forcis: An R package for accessing, handling and analysing the FORCIS database

Nicolas Casajus [1], **Sonia Chaabane** [1,2,3], **Xavier Giraud** [2], **Thibault de Garidel-Thoron** [2], **and Mattia Greco** [4]

1 Fondation pour la Recherche sur la Biodiversité (FRB-CESAB), Montpellier, France **2** Aix-Marseille Université, CNRS, IRD, INRAE, CEREGE, Aix-en-Provence, France **3** Department of Climate Geochemistry, Max Planck Institute for Chemistry, Mainz, Germany **4** Institute of Marine Sciences (ICM), CSIC, Barcelona, Catalonia, Spain

Summary

`forcis` is an R package designed to streamline access to the recently published FORCIS (Foraminifera Response to Climatic Stress) database (Chaabane et al., 2023). This package enables users to easily download the database directly into an R environment, filter and select relevant data, convert species counts across formats, and visualise the results.

Statement of need

The recently developed FORCIS (Foraminifera Response to Climatic Stress) database provides one of the most comprehensive collections of global planktonic foraminifera living census data, comprising over 163,000 samples collected via various sampling devices (Continuous Plankton Recorder — CPR —, plankton nets, pumps, and sediment traps). These samples span a wide temporal range (1910 to 2018), vertical depths (surface to 5,000 m), and spatial coverage (Chaabane et al., 2023; de Garidel-Thoron et al., 2022). FORCIS data are crucial for advancing insights into potential spatial and vertical migrations and understanding the impacts of global climate change on planktonic foraminifera biogeography and their seasonal and vertical distribution patterns observed in recent decades. Additionally, FORCIS's long temporal scope offers a valuable resource for investigating the influence of anthropogenic changes on planktonic foraminifera distribution and ecology (Chaabane et al., 2024).

However, working with the FORCIS database presents significant challenges due to the heterogeneity of the data, which has been compiled from 140 sources, each using its own taxonomic framework and reporting formats (Figure 1). This results in variability in data units, such as concentrations, frequencies, and raw counts, requiring extensive standardisation for meaningful comparison. Furthermore, the metadata associated with each sample — such as location, sampling depth, time, and environmental parameters — adds another layer of complexity, making data extraction and analysis challenging for users.
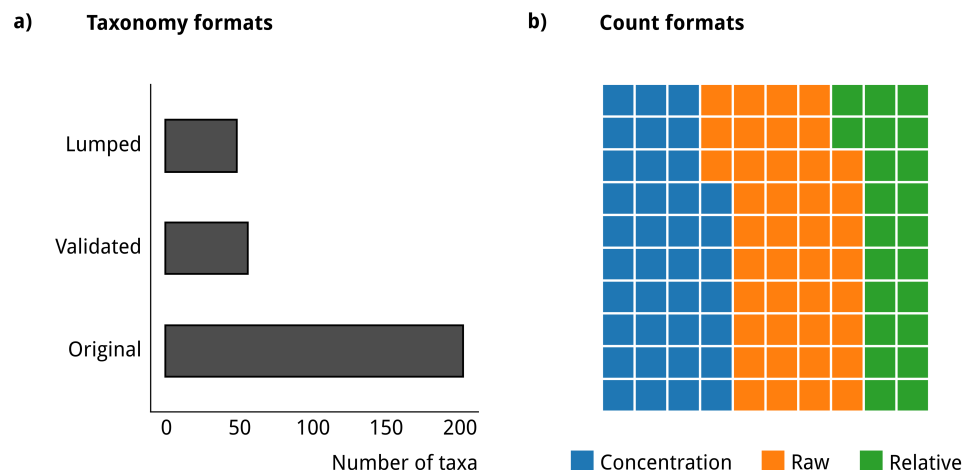
Figure 1: Heterogeneity of data within the FORCIS database. a) Number of taxa from different taxonomic frameworks present in the FORCIS database (net data). b) Different count formats from net samples included in the FORCIS database.

To overcome these obstacles, we developed the `forcis` package, an easy-to-use tool made especially for using the R programming environment to access, filter, harmonise, and visualise the FORCIS data. The `forcis` package enables users to download data directly from Zenodo the latest version of the FORCIS database, filter and select data according to user-specified criteria, harmonise taxonomic resolution, convert species counts into uniform units, and visualise patterns in diversity and abundance. By combining these features, the package enables researchers to access and analyse the data within the FORCIS database efficiently, streamlining their investigative efforts.

Main features

To facilitate efficient management and analysis of the FORCIS database, the `forcis` R package provides a comprehensive set of features fully described in the package vignettes, where users can find extensive documentation and tutorials on the major features of the package. The recommended workflow and the relevant main functions are illustrated in Figure 2.
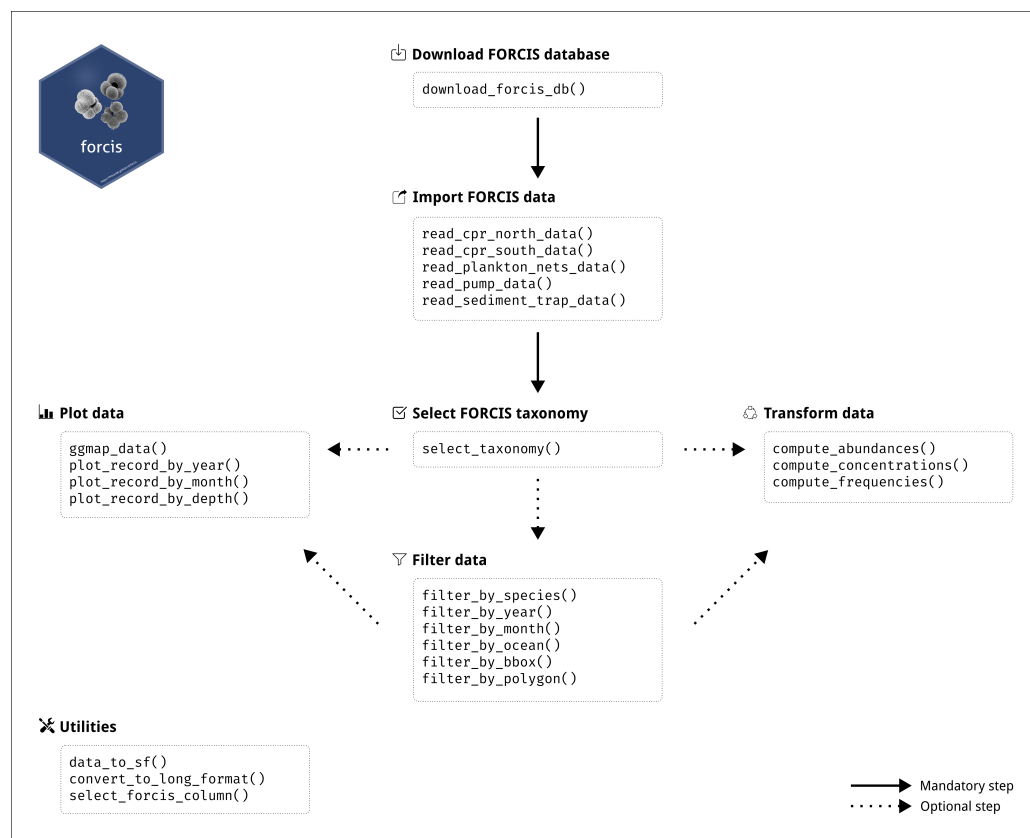
Figure 2: Recommended workflow and main features of the `forcis` R package.

Download and import FORCIS database in R

The `forcis` R package contains functions that simplify downloading and importing FORCIS datasets from [Zenodo](). The FORCIS database's most recent version can be retrieved using the function `download_forcis_db()`.

```r
# Create a data/ directory in the current directory ----
dir.create("data")
```

```r
# Download the latest version of the FORCIS database ----
download_forcis_db(path = "data", timeout = 300)
```

The `read_*_data()` function family helps users in importing dataset specific to a particular sampling device, enabling focused analyses.

```r
# Import plankton nets data (previously downloaded) ----
net_data <- read_plankton_nets_data(path = "data")
```

Once the data is imported in R, users can reduce the dataset to include only the metadata they are interested in by using the function `select_forcis_columns()`.

Harmonising taxonomy

To utilise most features of the `forcis` R package, users need to specify the taxonomic framework they wish to apply ([Figure 2]()). The FORCIS database includes counts at three different taxonomic levels: Original Taxonomy (OT), Lumped Taxonomy (LT), and Validated Taxonomy (VT). For a detailed explanation of the differences between these three taxonomic levels, we refer the reader to the FORCIS data descriptor ([Chaabane et al., 2023]()). For selecting

the taxonomic framework of choice, the users can use the function `select_taxonomy()` following the example below:

```
# Select a taxonomic framework ----
net_data_vt <- net_data |>
  select_taxonomy(taxonomy = "VT")
```

Filter data

After selecting the taxonomic framework, the `forcis` R package offers multiple functions to efficiently subset the FORCIS datasets. Users may be interested in analysing community structure at a specific time, or location, or even examining the counts of species of interest. Given the wide range of potential research questions, we have implemented six filtering functions within the `filter_by_*()` function family, allowing users to customise data extraction according to their investigation needs ([Figure 2](#)).

```
# Filter data by year(s) ----
net_data_sub <- net_data_vt |>
  filter_by_year(years = 1992)

# Filter data by spatial bounding box ----
net_data_sub <- net_data_vt |>
  filter_by_bbox(bbox = c(45, -61, 82, -24))

# Filter data by ocean name ----
net_data_sub <- net_data_vt |>
  filter_by_ocean(ocean = "Indian Ocean")

# Filter data by species ----
net_data_sub <- net_data_vt |>
  filter_by_species(species = "n_pachyderma_VT")
```

Transform data

The `compute_*()` function family allows users to convert FORCIS data between raw abundance, number concentration, and relative abundance, enabling them to use the units that best suit their analyses and facilitating comparison between the FORCIS data and their own. These functions utilise sample metadata to perform unit conversions. Specifically, conversions between raw abundance and number concentration in the `forcis` R package are calculated for each taxon using the following equations:

$$C_{number} = \frac{N_{raw}}{V_{filtered}}$$

where C_{number} is the number concentration, N_{raw} is the raw abundance (count of individuals), and $V_{filtered}$ is the volume of water filtered (in m^3 or L, depending on the dataset).

$$Frequency = 100 \cdot \frac{N_{raw}}{N_{total}}$$

where $Frequency$ is the relative abundance (in percentage), N_{raw} is the raw abundance (count of individuals) of a given taxon, and N_{total} is the total raw abundance (sum of all individuals in the sample or subsample).

The users can decide whether to convert counts at a sample or subsample level (see Chaabane et al. ([2023](#))) as the `compute_*()` functions propose the `aggregate` argument. If aggregate

Casajus et al. (2025). forcis: An R package for accessing, handling and analysing the FORCIS database. *Journal of Open Source Software*, *10*(114), 9217. https://doi.org/10.21105/joss.09217.

= TRUE, the function will return the transformed counts of each species using the sample as the unit. If aggregate = FALSE, it will re-calculate the species' abundance by subsample.

Visualisation

The `forcis` package also includes multiple functions to visualise the spatial distribution of samples selected by users. The `ggmap_data()` function generates publication-ready maps, displaying sample locations at a global scale (Figure 3a). Additionally, users can visualise sample records by various time units (season, month, year) and by depth, using the functions from the `plot_record_by_*()` function family (Figure 3b-d). These functions can be seamlessly combined with the `filter_by_*()` family of functions, allowing users to customise their sample selections according to their specific research needs.
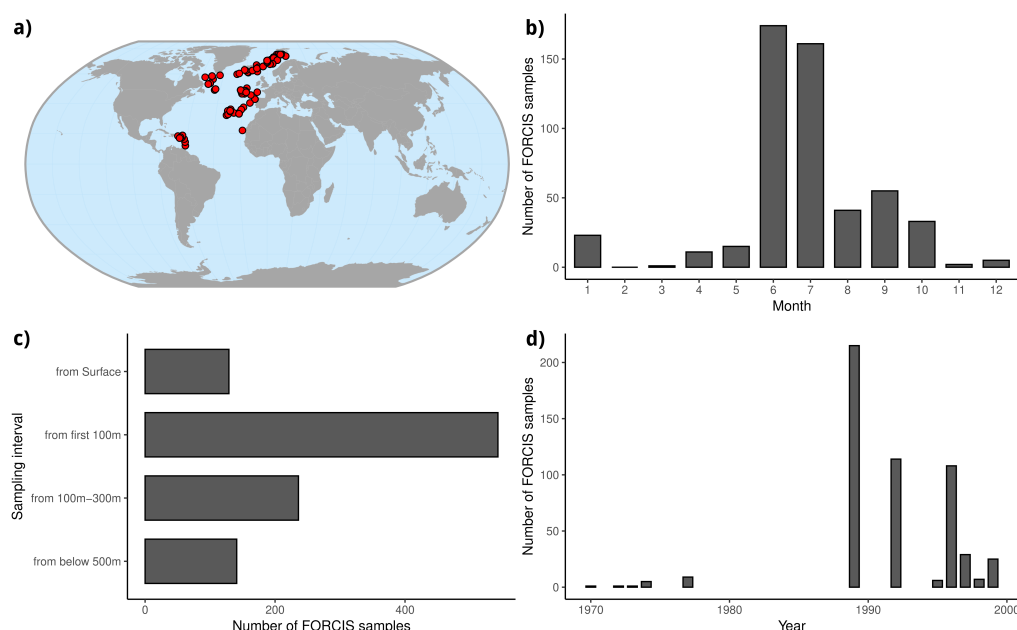

Figure 3: Overview of visualisations available in the `forcis` R package. a) World map produced by the function ggmap_data() to show the location of the data. b) Barplot of number of samples per month produced by the function plot_records_by_month(). c) Barplot of number of samples per depth class produced by the function plot_records_by_depth(). d) Barplot of number of samples per year produced by the function plot_records_by_year().

```
# Map raw net data ----
ggmap_data(net_data)

# Plot number of records by year of sampling ----
plot_record_by_year(net_data)

# Plot number of records by month of sampling ----
plot_record_by_month(net_data)

# Plot number of records by depth of sampling ----
plot_record_by_depth(net_data)
```

`forcis` provides five vignettes to learn more about the package:

- the Get started vignette describes the core features of the package
- the Database versions vignette provides information on how to deal with the versioning of the database

- the Select and filter data vignette shows examples to handle the FORCIS data
- the Data conversion vignette describes the conversion functions available in `forcis` to compute abundances, concentrations, and frequencies
- the Data visualization vignette describes the plotting functions available in `forcis`

Acknowledgements

References

Chaabane, S., de Garidel-Thoron, T., Giraud, X., Schiebel, R., Beaugrand, G., Brummer, G.-J., Casajus, N., Greco, M., Grigoratou, M., Howa, H., Jonkers, L., Kucera, M., Kuroyanagi, A., Meilland, J., Monteiro, F., Mortyn, G., Almogi-Labin, A., Asahi, H., Avnaim-Katav, S., … Ziveri, P. (2023). The FORCIS database: A global census of planktonic foraminifera from ocean waters. *Scientific Data*, *10*, 354. https://doi.org/10.1038/s41597-023-02264-2

Chaabane, S., de Garidel-Thoron, T., Meilland, J., Sulpis, O., Chalk, T. B., Brummer, G.-J. A., Mortyn, P. G., Giraud, X., Howa, H., Casajus, N., Kuroyanagi, A., Beaugrand, G., & Schiebel, R. (2024). Migrating is not enough for modern planktonic foraminifera in a changing ocean. *Nature*, *636*, 390–396. https://doi.org/10.1038/s41586-024-08191-5

de Garidel-Thoron, T., Chaabane, S., Giraud, X., Meilland, J., Jonkers, L., Kucera, M., Brummer, G.-J. A., Grigoratou, M., Monteiro, F. M., Greco, M., Mortyn, P. G., Kuroyanagi, A., Howa, H., Beaugrand, G., & Schiebel, R. (2022). The foraminiferal response to climate stressors project: Tracking the community response of planktonic foraminifera to historical climate change. *Frontiers in Marine Science*, *9*, 827962. https://doi.org/10.3389/fmars.2022.827962