Niche partitioning between planktivorous fish in the pelagic Baltic Sea assessed by DNA metabarcoding, qPCR and microscopy

README

Andreas Novotny, Kinlan Mehdi Goulwen Jan, Jan Dierking and Monika Winder

Contents

---

1 R scripts

3 Scripts were used for data analysis:

1. AdultFish_DataFiltration.Rmd
2. AdultFish_DataAnalyses.Rmd
3. Larvae_Analyses.Rmd

Note: The figure numbering in the scripts may not follow the same numbering than the manuscript

---

1.1 AdultFish_DataFiltration.Rmd

In this script the *18S* (**ps_18S_2020.rds**) and *COI* (**ps_COI_2020.rds**) data were:

- Imported as phyloseq objects
- Filtration to only keep adult fish and to remove
  - Teleostei sequences for *18S*
  - Host sequences for *COI*
- Rarefaction curves were made showing that all samples reached a plateau of alpha diversity
  - This step was facilited by the function found on: https://github.com/joey711/phyloseq/issues/143
- Filtered data were then rarefied at even depth to have the same sampling depth among samples
  - The rarefied phyloseq objects were saved under the name **ps_barcode_r.rds**, were *r* is for *rarefied*
  - It is important to save a file and to stick to it, as the rarefaction is based on a random subsampling
- Based on the **ps_barcode_r.rds** file, prevalent species (contributing to more than 0.1% in 70% of the sample group) is kept
  - The filtered files were saved as **ps_barcode_r_f.rds**, were *f* is for *filtered*
- A summary table is made to show the data that passed the quality control (Table 1 in the manuscript)

1.2 AdultFish_DataAnalyses.Rmd

In this script the main part of the data analysis is made:

1. DNA metabarcoding
2. RT-qPCR
3. Microscopic counts
4. Dissolved oxygen ("ICES Data Portal, Dataset on Ocean Hydrochemistry, 2022. ICES, Copenhagen")
5. Zooplankton abundance (https://sharkweb.smhi.se/hamta-data/)

1.2.1 DNA Metabarcoding

- In this part data were prepared for further analyses:
  - Reads were transformed as relative abundance
- Different plot were made
  - Barplot
  - Bipartite plot
  - Heatmap of the BCI
  - NMDS plot
  - Boxplot of the alpha diversity (Shannon Index)
- Statistical analyses were performed
  - PermANOVA
  - Pairwise PermANOVA
  - Beta regression on the prey contributing to most of the difference between fish species
  - ANOVA on the Shannon Index between barcode **and** fish species

1.2.2 RT-qPCR

- Boxplot of the different abundance of rotifer/fish gut
- ANOVA of log-transformated abundance is performed

1.2.3 Microscopic counts

- Data are shown by a barplot
- Student t-test was performed to test the difference between qPCR and microscopic for quantifying rotifers in sticklebacks gut
  - *Normality* was verified with **qqplots** and **Shapiro-Wilk normality test**
  - *Variance* was verified with a **F test of variance**

1.2.4 Zooplankton abundance

- Data were mined from sharkweb
- Data were visualized using the relative abundance from *ggplot*
  - *geom_bar(aes(fill=Genus), stat="identity", position="fill", width = 0.7)*

1.3 Larvae_Analyses.Rmd

This script was used to visualise and analyses sequences reads associated to Sprat and Flounder larvae

- Data filtration
  - First filtration to keep only fish larvae and to remove host sequences
  - Second filtration to keep only reads that reprensent more than 5% of the total reads
- Data visualisation
  - Barplot for each unique sample were made and grouped by *fish species* and *Baltic Sea Basin*
- Statistics

– PermANOVA based on Bray-Curtis distance between fish species were performed for both barcode
– Species contributing to more than 70% of the difference between species were found using *simper*

---

2 Data

Different kind of data were used:

1. Metabarcoding Data
2. Count data
3. qPCR data
4. Data from the SharkWeb database (https://sharkweb.smhi.se/hamta-data/)
5. Station coordinates

2.1 Metabarcoding Data

These data are in 2 folders ...

1. 18S

```
setwd("./Data/18S")
```

2. COI

```
setwd("./Data/COI")
```

... and contain the same kind of data *phyloseq object* named *ps*

phyloseq objects are a combination of:

1. **ASV Table** (ASV_Table.csv)

- ASV created with the DADA2 pipeline

2. **Sample Data** (Metadata.csv)

- **rowname**: Unique identification code from NGI
- **Well**: Well in the plate for the sequencing
- **Verify**:
- **Your.sample.name**: Unique identification in the lab
- **LIBRARY_ID**: Amplificom library identifier
- **LIBRARY_type**: COI or 18S
- **LIBRARY_gene**: COI or 18S
- **SORTED_genus**: Fish species
- **SAMPLE_type**: Fish/Fish Larvae
- **SORTED_number**: Number of individual used to extract the DNA
- **CRUISE_ID**: sampling cruise identifier
- **SAMPLE_date**: Date
- **STATION_ID**: Name of the station
- **STATION_comment**: Basin in which the samples were taken
- **STATION_lat**: Latitude of sampling
- **STATION_long**: Longitude of sampling
- **STATION_depth**: Depth of the station
- **STATION_helcom**: NA
- **SAMPLE_event**: Identifier of the sampling event

- **SAMPLE_time**: time of the sampling
- **SAMPLE_method**: Gear used for sampling
- **FILTER_poresize_µm**: NA
- **FILTER_volume**: NA
- **SAMPLE_preservation_time**: Preservation time
- **FISH_species**: Fish species
- **FISH_length**: Length of the fish (categorical variable)
- **FISH_amount**: Number of fish catched of this species used for further analyses
- **FISH_dissection**: Stomach / whole fish
- **LIBRARY_date**: Date of library preparation
- **LIBRARY_plate**: Identifier of the plate used for library preparation
- **LIBRARY_well**: Position in the plate for library preparation
- **WATER**: Volume of water used for final elution (ul)
- **SAMPLE**: Volume of sample used for final elution (ul)
- **Rel_Concentration**: Relative final concentration
- **LIBRARY_concentration**: DNA concentration after library preparation (ng/ul)
- **DNA_ID**: Unique identifier (same for same biological replicate)
- **DNA_date**: Date of the DNA extraction
- **DNA_person**: Person who extracted the DNA
- **DNA_elution_volume**: Final elution volume after DNA extraction
- **DNA_protocol**: Name of the protocol used
- **SORTED_type**: Fish / Fish larvae
- **SAMPLE_weather**: Weather during the sampling event
- **Biologist**: Person that have taken the samples
- **Sample_comment**: Comments

3. **Taxonomy Table** (Tax_table.csv)

---

1. Data in this form: **ps_barcode_2020.rds** are the *unfiltered* data
2. Data finishing by **r** are the *rarefied* data
3. Data finishing by **r_f** are the *rarefied* and *filtered* ones

---

2.2 Count data

```
readxl::read_excel("./Data/Fish_guts_Counting.xlsx")
```

It is an excel file containing:

1. **ID**, chr: Sample ID
2. **Species**, chr: Fish species (Sprat, Herring or Stickleback)
3. **Size**, chr: Fish size (small large or mixed)
4. **Station**, chr: Sea basin (Arkona, Gottland Basin, Northern Bornholm Basin or Southern Bornholm Basin)
5. **Copepoda**, num: Copepods count in gut content
6. **Cladocera**, num: Cladocerans count in gut content
7. **Rotifera_egg**, num: Rotifera eggs count in gut content
8. **Fish_egg**, num: Fish eggs count in gut content
9. **Analyzed_prop** num: The proportion of the gut analysed (0-1)

2.3 qPCR data

```
read.delim("./Data/2019-12-10_Rot18S_allSamples_both_plates.txt")
```

It is a text file containing:

1. **Include**, logi
2. **Color**, int
3. **Pos**, chr: Position on the plate
4. **Name**, chr
5. **Tm1**, num
6. **Tm2**, num
7. **Cp**, num
8. **Concentration**, num
9. **Standard**, int
10. **Status**, chr
11. **Plate**, int
12. **ID**, chr
13. **Species**, chr
14. **Size**, chr
15. **Station**, chr: Sea basin

2.4 Data from databases

Data for ploting zooplankton abundance were retrived in the *SharkWeb* database (https://sharkweb.smhi.se/hamta-data/):

2 options are possible to download the data:

1. Using the SMHI interface
2. Running the lines 1344-1349 of the AdultFish_DataAnalyses.Rmd script

2.5 Station coordinates

A small dataset with the decimal coordinate of the sampling stations:

```
read.csv("./Data/Station.coordinates.csv")
```

1. **Station**, chr
2. **Lat.dec**, dbl: Latitude
3. **Long.dec**, dbl: Longitude

---

3 Output

Note that all folders in the Output folder are intentionally left empty!

- All the outputs are in this folder
  - Data
  - Figures
  - Statistics

---

4 Reference

SMHI and Hav (2020) Svenskt Havarkiv, www.sharkdata.smhi.se.