

Monitoring the environment with artificial intelligence

December 13 2018



Dioindication of the health status of ecosystems by micro-organisms. Credit: Yvain Coudert

Microorganisms perform key functions in ecosystems, and their diversity reflects the health of their environment. However, they are still largely under-exploited in current biomonitoring programs because they are difficult to identify. Researchers from the University of Geneva (UNIGE), Switzerland, have recently developed an approach combining two cutting edge technologies to fill this gap. They used genomic tools to sequence the DNA of microorganisms in samples, and then exploited this considerable amount of data with artificial intelligence. They built predictive models capable of establishing a diagnosis of the health of ecosystems on a large scale, and identified species that perform important functions. This new approach, published in the journal *Trends in Microbiology*, will significantly increase the observation capacity of large ecosystems and reduce the time of analysis for very efficient routine biomonitoring programs.

Monitoring the health status of ecosystems is of crucial importance in a context of sustainable development and increasing human pressure on the environment. Different species of micro-organisms sensitive to changes in their surroundings are used as bio-indicators for monitoring [environmental quality](#). However, their morphological identification requires a lot of time and expertise. "A year ago, we were able to establish a water quality index based solely on the DNA sequences of unicellular algae present in the samples, without needing to visually identify each species," explains Jan Pawlowski, professor at the Department of Genetics and Evolution of the UNIGE Faculty of Science.

Use DNA sequences without having to identify them

Genomic tools make it possible to quickly and very accurately describe the biological communities inhabiting an environment. However, a large proportion of the data cannot be used to conduct environmental health diagnoses because many DNA sequences are not referenced in existing databases. The species that possess these sequences are therefore unknown, as well as their ecological role. "In order to exploit all environmental genomics data, namely all the biodiversity of the samples, we used a machine learning algorithm," notes Tristan Cordier, a member of the Geneva group and first author of the study.

The biologists used samples of different known ecological quality status, ranging from good to bad, from which they sequenced the DNA. The combination of this information allowed them to build a reference system with the data from each sample. "A predictive model was then developed with this algorithm, based on our training data. These include data from reference diagnoses and data from the sequencing of unknown [species](#)," says Jan Pawlowski. This model is refined and validated over time by including new reference samples to the existing training dataset.

Discover new bio-indicators

The combination of these two cutting edge technologies makes it possible to obtain ecological values for DNA sequences without having to identify them. Species of microorganisms, already described or not, performing important functions can be discovered through this approach, as well as new bio-indicators. "Our research shares some similarities with the research on the human microbiome. Both aim to unravel microbial communities and identify biomarkers that can be used as powerful diagnostic tools to detect environmental pollution or human disease," concludes Tristan Cordier.

More information: *Trends in Microbiology*, [DOI](#):

[10.1016/j.tim.2018.10.012](https://doi.org/10.1016/j.tim.2018.10.012)

Provided by University of Geneva

Citation: Monitoring the environment with artificial intelligence (2018, December 13) retrieved 1 May 2024 from <https://phys.org/news/2018-12-environment-artificial-intelligence.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.