

Appendix S1 - Data Collation

Nicholas Clark, James Kerry, Ceridwen Fraser

This appendix demonstrates how to download open-source data on fish species occurrences, functional traits and phylogenetic relationships, as well as recorded and projected climate data from the [FishMed](#) database (Albouy et al. 2015). First, we source functions from the **Functions** directory to download the data into our working environment

```
source("../Functions/extract_MedFish.R")
list2env(extract_MedFish(dataset = "Pres.abs.data"),
         .GlobalEnv)
list2env(extract_MedFish(dataset = "Climate.data"),
         .GlobalEnv)
list2env(extract_MedFish(dataset = "Phylogenetic.data"),
         .GlobalEnv)
list2env(extract_MedFish(dataset = "Functional.data"),
         .GlobalEnv)
```

Next, we create mean Summer and mean Winter Sea Surface Temp variables for each site and join these to the species occurrence data in the format necessary for Conditional Random Fields analysis in the **MRFCov** library (with the two covariate columns at the righthand side of the **dataframe**). Here we omit species observations that do not have associated SST data, and we remove species that were recorded in fewer than 10% or greater than 90% of observations (estimating occurrence probabilities for species that are either too rare or too common is difficult and will lead to imprecise estimates of network topology)

```
library(dplyr)
Medfish.mrf.dat = Obs.1980.Climate.dat %>%
  dplyr::group_by(Latitude, Longitude) %>%
  dplyr::mutate(mean.summ.temp = mean(July,
    August, September), mean.wint.temp = mean(January,
    February, March)) %>% dplyr::ungroup() %>%
  dplyr::select(Latitude, Longitude, mean.summ.temp,
    mean.wint.temp) %>% dplyr::right_join(Obs.1980.Medfish.dat) %>%
  na.omit() %>% dplyr::group_by(mean.summ.temp,
    mean.wint.temp, Latitude, Longitude) %>%
  dplyr::select(which(colSums(.) > nrow(.) *
    0.1)) %>% dplyr::select(which(colSums(.) <
    nrow(.) * 0.9)) %>% dplyr::ungroup() %>%
  dplyr::select(-mean.summ.temp, -mean.wint.temp,
    dplyr::everything())
```

Table 1: Preview of analysis data in binary format

Latitude	Longitude	Gymnammodytes semisquamatus	Lagocephalus lagocephalus
35.71	-5.89	0	1
35.81	-5.89	0	1
36.01	-5.89	0	1
36.11	-5.89	0	1
36.21	-5.89	0	1
35.81	-5.79	0	1

Extract GPS coordinates for each site, which will be used for generating smoothed spatial regression splines in the models

```
coords = Medfish.mrf.dat %>% dplyr::select(Latitude,
  Longitude)
Medfish.mrf.dat = Medfish.mrf.dat %>% dplyr::select(-Latitude,
  -Longitude)
```

Table 2: Preview of the coordinates dataframe

Latitude	Longitude
35.71	-5.89
35.81	-5.89
36.01	-5.89
36.11	-5.89
36.21	-5.89
35.81	-5.79
35.91	-5.79
36.01	-5.79
36.11	-5.79
35.81	-5.69

Finally, we remove species whose binomial names do not match those in the phylogenetic tree or the functional data matrix. This ensures we are only generating predictions for species that have associated trait data

```
unique.species <- Reduce(intersect, list(Medfish.tree$tip.label,
  Medfish.func.dat$Species))
Medfish.mrf.dat = Medfish.mrf.dat %>% ungroup %>%
  dplyr::group_by(mean.summ.temp, mean.wint.temp) %>%
  dplyr::select(matches(paste(unique.species,
    collapse = "|"))) %>% dplyr::ungroup() %>%
  dplyr::select(-mean.summ.temp, -mean.wint.temp,
    dplyr::everything())
```

Check that there are no names in the data that don't have functional or phylogenetic information (both calls should return FALSE)

```
mrf.species <- colnames(Medfish.mrf.dat[,
  1:(ncol(Medfish.mrf.dat) - 2)])
any((mrf.species %in% Medfish.tree$tip.label)) ==
  FALSE
```

```
## [1] FALSE
```

```
any((mrf.species %in% Medfish.func.dat$Species)) ==
  FALSE
```

```
## [1] FALSE
```

How many species are included in our final dataset?

```
length(mrf.species)
```

```
## [1] 215
```

Finally, we use the supplied functional trait data to generate a multivariate graphical dendrogram representing functional relatedness among species. This will be useful for estimating functional diversity in sites with different temperature profiles later on. First, we create **dataframes** of species habitat use (10 binary variables representing species' use of different habitat categories), diet (6 binary variables representing species' dietary preferences) and body length (a single continuous variable). These **dataframes** are used to create an unweighted Gower's distance matrix with functions in the **BBS.occurrences** and **ape** libraries

```
hab.dat.binary <- Medfish.func.dat %>% dplyr::mutate(yesno = 1) %>%
  tidyr::spread(Vertical_Distribution,
    yesno, fill = 0)
hab.dat.binary <- hab.dat.binary[, c("Littoral",
  "Continental_shelf", "Continental_slope",
  "Open_ocean", "Habitat_soft", "Habitat_rocky",
  "Habitat_posidonia", "Benthic", "Demersal",
  "Pelagic")]
rownames(hab.dat.binary) <- Medfish.func.dat$Species

diet.dat.binary <- Medfish.func.dat %>% dplyr::mutate(yesno = 1) %>%
  tidyr::spread(Diet_type_Larvae, yesno,
    fill = 0) %>% dplyr::mutate(yesno = 1) %>%
  tidyr::spread(Diet_type_Adult, yesno,
    fill = 0)

diet.dat.binary <- diet.dat.binary[, c("Leicitotrophic",
  "Planktotrophic", "Carnivorous", "Herbivorous",
  "Omnivorous", "Piscivorous")]
rownames(diet.dat.binary) <- Medfish.func.dat$Species

body.length <- data.frame(Medfish.func.dat$Common_length)
rownames(body.length) <- Medfish.func.dat$Species

if (!require(BBS.occurrences)) {
  devtools::install_github("nicholasjclark/BBS.occurrences")
}

func.matrix <- BBS.occurrences::prepDendrograms(func.datasets = list(hab.dat.binary,
  diet.dat.binary, body.length), prep.types = list("prep.binary",
  "prep.binary", "Q"))

Medfish.func.dendrogram <- ape::as.phylo(hclust(dist(func.matrix),
  method = "complete"))
```

Table 3: Preview of the functional distance matrix

	Abudefduf_vaigiensis	Acanthocybium_solandri	Acantholabrus_palloni
Abudefduf_vaigiensis	0.0000000	0.1583111	0.0077456
Acanthocybium_solandri	0.1583111	0.0000000	0.1506894
Acantholabrus_palloni	0.0077456	0.1506894	0.0000000

Save the datasets as a `.rda` file in the `Processed data` directory

```
save(Medfish.mrf.dat, Medfish.tree, Medfish.func.dendrogram,  
     Medfish.func.dat, mrf.species, coords,  
     Obs.1980.Climate.dat, Proj.2040.2059.Climate.dat,  
     Proj.2040.2059.Medfish.dat, Proj.2080.2099.Climate.dat,  
     Proj.2080.2099.Medfish.dat, file = "Processed data/MRF.data.rda")
```

References

Albouy, C., Lasram, F.B.R., Velez, L., Guilhaumon, F., Meynard, C.N., Boyer, S., Benestan, L., Mouquet, N., Douzery, E., Aznar, R., Troussellier, M., Somot, S., Leprieur, F., Le Loc'h, F. & Mouillot, D. (2015) FishMed: traits, phylogeny, current and projected species distribution of Mediterranean fishes, and environmental data. *Ecology*, 96, 2312-2313.