



Conference Abstract

The use of machine learning predictive models to assess rivers quality with molecular data

Maria João Feio ‡

‡ University of Coimbra, Department of Life Sciences, Marine and Environmental Sciences Centre (MARE), Coimbra, Portugal
University of Coimbra, Department of Life Sciences, Marine and Environmental Sciences Centre (MARE), Coimbra, Portugal

Corresponding author: Maria João Feio (mjf@ci.uc.pt)

Received: 02 Mar 2021 | Published: 04 Mar 2021

Citation: Feio MJ (2021) The use of machine learning predictive models to assess rivers quality with molecular data. ARPHA Conference Abstracts 4: e65380. <https://doi.org/10.3897/aca.4.e65380>

Abstract

Many tests have been made so far to assess the biological quality of rivers with molecular data. Most often HTS-related eDNA metabarcoding sequences clustered into Operational Taxonomic Units (OTUs) are assigned to taxa, using reference barcode databases. From there, the existing biotic indices, developed for morphological data are calculated. However, this approach has several drawbacks that may justify their lower performances compared to traditional ones, or not extracting the maximum potential from the molecular data. The first is the incompleteness of reference databases (despite their continuous evolution) - avoiding the conversion of molecular into taxonomic may overcome this issue. Yet, another likely source of bias in the assessments is at the basis of existing classification systems: a possible poor correspondence between the biological reference conditions developed based on species morphology and on molecular data. In other words, molecular-based assemblages from different rivers may not group similarly or respond to the same environmental variables as taxa. Correcting this would require rebuilding the whole systems and establishing new typological-based molecular reference values, which are then used to calculate Ecological Quality Ratios (EQR) and determine the Ecological Quality Status (EQS) of river sites. One alternative to the grouping step inherent to the typological approach, and that may be viewed as artificial (nature is a continuum), is the prediction of site-specific reference conditions based on abiotic characteristics of sites. Thus, we tested a combination of machine-learning modelling techniques to build a taxonomic-free site-specific index to assess rivers based on diatom assemblages, from 81

sites located in Portugal (Feio et al. (2020)Feio et al. 2020). The models are trained to predict diatom OTUs expected under reference conditions, from environmental data. Then, for each test site, an OTU EQR is calculated based on the deviation between the observed OTUs in the field samples and the expected OTUs under reference conditions, which is finally converted into a quality class (after the construction of a new classification system). The molecular-based model was accurate and sensitive to global anthropogenic disturbance (such as, changes in land use and habitat quality), which gives promising insights to its use for bioassessment of rivers. Further work consists of testing this approach with invertebrate data as well as investigating the potential of ISO or ESV in alternative to OTUs.

Keywords

rivers bioassment

metabarcoding

eDNA

predictive modelling

quality classification

Presenting author

Maria João Feio

Presented at

1st DNAQUA International Conference (March 9-11, 2021)

Hosting institution

University of Coimbra, Department of Life Sciences, Marine and Environmental Sciences Centre (MARE)

References

- Feio MJ, Serra SQ, Mortgua A, Bouchez A, Rimet F, Vasselon V, Almeida SF (2020) A taxonomy-free approach based on machine learning to assess the quality of rivers with diatoms. *Science of The Total Environment* 722 <https://doi.org/10.1016/j.scitotenv.2020.137900>