Concept Paper Draft 2 25 Jan 2021

Jesse H. Ausubel and Mark Y. Stoeckle

The Great Global Fish Count (GGFC): A Potential Project of the UN Ocean Decade

1. THE IDEA

a. What is the GGFC?

The GGFC is a citizen-science project to count fish and then other forms of marine life in coastal and later all waters using loose DNA in seawater (marine eDNA) shed by all animals.

The project would distribute filtration syringes or similar devices to millions of people. The devices would enable participants to take water samples of a standard fraction of a liter, filter the water, stably preserve the filter with the sediment containing DNA, and post the filters to qualified labs for eDNA analysis. The lab costs would not be borne by the citizen scientist.

Participants would use cellphones to record metadata including latitude, longitude, and time.

Each filter would have a unique barcode.

Labs would analyze samples, identify species, and upload data to a project database including Web-based Geographical Information Systems accessible to a mobile app.

The collector of each sample would learn the species and quantities of DNA in their sample(s) and have access to information from all other samples.

The totality of the data would open countless opportunities for analysts to discover patterns and trends.

Labs would analyze samples by high-throughput DNA sequencing methods, known as metagenomics or metabarcoding, to learn all fish species whose DNA is present in a sample.

One could also include a component focused on water collection and detection for presence of single species of particular interest, such as the coelacanth or particular rare or endangered species.

The project would begin by targeting fish species, because of their societal importance, public interest, availability of high-quality primers to grab relevant sequences, and richness of the DNA reference library of fish sequences.

Later in the decade the project could expand to all vertebrates (including marine mammals), molluscs, crustaceans, and other taxa, perhaps adding one major group each year.

The peak of sampling would occur every year on the same day, for example, World Ocean Day in June..

b. What about the idea is new?

The proposed scale of use of eDNA is unprecedented. The largest eDNA efforts so far involve a few hundred samples, while the GGFC would involve millions or tens of millions, an increase by a factor of ten or one hundred thousand times. The proposed scope, covering entire coastlines, basins, and the global ocean is also new.

c. Why is the idea important?

The GGFC could measure the baseline and changes in the totality of the distribution and abundance of most known fish in the Anthropocene oceans, and potentially other taxa. It could greatly advance knowledge of the impacts of fishing and aquaculture, urbanization of coast lines, pollution (chemical, noise, debris, radioactivity), offshore energy extraction, efforts to restore and protect habitat, climate change, and other factors, especially if eDNA surveys become a standard (essential) variable in ocean observing systems. eDNA can illuminate the invasive, elusive, or endangered as well as the common. The R&D component of the program would speed improvement and adoption of best practices. The program would foster scientific capacity and initiation of time-series.

d. What kind of opportunity is there to shape and frame the idea?

No US or international networks yet exist for the systematic advancement of marine eDNA. The program does not face a big problem of entrenched structures. The GGFC could create much of the national and global infrastructure for eDNA and marine genomics more broadly. It could bring down costs down and standardize practices.

e. Is there urgency or timeliness in pursuing the idea?

The rapid industrialization of the oceans, including urbanization of coastlines, plans for thousands of square kilometers of windfarms, and massive expansion of aquaculture, give urgency for more reliable, quick, affordable ways to monitor marine life. This, together with the stated goal of the UN Ocean Decade to leave no one behind, explains our initial focus on citizen scientists and coastal areas. The GGFC could also serve as a prototype and stimulus for the genomic dimension of Global Ocean Observing System and Marine Biodiversity Observing Network. It would also encourage prompt genome sequencing of more marine species and their inclusion in reference libraries. The convergence of genomics with information technologies and geographical information systems makes the program newly feasible and available for global participation.

In the USA, the GGFC meshes with NOAA's Genomics Road Map, marine genomic and biodiversity initiatives of the Smithsonian Institution, and archival initiatives such as Ocean Genome Legacy Project.

f. How does this relate to the Ocean Decade?

The GGFC would contribute to almost all the seven themes of the Ocean Decade: clean ocean, sustainable productive ocean, predicted ocean, healthy & resilient ocean, transparent and accessible ocean, and inspiring ocean. It is not strongly relevant to a safe ocean.

The GGFC could operate either as a **Decade** P**rogramme** (global in scale, fulfilling several Decade objectives, long-term (multi-year), interdisciplinary and multi-national with component projects and enabling activities. It could also operate as a **Decade Project** (a discrete and focused undertaking of shorter duration, standing alone but also contributing to an identified **Decade Programme**). People and institutions in every coastal nation could participate, and in fact those without coasts.

2. PROCESS

a. *How would an organization pursue the idea?*

The GGFC needs an initial financial patron and an initial operating organization to develop the concept between now and the kick-off conference of the Ocean Decade in Berlin in June 2022 where a program plan could be fully endorsed and the program launched. The financial patron could be governmental or private and need to provide, probably, $1-2 million. The operating organization would need to form a Secretariat of perhaps 3-4 people (they could be distributed or co-located) whose primary responsibility would be to elaborate, with wide consultation, the Decadal plan and to build the partnerships needed to support and implement it. The operating organization should have high credibility in marine genomics.

b. What project management concepts might work?

Models might be explored and adapted from previous successful international environmental research programs including the Census of Marine Life and those operated by the Scientific Committee on Oceanic Research (SCOR). The formula would probably involve an international scientific steering committee as well as national or regional committees, and working groups concerned with technology development, standards & protocols, data science, engagement, and other themes. The GGFC would benefit from producing an annual progress report as well as a final report at its conclusion.

c. Rough budget

The marginal direct costs in the USA for collecting and analyzing a water sample (or small set of samples) for eDNA, without labor costs of scientific personnel, are about $50. If the GGFC were to aim initially for one million samples, the cost to produce the data would come to about $50 million. We might reasonably assume an equal amount for the sum of planning activities and value added through analyses and outreach, for a total of about $100 million. If the GGFC were to extend to more taxa and time series over several years during the Ocean Decade and become a Programme, the expenditure might grow to between $250-$500 million. Promising means might reduce the cost. For example, companies might donate or subsidize the equipment (such as the syringes), mailing of filters, primers, and cloud storage of data.

d. How long would it take to make substantial progress?

Adoption of the GGFC as a Decade Programme or Project at the Berlin conference would be an initial milestone. 2024 might be the year of the Great Global Fish Count. While preliminary results could be available within weeks of the sampling, it would take about 6-12 months to pool and analyze the data in a careful way and visualize effectively.

e. *What would progress look like? What would be the main outputs*?

The first outputs would likely be standards and protocols, and placement of large-scale orders for sampling equipment. In parallel, software development would occur to set up a GGFC website (or network of them, including on social media) and to make appropriate interfaces for a range of participants and hardware. A small-scale demonstration project of sampling might be carried out during 2023 and then, following glitch reduction, the actual day or days of massive widespread sampling in 2024. The sampling would be followed by population of databases and preparation of various kinds of publications and maps. Data could be archived in the Ocean Biogeographical Information System (OBIS). Important results would be highlighted in press releases and other means to reach the broad public and various stakeholders, perhaps in conjunction with a scientific conference about the findings

in 2025. Aquariums and other informal education centers would be encouraged to mount GGFC displays.

f. *What might roadblocks might arise?*

No samples need cross international borders. However, issues of access and benefit-sharing in relation to genetic resources could arise. It would be wise to consult with experts on the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization and related agreements to make sure the GGFC adheres to appropriate practices. Because the short "barcode" sequences on which the GGFC would rely are in the public domain in databases such as GenBank, issues of intellectual property should not loom large.

The more extensive the participation and sampling, the more valuable will be the GGFC. However, the concept does not depend on the success of any one expedition or effort. The concept is thus robust. Also, if bad weather or other events delay sampling a few days, the consequences are very minor.

g. *Which are important potential partners beyond the academic ocean research community?*

Within the US government, most if not all participants in the National Ocean Partnership Program, should take an interest, including the Office of Naval Research, National Science Foundation, and DOI's Bureau of Ocean Energy Management as well as NOAA and Smithsonian. State coastal and environmental agencies might also have much to gain, for example, by enhancing efforts with regard to salmon in the Northwest or cod in the Gulf of Maine.

Suppliers of goods and services in the biotech industry such as New England Biolabs, Illumina, GeneWiz, and ThermoFisher/Applied Biosystems would be valuable partners. So would companies making and operating drones, which could extend sampling, especially far from shore and in deep water. Shipping companies such as FedEx and UPS might provide useful services, and IT companies with cloud storage.

Internationally, nongovernmental organizations, including POGO, SCOR, and SCAR (for Antarctica) could multiply success of the program by fostering eDNA science, standards and protocols, sampling, and big data analyses. In the North Atlantic ICES and in the Pacific PICES could use GGFC to develop eDNA standards for fisheries and other surveys. UN FAO (Fisheries Division) and other agencies could engage in capacity-building around GGFC and benefit from its findings, which should upgrade information for assessments of the State of Fisheries, Global Oceans, and marine sections of reports of IPCC and IPBES.

Aquariums and natural history museums would have much to contribute and to gain from the success of the GGFC. Cruise lines, resorts, others involved in marine and coastal leisure industries might also welcome the program, to which they could contribute samples and with which they could entertain their guests.

h. If the USA does not pursue this, are there others that might step up?

The USA and France have far the largest Exclusive Economics Zones (EEZ). For the GGFC to succeed, the USA and France must play leadership roles. Next in EEZ rank come Australia, Russia, UK, Indonesia, Canada, Japan, New Zealand, Brazil, and Chile. Ideally, people and organizations in most of these nations would take a strong interest in the GGFC and form a vigorous core, together with nations that have a mix of strong competence and shoreline, such as India, Mexico, Norway, Denmark, and Portugal.

g. *Is it possible to offer some cost/benefit estimate, or some kind of benchmarking against comparable efforts?*

NOAA's "Fisheries Data Collections, Surveys, and Assessments" budget was $168 million in FY19, when the USA accounted for about 1/6th of global gross domestic product. If one guesses that half of nations do such data collection, the global annual expenditure on national fisheries surveys could be pro-rated to about $500 million. All present surveys are in a position to collect and filter small bottles of water along with their net trawls and acoustic surveys, and many nations without vessels and equipment could implement eDNA surveys. If the GGFC could catalyze major improvement in such surveys, that alone would go far toward justifying its expense.

3. OUTCOMES

a. *What are likely longer term outcomes*?

As suggested, enhanced adoption of genomic techniques in GOOS and national agencies concerned with fisheries and aquaculture, harbors, marine sanctuaries, and coastal restoration could be a major long-term outcome, as well as enhancement of marine genomic reference libraries. Smarter boundaries in space and time for marine protected areas could be an outcome, as well as identification of new biodiversity hot spots and trouble spots. Detection and documentation of ecological recovery should become more reliable. The number of people qualified for careers in marine 'omics should grow. Entrepreneurial companies providing eDNA and other services in marine genomics should increase in number and scale. Technology should advance so that DNA results are affordably available to large numbers of people for identification of presence of marine species in near real-time, at least for single species. Of course, the GGFC should also produce surprises about distributions and abundance, which are by definition hard to anticipate.

b. How would the outcomes be measured or otherwise evaluated?

Changes in survey practices are easy to observe, as are changes in boundaries for fisheries management and for marine biodiversity protection. Inclusion of marine eDNA in educational curricula could also be monitored, and growth of marine genomics enterprises within the Blue Economy.

c. Who would the impacts benefit? Ethical issues?

No one need be left behind. Anyone with a cellphone should be able to benefit from the GGFC and its outcomes. People of all ages can make valuable collections and learn from their own samples. Not only children, but elderly and retired, usually excluded from science, can participate. Recreational boaters and fishers would have wonderful opportunities. Overall, the GGFC should develop global capacity to sense and know the oceans and encourage development of the next generation of ocean scientists.

An important ethical advantage of the program is that it does not require collection of living organism or stress animals.

d. What is the expected impact of these outcomes?

A clean ocean, sustainable productive ocean, predicted ocean, healthy & resilient ocean, transparent and accessible ocean, and inspiring ocean.