TaxonTableTools: A comprehensive, platform-independent graphical user interface software to explore and visualise DNA metabarcoding data

Till-Hendrik Macher | Arne J. Beermann | Florian Leese

Aquatic Ecosystem Research, University of Duisburg-Essen, Essen, Germany

Correspondence
Till-Hendrik Macher, Aquatic Ecosystem Research, University of Duisburg-Essen, Universitaetsstr. 5, 45141 Essen, Germany. Email: till-hendrik.macher@uni-due.de

Funding information
German Federal Environmental Agency, Grant/Award Number: FKZ 3719 24 2040; UNIVERSITAET DUISBURG-ESSEN

Abstract
DNA metabarcoding is increasingly used as a tool to assess biodiversity in research and environmental management. Powerful analysis software exists to process raw data. However, the translation of sequence read data into biological information and downstream analyses may be difficult for end users with limited expertise in bioinformatics. Thus, there is a growing need for easy-to-use, graphical user interface (GUI) software to analyse and visualise DNA metabarcoding data. Here, we present TaxonTableTools (TTT), a new platform-independent GUI that aims to fill this gap by providing simple, reproducible analysis and visualisation workflows. At its base, TTT uses a "TaXon table", which is a data format that can be generated easily within TTT from two input files: a read table and a taxonomy table obtained using various published metabarcoding pipelines. TTT analysis and visualisation modules include Venn diagrams to compare taxon overlap among replicates, samples, or analysis methods. TTT analyses and visualises basic statistics, such as read proportion per taxon, as well as more sophisticated visualisations, such as interactive Krona charts for taxonomic data exploration. Various ecological analyses can be produced directly, including alpha or beta diversity estimates, and rarefaction analysis ordination plots. Metabarcoding data can be converted into formats required for traditional, taxonomy-based analyses performed by regulatory bioassessment programs. In addition, TTT is able to produce html-based interactive graphics that can be analysed in any web browser. The software comes with a manual and tutorial, is free and publicly available through GitHub (https://github.com/TillMacher/TaxonTableTools) or the Python package index (https://pypi.org/project/taxonetabletools/).

KEYWORDS
biodiversity, bioinformatics, biomonitoring, data analysis, eDNA, OTUs
DNA metabarcoding is increasingly used to assess biodiversity of marine (Aylagas et al., 2018; Zaiko et al., 2018), freshwater (Bush et al., 2020; Elbrecht et al., 2017) and terrestrial ecosystems (Beng et al., 2016; Porter et al., 2019). It can be applied to a bulk sample containing multiple species (Elbrecht et al., 2017; Elbrecht & Steinke, 2019) or to environmental DNA (eDNA; Deiner et al., 2017; Zinger et al., 2019), and allows for rapid and cost-efficient assessments of taxonomic composition. While many different DNA metabarcoding laboratory protocols have been established (Leese et al., 2018), all DNA metabarcoding approaches involve the amplification of specific genetic markers for each target taxa. Common genetic markers for animals include mitochondrial cytochrome c oxidase subunit I (COI; Leray et al., 2013; Macher et al., 2018) and fragments of the small and large subunits of ribosomal RNA, such as the 12S (Hänfling et al., 2016; Miya et al., 2015) and 16S markers (Clarke et al., 2014; Elbrecht et al., 2016). Established barcodes for plants include the matK gene (CBOL Plant Working Group et al., 2009), the gene encoding the large subunit of ribulose-1,5-bisphosphate carboxylase-oxygenase (rbcL; Hollingsworth, 2011), the internal transcribed spacer (ITS; China Plant BOL Group et al., 2011; Hollingsworth, 2011) and the trnL P6 loop (Fahner et al., 2016), while the ITS is commonly used as a genetic marker in fungi (Blaïlaid et al., 2013; Mello et al., 2011).

PCR amplicons are sequenced using high-throughput sequencing platforms and translated into digital characters. Sequences are bioinformatically processed using several programs, such as JAMP (http://github.com/VascoElbrecht/JAMP), DADA2 (Callahan et al., 2016) and OBITOOLS (Boyer et al., 2016). While these pipelines are mostly command-line based, recent efforts have been made to transition to graphical user interfaces (GUIs), such as q2studio in qiime2 (Bolyen et al., 2019), the web-based applications SLIM (Dufresne et al., 2019) and mBRAVE (http://www.mbrave.net).

Generally, these pipelines follow a similar processing workflow, including quality control, primer trimming, quality trimming and clustering. Sequences are compared with a reference database for taxonomic assignment. The largest publicly usable database for COI data is the Barcode of Life Data Systems (BOLD) database, while ITS sequences are deposited in the UNITE database (Nilsson et al., 2019). GenBank (NCBI) holds the largest repository of sequences from a variety of markers and organisms; however, this database has only limited curation and control. Reference sequences are often downloaded from online databases, for example, using the blast+ software, to aid local taxonomic assignment (Camacho et al., 2009). Furthermore, several tools have been published to conduct automatic taxonomic assignment of sequences against online databases, for example, DADA2, JAMP, Ecotag (OBITOOLS pipeline) or BOLDigger (Buchner & Leese, 2020). The resulting taxonomic assignment tables are the final output of all bioinformatic pipelines, forming the basis for all downstream analyses in biodiversity research or biomonitoring.

Both laboratory protocols and bioinformatics tools have been published for each step in the DNA metabarcoding workflow. Nevertheless, only a few tools have been published to tackle comprehensive downstream analysis and visualisation of metabarcoding results. qiime2 and dada2 both allow for the filtering of tables, and the calculation and visualisation of diversity measurements and ordination analyses. The r package "vegan" (Oksanen et al., 2019) is another widely distributed community ecology package for ordination analyses and diversity measurement. Many of these applications are either command-line based, requiring basic bioinformatics skills, or lack crucial tools for processing metabarcoding data sets. Few web-based tools for the analysis of high-throughput sequencing data have been released. For microbiome studies, the online tool microbiomeanalyst (Dhariwal et al., 2017) is commonly used to analyse amplicon sequence data (in particular, 16S rRNA data) and shotgun metagenomics data in particular for microbial community data. Microbiomeanalyst offers sophisticated statistical analyses and visualisation of microbiome data. However, it lacks options common in DNA and eDNA metabarcoding biodiversity analyses of vertebrates and invertebrates, such as the processing of sample replicates, validating and filtering of taxonomy data, and the implementation of quality assessment analyses. Another web-based application used frequently for the analysis of eDNA metabarcoding data is mBRAVE (http://www.mbrave.net), which allows raw DNA metabarcoding sequencing data to be uploaded and processed. However, mBRAVE is limited to the production and analysis of BINs (barcode index numbers; Ratnasingham & Hebert, 2013), operational taxonomic units that closely correspond to individual species. BINs are extremely useful for analyses of COI data sets; however, the use of BINs for the analysis of other markers remains limited. Also, data privacy policies (e.g., European General Data Protection Regulation; EU 2016/679) of web-based tools can limit data analyses, particularly for federal or industrial data providers. Thus, the option to run software locally on one's own computer, or an especially protected or encrypted computer, is important.

Methodological advancements have allowed the upscaling of DNA metabarcoding analysis to hundreds of samples per week, facilitating comprehensive, large-scale biomonitoring programs. Consequently, the analysis of this growing amount of data and its translation into biological meaningful results can be time consuming, introducing a bottleneck in the analysis workflow and limiting the uptake of these methods by nonexperts and users inexperienced with bioinformatics. However, it is biologists that need to work with and interpret these data.

To meet the need for software that is able to analyse vast amounts of data in a user-friendly way, we developed TaxonTableTools (TTT). TTT was developed as part of the GeDNA project, which assesses the implementation of eDNA metabarcoding in regulatory biomonitoring. The program provides easy-to-use tools for biologists and nonbioinformaticians to analyse and visualise their metabarcoding data quickly and reproducibly via a GUI. It unites common steps in the processing of metabarcoding data with a set of modules used for taxonomic exploration, taxonomic verification tools and ecological analyses, while providing options to facilitate the use of data in regulatory biomonitoring applications.
### IMPLEMENTATION

TaxonTableTools is written in Python and available at GitHub [https://github.com/TillMacher/TaxonTableTools]. Python is currently supported by all three major operating systems: Windows, MacOs and Linux-based distributions (e.g., Ubuntu). Installation of the required Python packages uses the package-management system pip, requiring only minimal user input. To improve user-friendliness, TTT comes with a mouse-driven GUI, which includes a detailed manual and tutorial with a test data set, and allows the user to easily execute the various modules (Figure 1).

A key advantage of TTT is its comprehensive data management structure. New projects are created within a dedicated project folder, and all generated files are stored in the respective project directory, which drastically increases clarity and structure when working with multiple data sets or projects. Upon launch, TTT asks the user to either open an existing project or create a new project folder. This circumvents the explicit naming of output files because newly generated files are named according to their input file and the conducted application.

To improve user-friendliness, TTT exports interactive plots in html format as standard. They are computed using the open-source graphing library, plotly [https://www.plotly.com/]. The graphics allow the user to explore data in much greater detail. By hovering the mouse over scatter points, bar charts or lines, the respective data-points and further metadata are displayed. Additionally, the visibility of samples can be turned on and off in the legend. Another major goal of TTT was to offer a rapid and accessible tool to visualise data for reports and publications. Thus, in addition to production of html files, pdf versions of plots are exported, which retain vector graphics allowing the processing of plots with vector manipulation software. TTT allows the user to personalise plots by selecting themes, such as widely distributed layouts like “seaborn” or “ggplots2”, and the colour of each plot can be individualised by choosing either a pre-defined colour set or custom cascading styles sheet (CSS) colours.

### VERIFICATION DATA SET

For verification of TaxonTableTools, we provide a test data set that was generated by members of our working group in September 2019 at the Rhine-Main-Observatory (RMO), Hesse, Germany. Kick-net samples of benthic invertebrates were taken at five different stream sites at the RMO. Up to 300 samples per site were collected and sorted in the field. Samples were stored in ethanol and dried overnight in petri dishes. Subsequently, samples were transferred into
50 ml Turrax tubes and pulverised with two IKA Ultra-Turrax homogenisers. DNA was extracted using salt precipitation as described in Weiss and Leese (2016). Amplification was performed using a two-step PCR with the BF2/BR2 primer set (see Elbrecht & Leese, 2017 for more details). Amplicons were pooled and sequenced on a MiSeq v2 2 x 250 bp. The retrieved sequences were demultiplexed and processed using the JAMP Pipeline v.0.67. Taxonomic assignment was performed with BOLDigger (Buchner & Leese, 2020; pre-release version). The resulting read and taxonomy table were used as input data for the verification of TaxonTableTools and are available as tutorial data.

4 | INPUT FORMATS AND DATA CONVERSION

4.1 | Input format requirements

TTT requires two input files, a read table and a taxonomy table. Read tables are generally referred to as a data frame, which contains the read abundances for each operational taxonomic unit (OTU) or amplicon sequence variant (ASV) per sample and its respective sequence. Read tables are generated by various published DNA metabarcoding pipelines and wrappers, including jamp, dada2, qiime2 or obitools. The output layout differs between pipelines and must be adjusted to meet the specific layout requirements of TTT. Mostly, this requires only minor adjustments to the header. Taxonomy tables are defined as a data frame that holds taxonomic information for each OTU of the read table. The layout and informational content often differ drastically, as there is currently no consensus on a standard format. Taxonomy tables can, for example, be created using qiime2, slim, blast+ or BOLDigger, or can be compiled manually. The output format from the BOLDigger tools is used as the standard input format in TTT. As a requirement, the same OTUs must be present in both the taxonomy table from BOLDigger and the respective read table. Information on phylum, class, order, family, genus and species level are similarly required, while intermediate classifications are not permitted. Only one hit per OTU is accepted, requiring a prior filtering step, which can be performed with BOLDigger. The range of accepted input formats for read and taxonomy tables will be addressed in future versions of TTT to reduce the need for user-based reformatting prior to import into TTT.

4.2 | Merging read and taxonomy tables

A read table and a taxonomy table are merged creating one main table, referred to as the "TaXon table", which forms the base for all downstream applications in TTT. TaXon tables can also be manually created. Formatting is checked automatically within TTT to prevent downstream errors. The main advantage of the TaXon table format is that it combines all relevant data for DNA metabarcoding-derived taxonomic tables into one file, reducing redundancy and aiding rapid access to, and comparison of, data. For example, read distributions per OTU or sample can be investigated directly, and an OTU’s taxonomy and respective sequence can be checked immediately. This is advantageous for both manual investigations and bioinformatic processing.

5 | TAXON TABLE PROCESSING

5.1 | Replicate processing

Most data sets require some processing prior to analyses. TTT offers various tools to merge, filter and convert TaXon tables. All processing tools are optional and can be used as separate modules. The processing workflow generally starts with the processing of sample replicates. Replicates are generally recommended in DNA metabarcoding analyses to tackle PCR and sequencing stochasticity (e.g., Mata et al., 2019; Weigand & Macher, 2018). No specific replicate design is required; replicates are recognised via the sample names, which are marked with a trailing underscore and user-defined symbol (commonly "_rep1", "_rep2" or "_a", "_b"). The first module allows the filtering of OTUs, excluding OTUs that are not present in all replicates, setting the OTU read counts in these samples to zero. When research is focused on low abundant or rare OTUs, this module is not recommended, as it might lead to the exclusion of real OTUs. Subsequently, replicates can be merged into one representative sample by calculating the sum of reads for each OTU. This will drastically reduce redundancy in the TaXon table and is often useful for downstream analyses that do not require separate replicates.

5.2 | TaXon table filtering

Downstream filtering steps are often required when a user is focusing on specific taxa or a specific set of samples. The taxon-based filtering module allows the user to quickly create new tables containing a subset of taxon groups. Similarly, specific samples can be excluded. This function is particularly important as it also removes OTUs that are not present in all samples. This would not be possible when removing samples from the data set via other means, for example, when using Excel. TTT allows the user to filter TaXon tables by read number, excluding OTUs that fall below a user-defined relative (per sample) or absolute threshold.

5.3 | Adding metadata

Metadata are generally used as criterion to infer differences between samples by assigning specific values according to a sample’s location, environmental factors (such as pH or salinity) or classification (using, for example, the European Water Framework Directive ecological status). Sample-specific metadata can be used to aid downstream analyses. Differences in taxa distributions at a particular site can be
investigated, and statistical analyses, such as principal coordinate analysis (PCoA) or canonical-correlation analysis (CCA), can be performed using metadata to explore sample sets. TTT allows the user to create a separate metadata table that is automatically linked with the respective Taxon table. Metadata categories can be manually assigned as columns, where the rows stand for the respective value of the sample. Currently, only the inclusion of metadata for samples is implemented; however, future versions of TTT will include the analysis of metadata for OTUs.

### 5.4 Conversion to incidence data

With the development of metabarcoding, the use of read abundances as a proxy for specimen counts or an estimate of biomass has been scrutinised. Due to PCR stochasticity, varying primer binding efficiency and sequencing bias, there is often only a weak correlation between read abundances and specimen counts or biomass (Bista et al., 2018; Elbrecht & Leese, 2015; Elbrecht et al., 2017), although this can be improved with adjustments to primer settings (Schenk et al., 2019). Thus, often it is recommended to convert read abundance to incidence data for biodiversity analyses; however, incidence data limit the pool for diversity estimate analysis.

### 5.5 GBIF taxonomy check

Taxonomic synonyms and spelling mistakes can lead to false positive inflation of taxonomic richness within a data set or in the comparison of data sets. Accounting for this bias requires taxonomic

---

**FIGURE 2** Example graphical output produced with TaxonTableTools: Basic statistics provide an overview of a data set by plotting the number of reads, number of OTUs and number of OTUs at a species level for all samples (a). Read proportions can be plotted as a heatmap, where the colour scale represents the proportion of the respective taxon in a sample (b). Sample names on the x-axis have been removed for subplots a and b. Correlations between samples can be investigated by performing PCoA, which is based on Jaccard distances (c). An analysis of similarities (ANOSIM) and p-test is performed automatically. Taxa overlaps of up to three samples can be visualised on the Venn diagrams (d).
expertise and the manual checking of taxonomic assignments. This tedious work can be automatised using the Global Biodiversity Information Facility (GBIF) database (https://www.gbif.org/) application programming interface (API), accessed by TTT. This module automatically checks for the accepted name of a taxonomic group and corrects spelling mistakes. An additional log file is created to track the changes made to the TaXon table.

6 | TAXON TABLE ANALYSIS

6.1 | First insights

To get an overview of a data set, it is helpful to visualise as a plot the number of reads, number of OTUs and OTU assignment at the species level (Figure 2a). This plot allows the user to examine the overall quality of the data set. Generally, negative controls should represent only a fraction of the overall reads. Furthermore, samples that have lower than average read counts, which often translates to fewer OTUs, should be considered for removal from the data set, as they often create outliers in statistical analyses.

6.2 | Analysis of biological or technical replicates

The analysis of biological and/or technical replicates is important in most studies (Mata et al., 2019; Weigand & Macher, 2018). For routine analysis, technical replicates provide valuable information on the quality and validity of a data set (Elbrecht et al., 2017). TTT provides options to (i) calculate the percentage of shared OTUs between replicates of a sample, and (ii) calculate Spearman correlations between replicates (for two replicates of one sample only) for both reads and OTUs. This allows the user to, for example, infer whether the samples are sufficiently homogenised by testing extraction replicates, or whether PCR stochasticity has introduced bias by testing PCR replicates.

6.3 | Read proportions

Read proportions obtained from DNA metabarcoding data must be treated with caution (Elbrecht & Leese, 2015). Generally, it is not recommended to use read proportions as a proxy for biomass or specimen count; however, they may be useful when investigating primer bias or contamination of a data set, and have been shown to be biologically meaningful (Aylagas et al., 2018; Di Muri et al., 2020; Meyer et al., 2020). Read proportions can be illustrated as a bar chart, pie chart or heatmap (Figure 2b). Although bar and pie charts are arguably the most common way to visualise read proportions, OTUs that have been assigned to the same taxon are merged. The user has the option to choose how to display hits that are not identified on the analysis demanded taxonomic level, allowing the user to identify the taxa of low-resolution hits. For initial data exploration, it can be helpful to illustrate lower resolved hits on their best matching taxonomic level. For publication, it is recommended that these hits are merged and labelled as unidentified. Furthermore, the implementation of Krona tools (Ondov et al., 2011) allows the hierarchical illustration of read proportions (Beermann et al., 2021). Data can be explored by zooming through a multilayered, interactive pie chart that can be viewed in any web browser.

6.4 | Taxonomic richness and resolution

Measuring the taxonomic richness of a sample is an essential objective in every biodiversity analysis. In ecology, species richness is defined as the number of species in an ecological community, landscape or region. The most straightforward computation of species richness is to count the number of OTUs or species in a data set. The species richness can either be calculated for the whole data set or for each sample.

However, identification of individual species is often impossible. Many species remain undescribed, and reference sequences are lacking for a vast number of species. Higher taxonomic levels, such as genus or family, can provide information when assessing biodiversity. The overall taxonomic resolution of a data set can be visualised in a bar chart, by plotting the number of OTUs assigned to respective levels as lowermost rank. The taxonomic resolution can be used as an indicator for several potential sources of bias, such as varying binding efficiency of primers or bioinformatic processing bias. These can reduce taxonomic resolution and, consequentially, prevent the assignment of OTUs at a species level. To assess individual taxonomic groups, TTT plots the number of reads, number of OTUs and number of species per taxon. This allows the user to quickly infer which taxa are present in a data set.

6.5 | Rarefaction

Sample-based rarefaction is a commonly applied method to determine whether the number of samples taken was sufficient to capture complete species richness. This method computes the number of species by drawing samples at random without replacement from a set of samples. Replicating the drawing approximately 1,000 times substantially enhances the robustness of rarefaction and allows the calculation of standard deviations. Sample-based rarefaction curves can be calculated in TTT using the Rarefaction Curve tool. However, methodological limitations of DNA metabarcoding must be considered when interpreting rarefaction curves.

Read-based rarefaction is commonly used to examine the sequencing depth of high-throughput sequencing data. This module computes the average number of OTUs by drawing a relative number of reads for each sample, at random, iterating from 5% to 100% of the reads in 5% increments.
6.6 | Comparing samples via Venn diagrams

The taxon composition of up to three TaXon tables can be displayed in a Venn diagram (Figure 2d). Comparisons are performed at each taxonomic level in the TaXon table. To allow for more detailed investigations, the groupings are written to a separate Excel file, while only the number of taxa in each group is displayed on the Venn diagram. Venn diagrams of more than three data sets are not recommended as the plot becomes confusing.

6.7 | Diversity analyses and ordination methods

DNA metabarcoding data is often used for diversity analyses. TTT allows the calculation of alpha and beta diversity, and ordination analyses using tools dependent on the Python package scikit-bio (http://scikit-bio.org/). All diversity analyses require an incidence data TaXon table. The alpha diversity calculation is based on the number of OTUs per sample, which are displayed as a scatter plot. Beta diversity is calculated as Jaccard distances, which are illustrated in a distance matrix. Furthermore, a Jaccard-distance based PCoA can be performed (Figure 2c), while a CCA tool is also implemented. For both ordination analyses, it is possible to plot data against a range of different axes. All ordination analysis plots can be displayed in three formats: (i) a 2D scatter plot, displaying two user-defined axes, (ii) a third axis can be displayed using the 3D scatter plot option, and meshes between scatter points of metadata categories can be added to highlight clustering patterns; and (iii) four axes can be displayed in a plot matrix, where each row and column represents one axis.

6.8 | Taxon list

A taxon list can be created from the TaXon table. This list includes all taxa identified from the input TaXon table and reduces redundancy. Optionally, for each species level hit, a link to the GBIF website is created allowing for quick interrogation of the taxon list and the checking of unfamiliar taxa. Furthermore, statistics can be calculated for each taxon. These include the absolute number of reads per taxon and the relative proportion within the data set, the occupancy across all samples, the number of OTUs identified and the intraspecific distances for taxa with multiple OTUs. The calculation of intraspecific distances is dependent on sequences provided in the TaXon table. When multiple OTUs are assigned to a species level, the maximum, average and minimum genetic distance between OTU is calculated. This helps the user to identify highly diverse or cryptic species, and to identify errors in reference sequences, which may introduce errors into the intraspecific distance calculations. Conversely, low intraspecific distances can result from overclustering, suggesting that OTUs should be reclustered using a different threshold.

6.9 | Occurrence data

DNA metabarcoding-derived taxon lists often include several hundreds or thousands of OTUs and species. This makes it difficult for an individual to verify incorrectly assigned taxa in a data set. TTT offers a module to quickly assess potential erroneous taxonomic assignments by checking the biogeographic occurrence of assigned taxa against references in GBIF via an API. This search can be limited by geography to decrease runtime and prevent the download of unwanted data. Both the absolute number of occurrence references per country and relative proportions are downloaded. This data is used to calculate interactive occurrence plots, allowing the user to quickly identify species novel to a particular region.

6.10 | Conversion to formats compatible with regulatory assessment programs

One additional aim of TTT is to manipulate metabarcoding data sets into a format compatible with tools used by regulatory frameworks, such as the European Water Framework Directive (WFD, Directive 2000/60/EC). Monitoring activities by the WFD, and counterparts in other areas and other ecosystems, provides standardised assessments of ecological quality of waterbodies from their biota. The initial version of TTT allows the user to convert benthic invertebrate metabarcoding lists into a format that can be used by the German WFD analysis tool, with further biological quality elements added in future versions. This online analysis tool (www.gewaesser-bewertung-berechnung.de) assesses the ecological quality of German rivers from uploaded taxa lists generated by monitoring activities. In addition, many supporting metrics (such as feeding types or habitat preferences of macroinvertebrates) are calculated. The analysis tool requires the upload of a species-station table in Excel or ASCII format, with species in rows and stations in columns giving the abundance of the recorded taxa. Each taxon is linked to its specific autecological characteristics by a unique ID. The system standardises the taxonomy to an operational taxon list, defining for each taxon the taxonomic level that can be identified in routine water management. As an alternative to direct upload, the system offers a batch mode, allowing large data sets to be automatically read from water authority databases and the assessment results returned. TTT provides species abundance station tables in a format required by this system, including the taxon ID, which can be directly uploaded and used for river assessment. Beentjes et al., (2018) and Buchner et al., (2019) revealed that only minor deviations in the ecological status class occur between analyses of abundance data for benthic invertebrates. Nevertheless, taxon lists generated by DNA-based methods should be considered as a complementary approach and interpreted with caution, as many projects require the comparison of traditional, morphology-based taxon lists with DNA-based taxon lists when examining ecological status. The existing WFD analysis tools should be used until new metrics have been developed and the two approaches have been intercalibrated.
OUTLOOK AND FUTURE IMPLEMENTATIONS

TaxonTableTools is an evolving software, and further modules and functions will be included in future iterations. Updates to further improve user-friendliness are planned, such as an easier installation process and improved user guidelines on the processing and analysis of DNA metabarcoding data. Furthermore, additional functions to explore patterns between samples through clustering and correlation analysis will be implemented, and further tools for use in regulatory biomonitoring (by, for example, the WFD) will be included. Data conversion for further biological quality elements will be implemented, and the number of supported national quality assessment tools will be increased.

ACKNOWLEDGEMENTS

We thank members of the Leese laboratory, especially Dominik Buchner and Kamil Hupalo, for comments and feedback on the program. We thank Daniel Hering (University of Duisburg-Essen, Aquatic Ecology) for input and helpful contributions to the manuscript. We thank Jan Koschorreck and Jens Arle (German Federal Environment Agency) for their feedback on the manuscript. The software was developed as part of the GeDNA project funded by the German Federal Environmental Agency (FKZ 3719 24 2040).

Open access funding enabled and organized by Projekt DEAL.

AUTHOR CONTRIBUTIONS

T.M. conceived and designed the study and wrote the Python package. A.B. and F.L. provided input to the package and supervised the project. T.-H.M., F.L. and A.B. wrote the paper. All authors read and approved the final manuscript.

DATA AVAILABILITY STATEMENT

Project name: "TaxonTableTools - A comprehensive, platform-independent graphical user interface software to explore and visualise DNA metabarcoding data"; Project homepage: https://github.com/TillMacher/TaxonTableTools; Operating system(s): Platform independent; Programming language: Python 3; Other requirements: Krona tools (https://github.com/marbl/Krona); License: MIT licence; Any data set that supports this study is openly available in GitHub https://github.com/TillMacher/TaxonTableTools/tree/master/_tutorial_files.

ORCID

Till-Hendrik Macher https://orcid.org/0000-0001-6164-9557
Arne J. Beermann https://orcid.org/0000-0003-0403-0322
Florian Leese https://orcid.org/0000-0002-5465-913X

REFERENCES

Aylagas, E., Borja, Á., Muxika, I., & Rodriguez-Ezepeleta, N. (2018). Adapting metabarcoding-based benthic biomonitoring into routine marine ecological status assessment networks. Ecological Indicators, 95, 194–202. https://doi.org/10.1016/j.ecolind.2018.07.044

Beentjes, K. K., Speksnijder, A. G., Schilthuizen, M., Schaub, B. E., & van der Hoorn, B. B. (2018). The influence of macroinvertebrate abundance on the assessment of freshwater quality in The Netherlands. Metabarcoding and Metagenomics, 2, e26744.

Beermann, A. J., Werner, M.-T., Elbrecht, V., Ziska, V. M. A., & Leese, F. (2021). DNA metabarcoding improves the detection of multiple stressor responses of stream invertebrates to increased salinity, fine sediment deposition and reduced flow velocity. Science of the Total Environment, 750, 141969. https://doi.org/10.1016/j.scitotenv.2020.141969

Beng, K. C., Tomlinson, K. W., Shen, X. H., Surget-Groba, Y., Hughes, A. C., Corlett, R. T., & Silik, J. W. F. (2016). The utility of DNA metabarcoding for studying the response of arthropod diversity and composition to land-use change in the tropics. Scientific Reports, 6(1), 24965. https://doi.org/10.1038/srep24965

Bista, I., Carvalho, G. R., Tang, M., Walsh, K., Zhou, X., Hajibabaei, M., Shokralla, S., Seymour, M., Bradley, D., Liu, S., Christmas, M., & Creer, S. (2018). Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples. Molecular Ecology Resources, 18(5), 1020–1034. https://doi.org/10.1111/1755-0998.12888

Blalid, R., Kumar, S., Nilsson, R. H., Abarenkov, K., Kirk, P. M., & Kauserud, H. (2013). ITS1 versus ITS2 as DNA metabarcodes for fungi. Molecular Ecology Resources, 13(2), 218–224. https://doi.org/10.1111/1755-0998.12065

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alam, E. M., Armugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brisлавljen, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodriguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology, 37(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Boyer, F., Mercier, C., Bonin, A., Bras, Y. L., Taberlet, P., & Coissac, E. (2016). obitools: A unix-inspired software package for DNA metabarcoding. Molecular Ecology Resources, 16(1), 176–182. https://doi.org/10.1111/1755-0998.12428

Buchner, D., Beermann, A. J., Laini, A., Rolaffs, P., Vitecek, S., Hering, D., & Leese, F. (2019). Analysis of 13,312 benthic invertebrate samples from German streams reveals minor deviations in ecological status class between abundance and presence/absence data. PLoS One, 14(12), e0226547. https://doi.org/10.1371/journal.pone.0226547

Buchner, D., & Leese, F. (2020). BOLDigger – A Python package to identify and organise sequences with the Barcode of Life Data systems. Metabarcoding and Metagenomics, 4, e53355. https://doi.org/10.3897/mbrmg.4.53355

Bush, A., Monk, W. A., Compson, Z. G., Peters, D. L., Porter, T. M., Shokralla, S., Wright, M. T. G., Hajibabaei, M., & Baird, D. J. (2020). DNA metabarcoding reveals metacommunity dynamics in a threatened boreal wetland wilderness. Proceedings of the National Academy of Sciences, 117(15), 8539–8545. https://doi.org/10.1073/pnas.1918741117

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods, 13(7), 581–583. https://doi.org/10.1038/nmeth.3869

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. BMC Bioinformatics, 10(421), 1–21. https://doi.org/10.1186/1471-2105-10-421

CBOL Plant Working Group, Hollingsworth, P. M., Forrest, L. I., Spouge, J. L., Hajibabaei, M., Ratnasingham, S., van der Bank, M., Chase, M. W., Cowan, R. S., Erickson, D. I., Fazekas, A. J., Graham, S. W., James, K. E., Kim, K.-J., Kress, W. J., Schneider, H., van AlphenStahl, J., Barrett, S. C., van den Berg, C., … Little, D. P. (2009). A DNA barcode for land plants. Proceedings of the National Academy of Sciences
Schenk, J., Geisen, S., Kleinboelting, N., & Traunspurger, W. (2019). Metabarcoding data allow for reliable biomass estimates in the most abundant animals on earth. *Metabarcoding and Metagenomics*, 3, e46704. https://doi.org/10.3897/mbmg.3.46704

Weigand, A. M., & Macher, J.-N. (2018). A DNA metabarcoding protocol for hyporheic freshwater meiofauna: Evaluating highly degenerate COI primers and replication strategy. *Metabarcoding and Metagenomics*, 2, e26869. https://doi.org/10.3897/mbmg.2.26869

Weiss, M., & Leese, F. (2016). Widely distributed and regionally isolated! Drivers of genetic structure in Gammarus fossarum in a human-impacted landscape. *BMC Evolutionary Biology*, 16(1), 153. https://doi.org/10.1186/s12862-016-0723-z

Zaiko, A., Pochon, X., Garcia-Vazquez, E., Olenin, S., & Wood, S. A. (2018). Advantages and limitations of environmental DNA/RNA tools for marine biosecurity: Management and surveillance of non-indigenous species. *Frontiers in Marine Science*, 5, https://doi.org/10.3389/fmars.2018.00322

Zinger, L., Bonin, A., Alsos, I. G., Bálint, M., Bik, H., Boyer, F., Chariton, A. A., Creer, S., Coissac, E., Deagle, B. E., De Barba, M., Dickie, I. A., Dumbrell, A. J., Ficetola, G. F., Fierer, N., Fumagalli, L., Gilbert, M. T. P., Jarman, S., Jumpponen, A., ... Taberlet, P. (2019). DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Molecular Ecology*, 28(8), 1857-1862. https://doi.org/10.1111/mec.15060.

**How to cite this article:** Macher T-H, Beermann AJ, Leese F. TaxonTableTools: A comprehensive, platform-independent graphical user interface software to explore and visualise DNA metabarcoding data. *Mol Ecol Resour*. 2021;21:1705–1714. https://doi.org/10.1111/1755-0998.13358