Article

Harnessing RStudio package for ecological insights: Monitoring Diatoms with DiaThor

Arpita Srivastava[1], **Simoni Singhal**[1], **Anuradha Yadav**[1], **Rahat Zehra**[2], **Jyoti Verma**[1]

[1]Aquatic Ecology Lab, Department of Zoology, C.M.P Degree College, University of Allahabad, Prayagraj, U.P – 211001, India

[2]Geospatial Information Science and Engineering Hub, Indian Institute of Technology Bombay, Powai, Maharashtra, India

E-mail: diatombuster@gmail.com, jyoti.zoo@cmpcollege.ac.in

Received 1 July 2025; *Accepted 8 August 2025*; *Published online 20 August 2025*; *Published 1 March 2026*

Abstract

This study aims to integrate open-source computational tools with ecological biomonitoring by harnessing the capabilities of RStudio and the DiaThor package. The goal is to simplify the calculation of diatom-based indices and generate insightful visualizations for water quality assessment. Using RStudio as the platform and the DiaThor package as the computational engine, a case study was conducted on water quality datasets. Various ecological indices were computed including IPS, TDI, SLA, and others. The manuscript provides an end-to-end demonstration of data formatting, function usage, and graphical outputs for visual interpretation. The study successfully generated multiple plots such as bar plots, heatmaps, and radar diagrams, each representing the variation in diatom-based indices across multiple sample sites. The visualization outputs made it easier to detect ecological gradients and interpret water quality conditions efficiently. DiaThor in RStudio offers an efficient and reproducible method for ecological data processing. Its integration with R's visualization ecosystem enhances data clarity and enables automation in environmental monitoring workflows. This paper underscores the potential of open-source packages to revolutionize diatom-based assessment practices.

Keywords Diatoms; biomonitoring; ecological indices; visualization; open-source tools; water quality.

Computational Ecology and Software
ISSN 2220-721X
URL: http://www.iaees.org/publications/journals/ces/online-version.asp
RSS: http://www.iaees.org/publications/journals/ces/rss.xml
E-mail: ces@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

Diatoms (Bacillariophyta) are highly diverse, unicellular algae known for their sensitivity to changes in aquatic environments. They quickly respond to variations in water chemistry, such as nutrient levels, pH, and conductivity, as well as to physical conditions, making them excellent candidates for monitoring rivers and streams. Because each diatom species has specific tolerance levels and habitat preferences, analysing the composition of a diatom community can offer insights into the underlying water quality. The DiaThor R

package (Fig.1) (Nicolosi Gelis & Sathicq, 2020), is introduced as a valuable tool for extracting ecological insights from diatom assemblages in river systems. RStudio, the popular integrated development environment (IDE) for R, was developed by J.J. Allaire and released in 2011 by RStudio, Inc. (now known as Posit, PBC). The DiaThor package, designed for diatom-based water quality assessment, was developed by Martín Cejudo-Figueiras, Ignacio Hernández-Fariñas, and Francisco Rodríguez-Gallego. It provides automated tools to calculate ecological and biotic indices using diatom data and is freely available on CRAN and GitHub (Gelis et al., 2024; Jjallaire · GitHub, n.d.; "Posit," n.d.). Prioritizing ecological understanding over technical complexity, the study applies DiaThor to both real and simulated datasets to demonstrate how diatom- based indices and functional traits mirror environmental conditions.

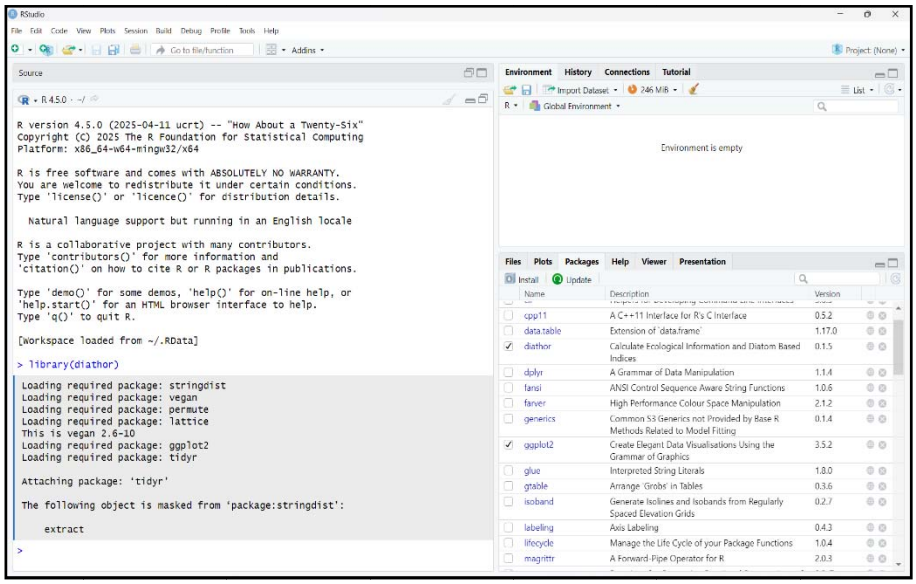

Fig. 0 Interface of RStudio with Diathor package.

This paper outlines the DiaThor workflow, from inputting species data to calculating a variety of biotic indices and ecological metrics. Findings from a random sample study show that DiaThor effectively tracks both spatial and temporal variations in water quality, successfully distinguishing less-impacted sites from polluted ones.

Notably, the trophic and pollution sensitivity indices generated through DiaThor showed strong congruence with measured nutrient concentrations and established water quality parameters and further confirmed that these indices track consistently along pollution gradients, reinforcing their reliability for biomonitoring. The findings are interpreted ecologically, with DiaThor capturing shifts in species diversity, community composition, and functional traits, highlighting the impacts of nutrient loading and pollution in freshwater systems.

Limitations such as the need for regional calibration and comprehensive species databases are acknowledged. Ultimately, DiaThor equips researchers and water managers with a powerful, open-source tool to translate diatom data into meaningful ecological insights. It modernises diatom monitoring, transforming a labour-intensive process into a streamlined, reproducible workflow. Diatoms are highly responsive to environmental change making them ideal Bio indicators and central to frameworks like the Water framework directive (WFD). Traditional diatom analysis based on manual identification and index calculation required significant expertise and time.

Early software like a Omnidia helped automate parts of this process index calculations and data management. These tools laid the foundation for more advanced platforms like DiaThor, which push Diatom biomonitoring into a more scalable, efficient, and data driven future.

More recently, dedicated software packages tailored for ecological analysis, like DiaThor, have emerged as powerful tools for processing diatom data (Lecointe et al., 1993).

2 Material and Methods

2.1 Evolution of DiaThor package in RStudio and new deep learning models for ecological analysis:

Diatoms, a prominent group of microalgae with silica-based cell walls, are known for their sensitivity to ecological conditions (Rimet & Bouchez, 2012). Diatoms respond predictively to environmental stress, making them powerful bioindicators for water quality assessment (Kelly & Whitton,1995; Wood et al., 2019). Over time, numerous syntheses have emerged like Trophic Diatom Index (TDI), Specific Pollution Sensitivity Index (IPS), and SPEAR-herbicide Index, among others (Kelly & Whitton, 1995; Wood et al., 2019). Additions like the Duero Diatom Index (DDI) and metrics based on ecological guilds and life form strategies allow more nuanced assessments (Álvarez-Blanco et al., 2013; Rimet and Bouchez, 2012). DiaThor integrates many of these within RStudio (Fig. 2), offering a flexible and reproducible analysis platform.

Fig. 2 R Studio's latest version launched.

During the post monsoon season, samples from 5 urban and 5 rural stream sides were collected. Epipelic diatoms were identified to species level via microscopy. These data were analysed in RStudio (v42) using DiaThor, calculating matrices like TDI, IPS, ecological guilds, and life-form traits, with visualizations via ggplot2. DiaThor efficiently generated ecological scores across all sites. Urban streams showed higher TDI values and dominance of motile taxa that were the clear signs of pollution (Nicolosi Gelis et al., 2020). In contrast, rural sites had more diverse communities, dominated by low- and high-profile guilds, indicative of healthier conditions (Tettinen et al., 2015). These patterns aligned well with known environmental gradients.

By combining indices, guilds and life form data DiaThor offers a holistic picture of stream health. It's seamless compatibility with other are packages enhances its use in ecological modelling (Rimet et al., 2019; Keck et al., 2017). Emerging methods using deep learning like DiatomNet and YOLOv5 based classifiers and automatic diatom identification from microscopic images, cuts down analysis time and minimises observer bias. However, species level accuracy remains a challenge and is an area of ongoing research. Simultaneously, DNA metabarcoding is gaining traction for aquatic bio-assessment, including in national programs like France's WFD and services like the Joint Danube study. This approach generates vast genetic datasets, requiring robust bioinformatics pipelines and reference libraries.

Tools like Diat.Barcode handle diatom-specific barcode data, though integration with index calculators like

DiaThor, is still evolving. The future of diatom biomonitoring lies in image-based AI and DNA based data into cohesive scalable systems. As software continues to evolve, high throughput, standardised and accurate ecological monitoring is moving from aspiration to reality.

Traditional index computation was slow and error prone, but open-source tools like DiaThor have streamlined ecological assessment workflows (Nicolosi Gelis et al., 2022).

The vision for a "Next-Generation of Biomonitoring to Detect Global Ecosystem Change" involves leveraging such technological advancements to provide more rapid, sensitive, and comprehensive assessments. Furthermore, software will continue analysis outputs, metabarcoding data, and traditional index calculation functionalities (like to support more complex ecological modelling and data analysis, allowing deeper insights into the factors influencing diatom communities and water quality. The potential for using advanced AI tools like ChatGPT to interact with and extract ecological insights using packages like DiaThor is essential in DNA metabarcoding is even being explored. This molecular technique allows for the identification numerous species from a single sample by sequencing specific genetic markers. Diatom DNA software tools, from initial inventory managers and index calculators to sophisticated deep learning models for automated identification and bioinformatics pipelines for metabarcoding, have fundamentally changed how diatoms are used to monitor pollution. Software like DiaThor provides essential capabilities for calculating and interpreting ecological indices, standardising the assessment process for morphological data. Combined with advancements in automated image analysis and DNA metabarcoding facilitated by dedicated software and bioinformatics, these tools are making diatom-based biomonitoring more efficient, objective, and capable of addressing complex environmental challenges. The future promises further integration of these technologies, leading to more comprehensive and powerful systems for assessing and protecting aquatic ecosystems (Fig. 2). Future platforms are likely to integrate automated image those in DiaThor into seamless, comprehensive systems.

Table 1 Evolution of Diatom based water quality monitoring leading to development of DiaThor package.

Evolution of Diatom-Based Water Quality Monitoring Leading to DiaThor			
Year	Milestone/ Event	Details & Impact	Citation
1961	Introduction of diatom-based biotic index	Developed a formula for calculating diatom indices based on species abundance, indicator values, and pollution sensitivity, laying the foundation for modern diatom bioassessment.	(Zelinka and Marvan, 1961)
1982	Development of biological methods for water quality assessment	Contributed to quantitative approaches using diatom assemblages, leading to indices like the Indice Biologique Diatomées (IBD) used in France.	(Coste, 1982)
1991–1993	Development of OMNIDIA software	Introduced proprietary software widely used for calculating 6 diatom indices globally, becoming a standard tool in laboratories for	(Lecointe et al., 1993)

		diatom analysis.	
1995	Development of the Trophic Diatom Index (TDI)	Published in the Journal of Applied Phycology; TDI assesses eutrophication levels in UK rivers, influencing the implementation of the Water Framework Directive (WFD) in Europe.	(Kelly and Whitton, 1995)
1999–2005	Implementation of the EU Water Framework Directive (WFD)	Mandated the use of biological elements, including diatoms, for ecological status classification, triggering a boom in index development and standardisation.	(European Union, 2000). Directive 2000/60/EC of the European Parliament and of the Council establishing a framework for Community action in the field of water policy.
2012	Classification of diatom ecological traits	Introduced classifications linking diatom morphology and function to environmental conditions, later integrated into trait-based assessments.	(Rimet and Bouchez, 2012)
2013	Design of the Duero Diatom Index (DDI)	Developed for rivers in northwestern Spain, integrating local species responses and environmental gradients.	(Álvarez-Blanco et al., 2013)
2015	Study on diatom communities in boreal urban streams	Explored how land use and urbanisation affect diatom assemblages, reinforcing the interaction between land use and diatom communities in bioassessment.	(Teittinen et al., 2015)
2017	Emphasis on digital tools in biomonitoring	Published "Freshwater biomonitoring in the information age," emphasising big data, automation, and open-source tools like R.	(Keck et al., 2017)
2019	Introduction of the SPEAR-Herbicides Index	Proposed a diatom-based tool to detect herbicide impacts on benthic communities, expanding diatom bioindication beyond nutrients and saprobity.	(Wood et al., 2019)

2019	Publication of the Diat.barcode database	Released an open-access DNA barcode and trait database for diatoms, central to many tools including DiaThor, and regularly updated.	(Rimet et al., 2019)
2020	Study on ecological guilds in Argentine urban streams	Investigated nuclear alterations and guilds in urban streams, providing real-world data for testing DiaThor.	(Nicolosi Gelis et al., 2020)
2021	Development of the DiaThor R package	Created an open-source R package designed to calculate over 15 indices (e.g., TDI, IBD, IPS, SPEAR), highly modular and user-extensible.	(Nicolosi Gelis et al., 2021)
2022	Publication of DiaThor's functionality	Published in Ecological Modelling, introducing 19 core functions, trait-based analysis, guilds, and an integrated GUI via Shiny.	(Nicolosi Gelis et al., 2022)
2023–2024	Ongoing updates to DiaThor	Linked with the diatbarcode package for auto-updating species traits, supporting integration with machine learning pipelines.	(Pu et al., 2023; UDE DIATOMS in the Wild 2024)
2025 (anticipated)	Research on self-supervised diatom identification	Publication on using AI for image-based diatom classification, potentially integrating into DiaThor's digital identification workflows.	(Bohan et al., 2020; Keck, 2019/2025)

2.2 Installation of RStudio

To begin installing R version 4.5.0 on a Windows system, open your preferred web browser and use Google to search for “Download R for Windows.” From the search results, select the official CRAN (Comprehensive R Archive Network) link. Once on the CRAN website, navigate to the section for Windows downloads and choose the latest version of R, which in this case is 4.5.0 (Fig. 3). Click the appropriate link to download the installer, then run the file and follow the on-screen instructions to complete the installation process (Nicolosi Gelis & Sathicq, 2020).

After successfully installing R (Fig. 3), the next step is to set up RStudio, a user-friendly integrated development environment (IDE) for R. Go back to Google and search for “Download RStudio.” Visit the official website at *https://posit.co/download/rstudio-desktop/* , and download the free desktop version available for Windows. Install RStudio by running the downloaded file and following the installation prompts. Once installed, open RStudio; you’ll see a workspace with several panes, including a Console, Environment, and Script Editor.

2.3 Setting up DiaThor

To install and run DiaThor, an R package used for computing diatom metrics and ecological indices, focus on the Console pane within RStudio. Type the command *install.packages("diathor")* and press Enter. After the installation is complete, activate the package by typing library*(diathor)* and pressing Enter again. This will load DiaThor and allow you to begin using its functions for diatom-based water quality assessment. If installed correctly, DiaThor will also appear in the "Packages" tab with a checkmark beside it, indicating it's ready to use (Fig. 4).

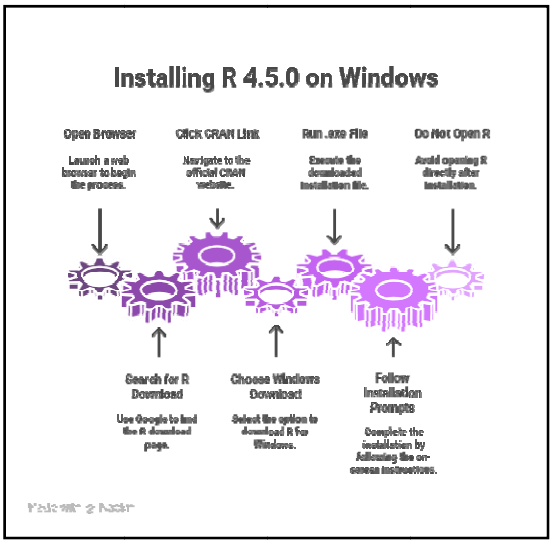

Fig. 3 Steps to install R 4.5.0 on Windows.

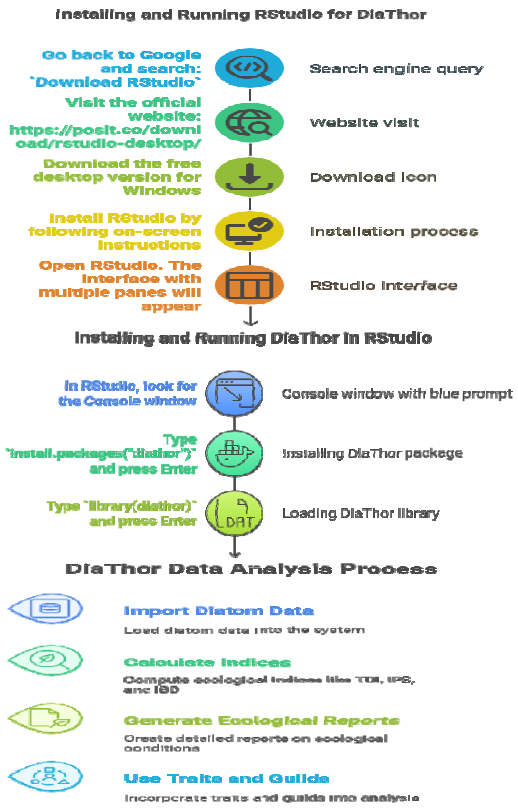

Fig. 4 Complete Steps of installing RStudio and DiaThor package.

2.4 Input data format

The primary input for DiaThor is a table of diatom species abundance or relative abundance per sampling site. Each row typically represents a diatom taxon (identified at the genus or species level), and each column corresponds to a sample site. The first column must contain species names, which the software uses to reference ecological trait data (Nicolosi Gelis & Sathicq, 2020). You can also use the function *diat_loadData()* to structure the data in a format compatible with DiaThor. This function organizes the dataset and can convert raw counts into relative abundances.

2.5 Getting started in RStudio

In the top-left corner, there is the code-editing pane (with syntax highlighting). Moving clockwise from there will take you to the environment pane (in which you can see the different objects that are loaded into the session), which is the viewing pane containing various options such as Files, Plots, Build, Help, and finally, at the bottom left, the Console. In the middle, there is one of the most useful features of RStudio, the ability to view data frames. This view can be created by clicking a data frame in the Environment panel at the top right. This function also enables sorting and filtering by column. All of the packages are installed using install.packages(""), and calling library () loads a subset of the packages. The R command, *result <- diaThorAll (resultsPath = "")* is used to execute all available diatom metrics provided by the DiaThor package. In this empty string ("") for the resultsPath argument means that the results will be saved to the current working directory by default. This can be done by right clicking the folder and choosing "copy as path" option and paste it in between the double inverted commas. This ensures that all generated files, such as CSVs, PDF summaries, and ecological index outputs, are saved in the desired directory without confusion or file path errors. Always remember to use either forward slashes (/) or double backslashes (\\) in your path, when working in Windows to avoid path-related errors. Specifying a clear results path also makes it easier to locate and manage your output files, especially when running multiple analyses or sharing your work with collaborators. For example, if your project is located in a folder named "Diatom Project" on your computer, you should specify an output folder for your results, such as *resultsPath ="C:/Users/YourName/Documents/DiatomProject/Results".* After this a small pop-up window will appear and will ask you to select the csv file of your data to be processed (Nicolosi Gelis & Sathicq, 2020).

You can also use the function *diat_loadData()* to structure the data in a format compatible with DiaThor. This function organizes the dataset and can convert raw counts into relative abundances if needed (using the isRelAb argument). It also connects to the Diat.barcode database to retrieve morphological and ecological information for each species, such as indicator values needed for index calculations. The update option *(updateDBC = TRUE)* was enabled to ensure the most up-to-date species data were used. (Nicolosi Gelis & Sathicq, 2020; Nicolosi Gelis et al., 2022). It also connects to the Diat.barcode database to retrieve morphological and ecological information for each species, such as indicator values needed for index calculations. The update option *(updateDBC = TRUE)* was enabled to ensure the most up-to-date species data were used. (Nicolosi Gelis & Sathicq, 2020) (Nicolosi Gelis et al., 2022). The primary input for DiaThor is a table of diatom species abundance or relative abundance per sampling site. Each row typically represents a diatom taxon (identified at the genus or species level), and each column corresponds to a sample site. The first column must contain species names, which the software uses to reference ecological trait data (Nicolosi Gelis & Sathicq, 2020). Begin by importing a matrix of diatom species abundance into R. This data frame should contain a column listing species names and separate columns for each sampling site with the corresponding species counts. If raw counts are being used, initial preprocessing such as, filtering out extremely rare species or aligning species names with accepted synonyms, may be required for consistency.

Calculate Ecological Information and Diatom Based Indices

Documentation for package 'diathor' version 0.1.5

- DESCRIPTION file.

Help Pages

cemfgs_rb	CEMFGS_RB
dbc_offline	DBC (offline)
ddi	DDI
des	DES
diaThor	DiaThor: A package to calculate multiple diatom-based biotic indices
diathor	DiaThor: A package to calculate multiple diatom-based biotic indices
diaThorAll	Runs all the DiaThor functions in a pipeline
diat_cemfgs_rb	Calculate the combined classification of ecological guilds and size classes for diatoms
diat_checkName	Searches all the taxa database for the input name
diat_ddi	Calculates the Duero Diatom Index (DDI)
diat_des	Calculates the Descy Index (DES)
diat_disp	Calculates the Diatom Index for Soda Pans (DISP)
diat_diversity	Calculate diversity parameters for diatoms using the vegan package
diat_edi	Calculates the Ecuador Diatom Index (EDI)
diat_epid	Calculates the EPID index (EPID)
diat_getDiatBarcode	Loads the 'Diat.Barcode' database into DiaThor in the correct format
diat_guilds	Calculate ecological guilds for diatoms
diat_idap	Calculates the Indice Diatomique Artois-Picardie (IDAP)
diat_idch	Calculates the Swiss Diatom Index (IDCH)
diat_idp	Calculates the Pampean Diatom Index (IDP)
diat_ilm	Calculates the ILM Index (ILM)
diat_ips	Calculates the Specific Polluosensitivity Index (IPS) index
diat_loadData	Loads the Data into DiaThor in the correct format
diat_lobo	Calculates the Lobo Index (LOBO)
diat_morpho	Calculate morphological parameters for diatoms
diat_pbidw	Calculates the PBIDW Index (PBIDW)
diat_pdise	Calculates the Swedish Phosphorus Diatom Index (PDIse)
diat_sampleData	Sample Data
diat_size	Calculate size classes for diatoms
diat_sla	Calculates the Sladecek Index (SLA)
diat_spear	Calculates the SPEAR(herbicides) Index (SPEAR)
diat_taxaList	Creates a single list with taxa names from all indices within DiaThor
diat_tdi	Calculates the Trophic (TDI) index
diat_vandam	Calculates ecological information for diatoms based on the Van Dam classification
disp	DISP
edi	EDI
epid	EPID
idap	IDAP
idch	ID-CH
idp	IDP
ilm	ILM
ips	IPS
lobo	LOBO
pbidw	PBIDW
pdise	PDISE
sla	SLA
spear	SPEAR(h)
taxaList	taxaList
tdi	TDI

Fig. 5 List of Packages and Functions in RStudio (Nicolosi Gelis & Sathicq, 2020).

To support trait and index calculations, DiaThor uses the open-access Diat.barcode database, a curated collection of diatom specie's ecological information. This allows the software to match each input species to its known environmental tolerances and indicator values. Before running analyses, users can use functions like *diat_checkName()* to verify and standardize species names according to the Diat.barcode reference list.

The master function *diaThorAll(),* is then run to compute all available indices and ecological metrics in a single process. This function automates the step-by-step calculation of diversity metrics (including richness, Shannon index, and evenness), distribution by size classes, guild composition, and all biotic index values for each sampling site. Internally, it uses sub-functions (Fig. 5) like *diat_loadData(), diat_morpho(), diat_size(), diat_diversity(), diat_guilds(), diat_vandam(), diat_loadData(), diat_ips(), diat_tdi(), diat_idp(), diat_des(), diat_epid(), diat_idch(), diat_ilm(), diat_lobo(), diat_sla(), diat_spear(), diat_pbidw(), diat_disp(), diat_idap(), diat_edi(), diat_ddi(), diat_pdise(), diat_cemfgs_rb(), diat_checkName(), diat_getDiatBarcode(), diat_taxaList(),* to generate and compile the full set of outputs. For our dataset, *diaThorAll()* was applied to the diatom abundance data from each site to retrieve the complete range of results (Nicolosi Gelis & Sathicq, 2020; Nicolosi Gelis et al., 2022) (Fig. 5).

2.6 Visual and Outputs

If plotAll = TRUE, the package automatically generates-

- Bar plots of index values across samples
- Guild and trait composition charts
- Boxplots of size classes and chloroplast forms
- Biovolume distribution graphs

All results are exported as:

- CSV tables with numerical index values
- PDF reports with summary plots

3 Results

3.1 Analysing Diatom species abundance data with DiaThor: Results from a case study of a random sample file

The ecological analysis conducted using the DiaThor package (Nicolosi Gelis & Sathicq, 2020) provides a comprehensive understanding of diatom community structure and functional traits across 18 freshwater sampling sites (labelled A1 through I2) as can be seen in plots formed using DiaThor functions (Fig. 6) (The dataset used for ecological index calculation is available in Supplementary Material 1). These samples represent varied ecological conditions, likely ranging from relatively undisturbed to moderately or highly impacted habitats. (Index values for IPS, TDI, ILM, IDP, SPEAR, and others are summarized in Supplementary Material 2). Sites such as E2, D1, and G1 have high IPS Scores (0.15), suggesting relatively clean or moderately impacted environments. A high IPS Score reflects communities dominated by sensitive species. These sites may have better flow regimes, lower nutrient levels, and fewer anthropogenic inputs. Lower IPS values (e.g., F1, I2) signal organic pollution or eutrophic conditions, possibly due to agricultural runoff or urban discharges.

Sites like C2, G2, and I2 show elevated TDI percentage (>70%). This implies eutrophic tendencies, high nutrient availability (especially phosphorus). Likely scenarios include sewage effluents, fertilizer inputs, or stagnant water conditions. I1 and G2 indicates disturbed habitats, likely subject to sedimentation or turbulence. B1, D2 suggests stable benthic communities with structural homogeneity. H2 shows planktonic dominance often found in deep or lentic systems, or where flow has been reduced. A dominance of two chloroplasts per cell across sites is typical for healthy pennate diatoms. Any increase in cells with >3 chloroplasts or altered

shapes (e.g., lobed, H-shaped) could imply adaptation to low light, turbidity, or nutrient stress. Sites like F1, G1 show increased biovolume, possibly due to larger centric diatoms. These may dominate under eutrophic, warmer, and lower-flow conditions. High biomass can indicate nutrient pulses or algal blooms. I2 and H2 shows small sized dominance (Class 1-2), indicating high disturbance, possibly with toxic or thermal stressors. C1 and D2 shows presence of larger size classes (Class 4-5), more typical of stable or oligotrophic systems. β-Mes saprobic to polysaprobic dominance (e.g., E2, G2): Signals moderate to poor water quality. I2, F1 suggest low oxygen availability possibly stagnation or organic overload. C1, D1 reflects oxygen-rich conditions, often in fast-flowing or aerated habitats. Higher diversity (D1, E2) reflects stable, heterogenous environments with niche diversity as indicated by diversity metrics (Shannon, Richness, Evenness). Low evenness (G1, I2) indicates dominance by a few pollution-tolerant taxa, often tied to stress.

D1, E1, C2 show high IPS, low TDI, rich diversity, balanced guilds indicating pristine or moderately impacted sites. G2, I2, F1 show low IPS, high TDI, dominance of motile/planktonic guilds, indicating moderately disturbed sites with organic/nutrient load. B2, H1, H2 shows transitional mixed impact sites possibly showing seasonal or anthropogenic fluctuation. Trait-based classification extended to ecological preferences. Saprobity levels ranged widely, with some sites dominated by β-Mes saprobic and polysaprobic species as clear indicators of moderate to poor water quality.

Oxygen preference classes suggest a spectrum from aerophilic to anoxic-tolerant species, with stressed sites likely showing a shift towards facultative anaerobes. The trophic state indicators, including the TDI (Trophic Diatom Index) and its percentage variants, support these observations, some samples lie in the eutrophic to mesotrophic range, confirming nutrient enrichment, potentially from agricultural runoff or urban influence. Using the CEMFGS system (which categorizes diatoms by cell shape, ecology, motility, formation, guild, and size), your data shows a rich mosaic of community types. Sites like B1, D2, and H1 show diverse combinations of morphological and functional classes, while others display more homogeneity often a sign of environmental filtering or disturbance. The presence of heterotrophic diatoms and low-oxygen-tolerant forms in some samples aligns with degradation trends or habitat alteration. A suite of standardized ecological indices strengthens the ecological diagnosis. IPS (Indice de Polluosensibilité Spécifique) scores, used widely in European bioassessment, range from low to high across the sites, marking a clear gradient of ecological integrity. Sites with higher IPS values (e.g., C2, E1) likely represent healthier systems. Conversely, TDI, IDP, ILM, DES, and EPID indices indicate varying levels of trophic enrichment, organic matter load, and human impact. For instance, high TDI % values in some samples confirm eutrophic tendencies, possibly from wastewater discharge or fertilizer runoff. Standardized indices like DDI (Deuro Diatom Index), EDI, SLM, and SPEAR offer further resolution, enabling cross-sample comparison and ranking of site condition. Together, the results from SAMPLE.csv underscore a complex ecological landscape. Some freshwater habitats are relatively pristine or moderately impacted, while others show clear signs of anthropogenic disturbance. The multi-metric, trait-based approach employed through DiaThor integrates community structure, functional traits, and environmental tolerance, making it a robust framework for diatom-based bioassessment. These insights aid not only in tracking ecosystem health but also in guiding restoration priorities informing management strategies, and supporting long-term ecological monitoring. (Plots of species richness, biovolume, and trait-based indices are shown in Supplementary Material 2). To perform these calculations, DiaThor first requires diatom abundance data in a "tidy" format, a data structure where each variable forms a column, each observation forms a row, and each type of observational unit forms a table. A distinctive feature of DiaThor is its internal linkage to the Diat.Barcode database, an authoritative and curated source of ecological and taxonomic information about diatom species. When you run a function like diat_ips() in R, DiaThor matches input species names with this barcode using fuzzy matching algorithms (i.e., approximate name matching), even if your

spelling differs slightly from the standardized taxonomy. This greatly improves data coverage and reduces user error.

Though DiaThor does not directly generate plots, it outputs results in clean, tabular data frames (R's version of a spreadsheet or matrix), which are easily exported as CSV files or piped into visualization workflows. By doing so, it empowers researchers to create informative and publication-ready graphs that highlight ecological gradients, community shifts, or pollution status. From a practical standpoint, DiaThor also supports batch processing, meaning researchers can run multiple indices across many sites or time points with a single command. This is particularly useful for long-term monitoring projects or regional ecological assessments. n a world increasingly reliant on reproducible research, DiaThor offers transparency and consistency. Because the entire process from data import to index computation is script-based, it can be documented, rerun, and shared with collaborators or reviewers. This fosters scientific rigor and collaboration.

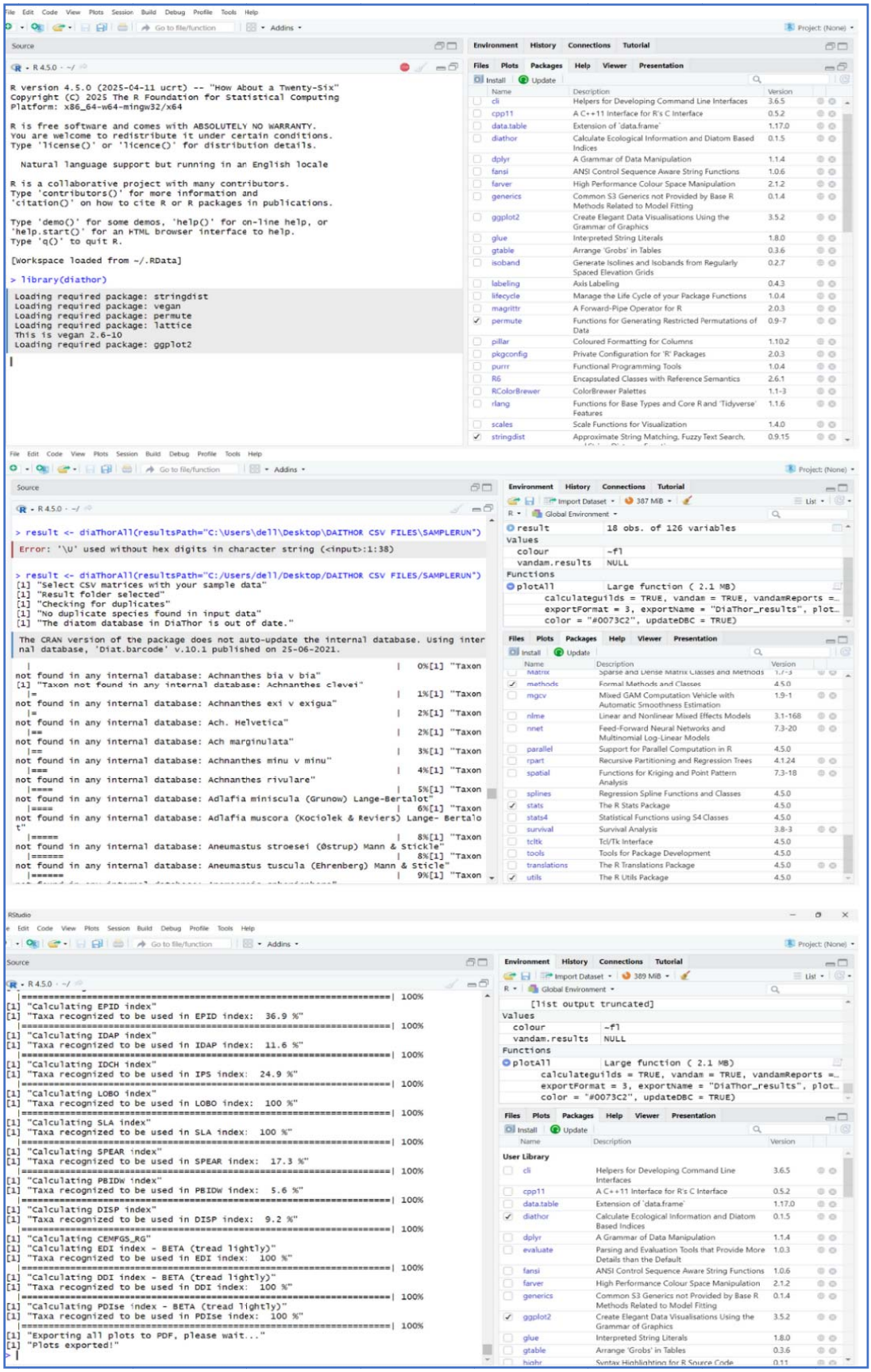

Fig. 6(a) Screenshot attached from the RStudio interface, processing the sample data using diaThorAll() function.

Richness	158	166	168	165	152	154	159	170	157	168	154	160	147	164	168	149	160	154
Shannon.H	4.53218	4.62278	4.53611	4.52248	4.52837	4.49883	4.50337	4.55537	4.50028	4.54095	4.50115	4.51969	4.47925	4.54594	4.52198	4.41357	4.47688	4.42326
Evenness	0.89523	0.9043	0.88527	0.88573	0.90137	0.89316	0.88843	0.88698	0.89004	0.88622	0.89363	0.89055	0.89757	0.89139	0.88252	0.88202	0.88211	0.87816
chloroplasts...0	1.47	0.53	0	1.84	0	1.26	1.16	0.56	0	1.39	0	1.24	0.58	0	2.48	0.48	0.58	1.22
chloroplasts...1	36.75	38.49	31.73	39.45	33.33	35.87	35.64	29.07	40	36.12	41.63	34.33	40.45	35.46	38.15	39.41	37.58	36.6
chloroplasts...2	50.96	53.47	60.44	53.21	58.8	58.51	57.85	63.7	52	53.21	53.2	59.38	50.89	58.29	52.52	55.76	58.41	59.16
chloroplasts...nb	4.9	5.34	5.96	3.21	4.58	1.89	3.5	5.59	5	6.49	1.74	3.74	5.78	2.92	4.96	2.88	1.74	1.22
chloroplasts...2..rarely.4.	1.47	1.07	0.5	1.83	0.65	1.26	0.58	0.56	1.5	1.85	2.89	0.62	0.58	1.46	0.5	0.96	0.58	0.61
chloroplasts...2..or.several.	4.41	1.07	1.49	0.46	2.61	1.26	1.17	0.56	1.5	0.93	0.58	0.62	1.73	1.94	1.49	0.48	1.16	1.22
shape.chloroplasts...0	1.47	0.53	0	1.84	0	1.26	1.16	0.56	0	1.39	0	1.24	0.58	0	2.48	0.48	0.58	1.22
shape.chloroplasts....	1.47	1.07	0.5	0.46	0.65	1.89	1.17	0.56	0.5	0.46	1.73	0.62	0.58	0.97	0.5	0.48	1.16	0.61
shape.chloroplasts...H.shaped	1.96	1.07	0.5	2.29	3.92	1.26	0.58	0.56	2.5	2.78	1.16	0.62	2.31	2.43	0.99	0.48	2.31	3.05
shape.chloroplasts...H.shape	21.07	24.59	23.29	25.23	17	20.76	22.21	19.56	25.5	21.77	23.71	21.23	24.27	21.86	25.27	29.32	21.97	19.52
shape.chloroplasts....1	2.45	2.67	1	3.21	1.95	3.78	3.5	0.56	4	4.63	5.79	1.24	2.31	2.92	1.99	0.96	4.62	4.27
shape.chloroplasts...C.shape	9.31	7.49	5.95	7.34	9.8	7.55	8.18	7.27	7.5	7.87	10.98	10.62	8.67	7.28	6.93	8.17	6.94	7.93
shape.chloroplasts...plate.like	54.39	52.4	60.94	53.21	60.11	58.51	58.44	64.26	53	52.75	52.62	58.12	52.04	58.77	53.51	55.76	58.99	59.16
shape.chloroplasts...discoid	2.94	2.67	4.96	1.83	2.62	1.26	2.92	3.91	4	5.1	1.16	0.62	5.2	1.46	4.46	1.92	0.58	0.61
shape.chloroplasts...flat..divided.into.2.lobes	0.98	1.07	0.5	0.46	1.31	0.63	0	1.12	1	0.46	0	2.5	0.58	0.49	0	0.96	0.58	0
shape.chloroplasts...long	0.98	2.14	0.99	0	0.65	1.26	0.58	0	0.5	0.93	0.58	1.88	0.58	0.97	0.5	0.48	0	0.61
shape.chloroplasts...lobed..small.plate.like	0.98	1.6	0.5	0.92	0.65	0	0.58	0.56	0	0.93	0.58	0.62	0	0.97	0.5	0	0.58	0.61
shape.chloroplasts...lobed..plate	1.96	2.67	0.99	3.21	1.31	1.89	0.58	1.12	1.5	0.46	1.73	0.62	2.89	1.46	2.97	0.96	1.16	1.83
shape.chloroplasts...ribbon	0	0	0	0	0	0	0	0	0	0.46	0	0	0	0.49	0	0	0.58	0.61
Total.Biovolume	190086	172739	107350	126721	128718	210900	159071	135232	110969	139519	184193	136947	125593	139692	128305	117988	147267	140097
Size.class.1	2.45	1.07	2.48	2.29	3.92	2.52	1.16	0.56	2.5	4.17	2.32	0.62	2.89	3.89	1.49	0.48	3.47	3.66
Size.class.2	26.95	29.41	24.28	23.4	20.25	31.45	28.64	26.82	25	23.62	26.02	21.87	23.7	25.25	26.26	24.52	24.28	21.35
Size.class.3	28.91	28.33	30.73	30.72	30.06	25.8	29.81	29.05	30	28.24	27.18	31.23	29.47	31.58	30.22	36.04	28.91	28.06
Size.class.4	19.6	19.24	26.25	22.93	22.88	18.25	19.86	22.36	20.5	20.82	23.13	20.62	23.13	20.41	25.27	19.71	23.14	24.39
Size.class.5	21.56	20.85	15.88	18.82	21.55	22.03	19.27	19.57	21	19.45	21.39	21.22	20.24	16.99	15.87	16.34	18.51	19.52
Size.class.Indet	0.5	1.1	0.4	1.8	1.3	0	1.3	1.6	1	3.7	0	4.4	0.6	1.9	0.9	2.9	1.7	3
Size.Taxa.used	35	36	32	36	35	39	32	37	35	33	37	36	40	33	35	36	38	41
Guild..HP	43.61	51.86	45.09	44.95	36.61	42.14	47.93	46.93	46.5	44.01	42.8	41.23	48.54	41.78	48.06	47.59	41.05	37.21
Guild..LP	14.21	11.23	7.45	10.55	14.37	13.22	11.09	8.95	11.5	12.5	14.45	13.1	12.14	11.65	9.41	11.05	10.99	12.2
Guild..Mot	41.65	35.81	46.08	42.66	47.68	42.8	39.72	42.48	40	40.25	42.79	39.98	38.75	44.2	41.64	37.97	46.27	47.57
Guild..Plank	0	0	1	0	0	1.89	0	0	1	0	0	1.25	0	0.98	0	0.48	0.58	0.61
Guild..Indet	0.53	1.1	0.38	1.84	1.34	0	1.26	1.64	1	3.24	0	4.44	0.57	1.39	0.89	2.91	1.11	2.41
Guilds.Taxa.used	36	35	37	37	34	37	35	33	36	38	35	36	33	40	36	35	39	38
VD.Salinity.1	4.41	1.6	1.49	0.92	2.61	1.26	1.17	1.12	1.5	0.93	1.16	1.24	1.73	2.43	1.49	0.48	1.16	1.22
VD.Salinity.2	82.81	85.02	84.72	88.06	86.25	89.97	85.32	84.38	85	81.93	90.78	87.47	80.94	89.36	85.21	84.12	89.04	89.05
VD.Salinity.3	3.43	2.67	1.5	2.3	3.91	1.89	1.16	2.8	2	1.85	1.74	2.5	2.89	1.95	2.48	2.88	2.32	1.22
VD.Salinity.4	1.96	0.53	3.48	0.92	1.96	3.15	0.58	1.12	2.5	1.39	1.73	0.62	1.16	1.47	1	1.92	1.16	2.44
VD.Salinity.Indet	7.4	10.2	8.8	7.8	5.3	3.7	11.8	10.6	9	13.9	4.6	8.2	13.3	4.8	9.8	10.6	6.3	6.1
VD.Salinity.Taxa.used	31	33	29	31	31	35	29	33	31	30	33	32	35	30	32	33	33	35
VD.N.Het.1	1.96	4.27	1	2.76	0.65	3.78	5.25	2.8	4	3.24	4.64	3.74	2.31	5.35	2.49	1.92	5.2	4.88
VD.N.Het.2	46.06	46	44.6	44.48	50.31	49.08	42.07	46.39	43.5	39.81	45.09	47.48	45.68	42.74	41.12	41.82	43.96	46.35
VD.N.Het.3	11.76	10.15	12.88	11.47	9.8	11.32	8.76	10.05	11	8.33	9.25	9.99	10.41	12.14	10.41	11.05	8.1	9.15
VD.N.Het.4	9.31	9.08	12.38	11.47	11.75	11.96	12.85	12.86	7.5	9.71	10.99	13.13	10.41	13.6	11.4	10.57	13.87	14.03
VD.N.Het.Indet	30.9	30.5	29.1	29.8	27.5	23.9	31.1	27.9	34	38.9	30	25.7	31.2	26.2	34.6	34.6	28.9	25.6
VD.N.Het.Taxa.used	25	26	22	26	25	30	23	25	25	25	28	26	28	24	27	25	25	30
VD.Oxygen.1	0.49	1.6	0	1.84	0	0	1.75	2.8	0.5	0	1.74	2.5	0.58	2.92	1	0.48	0.58	1.22
VD.Oxygen.2	34.79	30.48	30.74	31.64	35.94	35.25	27.45	27.95	35	28.7	29.49	31.21	31.79	31.57	31.22	29.8	34.12	35.37
VD.Oxygen.3	14.21	16.02	14.87	12.4	15	15.74	12.83	14.54	11.5	15.27	19.08	18.12	12.74	15.06	9.44	11.53	13.89	15.25
VD.Oxygen.4	16.66	12.84	18.82	18.34	17.64	16.98	16.96	16.2	15.5	12.03	13.3	16.88	16.76	18.45	18.32	17.78	17.34	17.69
VD.Oxygen.5	0	0	0	0.46	0	0	0	0	0	0.46	0	0	0	0.49	0	0	0	0.61
VD.Oxygen.Indet	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VD.Oxygen.Taxa.used	25	26	23	26	25	30	23	26	25	25	28	26	29	24	27	26	26	30
VD.Saprobity.1	1.96	2.67	1	2.3	0.65	3.15	4.67	2.8	4	3.24	3.48	3.74	0.58	5.35	1.99	1.92	3.47	4.27
VD.Saprobity.2	60.76	60.44	55.49	58.24	62.74	59.14	56.7	59.79	59	51.39	58.39	60.6	59.53	54.87	57.95	58.17	58.4	59.77
VD.Saprobity.3	7.35	7.47	9.92	5.51	7.16	7.56	5.23	5.03	6	8.79	10.41	5.61	4.64	10.21	3.99	4.32	8.11	7.32
VD.Saprobity.4	9.8	8.55	9.91	10.55	8.49	10.69	6.43	8.94	9.5	5.09	6.94	10.63	9.83	8.26	8.92	9.61	6.94	7.93
VD.Saprobity.5	7.35	6.42	8.91	9.63	9.8	7.55	11.11	9.5	6.5	7.4	6.94	9.38	8.09	10.68	9.9	9.13	10.4	10.98
VD.Saprobity.Indet	12.8	14.5	14.8	13.8	11.2	11.9	15.9	13.9	15	24.1	13.8	10	17.3	10.6	17.2	16.9	12.7	9.7
VD.Saprobity.Taxa.used	28	29	26	29	28	33	26	29	28	28	31	29	32	27	30	29	29	33
VD.Aero.1	19.11	12.3	15.87	12.39	15.67	15.74	14.02	15.1	16	14.82	11.57	16.23	13.3	17.49	16.36	15.38	20.83	20.12
VD.Aero.2	9.8	12.29	7.94	9.18	9.79	11.33	9.92	9.51	9	9.72	15.61	11.86	9.83	9.22	6.95	8.65	11.56	11.59
VD.Aero.3	35.77	34.75	40.12	41.28	41.82	39.64	34.47	35.76	36	30.07	33.54	38.75	38.16	40.32	36.17	34.12	32.96	37.82
VD.Aero.4	1.47	1.6	0.5	1.83	1.3	1.26	0.58	1.12	1.5	1.85	2.89	1.87	0.58	1.46	0.5	1.44	0.58	0.61
VD.Aero.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VD.Aero.Indet	33.8	39.1	35.6	35.3	31.4	32	41	38.5	37.5	43.5	36.4	31.3	38.1	31.5	40	40.4	34.1	29.9
VD.Aero.Taxa.used	25	26	23	26	25	30	23	26	25	25	28	26	29	24	27	26	26	30
VD.Trophic.1	0	0.53	0	0.46	0	0	0	0.56	0	0	0.58	0.62	0	0.49	0	0	0	0
VD.Trophic.2	0.49	1.07	0.5	0	0	1.26	0.58	0	1	0.46	0.58	0.62	0	0.97	0	1.44	0.58	0
VD.Trophic.3	14.7	10.7	9.42	11.01	10.45	11.32	12.87	12.29	13	10.19	10.99	12.5	9.82	12.63	11.89	11.06	10.41	10.98
VD.Trophic.4	12.74	13.91	16.35	13.76	17	16.99	14.61	13.97	17.5	12.96	12.72	15.62	12.72	13.6	12.39	12.5	12.72	13.42
VD.Trophic.5	44.1	45.43	41.14	40.37	43.76	43.42	37.37	39.68	37.5	36.58	45.1	41.22	44.52	41.29	39.16	38.92	39.91	40.26
VD.Trophic.6	7.35	6.42	8.91	9.63	9.8	7.55	11.11	9.5	6.5	7.4	6.94	9.38	8.09	10.68	9.9	9.13	10.4	10.98
VD.Trophic.7	10.78	8.02	9.91	12.38	12.41	8.81	8.18	11.74	10.5	12.95	13.3	10.62	8.1	11.66	10.9	11.54	14.46	15.24
VD.Trophic.Indet	9.8	13.9	13.8	12.4	6.6	10.6	15.3	12.3	14	19.5	9.8	9.4	16.8	8.7	15.8	15.4	11.5	9.1
VD.Trophic.Taxa.used	30	31	28	30	29	34	28	31	30	29	32	31	33	29	31	31	31	34
IPS	3.46468	3.47233	3.33587	3.36998	3.3874	3.3947	3.48617	3.48288	3.47276	3.56196	3.31556	3.40462	3.50391	3.39269	3.44354	3.39818	3.42916	3.29898
IPS20	12.7072	12.7435	12.0954	12.2574	12.3402	12.3748	12.8093	12.7937	12.7456	13.1693	11.9989	12.422	12.8936	12.3653	12.6068	12.3913	12.5385	11.9202
TDI20	9.00429	9.71799	8.77885	8.89515	8.49136	8.8287	10.4563	9.35772	8.8526	9.29679	8.95794	8.38504	9.09922	8.7009	9.41146	8.59266	8.90536	8.38606
TDI100	57.8721	54.1159	59.0587	58.4466	60.5718	58.7963	50.2299	56.012	58.6705	56.3327	58.1161	61.1314	57.3725	59.4689	55.7292	60.0386	58.3929	61.126
IDP	2.04839	2	2.20313	2.11364	2.17342	2.06982	2.01969	2.07857	2.01342	1.91026	2.01535	2.14583	2.13105	2.1745	2.11164	2.00536	2.05385	2.13034
IDP20	10.2702	10.5	9.53516	9.96023	9.67624	10.1684	10.4065	10.1268	10.4362	10.9263	10.4271	9.80729	9.87752	9.67114	9.96973	10.4746	10.2442	9.88088
ILM	3.40344	3.46389	3.3131	3.36142	3.36015	3.34425	3.46154	3.48484	3.45139	3.49019	3.34029	3.3937	3.39832	3.36936	3.42509	3.42931	3.36253	3.31901
ILM20	12.4163	12.7035	11.9872	12.2168	12.2107	12.1352	12.6923	12.803	12.6441	12.8284	12.1164	12.3701	12.392	12.2545	12.5192	12.5392	12.222	12.0153
DES	3.89076	3.9195	3.80802	3.83544	3.89007	3.8717	3.87261	3.88252	3.94118	4.00269	3.9316	3.95406	3.81169	3.81452	3.81016	3.92391	3.96285	3.96575
DES20	14.7311	14.8676	14.3381	14.4684	14.7278	14.6406	14.6449	14.692	14.9706	15.2628	14.9251	15.0318	14.3555	14.369	14.3483	14.8886	15.0735	15.0873
EPID	1.7269	1.63788	1.70147	1.64928	1.78373	1.75901	1.53065	1.64208	1.62738	1.63115	1.74473	1.71696	1.62137	1.68676	1.66757	1.62859	1.68479	1.77059
EPID20	11.7972	12.2201	11.918	12.1659	11.5273	11.6447	12.7294	12.2001	12.2699	12.2521	11.7125	11.8445	12.2985	11.9879	12.079	12.2642	11.9973	11.5897
IDAP	3.07561	2.99868	3.01064	3.07802	3.1597	3.01319	3.02458	2.94868	3.08765	3.22849	2.97852	2.91295	3.13667	3.04821	3.03243	3.08412	3.15368	3.06519
IDAP20	10.8591	10.4937	10.5505	10.8706	11.2586	10.5627	10.6167	10.2563	10.9163	11.5853	10.398	10.0865	11.1492	10.729	10.6541	10.8996	11.23	10.8096
IDCH	4.13704	4.14831	4.24096	4.36905	4.54496	4.48303	3.95008	4.28696	4.15395	3.87018	4.3828	4.49353	4.0961	4.48671	4.34799	4.30791	4.42951	4.39421
IDCH20	11.4861	11.4555	11.204	10.8564	10.379	10.547	11.9935	11.0792	11.4402	12.2103	10.8191	10.5186	11.5972	10.5371	10.9135	11.0223	10.6923	10.7881
LOBO	2.68819	2.68462	2.83197	2.83764	2.77411	2.70745	2.79637	2.86885	2.81055	2.71756	2.57317	2.81818	2.76395	2.88961	2.73864	2.76506	2.67784	2.78533
LOBO20	11.6913	11.6687	12.6018	12.6378	12.2354	11.8133	12.3764	12.8354	12.4662	11.8773	10.9629	12.5145	12.1711	12.9669	12.0108	12.1781	11.6257	12.3065
SLA	1.84043	1.82838	1.85139	1.82841	1.86587	1.83083	1.8018	1.83408	1.81379	1.78542	1.86426	1.83538	1.8	1.82778	1.8512	1.84434	1.84179	1.8584
SLA20	11.258	11.3152	11.2059	11.3151	11.1371	11.3036	11.4415	11.2881	11.3845	11.5193	11.1448	11.2819	11.45	11.3181	11.2068	11.2394	11.2515	11.1726
SPEAR	8.01825	9.3338	7.98688	8.8414	6.78039	7.86853	10.0144	8.41474	8.16494	8.28937	8.06071	7.65853	7.08446	8.14231	8.70289	8.96179	7.66824	7.26467
PBIDW	5.533	4.527	5.533	4.527	4.9797	6.22463	4.527	4.527	5.533	5.533	4.9797	4.9797	4.9797	4.9797	4.527	5.533	4.527	4.14975
DISP	2.14697	2.22558	2.25697	2.20927	2.15732	2.21204	1.99957	2.09887	2.22367	2.02979	2.04226	2.29386	2.23853	2.10097	2.15949	2.25509	2.04639	2.10561
HS1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HS2	1.23	1.1	1.53	1.92	0.59	2.13	1.19	0.61	1.38	2.06	1.83	0.37	1.47	1.23	1.94	1.55	1.85	1.05
HS3	3.99	3.45	4.75	5.11	4.31	3.29	2.72	2.77	4.6	3.23	4.04	4.26	3.76	3.84	4.32	5.59	3.88	4.6
HS4	2.45	2.19	2.6	2.08	1.37	0.96	2.89	2.91	2.3	2.36	1.29	1.31	3.43	1.08	2.83	1.55	1.83	1.26
HS5	0.31	0.31	0.15	0.16	0.39	0.19	0	0.31	0.31	0.15	0	0.74	0.16	0.15	0	0.31	0.18	0
LS1	0.61	0.31	0.15	0.8	1.18	0.39	0.17	0.15	0.77	0.88	0.37	0.19	0.65	0.77	0.3	0.16	0.74	1.05
LS2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LS3	0.77	1.1	0.31	1.28	0.39	0.19	1.36	0.46	0.77	0.59	0.55	0.37	0.16	0.15	0.75	0.31	0.18	0.21
LS4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LS5	3.37	2.5	1.99	2.72	3.14	2.9	2.72	2.15	2.45	2.65	4.03	3.34	2.61	2.62	2.24	2.8	2.59	2.93
MS1	0.15	0	0.61	0	0	0.39	0.17	0	0	0.44	0.37	0	0.16	0.46	0.15	0	0.37	0.21
MS2	6.45	4.54	4.44	4.48	5.3	4.64	3.91	3.99	4.75	4.42	4.02	4.64	4.58	4.93	3.88	4.51	4.62	5.65
MS3	2.76	2.51	3.07	3.35	3.53	2.7	3.75	3.39	2.3	2.95	2.38	3.71	2.45	4.15	3.28	3.58	3.51	3.77
MS4	3.98	3.45	3.52	5.11	4.12	2.7	2.72	4	3.36	3.98	4.4	3.16	3.1	3.99	4.62	4.19	4.61	5.44
MS5	2.92	2.36	2.9	3.68	4.33	3.09	3.23	3.38	3.99	4.13	2.75	3.54	2.77	3.38	2.54	3.43	2.57	4.19
PS1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PS2	0.15	0	0	0.32	0	0	0.17	0	0	0.15	0	0.19	0	0	0.3	0	0.18	0.21
PS3	0	0	0.15	0	0	0.39	0	0	0.15	0	0	0	0	0.15	0	0	0.18	0.21
PS4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PS5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CEMFGS.class.Indet	70.9	76.2	73.8	69	71.3	76	75	75.9	72.9	72	74	74.2	74.7	73.1	72.8	72	72.7	69.2
CEMFGS.Taxa.used	34	29	34	34	35	30	30	35	30	33	29	32	29	32	33	31	34	34
EDI	2.49467	2.43713	2.17686	2.27791	2.51568	2.16944	2.28621	2.46814	2.40269	2.22519	2.20102	2.35223	2.30053	2.14603	2.26204	2.26708	2.24664	2.2943
DDI	9.15398	9.19346	9.15728	9.15693	9.12886	9.16449	9.1533	9.15143	9.19465	9.17114	9.21159	9.14542	9.1693	9.13472	9.13987	9.15842	9.15741	9.12621
PDIse	4.36823	4.30919	4.25682	4.3548	4.35814	4.36667	4.27078	4.31806	4.29745	4.3147	4.37741	4.3803	4.22475	4.36523	4.24878	4.3123	4.43891	4.42851

Fig. 6(b) Result of all the calculated index using diaThorAll() function in diaThor (Refer Supplementary Material 2).

S No.	Sample	IPS	TDI	IDP	ILM	DES	EPID	IDAP	IDCH	LOBO	SLA	SPEAR	PBIDW	DISP	EDI	DDI	PDIse
1	A1	76	69	33	45	32	62	23	46	32	59	8	9	19	20	25	28
2	A2	82	70	27	40	28	63	21	43	33	57	10	11	16	19	23	28
3	B1	83	69	34	47	34	66	25	46	33	60	10	9	16	20	21	29
4	B2	81	72	30	40	31	61	21	46	32	59	9	11	17	20	24	31
5	C1	76	65	33	43	35	62	21	45	33	56	9	10	17	22	22	28
6	C2	77	60	29	41	28	59	21	40	30	56	8	8	15	18	22	27
7	D1	77	64	28	40	31	59	20	41	34	57	10	11	14	21	22	28
8	D2	81	71	31	43	35	62	21	43	32	57	10	11	17	23	23	28
9	E1	74	66	28	44	29	59	19	44	30	54	9	9	16	16	21	31
10	E2	83	67	34	43	35	64	22	47	34	61	9	9	17	22	25	29
11	F1	80	70	28	39	29	57	21	40	31	55	9	10	15	21	23	28
12	F2	79	66	30	42	34	63	22	46	32	59	11	10	15	21	23	29
13	G1	72	59	29	36	29	56	19	39	31	53	8	10	13	19	22	24
14	G2	82	70	30	41	33	59	23	47	29	59	11	10	16	20	24	29
15	H1	82	70	32	46	32	64	23	43	38	60	11	11	18	21	21	30
16	H2	77	65	26	42	29	61	19	42	31	56	10	9	16	20	24	29
17	I1	81	67	34	43	35	60	21	47	31	59	8	11	15	20	21	29
18	I2	83	68	31	42	33	63	20	45	37	58	9	12	15	22	22	32

Fig. 6(c) Result exported on using diaThorAll function (Refer Supplementary Material 3).

4 Discussion

4.1 Insights into DiaThor outputs and packages

DiaThor generates outputs as an R list object and also exports summary tables as CSV files. A central output was a table presenting index values for each site across all computed metrics. By enabling *plotAll=TRUE* in the *diaThorAll* function, we generated diagnostic plots foe each metric category, including guild and size class distributions, as well as bar charts displaying index values alongside water quality thresholds.

These visualizations made it easy to spot site-specific trends, highlighting locations with notably high or low ecological scores.

All numerical results were exported to a spreadsheet for deeper inspection and interpretation. While no advanced statistical modelling was applied in this case, DiaThor outputs are fully compatible with broader environmental datasets, allowing seamless integration for analysis like nutrient-index correlations using regression models. The use of scripted R code throughout ensured that the workflow was fully reproducible and easily transferable to other datasets. DiaThor is presented as an open-source tool, with its source code available on GitHub, encouraging collaborative development and the incorporation of new functionalities. This open platform allows researchers to suggest and integrate new statistics and indices. This contrasts with proprietary software like OMNIDIA (Lecointe et al., 1993), which is also widely used for diatom counting, inventory management, and index calculation. OMNIDIA calculates indices based on the formula of Zelinka & Marvan (1961) using species abundance, indicator value, and pollutosensitivity (Supplementary Material 4).

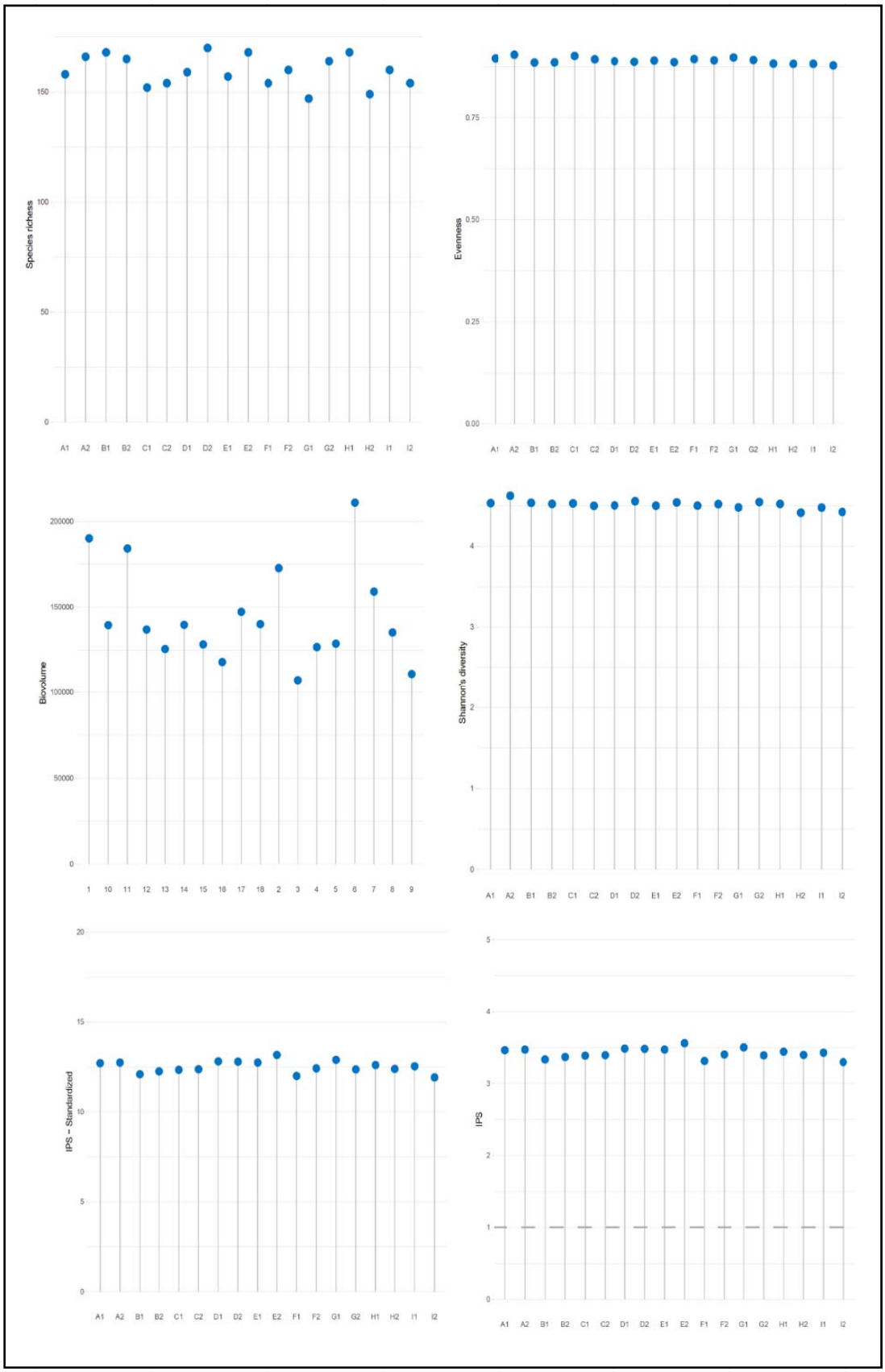

Fig. 6(d) Plots and Graphs generated of various indices using diaThorAll function from random sample file

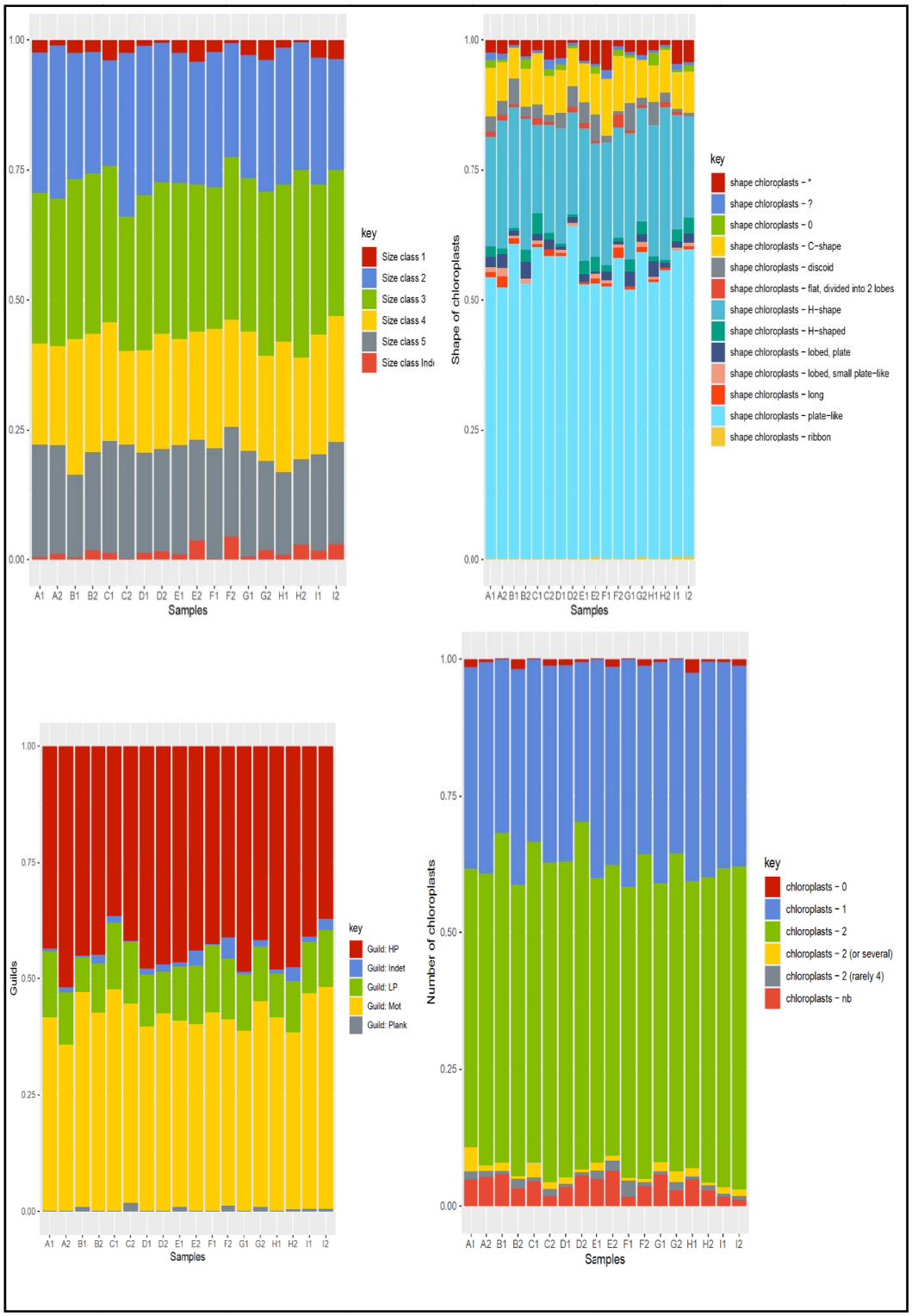

Fig. 6(e) Plots and Graphs of various indices generated using diaThorAll function from random sample file.

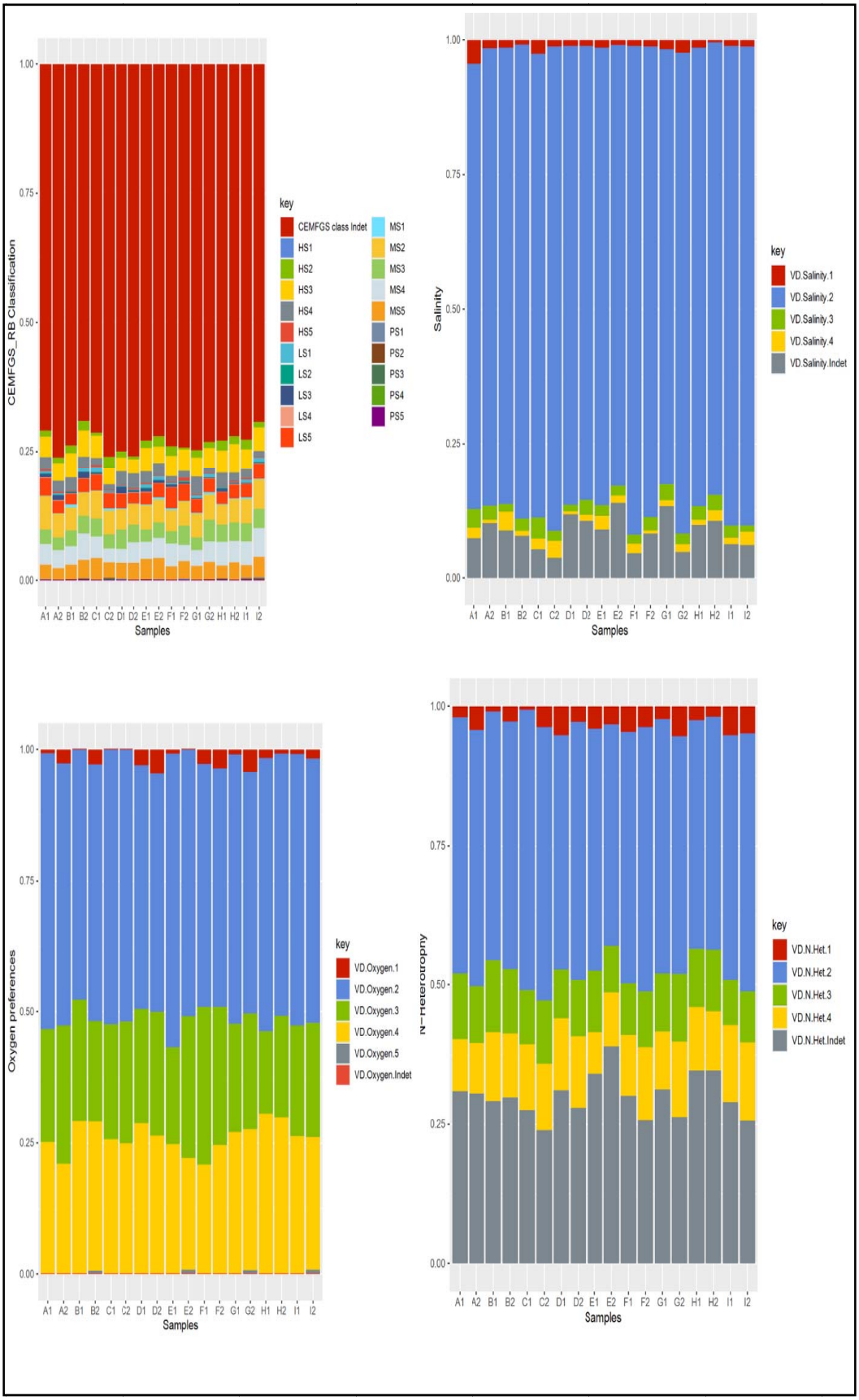

Fig. 6(f) Plots and Graphs of various indices generated using diaThorAll function from random sample file.

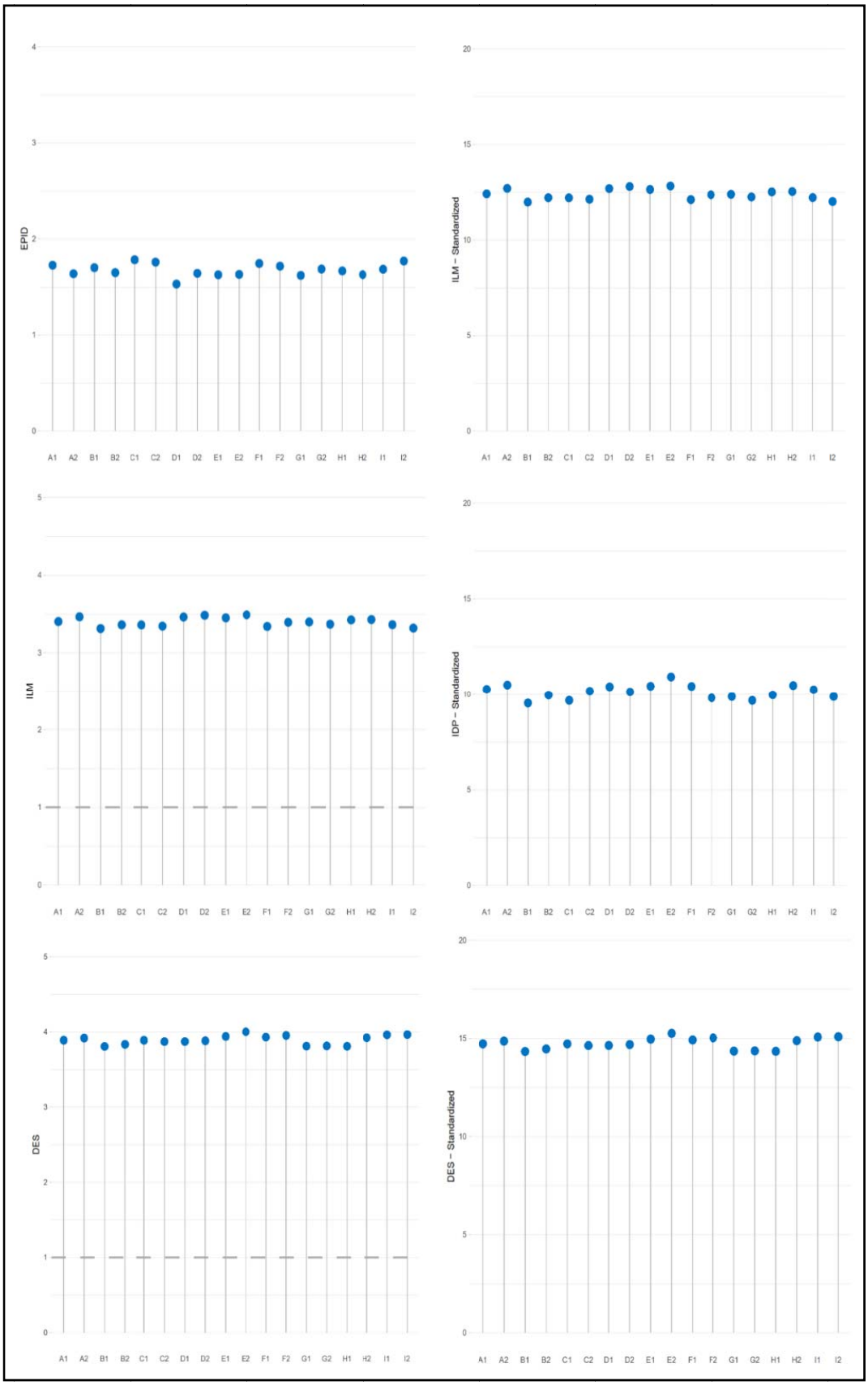

Fig. 6(g) Plots and Graphs of various indices generated using diaThorAll function from random sample file.

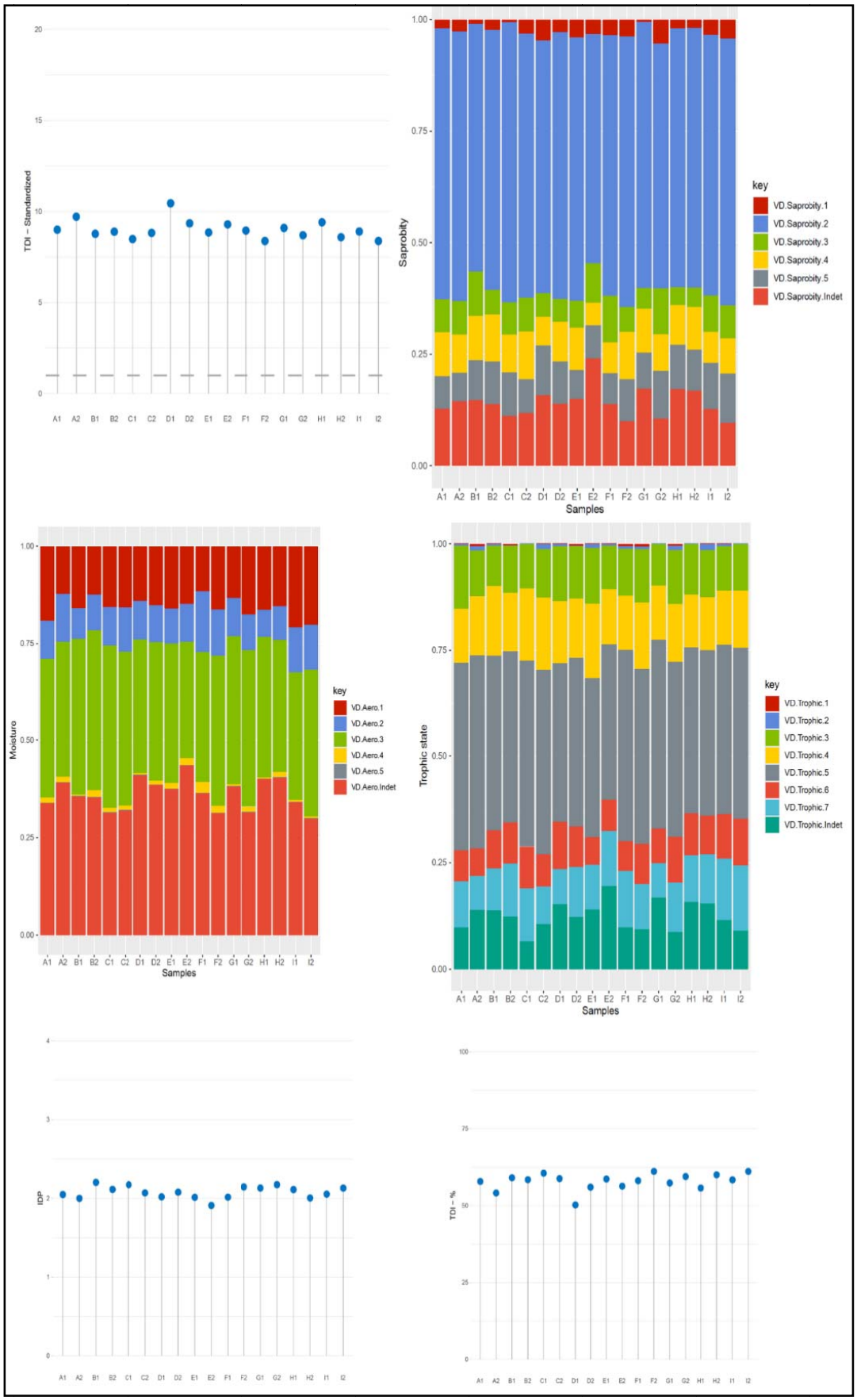

Fig. 6(h) Plots and Graphs of various indices generated using diaThorAll function from random sample file.

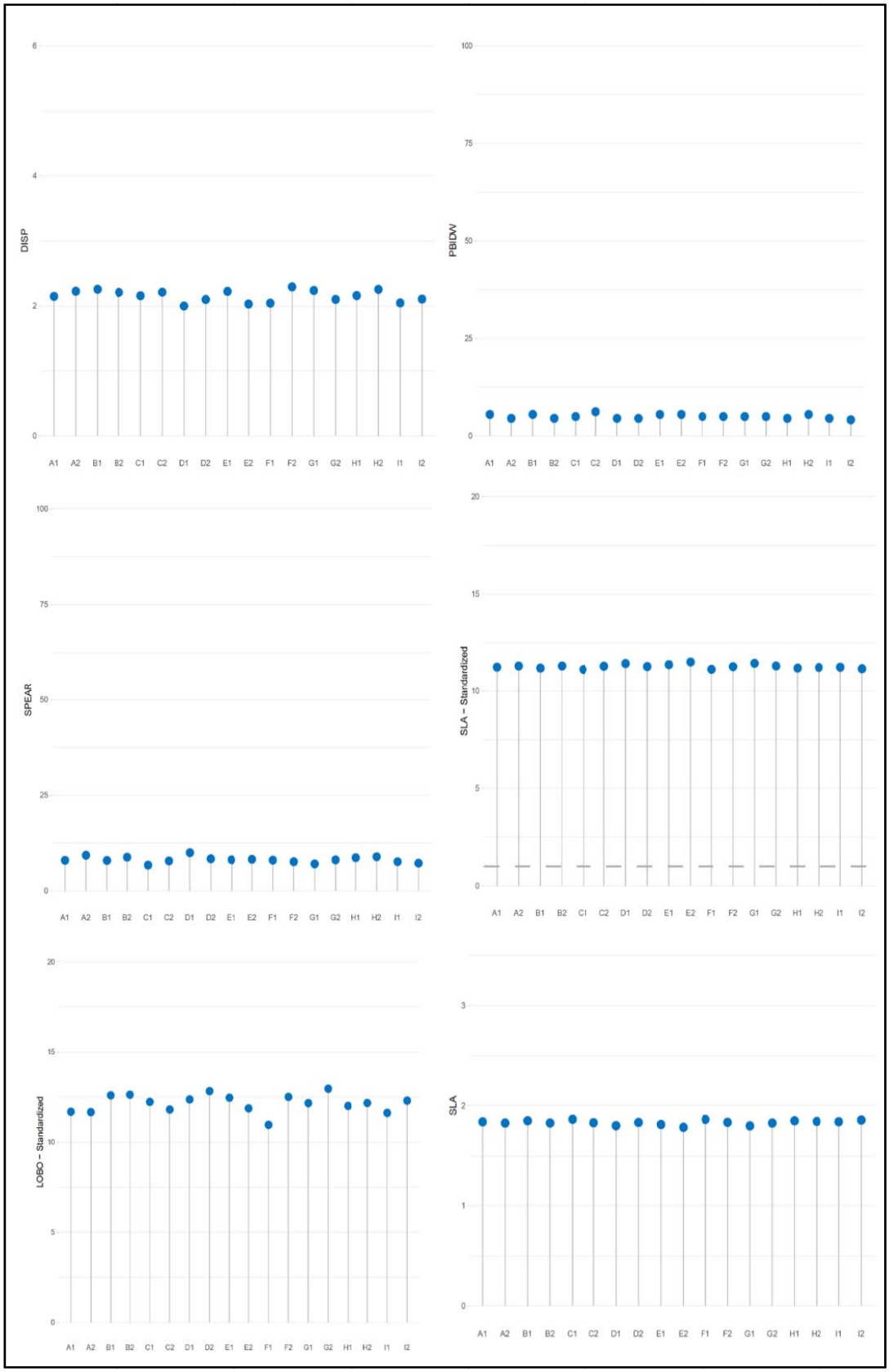

Fig. 6(i) Plots and Graphs of various indices generated using diaThorAll function from random sample.

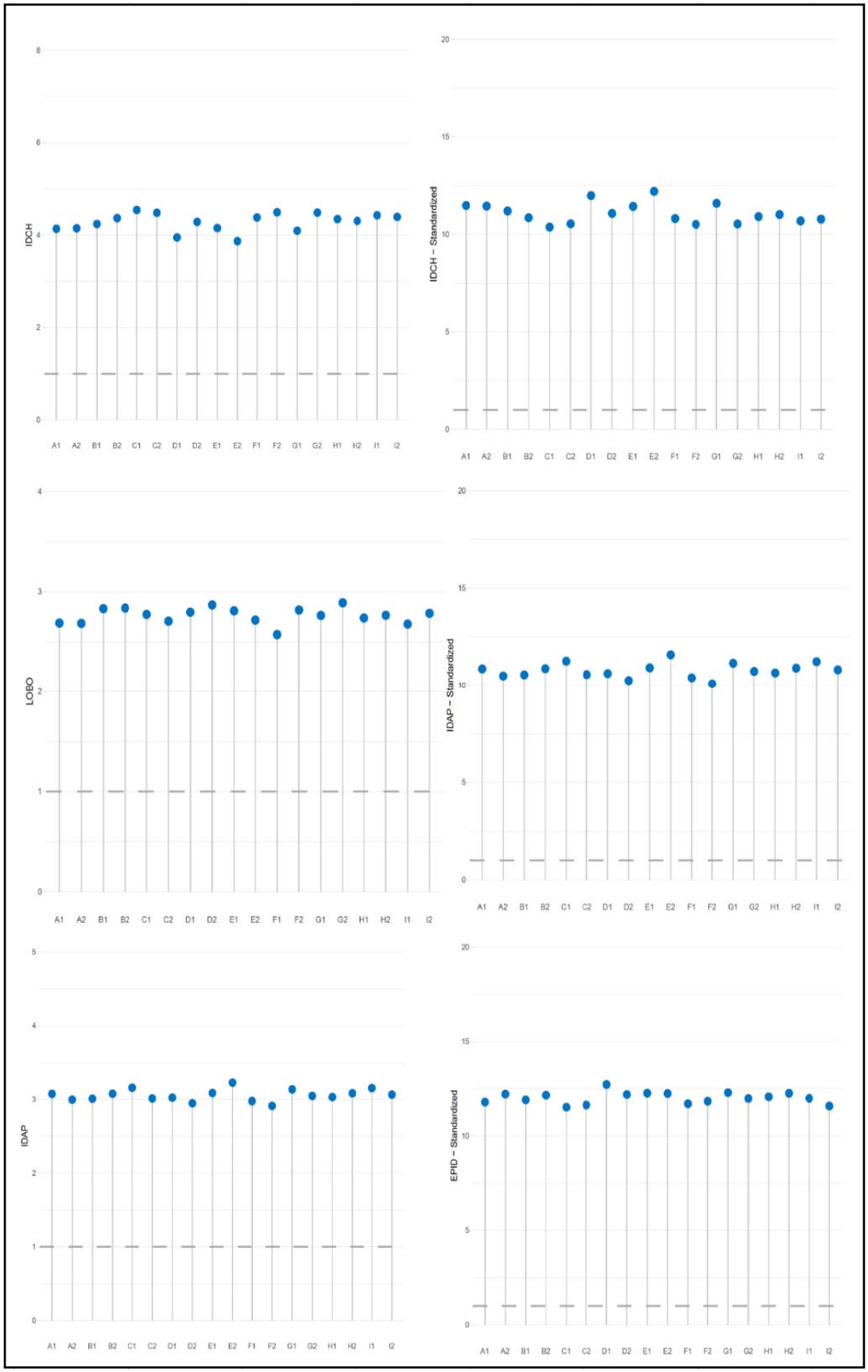

Fig. 6(j) Plots and Graphs of various indices generated using diaThorAll function from random sample file.

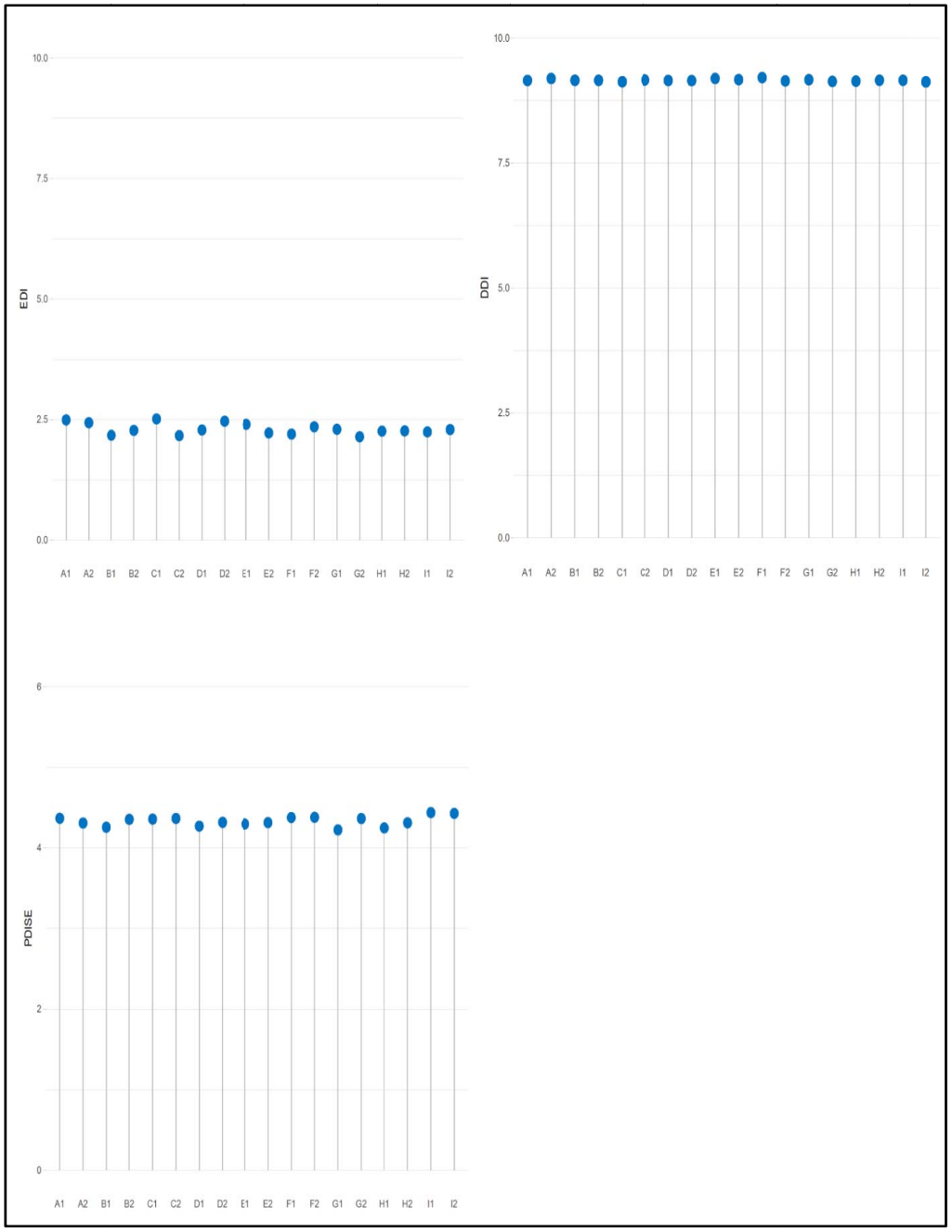

Fig. 6(k) Plots and Graphs of various indices generated using diaThorAll function from random sample file.

In the broader context of diatom monitoring protocols, computational tools facilitate crucial steps beyond metric calculation. This includes data management, such as database design, data entry, and quality control [Protocol generally]. Diatom analysis routinely involves multivariate statistical methods to relate diatom assemblages to environmental variables. Techniques like Principal Components Analysis (PCA), Correspondence Analysis (CA), Detrended Correspondence Analysis (DCA), Canonical Correspondence Analysis (CCA), and Detrended Canonical Correspondence Analysis (DCCA) are employed, requiring multivariate analysis software. Calibration sets, used to statistically predict environmental variables from diatom data, are also analysed with computational tools like the C2 software. Dissimilarity measures, such as the chord distance, are computed to quantify changes in diatom community composition over time or between

samples. Even field aspects utilize computational tools like GPS for precise sampling site location. (Nicolosi Gelis et al., 2022). Thus, DiaThor serves as a powerful, free, and open-source computational tool specifically for calculating a wide range of diatom metrics and indices, streamlining a critical part of the analysis workflow. Its integration into the R environment allows its outputs to be readily used in subsequent, more complex computational analyses common in diatom monitoring protocols, such as multivariate statistics and environmental reconstructions, thereby enhancing the ecological insights derived from diatom data.

We can extract or calculate specific component with individual functions; combine outputs with other environmental datasets; perform statistical modelling, such as regression or ordination; Integrate with packages like ggplot2, vegan, or sf for spatial and ecological analysis.

The tidy structure is essential for seamless integration with popular R packages such as in Table 2.

Table 2 Popular R Packages for plotting graphs.

dplyr:	A package used for data manipulation. It helps filter, summarize, arrange, and group data effectively before analysis.
tidyr:	Complements dplyr by reshaping data from wide to long format (and vice versa), which is often required for plotting or statistical modelling.
ggplot2:	The most widely used R package for creating beautiful and customizable data visualizations. It's based on the "grammar of graphics" and is ideal for making bar plots, boxplots, line charts, and more.
plotly:	An interactive plotting library that extends ggplot2 or works independently to create graphs where users can hover over data points and explore results dynamically.
vegan:	A specialized ecology-focused package for community ecology analyses, including ordination (PCA, NMDS) and diversity metrics.
igraph and networkD3:	Packages for creating network graphs or Sankey diagrams, which help visualize how species, traits, or guilds are functionally related across sites.

For researchers working in aquatic ecology or biomonitoring, DiaThor is an incredibly powerful R package. It processes diatom abundance data to calculate a variety of ecological indices. However, one major challenge emerges for those who are unfamiliar with R programming. DiaThor does not automatically generate plots or visual summaries of the results. While it excels in numerical computation, the responsibility of visualizing and interpreting those numbers still rests heavily on the user and that's where coding becomes essential.

If you lack experience with R, you may quickly find yourself hitting a wall. DiaThor outputs results in tabular form (as .csv files or data frames), but these tables are not immediately digestible for ecological interpretation. You won't see any trend lines, bar charts, or radar plots by default. In order to visualize patterns such as the change in IPS across sites or how trophic levels differ between regions you need to manually write plotting code using tools like ggplot2, plotly, or base R functions. Without this step, the data can feel static and overwhelming.

Another limitation is the modular structure of DiaThor functions. If you're running single indices one at a time (say, just diat_ips() or diat_tdi()), each result gets saved separately in a unique data frame. These aren't merged or compared unless you write additional code to join them. For non-coders, combining multiple outputs into a single comparative visualization becomes tedious or even impossible without outside help. It can feel like trying to piece together a puzzle with no clear picture on the box.

Moreover, DiaThor doesn't come with a graphical interface or dashboard, there's no drag-and-drop GUI, nor do you get intuitive buttons for plotting. Everything operates from the console, which can be intimidating. If you make a mistake, like loading the wrong file or having species names that don't match the internal database, you'll encounter cryptic error messages. R won't explain these errors in plain English. Unless you know how to troubleshoot, you'll spend more time stuck than analysing. From a workflow perspective, not knowing R restricts your automation potential. You'll end up repeatedly copy-pasting, reformatting files in Excel, or asking someone else for help. This not only slows down your research but also increases the chances of inconsistencies or errors in the process.

DiaThor is a highly capable engine, but without coding, you can't unlock its full power, especially for visualization. But, even without deep programming knowledge, a biologist can still gain tremendous value from R and DiaThor, especially with a bit of guidance and the right resources. DiaThor acts as a scientific calculator tailored for diatom based environmental assessment. Once a biologist has collected species abundance data, typically from microscopy counts, it can be formatted into a simple CSV spreadsheet. This is the main input for DiaThor. Even without writing complex code, basic steps like loading data *(diat_loadData())* and running a single index function *(e.g., diat_ips())* are relatively straightforward. Many of these commands can be reused with minimal changes, like filling in your own file path or adjusting species names to match the DiaThor database. For researchers new to R, these initial steps are often manageable with a bit of initial handholding or templated scripts.

Moreover, R and DiaThor support open, reproducible science. Even if a biologist can't code fluently, collaborating with someone who can (a data-savvy student or colleague) enables them to automate repetitive tasks, generate consistent plots, and document their entire analysis pipeline. Once the first round of code is written, the biologist can reuse it again and again on different datasets, with minimal editing. This transforms months of spreadsheet-based work into a few minutes of R code execution.

Still, if coding appears out of reach, there's an emerging solution known as Shinny Apps. These are user friendly dashboards built in R that allow biologists to upload their data, select the indices they want, and receive instant plots, all through point and click interfaces. It's like using R without writing a single line of code. If needed, such a shiny tool can be custom built for a lab or project, removing the last barrier between biologist and data driven insight (Jia et al., 2022).

The graph below (Fig. 7) "Functions Categorized by Ecological Role" provides a visual summary of the various analytical functions embedded within the DiaThor R package, each grouped by its ecological purpose. Along the x-axis, the functions are categorized under ecological themes such as Pollution Indices, Eutrophication, Toxicity, Functional Traits, and Biodiversity, while the y-axis lists the actual function names (e.g., diat_ips, diat_tdi, diat_guilds). Each point represents a unique DiaThor function, and the adjacent italicized text describes what each function measures, its index range, and how the values should be interpreted This layout makes it easy to identify which DiaThor functions are relevant for specific types of environmental assessments, such as nutrient pollution, morphological traits, or soil activity. It helps to quickly understand the role of DiaThor's & its capabilities.

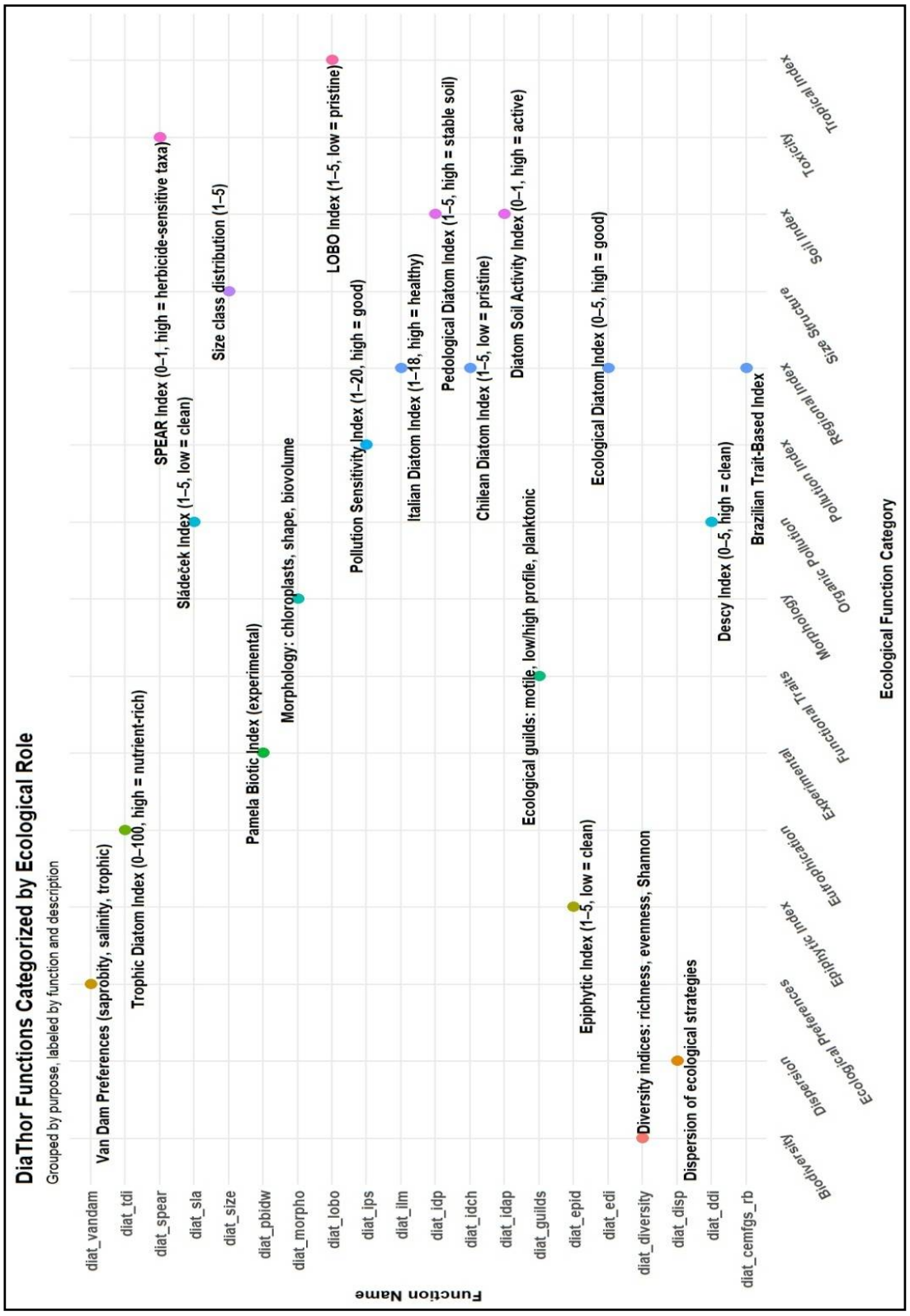

Fig. 7 DiaThor functions categorized by ecological role.

4.2 Interpreting ecological parameters derived from 'diathorAll()' function

a) Pollution and Water Quality Assessment Indices

The Pollution Sensitivity Index (IPS) or diat_ips(), reflects the sensitivity of freshwater diatom communities to organic pollution and general anthropogenic disturbances. With a value range from 1 to 20, higher scores (especially >15) indicate pristine conditions with low organic load, while scores below 9 signify heavy pollution. This index integrates the sensitivity and frequency of diatom taxa to environmental stressors, making it valuable for monitoring riverine health, particularly in temperate regions. Next, the Trophic Diatom Index (TDI) or diat_tdi(), assesses eutrophication through a scale from 0 to 100. It quantifies nutrient enrichment, especially phosphorus, which is a key driver of algal blooms. Water bodies with TDI scores below 30 are oligotrophic (nutrient-poor), those between 30–50 are mesotrophic (moderate), and those exceeding 50 are eutrophic (nutrient-rich), indicating potential risks of ecological imbalance. The Sládeček Index (diat_sla()) and Descy Diatom Index (DDI) (diat_ddi()) are similar in their focus but operate on a 1–5 or 0–5 scale. Both indices estimate organic load and wastewater pollution. Clean systems yield low scores (1–2 or >4), whereas polluted systems push the values toward 4–5 or below 2, respectively. These indices are frequently used in national monitoring programs. The LOBO Index, tailored for tropical systems (especially in Brazil), simplifies the assessment by labelling 1 as clean and 5 as polluted. This makes it user-friendly for field applications in tropical and subtropical climates where species composition differs significantly from temperate zones. The Ecological Diatom Index (EDI) or diat_edi(), is specifically designed for Mediterranean water systems. Ranging from 0 to 5, it categorizes ecological quality into four bands: very good (>4), good (3–4), moderate (2–3), and bad (<2). It's particularly useful in dry, variable-flow rivers that characterize Mediterranean biogeography, helping track the combined effects of drought and nutrient load.

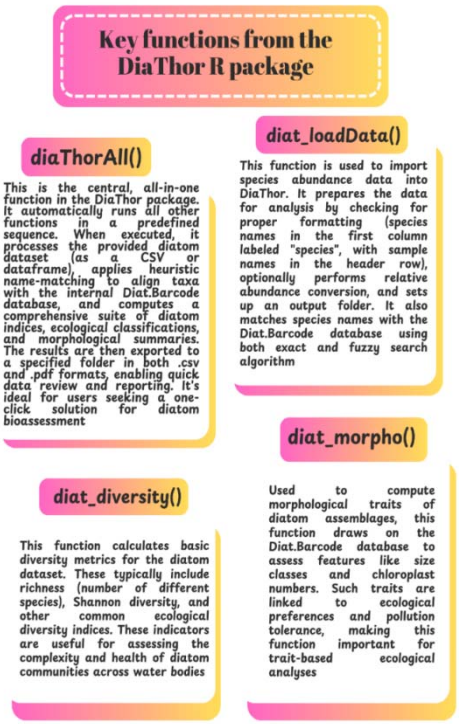

Fig. 8(a) Key Functions in DiaThor.

Likewise, the Italian Diatom Index (ILM) or (diat_ilm()), which ranges from 1 to 18, is a more region-specific adaptation reflecting the sensitivity of diatoms in Italian rivers. It shares a similar interpretative scale to the IPS and allows regionally fine-tuned evaluations of pollution. The Chilean Diatom Index (IDCH) or (diat_idch()) follows the same logic, scoring from 1 to 5 to evaluate South American rivers' water quality, particularly in volcanic and glacier-fed systems. The EPID Index or (diat_epid()), meanwhile, evaluates the health of epiphytic communities (diatoms growing on macrophytes or other surfaces). It is often used in wetlands or shallow, vegetated lakes. Scores of 1 imply a clean, undisturbed state, whereas 5 indicates pollution stress and a disrupted community.

b) Soil quality and land use indices

The Pedological Diatom Index (IDP)or (diat_idp()) gives insight into the degree of soil disturbance. Natural, undisturbed soils show higher values, while degraded, compacted, or urban soils exhibit low scores. This index is particularly valuable in environmental impact assessments, especially in agricultural and construction zones. The Diatom Soil Activity Index (IDAP) or (diat_idap()) is another soil-specific tool with a 0–1 range. Values above 0.7 represent highly active and biologically rich soils, while those below 0.3 indicate biologically poor soils with limited microbial or diatom activity—often a sign of land degradation or heavy human interference.

c) Region-specific and emerging indices

Some indices cater to regional needs or are still under development. The Brazilian Regional Index (CEMFGS_rb) or (diat_cemfgs_rb()) is calibrated to the local flora and environmental stressors in Brazil. Its value ranges is variable and require region-specific interpretation bands that account for endemic and specialized taxa. The Pamela Biotic Index Weighted (PBIDW) or (diat_pbidw()) is an experimental index still under refinement. Though its structure is not fully standardized, it aims to combine weighted ecological traits and sensitivity values for a more nuanced pollution response. This index may evolve as new datasets are incorporated into the DiaThor framework.

d) Trait-Based and Functional Indices

The Van Dam Ecological Preferences (diat_vandam()) index categorizes species based on ecological tolerance to saprobity (1–4), salinity (1–5), and trophic level (1–5). It helps in understanding broader ecological trends, such as salinization or nutrient influx. The Ecological Guilds Index or (diat_guilds()) evaluates community structure by calculating the relative abundance of guilds—motile, high-profile, low-profile, and planktonic forms. A dominance of motile or planktonic taxa often indicates disturbance, while a balanced presence of high- and low-profile taxa suggests ecosystem stability. The Morphological Groups Index or (diat_morpho()) examines diatom shape, size, and mobility traits. For instance, a high percentage of motile forms may indicate sediment instability or nutrient stress. Conversely, an abundance of small, adnate species is usually tied to stable, low-nutrient environments. The Size Classes Index or (diat_size()) measures the percent composition of diatom taxa across size gradients. Larger diatoms, often with lower reproduction rates, tend to disappear in disturbed or polluted environments, so a shift toward smaller taxa may signal ecological stress. High richness typically suggests a diverse, stable ecosystem; low richness may indicate disturbance or pollution. Biovolume reflects not just presence, but ecological biomass. A few large-celled taxa might dominate biovolume despite low numerical abundance. Shannon’s index accounts for both specie’s richness and evenness. Higher values mean more evenly distributed taxa, indicating a more stable, diverse community. It’s key in evaluating resilience and functional diversity. Diatom cell counts per unit volume (e.g., cells/mL or cells/cm²). Sudden spikes may indicate blooms; low values might suggest stress or recent disturbance. Often used alongside chlorophyll-a, for productivity assessments.

Diatom chloroplast shape (e.g., H-shaped, lobed, ribbon-shaped) relates to taxonomy and sometimes light preferences. Functional trait diversity here can indicate community adaptation to light regime or turbidity.

Fewer chloroplasts may indicate adaptations to low light, while many may suggest high metabolic activity. It's a subtle but informative trait. Shifts in salinity guilds signal hydroclimatic or land-use changes. Categorizes taxa into freshwater to brackish preference groups. Very useful in coastal, estuarine, or saline intrusion studies. Oxygen preference plot shows oxygen-rich, variable, or anoxic conditions. Strongly linked to organic pollution, eutrophication, and decomposition. A dominance of heterotrophs may indicate low light, high organic load, or high DOC (dissolved organic carbon) environments. Reflects the mix of autotrophic (photosynthetic) vs heterotrophic (non-photosynthetic) diatoms (Fig. 8).

Fig. 8(b) Key Functions in DiaThor.

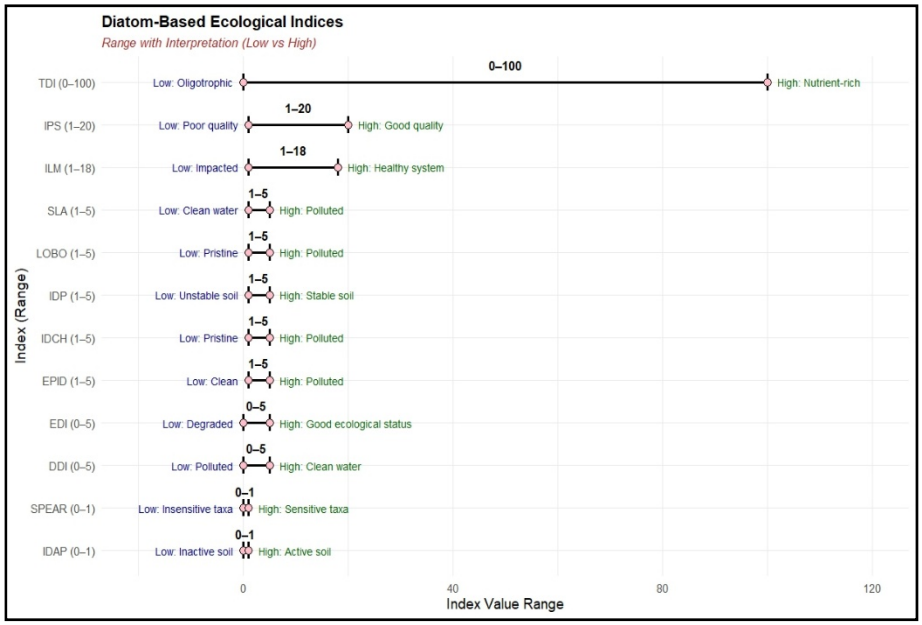

Fig. 9 Diatom indices range.

5 Limitations of using DiaThor and its plotting functions

DiaThor requires exact species names that match its internal or linked databases (e.g. Diat.Barcode). Slight spelling errors, synonyms or outdated nomenclature led to exclusion from analysis. The CSV or Excel file format-sample names in rows, taxon names in columns, or vice-versa otherwise, functions may crash without clear error messages. Missing values, extra spaces, or mismatched column names can result in execution errors or blank outputs without sufficient warnings. Some functions require Despite its strengths, automated work-flows reproducibility, and broad index coverage, DiaThor is not without limitations. Challenge lies in its reliance on an internal taxonomic database. Input taxa must align precisely with, accepted names; deviations or synonyms can result in unmatched entries, skewing index results and trait interpretations.

Additionally, the tool's taxonomic database is largely Eurocentric, leading to under representation of region-specific taxa particularly from tropical and under-studied regions like South Asia or Africa. This can result in incomplete or in accurate assessments in non-European ecosystems. Usability is another concern. The key challenge is package is sensitive to input formatting, which may frustrate users unfamiliar with R. While visual outputs are informative, they are static. and lack interactivity limiting exploration of complex patterns. Furthermore, interpreting multi-index results such as IPS, IDP, or SLAM requires ecological expertise; without proper guidance, users may misapply results. The screenshots of the results have been attached to understand output files that DiaThor generates as a result. DiaThor's output can be effectively visualized for ecological interpretation through a combination of its automated plotting features, and the use of external r packages like **ggplot2, reshape2, pheatmap, and fmsb** for more customised and advanced visualizations. While DiaThor is primarily designed as a back-end engine for calculating biotic indices and metrics rather than for interactive visuals, its outputs are fully compatible with broader environmental datasets and can be integrated with other R packages for detailed analysis and plotting. relative abundances; if absolute counts are input without normalization, the outputs may be misleading.

6 Conclusion

Although literature on DiaThor remains limited, this study demonstrates its utility in computing a wide range of diatom-based indices and ecological traits. By applying the tool to both real and simulated data, shifts in community composition whether natural or stress-induced are reflected in index patterns. High scores and diverse, sensitive taxa indicated healthy conditions; in contrast, low scores and dominance by tolerant species signalled ecological degradation. These findings reaffirm the role of diatoms as sensitive indicators of freshwater health and highlight how computational tools can enhance bioassessment. As an open-source R package, DiaThor reduces barriers to diatom analysis making advanced ecological assessments accessible to researchers and water managers particularly in resource- limited regions. This paper aims to introduce DiaThor's features, its user-friendly interface which makes it an essential tool for researchers and environmental managers. It facilitates the diatom analysis into actionable insights where diatom species are identified by their species names through a heuristic search and has master function diaThorAll, and other 17 functions like diat_tdi, diat_ips, diat_morpho, etc. It contributes to water quality assessment based on diatom assemblages and provides researcher with an open platform to suggest new statistics and functionalities that can be integrated into future builds and hence, considered as a game changer in computing diatom metrics for ecological monitoring and water quality assessment. Its modular design encourages continuous improvement of new indices, regional calibrations, and trait expansions can be incorporated, keeping the platform current and globally relevant.

Looking ahead, DiaThor has the potential to evolve when to a more adaptive and user-friendly tool. Future development might include taxonomic harmonization using live links to Diat.barcode or Algalbase, machine learning to predict missing traits, special mapping integration via packages like leaflet or sf, a graphical user interface (GUI) or Shiny app for broader accessibility, expansion of the internal trait database to better represent tax are from across the globe, and a community driven depository of case studies and feedback to enhance user learning and tool refinement.

In summary, DiaThor exemplifies the synergy between ecological expertise and open-source computation. It transforms raw biological data into actionable insights enabling accurate reproducible assessments of riverine health. As global water quality challenges intensify, tools like DiaThor will play a pivotal role in safeguarding aquatic ecosystems through science-based monitoring and management. The advancement of ecological modelling depends not only on biological insight but on the ingenuity of data scientists and engineers who translate complex ecosystems into code; build tools like DiaThor, and empower us to monitor nature with clarity, consistency, and scale, that is transforming diatom data into code & algorithms to build dynamic models, scaling conservation, with fusion of ecology, data engineering and machine- learning, because safeguarding water isn't just research, it's a global responsibility (Refer Supplementary Material 6 for all the plots generated) (Refer Supplementary Material 7 for file generated for Van Dam classification) (Refer Supplementary Material 8 for file generated listing included/excluded taxa) (Refer Supplementary Material 9 for pictures/screenshots attached).

Reference

Acosta-Arreola J, Domínguez-Hüttinger E, Aguirre P, González N, Meave JA. 2023. Predicting dynamic trajectories of a protected plant community under contrasting conservation regimes: insights from data-based modelling. Ecological Modelling, 484: 110449. https://doi.org/10.1016/j.ecolmodel.2023.110449

Beebe NHF, Rm E. A complete bibliography of publications in ecological modelling (2020–2029).

Blanco S, Álvarez-Blanco I, Cejudo-Figueiras C, Bécares E. 2012. The Duero Diatom Index (DDI) for river water quality assessment in NW Spain: design and validation. Environmental Monitoring and Assessment, 185: 969-981. https://doi.org/10.1007/S10661-012-2607-Z

Blanco S, Álvarez-Blanco I, Cejudo-Figueiras C, Recio JM, Borja C, Bécares E, Díaz F, Cámara R. 2013. The diatom flora in temporary ponds of Doñana National Park (southwest Spain): five new taxa. Nordic Journal of Botany, 31: 489-499. https://doi.org/10.1111/J.1756-1051.2013.01691.X

Bohan DA, Gravel D, Tamaddoni-Nezhad A, Vacher C, Robin S. 2020. A next-generation of biomonitoring to detect global ecosystem change. Frontiers Media SA. https://doi.org/10.3389/978-2-88966-027-8

Calculating autoecological data (optima and tolerance range) for multiple species with the 'optimos.prime' R package—Sathicq—2020—Austral Ecology—Wiley Online Library. (n.d.). Retrieved May 15, 2025, from https://onlinelibrary.wiley.com/doi/10.1111/aec.12868

Community characteristics analysis of eukaryotic microplankton via ITS gene metabarcoding based on environmental DNA in lower reaches of Qiantang River, China. (n.d.). Retrieved May 14, 2025, from https://www.scirp.org/(S(dt0vxmy1blcwx245otj11h3a))/journal/paperinformation?paperid=108225

Coste M, Ricard M. 1983. "1982": AlgaeBase. (n.d.). Retrieved May 14, 2025, from https://www.algaebase.org/search/bibliography/detail/?biblio_id=59993

Directive 2000/60/EC of the European Parliament and of the Council establishing a framework for Community action in the field of water policy—European Environment Agency. (n.d.). Retrieved May 10, 2025, from https://www.eea.europa.eu/policy-documents/directive-2000-60-ec-of

Diatom eDNA metabarcoding and morphological methods for bioassessment of karstic river. (2022). Science of The Total Environment. https://doi.org/10.1016/j.scitotenv.2022.154536

Gelis MMN, Sathicq MB, Cochero J. 2024. diathor: Calculate ecological information and diatom based indices (version 0.1.5) [computer software]. https://cran.r-project.org/web/packages/diathor/index.html

Jia L, Yao W, Jiang Y, Li Y, Wang Z, Li H, Huang F, Li J, Chen T, Zhang H. 2022. Development of interactive biological web applications with R/Shiny. Briefings in Bioinformatics, 23(1): bbab415. https://doi.org/10.1093/bib/bbab415

Jjallaire. (n.d.). GitHub. Retrieved May 16, 2025, from https://github.com/jjallaire

Kang W, Anslan S, Börner N, Schwarz A, Schmidt R, Künzel S, Rioual P, Echeverría-Galindo P, Vences M, Wang J, Schwalb A. 2021. Diatom metabarcoding and microscopic analyses from sediment samples at Lake Nam Co, Tibet: the effect of sample-size and bioinformatics on the identified communities. Ecological Indicators, 121: 107070. https://doi.org/10.1016/j.ecolind.2020.107070

Kelly MG, Whitton BA. 1995. The Trophic Diatom Index: a new index for monitoring eutrophication in rivers. Journal of Applied Phycology, 7(4): 433-444. https://doi.org/10.1007/BF00003802

Keck F. 2025. Fkeck/diatbarcode [R]. https://github.com/fkeck/diatbarcode (Original work published 2019)

Keck F, Vasselon V, Tapolczai K, Rimet F, Bouchez A. 2017. Freshwater biomonitoring in the information age. Frontiers in Ecology and the Environment, 15(5): 266-274. https://doi.org/10.1002/fee.1490

Lecointe C, Coste M, Prygiel J. 1993. "Omnidia": software for taxonomy, calculation of diatom indices and inventories management. Hydrobiologia, 269(1): 509-513. https://doi.org/10.1007/BF00028048

Liu B, Chen S, Liu H, Guan Y. 2020. Modeling cyanobacteria biomass by surface sediment diatoms in lakes: problems and suggestions. Ecological Modelling, 430: 109056. https://doi.org/10.1016/j.ecolmodel.2020.109056

Maria Mercedes Nicolosi Gelis. (n.d.). ORCID. Retrieved May 15, 2025, from https://orcid.org/0000-0001-6324-7930

María Belén Sathicq. (n.d.). ORCID. Retrieved May 15, 2025, from https://orcid.org/0000-0002-3534-8950

Nicolosi Gelis MM, Cochero J, Donadelli J, Gómez N. 2020. Exploring the use of nuclear alterations, motility and ecological guilds in epipelic diatoms as biomonitoring tools for water quality improvement in urban impacted lowland streams. Ecological Indicators, 110: 105951. https://doi.org/10.1016/j.ecolind.2019.105951

Nicolosi Gelis MM, Sathicq MB. 2020. diathor: Calculate ecological information and diatom based indices (p. 0.1.5) [dataset]. https://doi.org/10.32614/CRAN.package.diathor

Nicolosi Gelis MM, Sathicq MB, Jupke J, Cochero J. 2022. DiaThor: R package for computing diatom metrics and biotic indices. Ecological Modelling, 465: 109859. https://doi.org/10.1016/j.ecolmodel.2021.109859

Posit. (n.d.). Posit. Retrieved May 16, 2025, from https://www.posit.co/

Pu S, Zhang F, Shu Y, Fu W. 2023. Microscopic image recognition of diatoms based on deep learning. Journal of Phycology, 59(6): 1166-1178. https://doi.org/10.1111/jpy.13390

Rimet F. (n.d.). Larras F, Bouchez A, Rimet F & Montuelle B. 2012. Using bioassays and species sensitivity distributions to assess herbicide toxicity towards benthic diatoms. PLoS ONE, 7(8). https://doi.org/10.1371/journal.pone.0044458

Spaulding SA, Potapova MG, Bishop IW, Lee SS, Gasperak TS, Jovanoska E, Furey PC, Edlund MB. 2021. Diatoms.org: supporting taxonomists, connecting communities. Diatom Research, 36(4): 291-304. https://doi.org/10.1080/0269249X.2021.2006790

Tapolczai K, Chonova T, Fidlerová D, Makovinská J, Mora D, Weigand A, Zimmermann J. 2024a. Molecular metrics to monitor ecological status of large rivers: implementation of diatom DNA metabarcoding in the Joint Danube Survey 4. Ecological Indicators, 160: 111883. https://doi.org/10.1016/j.ecolind.2024.111883

Tapolczai K, Chonova T, Fidlerová D, Makovinská J, Mora D, Weigand A, Zimmermann J. 2024b. Molecular metrics to monitor ecological status of large rivers: implementation of diatom DNA metabarcoding in the Joint Danube Survey 4. Ecological Indicators, 160: 111883. https://doi.org/10.1016/j.ecolind.2024.111883

Tree of Science with Scopus: A Shiny Application. (n.d.). Issues in Science and Technology Librarianship. Retrieved May 14, 2025, from https://journals.library.ualberta.ca/istl/index.php/istl/article/view/2698

UDE Diatoms in the Wild 2024. 2024. A new image dataset of freshwater diatoms for training deep learning models. https://doi.org/10.1093/gigascience/giae087

Valentin V, Frédéric R, Isabelle D, Olivier M, Yorick R, Agnès B. 2019. Assessing pollution of aquatic environments with diatoms' DNA metabarcoding: experience and developments from France Water Framework Directive networks. Metabarcoding and Metagenomics, 3: e39646. https://doi.org/10.3897/mbmg.3.39646

Villar P, Casado J, Fernández D, Sánchez P, Tabik S. 2025. DiatomNet: an automatic diatom genus identification system through microscopic images and deep learning. bioRxiv, 2025.02.10.635050. https://doi.org/10.1101/2025.02.10.635050

Wood RJ, Mitrovic SM, Lim RP, Warne MSJ, Dunlop J, Kefford BJ. 2019. Benthic diatoms as indicators of herbicide toxicity in rivers – a new SPEcies At Risk (SPEARherbicides) index. Ecological Indicators, 99: 203-213. https://doi.org/10.1016/j.ecolind.2018.12.035

Yu W, Xiang Q, Hu Y, Du Y, Kang X, Zheng D, Shi H, Xu Q, Li Z, Niu Y, Liu C, Zhao J. 2022. An improved automated diatom detection method based on YOLOv5 framework and its preliminary study for taxonomy recognition in the forensic diatom test. Frontiers in Microbiology, 13. https://doi.org/10.3389/fmicb.2022.963059

Supplementary Material

Supplementary Material No.	File Name	File Format	Description
Supplementary Material 1	RANDOM SAMPLE FILE.csv	CSV	Diatom species abundance sample dataset used for DiaThor analysis.
Supplementary Material 2	DiaThor_results - Results.csv	CSV	Ecological index values calculated for each site using the diaThorAll() function.
Supplementary Material 3	num_taxa.csv	CSV	Sample-wise summary of taxon richness derived from the input file
Supplementary Material 4	Plots.pdf	PDF	Full plots and ecological index visualizations generated using DiaThor
Supplementary Material 5	VanDam Taxa used.csv	CSV	Van Dam classification file
Supplementary Material 6	Taxa included.csv / Taxa excluded.csv	CSV	Lists of included/excluded taxa
Supplementary Material 7	Pictures	Folder	Diagrams made using Canva App / R Studio/ Screenshots taken while performing the functions in RStudio interface in .jpeg format