tRophicPosition one baseline example: variation in Fish TP from Lakes Inari and Kilpis (Finland)

Claudio Quezada-Romegialli, Kimmo K. Kahilainen & Chris Harrod

October 5 2017

In this vignette we show how to calculate trophic position with a one baseline Bayesian model. We will use an example of two food webs datasets gathered and compiled by Kimmo Kahilainen's research team about several fish species and whitefish morphs captured in two different subarctic lakes: Inarijärvi (Inari; 68.58ºN, 27.4ºE) and Kilpisjärvi (Kilpis; 69.03ºN, 20.33ºE), located in northern Finnish Lapland.

Introduction

Data have been initially collected to understand the role of ecological divergence of whitefish morphs in the food webs (Thomas et al. 2017). Both lakes are large (Inari 1043 km2, Kilpis 37 km2) and deep (mean depth 14.5 and 19.4 m) oligotrophic lakes (totP 4-5 µg/L). In both lakes the water column is well oxygenated year-round providing deep water habitat for various fish species. Three principal habitat types, litoral, pelagic and profundal, are well developed in both lakes. Invertebrates from all three habitat types were collected for stable isotope baseline purpose. Litoral (BMI_litoral) and profundal (BMI_profundal) macroinvertebrate samples were collect using a benthic grab, whereas pelagic zooplankton (zpl) was collected using a plankton net.

Fish were sampled using multi-mesh gill nets, beach seine, pelagic trawl, long-line and lure from all habitat types to assure capture of different species in community (Thomas et al. 2017). Both subarctic lakes are dominated by salmonid fish, of which whitefish (*Coregonus lavaretus*) is clearly the most abundant species (Siwertsson et al. 2010). The region is famous on postglacial adaptive radiation and ecological speciation of whitefish, where the most divergent lakes, such as Lake Inari here, can harbor up to four sympatric morphs of whitefish (Thomas et al. 2017). Whitefish morphs are ecomorphologically and genetically differentiated to use of pelagic, littoral or profundal habitat (Harrod et al. 2010; Præbel et al. 2013). Whitefish morphs are named according to their body size and heritable trait, number of gill rakers, both showing pronounced differences among the morphs (Kahilainen et al., 2011, 2017). Here, small sparsely rakered (ssr) whitefish is a small size profundal benthivore with a very low amount of short gill rakers, while large sparsely rakered (lsr) whitefish is a large sized littoral benthivore in polymorphic lakes and a generalist in monomorphic lakes with intermediate number of gill rakers in both cases (Harrod et al. 2010; Hayden et al. 2014). Large densely rakered (ldr) is also large sized morph, but use pelagic resources and has a high number of long gill rakers. Finally, densely rakered (dr) whitefish is a small sized pelagic zooplanktivore having the highest number of long, fine and densely spaced gill rakers.

The Lake Inari also includes two introduced species, the zooplanktivorous vendace (*Coregonus albula*) and piscivorous lake trout (*Salvelinus namaycush*), of which the former may hybridize with whitefish (Kahilainen et al. 2011). The other piscivorous species in the lakes are Arctic charr (*Salvelinus alpinus*), pike (*Esox lucius*), burbot (*Lota lota*) and brown trout (*Salmo trutta*), whereas littoral grayling (*Thymallus thymallus*) and perch (*Perca fluviatilis*) may shown an ontogenetic shift to piscivory (Thomas et al. 2017). The small sized forage fish alpine bullhead (*Cottus poecilopus*), 9-spined stickleback (*Pungitius pungitius*) and minnow (*Phoxinus phoxinus*) are mainly littoral species (Thomas et al. 2017). In relatively simple food web structure lakes e.g. Lake Kilpis, parasites, such as tapeworms *Diphyllobothrium* spp. may reach considerable densities and have thus significant role in various levels of food web (Hayden et al. 2014).

Installing tRophicPosition

First of all, you need to install JAGS for your platform, and then install the stable version of `tRophicPosition` from CRAN:

```
install.packages("tRophicPosition")
```

After that, you have to load the package with:

```
library(tRophicPosition)
```

```
## This is tRophicPosition 0.7.2
```

Inari and Kilpis Lakes - one baseline example

We start loading the dataset into R. We have included the dataset for convenience:

```
data("Finnish_Lakes")
```

As usual, the first thing to do is to check its structure:

```
dplyr::glimpse(Finnish_Lakes)
```

```
## Observations: 1,258
## Variables: 7
## $ Lake          <fctr> kilpis, kilpis, kilpis, kilpis, kilpis, kilpis,...
## $ Species.group <ord> whitefish lsr, whitefish lsr, whitefish lsr, whi...
## $ d13C          <dbl> -25.599, -25.208, -25.730, -17.053, -27.239, -25...
## $ d15N          <dbl> 8.731, 8.848, 8.210, 7.035, 8.663, 8.135, 8.831,...
## $ C             <dbl> 46.781, 47.337, 47.083, 47.701, 47.471, 47.499, ...
## $ N             <dbl> 14.990, 14.891, 14.810, 14.750, 14.873, 14.790, ...
## $ C.N           <dbl> 3.120864, 3.179004, 3.179180, 3.233940, 3.191848...
```

Here we see that the dataset is organised in several columns:

- `Lake` is a factor with two levels, each representing one Lake;
- `Species.group` is a factor with 20 levels, each representing one species;
- `d13C`, `d15N`, `C`, `N` and `C.N` are numerical variables, representing δ^{13}C and δ^{15}N isotope values, ammount of C and N, and ratio C/N, respectively.

In this dataset, `Species.group` represents all fish species found, plus litoral and profundal benthic macroinvertebrates, zooplankton and a parasite - *Diphyllobothrium* spp. We have to create a new variable to describe the functional group of fish species and these groups:

```
Lakes <- dplyr::mutate(Finnish_Lakes,
                       FG = ifelse(Species.group %in%
                                     c("BMI_litoral", "BMI_profundal",
                                       "zooplankton"), "Baseline",
                                   ifelse(Species.group == "diphyllobothrium",
                                          "Parasite", "Fish")))
```

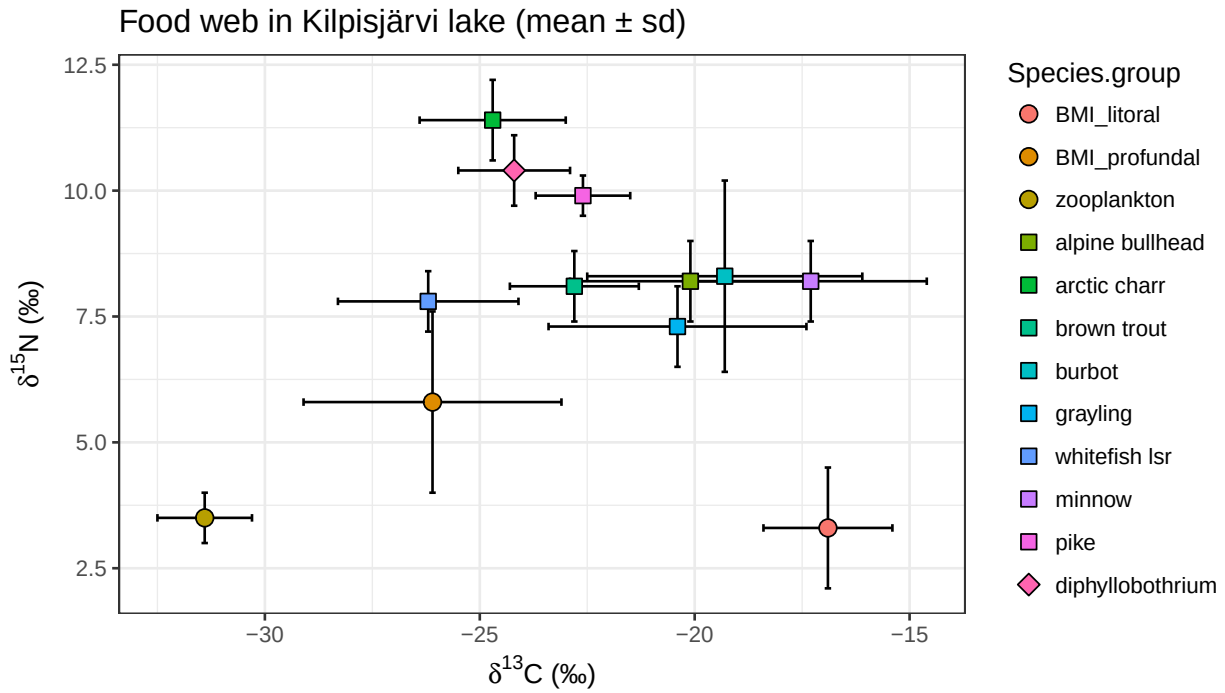
```
## Warning: package 'bindrcpp' was built under R version 3.4.1
```

```
# And also we have to convert the FG variable into an ordered factor
Lakes$FG <- factor(Lakes$FG, ordered = TRUE)
```

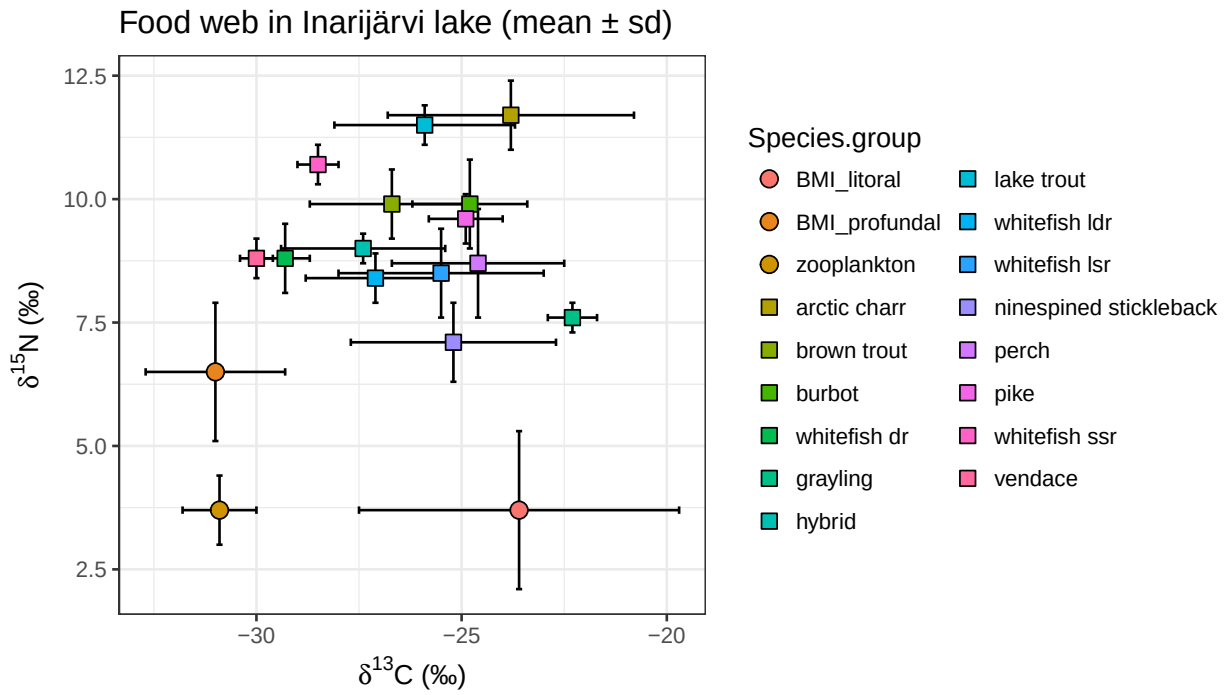
You see that we use `dplyr::mutate`. Using this code avoids to load `dplyr` into memory, and use the function `mutate` from the package `dplyr` directly (obviously, `dplyr` needs to have been installed before).

Now we are ready to screen the dataset

```
# For Kilpisjärvi, we subset Lakes with Lakes[which(Lakes$Lake == "kilpis"),]
Kilpis_plot <- screenFoodWeb(Lakes[which(Lakes$Lake == "kilpis"),],
                             grouping = c("Species.group", "FG"),
                             title = "Food web in Kilpisjärvi lake (mean \u00B1 sd)",
                             order = TRUE)
```




```
# And Lakes[which(Lakes$Lake == "inari"),] for Inarijärvi
Inari_plot <- screenFoodWeb(Lakes[which(Lakes$Lake == "inari"),],
                            grouping = c("Species.group", "FG"),
                            title = "Food web in Inarijärvi lake (mean \u00B1 sd)",
                            order = TRUE)
```



In both food webs we noted that zooplankton (pelagic baseline) and benthic macroinvertebrates (BMI_litoral, i.e. benthic baseline) have roughly the same δ^{15}N mean values, at least graphically. We will calculate a summary to check if this is correct:

```
# For Inarijärvi
summariseIsotopeData(Lakes[which(Lakes$Lake == "inari"),],
                     grouping = c("FG", "Species.group"))
```

```
##          FG          Species.group n_d13C mean_d13C SD_d13C n_d15N
## 1  Baseline            BMI_litoral     51     -23.6     3.9     51
## 2  Baseline          BMI_profundal     25     -31.0     1.7     25
## 3  Baseline             zooplankton      3     -30.9     0.9      3
## 4      Fish           arctic charr     79     -23.8     3.0     79
## 5      Fish            brown trout    115     -26.7     2.0    115
## 6      Fish                 burbot     31     -24.8     1.4     31
## 7      Fish           whitefish dr     50     -29.3     0.6     50
## 8      Fish               grayling     27     -22.3     0.6     27
## 9      Fish                 hybrid      4     -27.4     2.0      4
## 10     Fish             lake trout     21     -25.9     2.2     21
## 11     Fish          whitefish ldr     48     -27.1     1.7     48
## 12     Fish          whitefish lsr     50     -25.5     2.5     50
## 13     Fish ninespined stickleback     30     -25.2     2.5     30
## 14     Fish                  perch     50     -24.6     2.1     50
## 15     Fish                   pike     25     -24.9     0.9     25
## 16     Fish          whitefish ssr     50     -28.5     0.5     50
## 17     Fish                vendace     50     -30.0     0.4     50
##    mean_d15N SD_d15N
## 1        3.7     1.6
## 2        6.5     1.4
## 3        3.7     0.7
## 4       11.7     0.7
## 5        9.9     0.7
```

```
## 6        9.9     0.9
## 7        8.8     0.7
## 8        7.6     0.3
## 9        9.0     0.3
## 10      11.5     0.4
## 11       8.4     0.5
## 12       8.5     0.9
## 13       7.1     0.8
## 14       8.7     1.1
## 15       9.6     0.5
## 16      10.7     0.4
## 17       8.8     0.4
```

```r
# And for Kilpisjärvi
summariseIsotopeData(Lakes[which(Lakes$Lake == "kilpis"),],
                     grouping = c("FG", "Species.group"))
```

```
##          FG    Species.group n_d13C mean_d13C SD_d13C n_d15N mean_d15N
## 1  Baseline      BMI_litoral     31     -16.9     1.5     31       3.3
## 2  Baseline    BMI_profundal     32     -26.1     3.0     32       5.8
## 3  Baseline      zooplankton     14     -31.4     1.1     14       3.5
## 4      Fish  alpine bullhead     17     -20.1     2.8     17       8.2
## 5      Fish     arctic charr    115     -24.7     1.7    115      11.4
## 6      Fish      brown trout     21     -22.8     1.5     21       8.1
## 7      Fish            burbot     65     -19.3     3.2     65       8.3
## 8      Fish         grayling     19     -20.4     3.0     19       7.3
## 9      Fish    whitefish lsr    132     -26.2     2.1    132       7.8
## 10     Fish           minnow     12     -17.3     2.7     12       8.2
## 11     Fish             pike     35     -22.6     1.1     35       9.9
## 12 Parasite diphyllobothrium     56     -24.2     1.3     56      10.4
##    SD_d15N
## 1      1.2
## 2      1.8
## 3      0.5
## 4      0.8
## 5      0.8
## 6      0.7
## 7      1.9
## 8      0.8
## 9      0.6
## 10     0.8
## 11     0.4
## 12     0.7
```

Now we have seen each food web and a numerical summary of their δ^{13}C and δ^{15}N isotope values, we will delete some groups we are not interested in:

```r
# Delete BMI_profundal, diphyllobotrium, whitefish lsr and grayling from the Kilpis dataset
Kilpis_edited <- Lakes[Lakes$Lake == "kilpis" &
                         Lakes$Species.group != "grayling" &
                         Lakes$Species.group != "whitefish lsr" &
                         Lakes$Species.group != "diphyllobothrium" &
                         Lakes$Species.group != "BMI_profundal",]

# Delete BMI_profundal, zpl and BMI_litoral from the Inari dataset
Inari_edited <- Lakes[Lakes$Lake == "inari" &
```

```
                        Lakes$Species.group != "BMI_profundal" &
                        Lakes$Species.group != "hybrid",]

# And combine both dataframes
fishes <- rbind(Kilpis_edited, Inari_edited)
```

`fishes` is the main object we will iterate through, extracting δ^{13}C and δ^{15}N isotope values for each species, and its respective baseline. But first we must define a trophic discrimination factor, taking values from the bibliography:

```
TDF_values <- TDF()
```

```
## You selected Post's (2002)
##               d15N: 56 values with 3.4 mean +- 0.98 sd
```

```
## d13C: 107 values with 0.39 mean +- 1.3 sd
```

In this case we selected Post's (2002) values, and we saved them into the dataframe `TDF_values`. Now `TDF_values` has 2 columns (i.e. variables), `deltaC` and `deltaN`, which we will use when we extract the isotope values from the `fishes` data frame:

```
InariKilpisList <- extractIsotopeData(fishes, b1 = c("BMI_litoral", "zooplankton"),
                                       baselineColumn = "Species.group",
                                       speciesColumn = "Species.group",
                                       communityColumn = "Lake",
                                       deltaC = TDF_values$deltaC,
                                       deltaN = TDF_values$deltaN)
```

And now we calculate trophic position for each species. We will use 20,000 iterations for the adaptive phase, 20,000 iterations as burnin (iterations discarded at the beggining of posterior sampling) and 20,000 actual iterations. Also we will use 5 chains to calculate trophic position, while using a one baseline Bayesian model.

We will save all results into `InariKilpis_TPmodels`. To make things faster, we will calculate trophic position with a parallel processing approach that maximises the efficient use of computing power available to the user:

```
# First we create a cluster with the cores available
cluster <- parallel::makePSOCKcluster(parallel::detectCores())

# Then we run the model in parallel, nested in system.time()
# in order to know how much time it takes to finish calculations
system.time(InariKilpis_TPmodels <- parallel::parLapply(cluster,
                                                        InariKilpisList,
                                                        multiModelTP,
                                                        adapt = 20000,
                                                        n.iter = 20000,
                                                        burnin = 20000,
                                                        n.chains = 5,
                                                        model = "oneBaseline"))
```

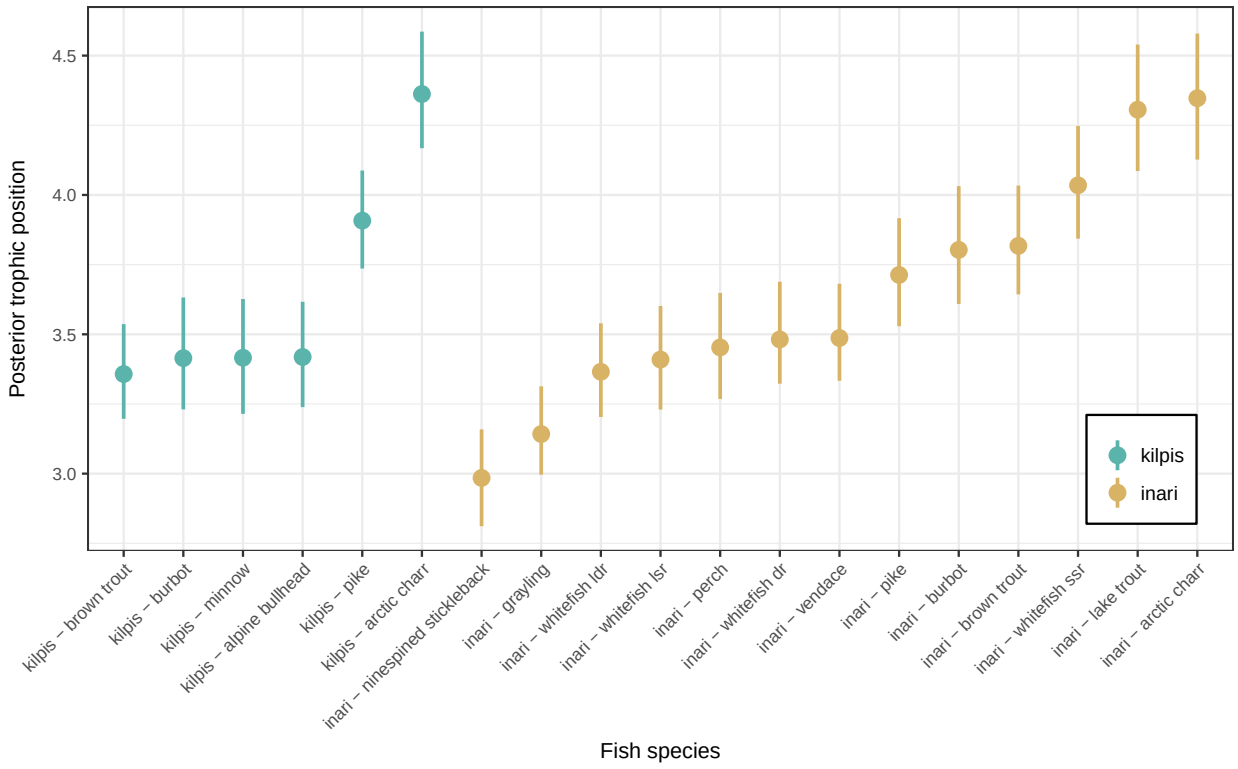
```
##    user  system elapsed
##    0.03    0.01   69.97
```

```
parallel::stopCluster(cluster)
```

We now analyse the posterior trophic position. As we saved results into `InariKilpis_TPmodels`, we have to get the data from it. We use the function `fromParallelTP()` to get the summary:

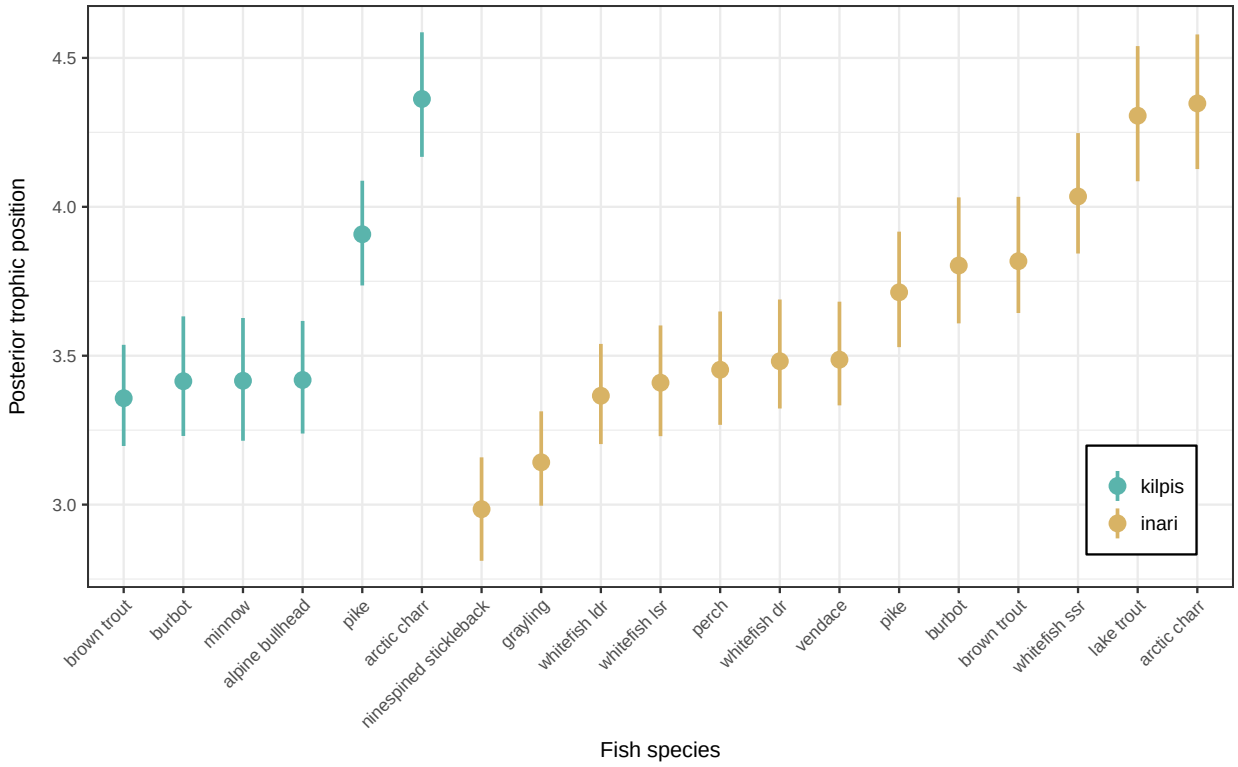
```
ggplot_df <- fromParallelTP(InariKilpis_TPmodels, get = "summary")
```

```
# Here we want to arrange the resulting plot first by community (Inarijärvi
#  first, Kilpisjärvi next) and then by the mode of each species
ggplot_df2 <- dplyr::arrange(ggplot_df, community, mode)

# As ggplot2 needs factors to be ordered, we prepare them with
values <- paste(ggplot_df2$community,"-",ggplot_df2$species)
ggplot_df2$species_ordered <- factor(values, levels = values, ordered = TRUE)

# And then we plot our data
plot_fishes <- credibilityIntervals(ggplot_df2, x = "species_ordered",
                                    plotAlpha = FALSE,
                                    legend = c(0.92,0.15),
                                    group_by = "community",
                                    xlab = "Fish species",
                                    scale_colour_manual = c("#5ab4ac",
                                                            "#d8b365"))
```



Above you see that labels of the x axis have the lake name concatenated to species name. This is needed, as each element on the x axis can not have the same name, otherwise each species would have two trophic positions plotted and the resultant plot would be confusing. As `credibilityIntervals()`, `screenFoodWeb()`, and also the S3 `plot()` method return an actual `ggplot2` object, we can modify it as any ggplot2 plot. In this case, we want to change the x axis labels to be only the species. We use the following code to do that:

```
plot_fishes <- plot_fishes +
  ggplot2::scale_x_discrete(labels = as.character(ggplot_df2$species))

# As we modified a ggplot2 object, we need to print it afterwards
print(plot_fishes)
```



Now, we need to get trophic position posterior estimations from the object `InariKilpis_TPmodels` (with parallel calculations). We use the same function `fromParallelTP()`, but this time we get "TP":

```
TPs <- fromParallelTP(InariKilpis_TPmodels, get = "TP")
```

Looking at the plot above, we noted that Pike and Arctic Charr (in Kilpisjärvi lake) and Lake Trout and Arctic Charr (in Inarijärvi lake) have the highest trophic positions. So, following Post, Pace & Hairston (2000), we are interested to know which species has the Maximum Trophic Position (MTP) in both lakes:

```
# First we select specifically those 4 species
MTP <- list("Kilpis_Pike" = TPs$`kilpis-pike`,
            "Kilpis_ArcticCharr" = TPs$`kilpis-arctic charr`,
            "Inari_LakeTrout" = TPs$`inari-lake trout`,
            "Inari_ArcticCharr" = TPs$`inari-arctic charr`)

# Then we calculate the mode of each posterior trophic position
(MTP_modes <- sapply(MTP, getPosteriorMode))
```

```
##        Kilpis_Pike Kilpis_ArcticCharr    Inari_LakeTrout
##              3.908              4.362              4.306
##  Inari_ArcticCharr
##              4.347
```

```
# And then we calculate pairwise comparisons to test if they are different
# First using the approach delineated in SIBER (see the procedure
# Comparing the posterior distributions in
# https://github.com/AndrewLJackson/SIAR-examples-and-queries/blob/master/learning-resources/siber-comp
pairwiseComparisons(MTP, test = ">=")
```

```
##                          [1]   [2]   [3]   [4]
## [1] Kilpis_Pike        0.000 0.000 0.004 0.001
```

```
## [2] Kilpis_ArcticCharr 1.000 0.000 0.660 0.539
## [3] Inari_LakeTrout    0.996 0.340 0.000 0.379
## [4] Inari_ArcticCharr  0.999 0.461 0.621 0.000
```

```
# Or calculating the overlapping proportion between each pair of posterior distributions
# using the Bhattacharyya coefficient (implemented in dispRity package)
pairwiseComparisons(MTP, test = "bhatt")
```

```
##                          [1]   [2]   [3]   [4]
## [1] Kilpis_Pike        0.000 0.065 0.169 0.098
## [2] Kilpis_ArcticCharr 0.065 0.000 0.949 0.993
## [3] Inari_LakeTrout    0.169 0.949 0.000 0.973
## [4] Inari_ArcticCharr  0.098 0.993 0.973 0.000
```

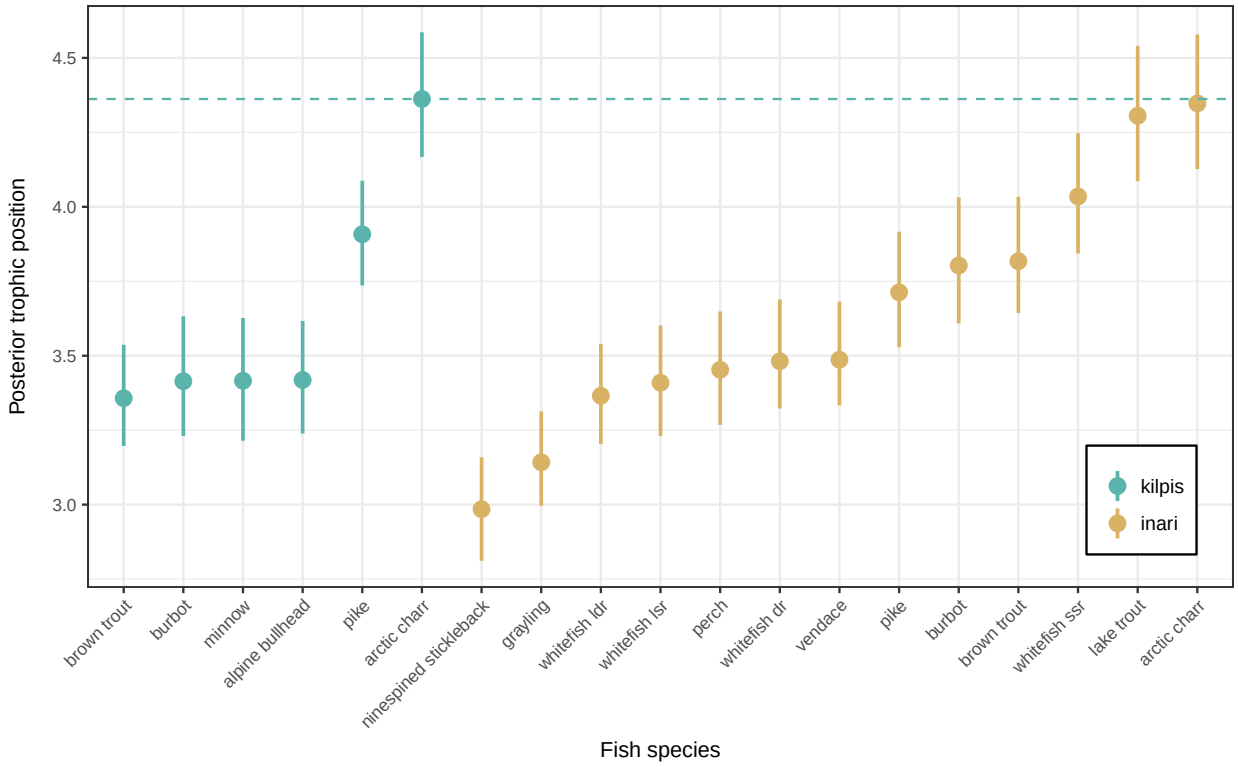
When we use a logical test (either of "<", "<=", ">" or ">=") to compare two posterior distributions, we are randomly taking a sample from each posterior distribution, comparing them using the logical test selected, and repeating the procedure until all posterior estimates were compared. Thus, if we select ">=" as a logical test, the matrix must be read as: what is the probability that the consumer in the row has a posterior trophic position higher than or equal to consumer in the column? As this is based on Bayesian inference, the higher the value, the more confidence we will have in that comparison. For example, how confident we are saying that lake trout in Inari Lake has a higher trophic position compared to Arctic charr in the same lake? As the value is 0.379, we are not very confident stating that. Conversely, how confident we are saying that Arctic charr has a higher trophic position compared to pike in Kilpis Lake? As the probability is quite high (1), we are pretty confident stating that.

The Bhattacharyya coefficient, however, calculates the probability of overlap between two distributions. Considering the same two examples in the paragraph above, the probability of overlap between lake trout and Arctic charr in Inari Lake is 0.973, thus they exhibit a high degree of overlap. Arctic charr and pike in Kilpis Lake show only 0.065 as probability of overlap.

Considering the results from above we see that Arctic charr in Kilpisjärvi has the highest mode of posterior trophic positions among these 4 species, but the probability that is higher or equal than lake trout and Arctic charr in Inari is rather low, being only higher to pike in Kilpis. Also, considering the Bhattacharyya coefficient, we see that these three species (Arctic charr in Kilpis, and lake trout and Arctic charr in Inari) exhibit a high degree of overlap. Thus we conclude that posterior trophic position estimations of Arctic charr in Kilpis, and lake trout and Arctic charr in Inari are not different each other.

Finnally, we add a horizontal line to the plot to indicate which is the maximum trophic position (MTP) in both lakes:

```
plot_fishes + ggplot2::geom_hline(yintercept = max(MTP_modes),
                                  linetype = "44",
                                  colour = "#5ab4ac")
```



References

1. Harrod, C., Mallela, J. & Kahilainen, K.K. 2010: Phenotype-environment correlations in a putative whitefish adaptive radiation. Journal of Animal Ecology 79: 1057-1068.
2. Hayden, B., Harrod, C. & Kahilainen, K.K. 2014: Dual-fuels: intra-annual variation in the relative importance of benthic and pelagic resources to maintenance, growth and reproduction in a generalist salmonid fish. Journal of Animal Ecology 83: 1501-1512.
3. Kahilainen, K.K., Østbye, K., Harrod, C., Shikano, T., Malinen, T. & Merilä, J. 2011: Species introduction promotes hybridization and introgression in Coregonus: is there sign of selection against hybrids? Molecular Ecology 20: 3838-3855.
4. Præbel, K., Knudsen, R., Siwertsson, A., Karhunen, M., Kahilainen, K.K., Ovaskainen, O., Østbye, K., Peruzzi, S. & Fevolden, S-E. & Amundsen P.-A 2013. Ecological speciation in postglacial European whitefish: rapid adaptive radiations into the littoral, pelagic and profundal lake habitats. Ecology and Evolution 3: 4970-4986.
5. Siwertsson, A., Knudsen, R., Kahilainen, K.K., Præbel, K., Primicerio, R. & Amundsen, P-A. 2010. Sympatric diversification as influenced by ecological opportunity and historical contingency in a young species lineage of whitefish. Evolutionary Ecology Research 12: 929-947.
6. Thomas, S.M., Harrod, C., Hayden, B., Malinen, T. & Kahilainen, K.K. 2017. Ecological speciation in a generalist consumer expands the trophic niche of a dominant predator. Scientific Reports. 7, 8765 doi: 10.1038/s41598-017-08263-9.