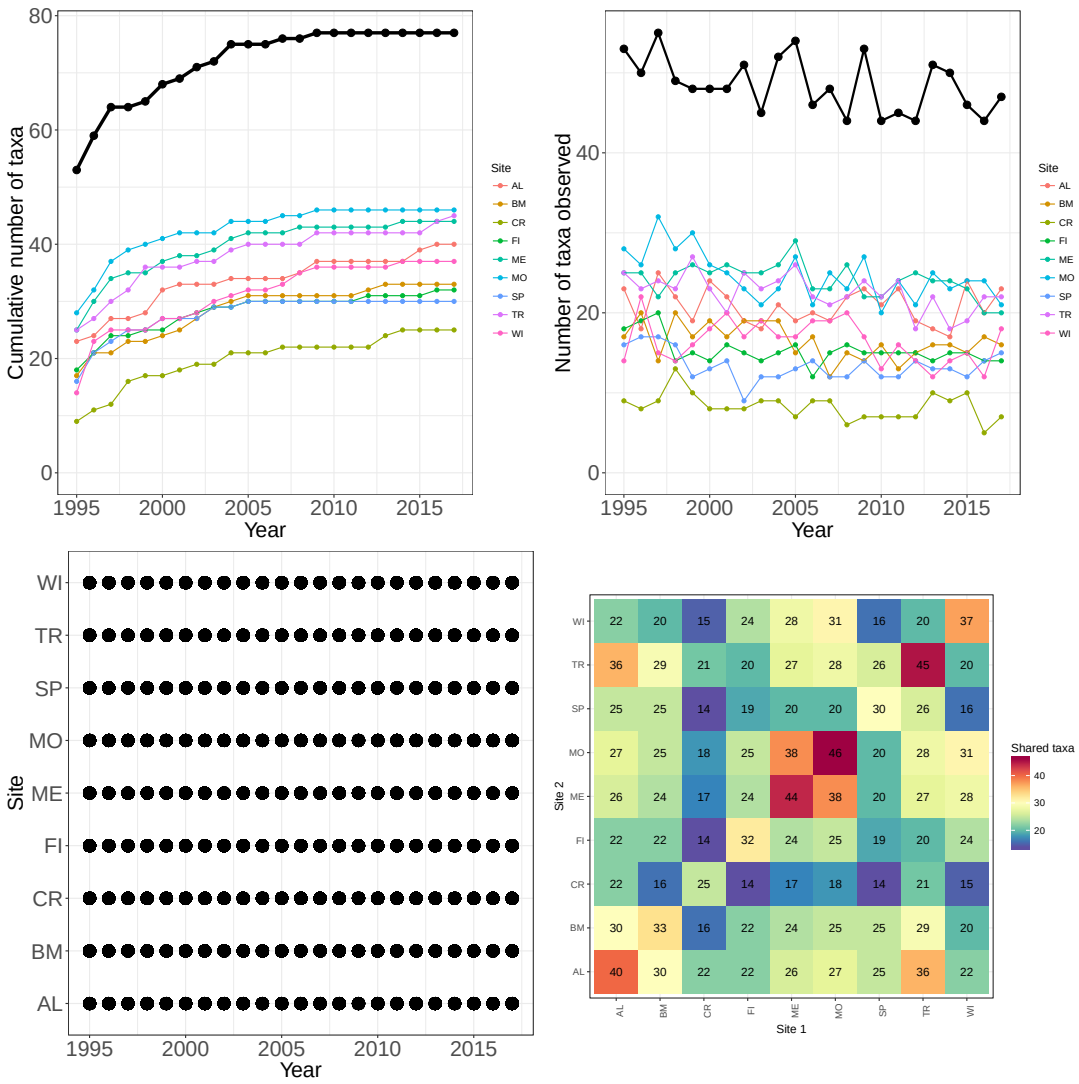

Figure 11: **NTL-fish:** Species accumulation curves (top left), annual richness (top right), and sampling effort (bottom) for 81 fish species observed at 9 lakes in the North Temperate Lakes LTER (1995-2016). The black lines represent total site-level values across all plots.

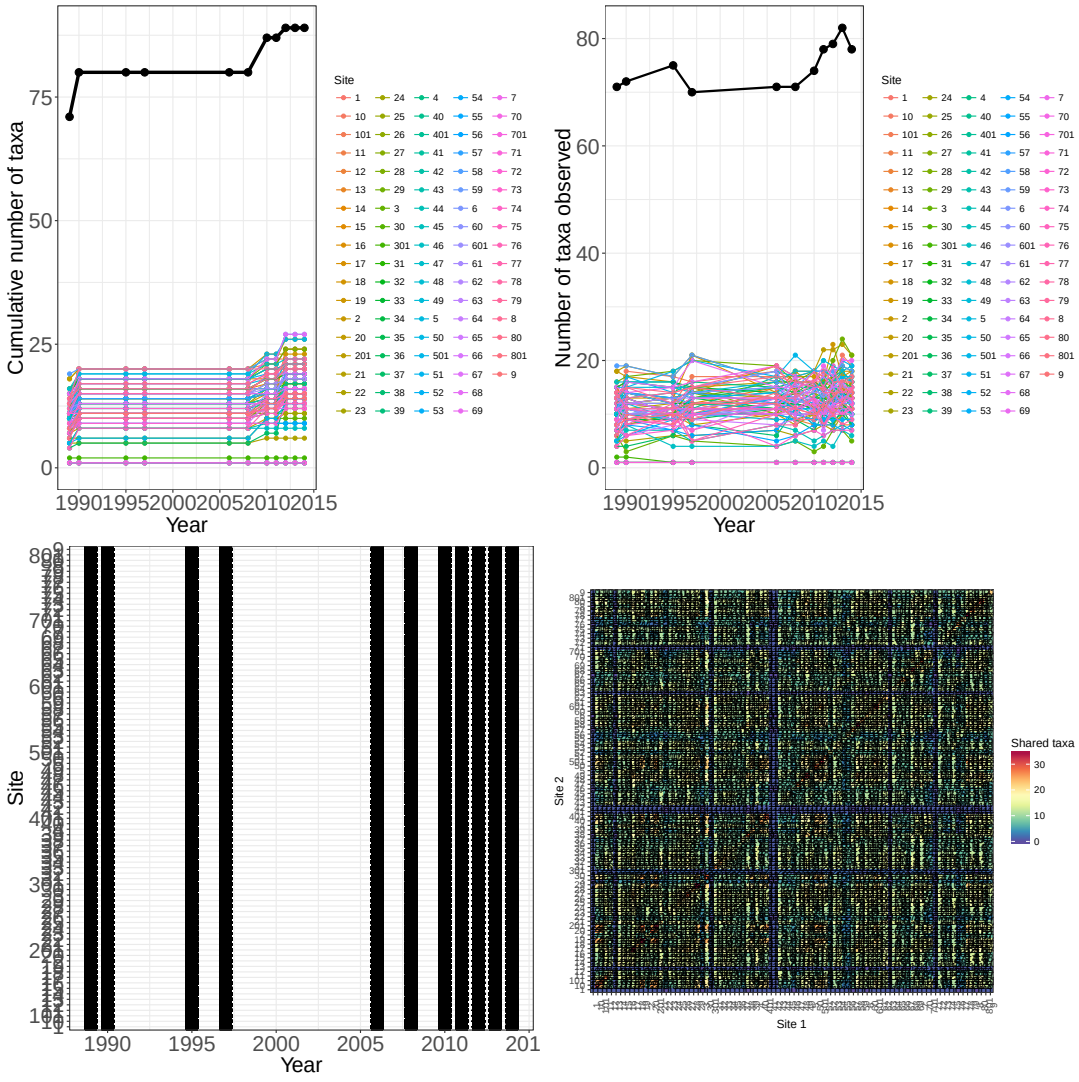

Figure 12: **NWT-plants:** Species accumulation curves (top left), annual richness (top right), and sampling effort (bottom) for 109 plant species observed at 88 plots in the Niwot Ridge LTER (1989-2014). The black lines represent total site-level values across all plots.

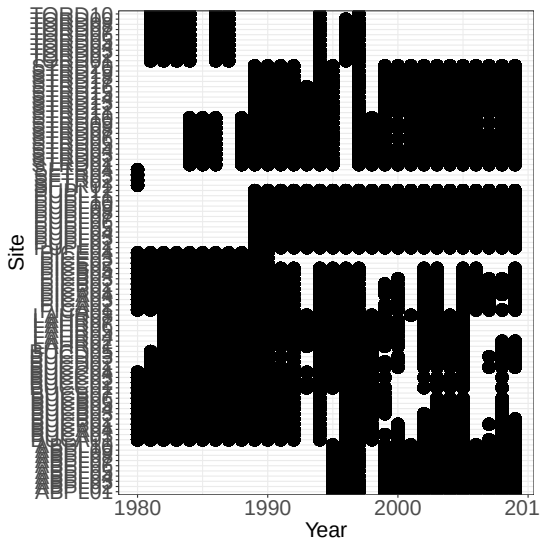

Figure 13: **AND-plants:** Species accumulation curves (top left), annual richness (top right), and sampling effort (bottom) for plant species observed at Mt. St. Helens. The black lines represent total site-level values across all plots.

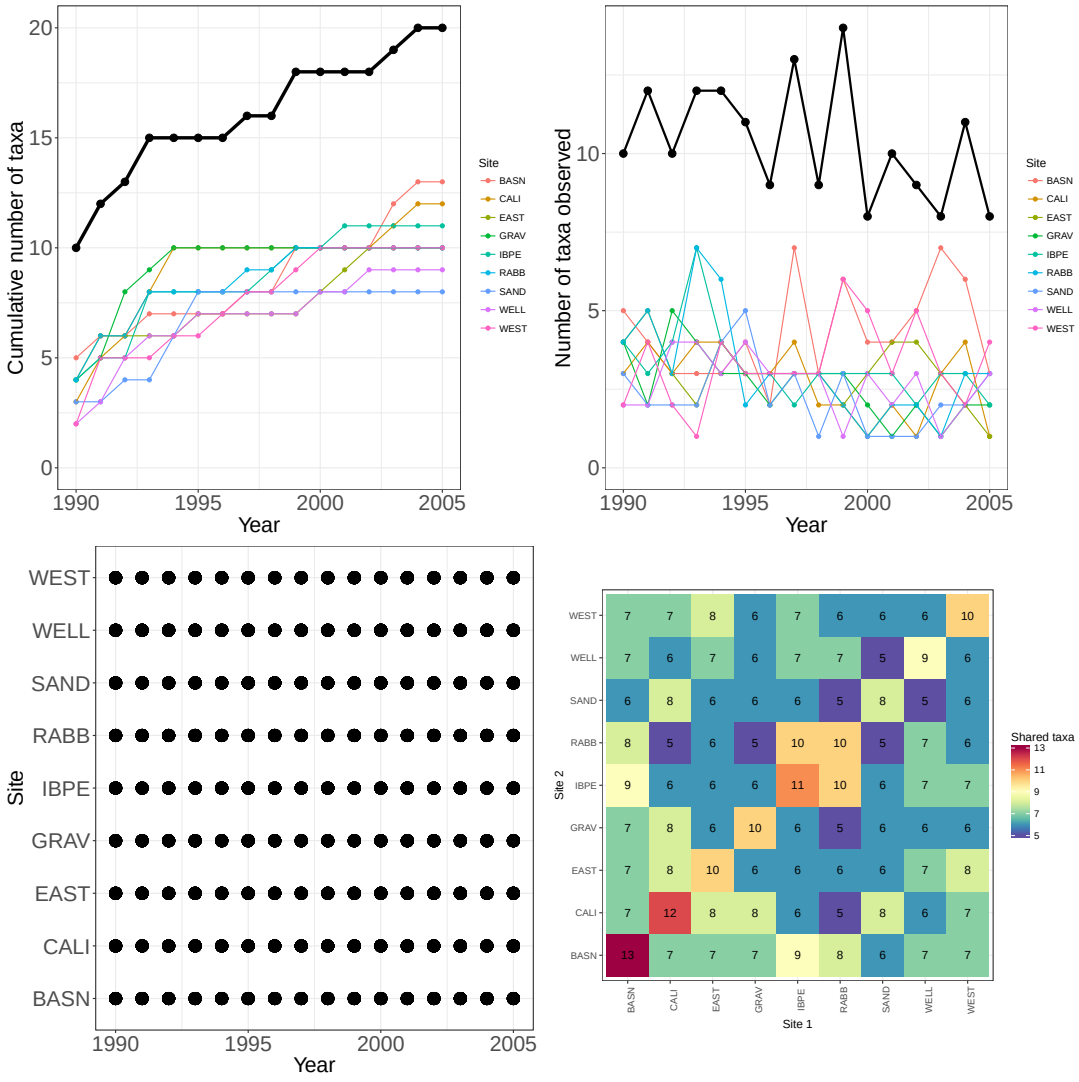

Figure 14: **JRN-lizards:** Species accumulation curves (top left), annual richness (top right), and sampling effort (bottom) for 20 lizard species observed at 9 plots in the Jornada LTER (1990-2005). The black lines represent total site-level values across all plots.

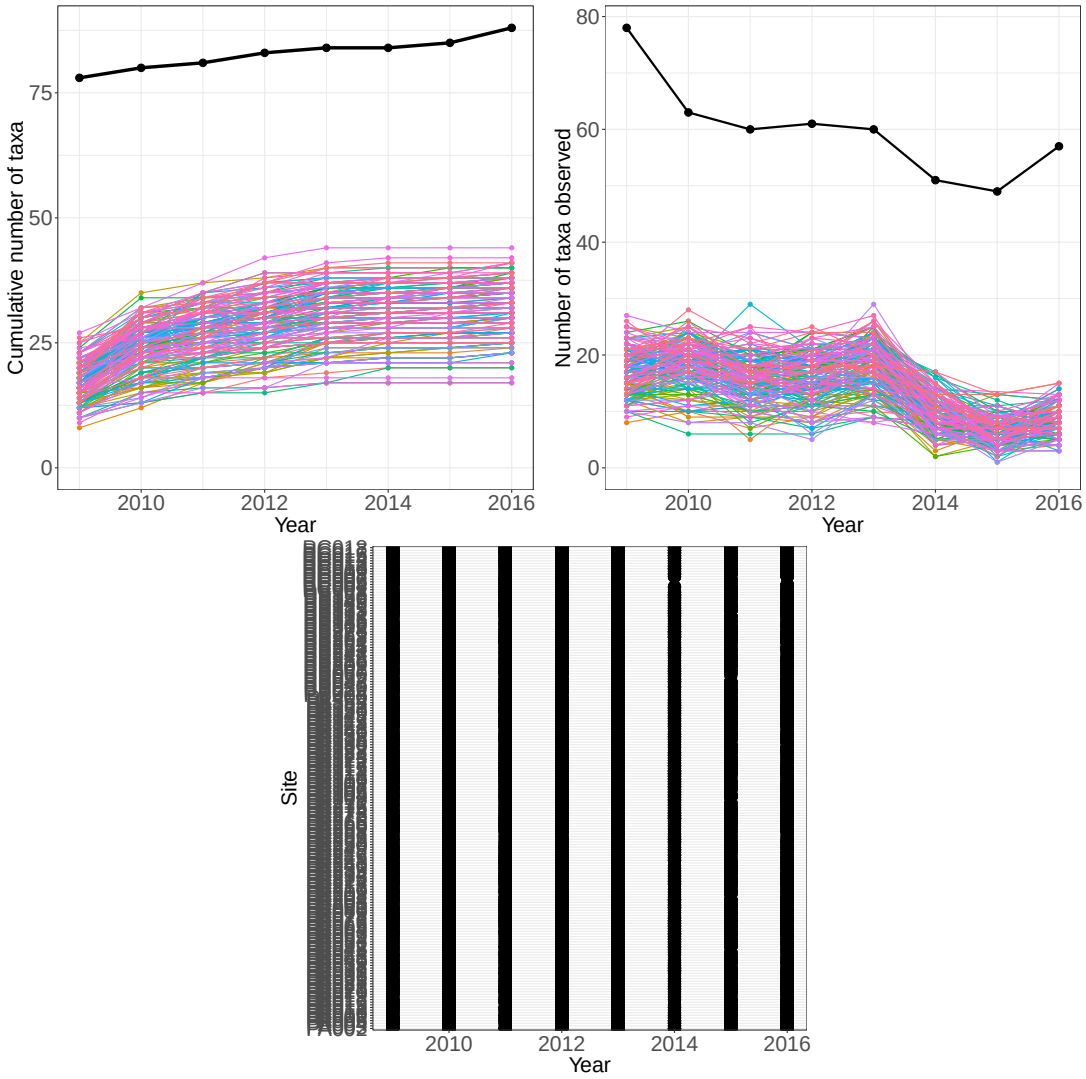

Figure 15: **AND-birds:**

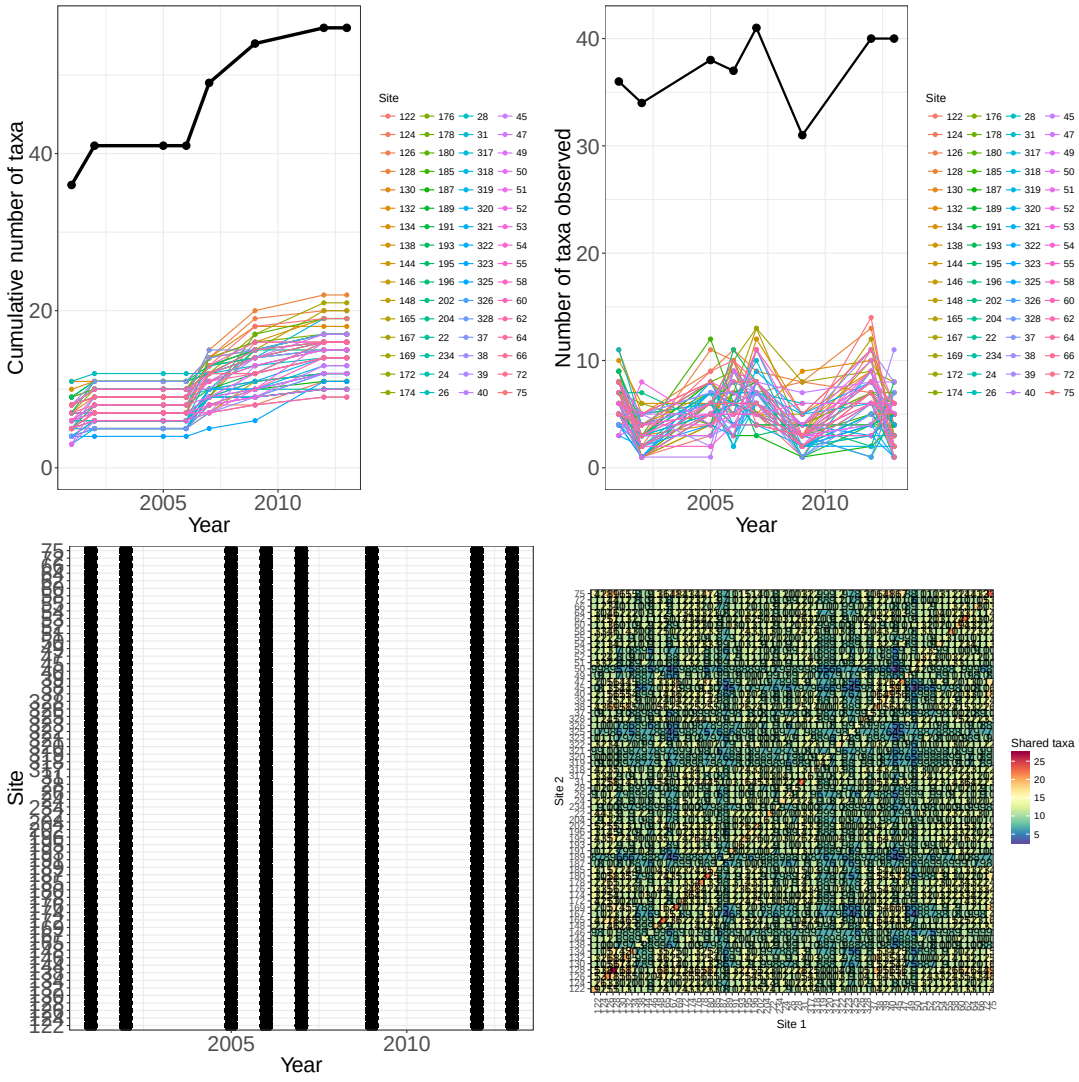

Figure 16: **HBR-birds:**

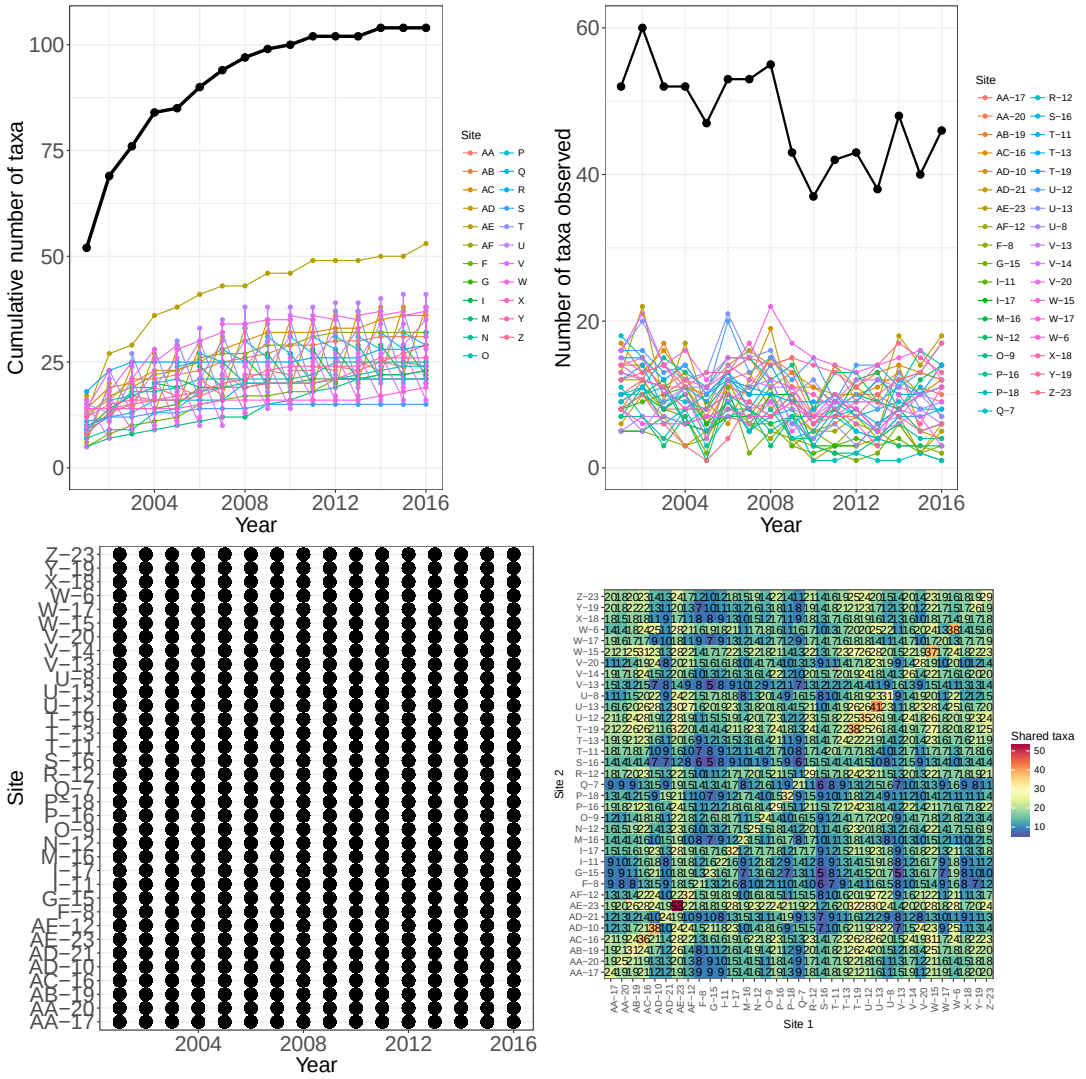

Figure 17: **CAP-birds:**