eDNA First: A Framework for Transition in Biodiversity Monitoring

Mauro de Freitas Rebelo

Bio Bureau Biotecnologia Ltda.

UFRJ Technology Park
Rua Sydiney Martins Gomes dos Santos 52, Module 9
Cidade Universitária
Rio de Janeiro, RJ 21941-859
Brazil

mauro.rebelo@biobureau.com.br

Biodiversity is vast, but the methods we use to observe it rely on capturing and identifying individual organisms. We are making decisions that carry legal, operational, and reputational consequences for both project developers and environmental agencies with precarious information

In 2017, I heard Pablos Holman describe the problem through Excel: if you use the spreadsheet, you cannot use all the data; if you use all the data, you cannot use the spreadsheet. Biodiversity presents that same mismatch between what we need to observe and the tools we routinely use. It is a big data problem with small data tools.

Capture based surveys observe isolated points while biodiversity varies across space and time.

Morphological taxonomy allows us to speak in great detail about what was collected, but it cannot tell us anything about what was not captured: the vast majority.

Confidence comes from long time series, which can only exist if we start measuring before demand arises. When a decision is required, the data will already exist and can be converted into actionable information that is comparable, traceable, and technically auditable.

Mind the gap

Building reference databases from the organisms we have captured is legitimate, but is it the priority when most organisms remain unknown to us [7]?

In 2025, Basecamp Research described the lack of data as a *data wall* limiting the training of AI models in biology [8]. The paper struck me for two reasons: the size of the gap and the business opportunity in biodiversity data.

UniRef held 454 million protein sequences accumulated over decades, while BaseData held 9.8 billion, generated through a collection network spanning 26 countries and autonomous regions. Even after accounting for redundancy, BaseData expanded known protein diversity more than tenfold relative to the protein databases compared in the paper [8].

Reference libraries help us name the sequences we collect; they cannot supply environmental observations we never made. For me, the Basecamp result makes postponing eDNA until those libraries are complete untenable.

For me, the priority is to collect and sequence environmental samples now, so decisions can draw on more data, even at lower taxonomic resolution.

Fill in the blanks

We can fill this gap if we accept measuring a proxy for individuals: the DNA they leave in the environment.

In 2011, a *Scientific American* title caught my attention: “DNA in a Cup of Water Reveals Lake Denizens” [9]. A cup of water could show us a lake’s inhabitants without capturing them.

Organisms leave DNA traces in water, soil, and air. Sequencing reads the order of bases in those traces, turning atoms into bytes. The resulting sequences can identify organisms and compare communities across places and time—without capturing each organism first.

eDNA is an informative proxy. It does not merely complement capture data: it fills the gap capture leaves—the organisms we could not capture. By changing the unit of observation from the individual to its DNA, we can sample more broadly.

We already have sequencing technology capable of working at this scale, and the science is mature for this transition. Environmental monitoring practice has not caught up.

Prêt-à-porter

Environmental monitoring needs to compare ecosystems across space and time, but it uses methods invented by naturalists to catalogue individuals in collections more than three centuries ago. In a difficult dinoflagellate identification task, 16 experts correctly labelled 72% of specimens; agreement between experts in earlier tests fell to 43% [16].

In Guanabara Bay, 28 water samples from ten localities produced 31,667 sequence groups with a single marker. Comparing their composition separated the bay into regions; the result changed when we compared unnamed sequences instead of named species. That is information about the ecosystem's structure that a list of captured individuals would miss.

Saturation curves show the additional diversity each sample contributes, allowing us to estimate how representative the sampling is in each region and whether we can compare regions [3]. If uncertainty remains, we can use targeted PCR on the same samples, or an independent method such as capture and morphological identification when higher resolution is needed. ISO 17805:2026 provides procedures for sampling, capturing, and preserving eDNA from water [6].

eDNA first

We propose eDNA first, not eDNA only: an automated way to approach the environment at large scale before deciding what needs closer inspection.

Think of a satellite image. You use it to find the tree you need to examine. You would not measure one tree first and then take a satellite image to find out whether that tree is beautiful.

eDNA first also means OTUs first. The technological bet is that OTU richness and abundance can carry more information about differences between environments than a list of BLAST-assigned taxa [2]. A named species brings its history—what we already know about where and how it lives—and that matters.

It also means, inevitably, shotgun sequencing. Whenever the eDNA data is insufficient for a conclusion, we can check, double-check, triple-check the conclusion with different complementary data and techniques, including capture and morphological taxonomy.

Because DNA sequencing is scalable and eDNA sampling is simpler, we will be able to survey more frequently and

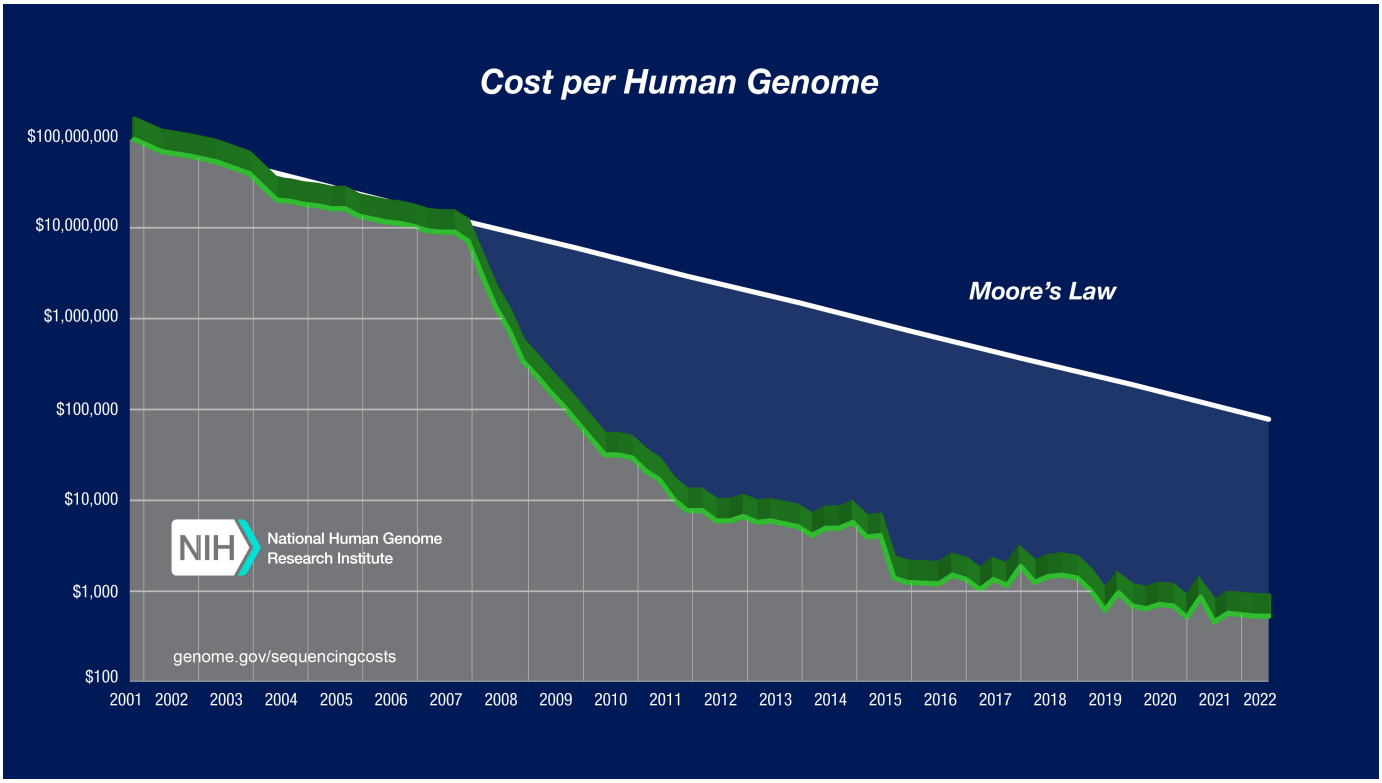


Figure 1. NHGRI/NIH estimated cost per sequenced human genome from 2001 to 2022 (logarithmic scale). This shows the decline in sequencing cost; it does not measure the total cost of eDNA monitoring [1].

collect more data, compare biodiversity structure and functional potential from individual sites to whole territories, and show how they change over time. Those comparisons will back regulatory reports, the biodiversity outcomes of forest restoration, the biodiversity claims of REDD+ carbon credits, and territorial and financial instruments such as CPR Verde in Brazil and biodiversity credits globally.

On shore and offshore

We wanted to test-drive the framework in two scenarios. One was easy to access and hard to conserve, to show the feasibility and importance of frequent monitoring. We chose Guanabara Bay, our backyard and a postcard of Brazil. With ten million people living around it, thousands of companies operating along its shores, and with city, port, industry, and mangrove meeting there, it was the perfect location. The other was hard to access and hard to sample: an FPSO in the Santos Basin, representing the access challenge of offshore oil operations. There we could test eDNA first under the conditions environmental monitoring actually faces.

In September 2025, we collected 28 samples at ten locations in Guanabara Bay over three days by sailboat. In April 2026, FIDO—the Filtering Instrument for DNA Observation—collected more than 100 samples autonomously at Marina da Glória over ten days. At the same time, the uncrewed Tupan II spent seven days surveying Santos Basin oil fields; around the FPSO, it collected 47 samples at 21 points. The April operations were a joint effort by Bio Bureau, TideWise, and MBARI under Equinor's ANP R&D project. The offshore methods and results are reported in our preprint [10].

In Guanabara Bay, Jaccard-based analyses separated the ten localities (p = 0.001), and DAPC found differences after Holm correction. In the OTU analysis, the coastal stations Marina da Glória, Jurujuba/Boa Viagem, Enseada de Botafogo, Praia da Glória, and Praia Vermelha formed one group; Museu do Amanhã, the two Vão Central sites, and Ilha do Fundão formed another; Paquetá stood apart. IndVal associated *Guinardia striata* with Jurujuba/Boa Viagem, *Perna viridis* with Vão Central north, and *Streblospio gynobranchiata* with Paquetá (unpublished data).

At Marina da Glória, 110 FIDO samples collected over about ten days yielded 122 assigned species. The accumulation analysis measured what each additional sample contributed. One sample a day retained all 22 recurring species; sampling about every two hours recovered at least 90% of the full species inventory. Those results let us choose frequency according to what the monitoring needs to detect (unpublished data).



Figure 2. FIDO (Filtering Instrument for DNA Observation), developed by the Monterey Bay Aquarium Research Institute (MBARI), during the April 2026 autonomous sampling campaign at Marina da Glória. Photo: Bio Bureau.

Around the FPSO, 47 samples from 21 points between 300 and 1,500 m showed no distinct radial groups with either 12S or 16S. Samples collected one or two hours apart at the same point were almost as different as samples from different points, while technical replicates agreed. That result supports testing fewer stations with more frequent collections.

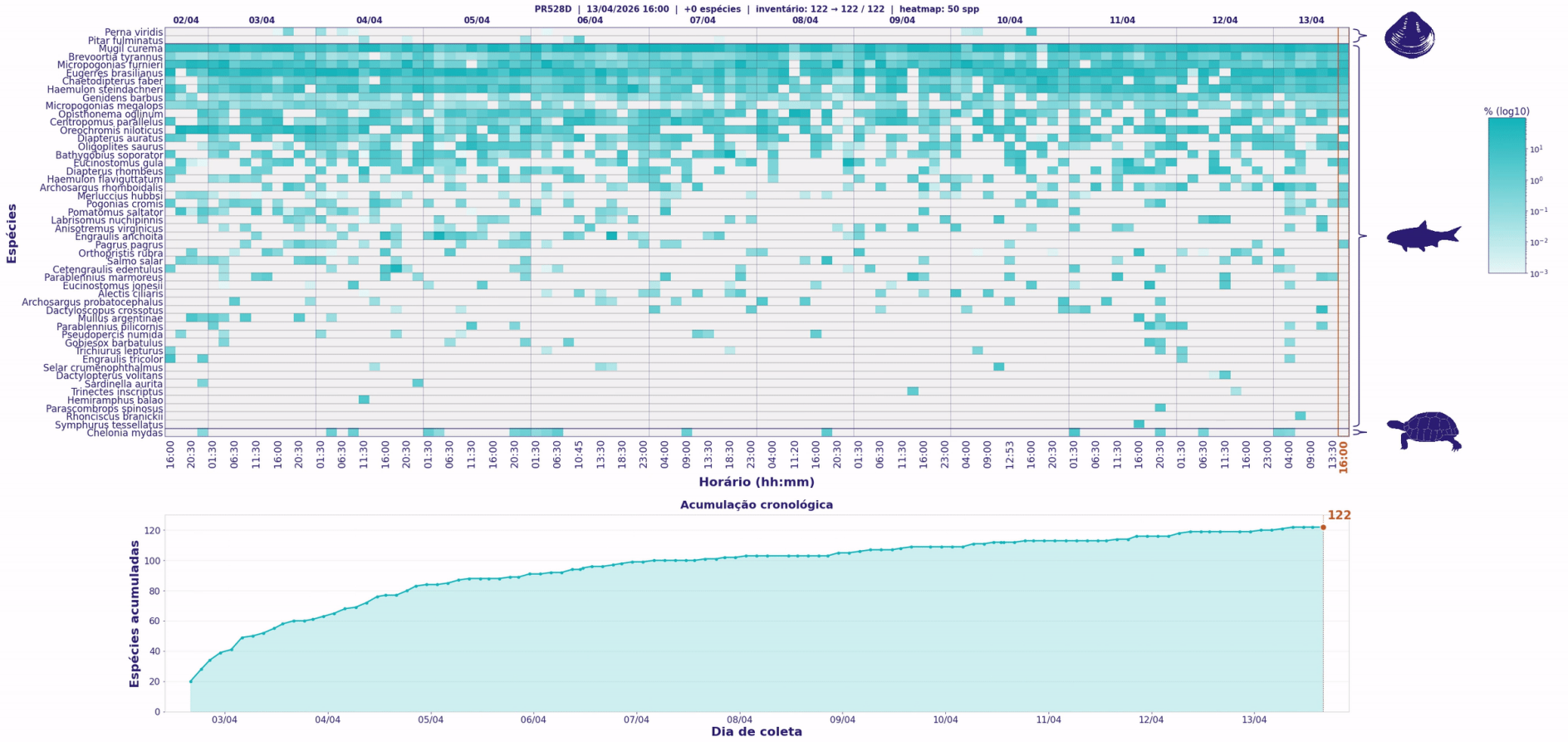


Figure 3. FIDO detections at Marina da Glória over ten days, including charismatic and invasive species, commercially important fish, and rare fish. The final frame shows the species accumulation curve reaching saturation, allowing us to estimate how often to sample this area—an unprecedented result for Guanabara Bay. View the animation. Analysis: Bio Bureau/MBARI (unpublished data).

The pilots met the scope we set for them. In Guanabara Bay, 28 water samples collected in three days yielded 1,894 assigned taxa—an inventory we found larger than a decade of PELD surveys in the bay [11]. Offshore, the uncrewed expedition collected 47 samples at 21 points around the FPSO at an estimated one tenth of the cost of a conventional capture campaign. FIDO's 110 samples over ten days, together with the offshore expedition, demonstrated that autonomous eDNA sampling can provide the repeated observations needed for long-term monitoring.

eDNA first is not eDNA only. We now have a course of action that can work with the existing requirements of the Brazilian Institute of Environment and Renewable Natural Resources (IBAMA), Brazil's federal environmental agency. The Environmental Characterization Project (PCA) characterizes an oil field block through a process established over decades, with legal certainty. The Environmental Monitoring Project (PMA) [15] is more flexible because its monitoring plan responds to the reality of each project. No one wants to dismantle an established process to introduce a new technique; eDNA would have to be added. But companies do not want another permanent expense either. How do we solve that dilemma? We propose adding eDNA to the PCA, at marginal cost within a campaign that already collects large amounts of sediment and water. When the PMA begins, that baseline can support eDNA first, with capture used where the molecular results leave a specific question to resolve. The added cost in the PCA could then be offset by a different use of sampling effort in the PMA, within IBAMA's requirements.

Trustless chain of custody

If there is a dispute between stakeholders, an environmental decision may eventually need to stand up in court. So must the evidence behind it.

An emergent property of the amount of genetic information in an environmental sample is its uniqueness and power to discriminate. A biodiversity profile works like a fingerprint of a place and time. A rich profile gives us more features with which to test its declared origin, provided samples from the same time and place share a recognizable identity and differ from samples from other environments. Someone could create a FASTQ file and say it came from a sequencer. To make it appear to come from a particular environment, they would have to reproduce that pattern—not merely include the species expected there [4].

We can test the declared origin by asking whether an unknown sample is classified with reference samples from that place and time—for example, using DAPC. Sequencing also turns biological information into digital files. A SHA-256 hash lets us check whether a file has changed since its hash was recorded: compare the two values, and any alteration becomes visible [5]. The hashes of the raw files and bioinformatics outputs could be combined in a Merkle-tree-like structure. Registering its final hash as a non-fungible asset on a public blockchain would provide dated proof that those files existed. These are separate proofs of origin, integrity, and existence [4].

We took eDNA, an almost digitally native asset, and made its record fully digital. More samples let us measure uncertainty and reduce it, while making it dramatically harder to recreate the profile of a specific place and time. This is the trustless chain of custody we were missing: an environmental agency does not have to rely only on someone's declaration of origin, and later tampering with the digital files can be detected. This is legal certainty.

In *Michigan v. United States Army Corps of Engineers*, 667 F.3d 765, 783–84 (7th Cir. 2011), the appeals court considered eDNA evidence about invasive carp alongside capture results [12]. The court did not grant the injunction the states sought, but it treated the eDNA results as evidence worth considering. We have an example of a court examining what eDNA can show. The next step is to collect such records before a dispute arises. Records that can be reexamined can support scrutiny across research, industry, public debate, regulation, and litigation.

A call for a molecular REVIZEE

If data is the new oil, eDNA is the new exclusive exploration zone. REVIZEE was Brazil's effort to inventory living resources across its Exclusive Economic Zone and estimate sustainable catches [13]. It mobilized researchers and vessels for a decade. Today, we are using eDNA largely to put names on sequences, measuring it against records built from the small fraction of biodiversity that capture-based taxonomy reached—including errors carried into GBIF and GenBank [14]. That keeps molecular bioprospecting inside the limits of the method it could surpass. The resource is already in our waters, and we know almost nothing about it. Whoever breaks free of that constraint and starts collecting it at scale will gain the prize: the information itself.

In Luciano Candisani's photograph, Tupan II leaves Guanabara Bay for Santos. The TideWise uncrewed vessel carried MBARI's ESP on a Bio Bureau expedition, traveled 200 miles to an FPSO in the Santos Basin, and collected around 50 eDNA samples over three days. The image captures the potential of uncrewed navigation, autonomous sampling, and eDNA sequencing working together. The technology for a molecular REVIZEE exists. We have deployed it.

The next step is opportunistic and speculative sampling: collect before a monitoring project or a bioprospecting question tells us what to look for, using the vessels and places



Figure 4. Tupan II leaving Guanabara Bay for the Santos Basin with MBARI's Environmental Sample Processor aboard. The uncrewed expedition brought autonomous eDNA sampling offshore. Photo: Luciano Candisani/Bio Bureau.

already within reach. Sampling can be active, as with Tupan's ESP, or passive where a sampler can remain in the water. Offshore support vessels, oceanographic buoys, and sailors could contribute across different places and times. Sequence the samples, keep their location and date, and connect them to oceanographic and satellite data. Together, those records would become a database we can question again as new uses emerge.

The database would serve two purposes. Repeated samples would let us compare biodiversity across places and time, including changes we did not know to look for when we collected them, as we are already doing with TODABIO. The same sequences could lead bioprospectors to genes, proteins, and enzymes. We do not need a species name for every sequence before either kind of work can begin; names and interpretations can improve as the database grows.

eDNA is a mature technology when we use it for the right purpose. It gives us a satellite image of biodiversity, not a magnifying glass: more information per sample, better suited to comparing environments than the precarious record built from captured individuals. Lower sampling costs make frequent observation possible, so uncertainty can be measured and reduced over time. Digital records make tampering detectable and let us examine biodiversity that capture never reached. eDNA first is not eDNA only. Capture and taxonomy remain available to verify results when we need greater resolution. The science is established. The technology has been deployed. With eDNA first, we can start now.

Darwin sailed on HMS Beagle without knowing what he would find. Today, we can use Tupan II to explore the molecular biodiversity of our waters. The next molecular REVIZEE should begin before we know every question its data will answer.

References

[1] National Human Genome Research Institute. DNA Sequencing Costs: Data. https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data

[2] Bouchet VMP et al. (2026). Taxonomy-free approach overcomes the gaps in ecological knowledge: The case of foraminiferal metabarcoding applied to environmental impact assessment. PLOS ONE 21: e0356357. https://doi.org/10.1371/journal.pone.0356357

[3] Cortez T et al. (2025). Insights into the representativeness of biodiversity assessment in large reservoir through eDNA metabarcoding. PLOS ONE 20: e0314210. https://doi.org/10.1371/journal.pone.0314210

[4] Rebelo MF et al. (2025). eDNA enables proof-of-origin of samples from reservoirs with rich and abundant biodiversity. bioRxiv, preprint. https://doi.org/10.1101/2025.05.13.653644

[5] National Institute of Standards and Technology. Hash Functions. https://csrc.nist.gov/projects/hash-functions

[6] International Organization for Standardization (2026). ISO 17805:2026. Water quality — Sampling, capture and preservation of environmental DNA from water. https://www.iso.org/standard/87980.html

[7] Mora C et al. (2011). How Many Species Are There on Earth and in the Ocean? PLOS Biology 9: e1001127. Estimates 8.7 ± 1.3 million eukaryotic species (standard error), rather than counting them directly. https://doi.org/10.1371/journal.pbio.1001127

[8] Vince O et al. (2025). Breaking Through Biology's Data Wall: Expanding the Known Tree of Life by Over 10x using a Global Biodiscovery Pipeline. bioRxiv, preprint. https://doi.org/10.1101/2025.06.11.658620

[9] Bushwick S (2011). DNA in a Cup of Water Reveals Lake Denizens. *Scientific American*, podcast, December 19. https://www.scientificamerican.com/podcast/episode/dna-in-a-cup-of-water-reveals-lake-11-12-19/

[10] do Amaral D. L. A. S. et al. (2026). Autonomous sampling and eDNA metabarcoding for offshore biodiversity monitoring: a pilot study around an FPSO in the Southwestern Atlantic. bioRxiv, preprint. https://doi.org/10.64898/2026.10.22.761088

[11] Valentin JL, Gouvêa GV, Gomes CL (2020). Mesozooplâncton e massas d'água na Baía de Guanabara: dez anos de monitoramento. *Oecologia Australis* 24(2). https://doi.org/10.4257/oeco.2020.2402.09

[12] *Michigan v. United States Army Corps of Engineers*, 667 F.3d 765, 783–84 (7th Cir. 2011), No. 10-3891. https://media.ca7.uscourts.gov/cgi-bin/OpinionsWeb/processWebInputExternal.pl?Path=Y2011%2FD09-13%2FC%3A10-3891%3AJ%3AWood%3Aaut%3AT%3Aop%3AN%3A769804%3AS%3A0&Submit=Display

[13] Brazilian Ministry of Science and Technology (2006). Management report describing the REVIZEE program. https://repositorio.mcti.gov.br/bitstream/mctic/4802/1/2006_relatorio_gestao_mct_jan_dez.pdf

[14] Smith BE, Johnston MK, Lücking R (2016). From GenBank to GBIF: Phylogeny-Based Predictive Niche Modeling Tests Accuracy of Taxonomic Identifications in Large Occurrence Data Repositories. *PLOS ONE* 11(3): e0151232. https://doi.org/10.1371/journal.pone.0151232

[15] Brazilian Institute of Environment and Renewable Natural Resources (IBAMA) (2019). Technical Note 4/2019/COEXP/CGMAC/DILIC: Guidelines for Environmental Characterization Projects (PCA) and Environmental Monitoring Projects (PMA) in offshore oil and gas drilling licensing. https://www.gov.br/ibama/pt-br/assuntos/laf/orientacoes-tecnicas/arquivos/20230524_SEI_Ibama_6580026_Nota_Tecnica_04_2019_Dilic.pdf

[16] Culverhouse PF, Williams R, Reguera B, Herry V, González-Gil S (2003). Do experts make mistakes? A comparison of human and machine identification of dinoflagellates. *Marine Ecology Progress Series* 247: 17–25. https://doi.org/10.3354/meps247017