ECO-ALPSWATER

Innovative Ecological Assessment and Water Management Strategy for the Protection of Ecosystem Services in Alpine Lakes and Rivers

Priority 3: Liveable Alpine Space. SO3.2 - Enhance the protection, the conservation and the ecological connectivity of Alpine Space

Work Package WPT1 - Coordination: PP6 INRAE
Activity A.T1.3 - Deliverable D.T1.3.2.
Version 1.0 - Date April 2019

D.T1.3.2. Validated protocols for smart bioinformatic tools in NGS/metabarcoding

Environmental DNA (eDNA) metabarcoding has recently been promoted as an attractive alternative to traditional methods of bioindication for aquatic ecosystems. One of the major advantages of eDNA metabarcoding for biodiversity monitoring, compared to classical morphological approaches, is its ability to provide methods with rather easy standardization and repeatability. Instead of using morphological features to differentiate individuals/species, which may lead to variability in the identification results depending on the observer's experience and the complexity of the biological group, automatized bioinformatics steps are applied to generate taxonomic lists from DNA reads. Though these bioinformatics steps are not exempt of biases that may affect the final species identification, the potential errors and biases are repeatable and can be more largely and easily detected and corrected.

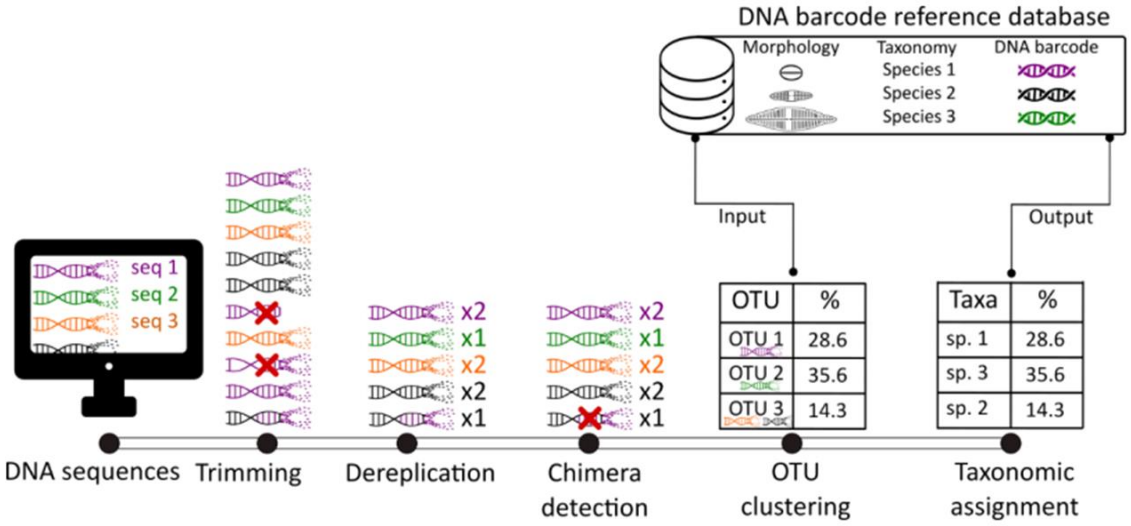

Figure 1 – Major steps of eDNA metabarcoding bioinformatics treatments applied to transform raw DNA sequences obtained after High-Throughput Sequencing (HTS) into species taxonomic list, example illustrated with the case of diatoms.

Using the raw DNA sequences produced by High-Throughput Sequencing (HTS) technologies (see deliverable D.T1.1.2), most of the bioinformatics pipelines include (figure 1): (i) a "trimming" step which corresponds to a curation step used to remove DNA sequences that does not meet minimum length and quality requirements, removing by this way potential

sequencing errors; (ii) as several DNA sequences might be identical, a “dereplication” step is applied to keep only one unique sequence for further bioinformatics steps while the quantitative information is saved independently, reducing the computing power and the time required for downstream bioinformatics steps; (iii) during the laboratory preparation of eDNA samples, the PCR amplification step may create artefactual DNA sequences, called chimera, corresponding to a mix of several “real” DNA sequences ; as those chimerical sequences may inflate the real biodiversity, the “chimera detection” step enables to detect and remove those erroneous DNA sequences; (iv) according to their genetic similarity, close related DNA sequences can be grouped together within the same Operational Taxonomic Units (OTU) during the “OTU clustering”, the OTUs are supposed to reflect species (the conversion from DNA reads to OTUs can be skipped and replaced by other alternatives in some pipelines) ; (v) comparing the DNA representative sequence of each OTU to a DNA barcode reference database (where each reference is associated to a taxonomic name), a taxonomy is assigned to each OTU, allowing to produce the final taxonomic list.

In the context of the Eco-AlpsWater project, several bioinformatics pipelines and strategies have been formalized in an initial version within the D.T1.1.3 deliverable. These protocols have then been tested in real in order to produce the validated required as deliverable DT1.3.2 of the EAW project.

Several technical documents are constitutive of the D.T1.3.2 due to the necessity to present distinct bioinformatics pipelines according to the targeted biological groups, i.e. bacteria & cyanobacteria, micro-algae & micro-eukaryotes, diatoms, fish. For each of these biological target, a specific molecular reference database is required to obtain the most relevant taxonomic affiliations. These protocols help to consolidate the approach of standardization of eDNA workflows for aquatic biomonitoring and should facilitate the future the implementation of these tools for routine survey of lakes and rivers.

The protocols are validated and accessible to all the Eco-AlpsWater partners, and will be made accessible through scientific publication or published protocols :

-Bioinformatics procedure for Prokaryotes using DADA2

-Bioinformatics procedure for Protists using DADA2

-Bioinformatics procedure for fish using Obitools

-Bioinformatics procedure for diatoms using Mothur

-Bioinformatics procedure for diatoms using DADA2