Environmental filtering and niche (mis)matching of riverine invertebrate communities

NRSA DisEQ-Data Management

Background

Raw invertebrate data were formatted for analysis by reshaping the file from a long to wide format. We also calculated the relative abundance of all invertebrates, retaining all taxa with a relative abundance $> 0.25\%$.

```
## Read in the raw invertebrate data
raw.invertebrate.data <- read_csv(
    "data/EPA-NRSA_0809-invertebrates-RAW.csv",
    show_col_types = FALSE
    )
```

```
## Reshape the data to a site-by-taxa matrix (rows = sites, columns = taxa)
reshaped.invertebrate.data <- dcast(
    raw.invertebrate.data,
    UID ~ TARGET_TAXON,
    sum,
    value.var = "TOTAL300"
    )
```

```
## Percent relative abundance by total abundance of all collected invertebrates
relative.abundance.data <- reshaped.invertebrate.data[, 2:975]/549426*100
# Total abundance of collected invertebrates = 549426
```

```
taxa.list <- relative.abundance.data[, colSums(relative.abundance.data, na.rm = TRUE) > 0.25]
# 0.25% cutoff = 73.86 total abundance of the CUS
```

```
## Revert relative abundance back to total abundance for later analyses
invertebrate.data <- taxa.list/100*549426

## Add UID
invertebrate.data$UID <- reshaped.invertebrate.data$UID
```

```
write_csv(invertebrate.data, file = "data/NRSA-invertebrates-rel_abund.25.csv")
# File name indicates that the abundance is from a relative abundance > 0.25%
```

R Session Information

Table 1: Packages required for data management.

Package	Loaded Version	Date
dplyr	1.0.8	2022-02-08
forcats	0.5.1	2021-01-27
ggplot2	3.3.5	2021-06-25
kableExtra	1.3.4	2021-02-20
knitr	1.38	2022-03-25
purrr	0.3.4	2020-04-17
readr	2.1.2	2022-01-30
reshape2	1.4.4	2020-04-09
stringr	1.4.0	2019-02-10
tibble	3.1.6	2021-11-07
tidyr	1.2.0	2022-02-01
tidyverse	1.3.1	2021-04-15