Data Flows for Metabarcoding-based Monitoring in the Project Automated Multisensor Stations for Monitoring of Biodiversity

Peter Grobe[‡], Birgit Gemeinholzer[§], Ameli Kirse[‡], Ammar Saeed[‡], Sarah J. Bourlat[‡]

‡ LIB - Leibniz Institute for the Analysis of Biodiversity Change, Bonn, Germany
§ Kassel University, Kassel, Germany

Corresponding author: Peter Grobe (p.grobe@leibniz-lib.de)

Abstract

Biodiversity monitoring is an important tool to document rapid ecosystem changes and on-going species loss by collecting high-resolution biodiversity data over long periods of time. Large-scale monitoring provides long-range information on species occurrence, interactions, and migrations. In the German project "Automated Multisensor Stations for Monitoring of BioDiversity" (AMMOD, https://ammod.de/), field stations equipped with different sensors (e.g. acoustic sensors, camera traps, volatile organic compound sensors, pollen and insect traps) are being developed that can be operated largely autonomously. These stations provide continuous data on the occurrence of birds, bats, mammals, insects and plants and associated environmental parameters.

Besides the autonomous operation of the stations, a challenge is to master the different requirements that the wide range of sensors bring with them (cf. Wägele et al. 2022). In addition to the detection of fauna and flora by means of sounds, images, films and smells, samples are also collected in malaise and pollen traps with automatically rotating multi samplers. The mixed samples are processed in the laboratory using metabarcoding methods and the resulting sequence data bioinformatically processed into amplicon sequence variants (ASV, Callahan et al. 2017).

Unlike data from acoustic sensors or camera traps, which can flow and be analyzed almost entirely automatically, metabarcoding requires a labor-intensive intermediate step in the laboratories at the University of Kassel (pollen) and the LIB (insects). In addition, metabarcoding requires good (preferably local) barcode reference databases, like Barcode Of Life Data system (BOLD) and the German Barcode of Life library (GBOL); long-term data storage for raw and processed data; reliable retrieval of ASV sequences; and versioning of supercharged data through dynamic assignment during taxon annotation. Here, we present the pipeline for acquisition, intersection, management, storage and publication of data from metabarcoding derived from bulk insect and pollen samples. The metabarcoding data pipeline is connected to the overall data flow developed in the

AMMOD project (Wägele et al. 2022, Fig. 12) and ultimately delivers processed taxon tables with linked information to the raw sequences published, methods used and other sources of information to the upcoming NFDI4Biodiversity Research Data Commons (RDC, https://www.nfdi4biodiversity.org, Glöckner et al. 2020) using the AMMOD data broker. Raw sequences (i.e., FASTQ files, sampling and laboratory metadata) will be stored and published the in the International Nucleotide Sequence Database Collaboration (GenBank or ENA) using the data flows from the German Federation for Biological Data (GFBio, https://gfbio.org) as proxy.

The three data sources for the AMMOD metabarcoding data pipeline are derived from the multisampler with connected malaise or spore traps plus environmental data loggers. Bioinformatic analysis and generation of ASVs is provided by the users, using their own pipelines. For storage, publication and transmission of results to the NFDI4Biodiversity RDC, an ASV table registry is provided along with the possibility of uploading log files, laboratory protocols and collection tables. In the ASV registry, the uploaded tables are taxonomically annotated through the connected reference databases and stored as versioned datasets. The processed and annotated datasets are transmitted to the AMMOD broker in a standardized format as JSON datasets.

Connection of the components: an identifier, which uniquely identifies the data from the respective components of field stations and laboratories, is the prerequisite for the unique assignment of data parts. This AMMOD ID consists of the project name, the case number, the sensor type, a time stamp with the number of the sample and the bottle or plate number. For example, the ID AMMOD-T1-M-R210511-B1 denotes a sample of the AMMOD project from trap 1 of type Malaise dated May 11, 2021 from the first bottle. All bottles and data entries from the laboratory, log files and subsequent processes are provided with such an ID.

The orchestra of different components was built under the aspect of sustainable software development and compatibility to existing systems such as the Leibniz Institute for the Analysis of Biodiversity Change (LIB) Biobank and the LIB Digital Collection Catalog.

All developed software is published at: https://gitlab.leibniz-lib.de/AMMOD/ and https://gitlab.leibniz-lib.de/GBOL/asv-table-registry

The portal for ASV registration and upload is available on: https://bolgermany.de/metabarcoding

Keywords

ASV sequences, data pipelines, sequence data, data duplication

Presenting author

Peter Grobe

Presented at

TDWG 2022

Conflicts of interest

References

- Callahan BJ, McMurdie PJ, Holmes SP (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. The ISME Journal 11 (12): 2639-2643. https://doi.org/10.1038/ismej.2017.119
- Glöckner FO, Diepenbroek M, Felden J, Güntsch A, Stoye J, Overmann J, Wimmers K, Kostadinov I, Yahyapour R, Müller W, Scholz U, Triebel D, Frenzel M, Gemeinholzer B, Goesmann A, König-Ries B, Bonn A, Seeger B (2020) NFDI4BioDiversity - A Consortium for the National Research Data Infrastructure (NFDI). Zenodo https://doi.org/10.5281/zenodo.3943644
- Wägele JW, Bodesheim P, Bourlat S, Denzler J, Diepenbroek M, Fonseca V, Frommolt K, Geiger M, Gemeinholzer B, Glöckner FO, Haucke T, Kirse A, Kölpin A, Kostadinov I, Kühl H, Kurth F, Lasseck M, Liedke S, Losch F, Müller S, Petrovskaya N, Piotrowski K, Radig B, Scherber C, Schoppmann L, Schulz J, Steinhage V, Tschan G, Vautz W, Velotto D, Weigend M, Wildermann S (2022) Towards a multisensor station for automated biodiversity monitoring. Basic and Applied Ecology 59: 105-138. https://doi.org/10.1016/j.baae.2022.01.003