

Ocean DNA Program ‘Environmental Profiles’ project

Scope of Work

The Smithsonian Institution, National Museum of Natural History’s (NMNH), Ocean DNA Program is in need of a Consultant that will provide professional, technical, non-personal services outlined below in the Statement of Work.

Statement of Work

The Consultant shall work to architect and implement a beta informatics solution to handle the growing aggregate of environmental sample derived sequence data with a scalable strategy. The Consultant shall work with the existing Ocean DNA Program (or related NMNH) environmental DNA profiles of COI amplicons as the test set. The Consultant shall create an in-house COI amplicon sequence variant (ASV) registry that will scale with projected growth of environmentally derived samples in the Ocean DNA Program. The Consultant will scope material sample metadata, mapping to existing standards where available and determine augmentation sets where applicable. The Consultant shall also build a taxonomic classification engine that can create versioned classification sets of the ASV registry based on a curated voucher-based reference library. The Consultant will work with the Ocean DNA team to create visualization toolkits to present biodiversity data occurrence data across taxa, time and space. The results of these efforts will be used to produce a position paper based on the generated “state of knowledge”. All results, products, and knowledge derived from this contract are the property of The Smithsonian, NMNH and the Ocean DNA program as outlined below in the rights in data clause.

A. Specific Tasks

1. Assemble and collate OceanDNA team-provided environmentally derived COI amplicon sequence variant (ASVs) datasets across marine collections (eDNA, ARMS, fish guts, plankton tows, etc.) from demonstration sites (Moorea, Curacao, Ft Pierce, Chesapeake) and the Global ARMS Program, consisting of up to 2000 profiles. Where possible, these data will be reprocessed from raw read data in order to standardize data quality and create consistent informatics workflows.
2. Determine minimum metadata standards for these derived libraries in order to enable dynamic queries based on sampling method, place or time to enable large-scale questions, visualizations and encourage data reuse and cross-comparisons. These standards should be mapped to existing standards including Darwin Core and MIxS community standards as best possible. The Consultant shall work with the Ocean DNA team to liaison with ongoing GLOMICON and ‘Omics-BON members to develop augmented fields for material samples and processing pipelines.
3. Create a mtDNA COI ASV registry that can be augmented with newly generated datasets. This product is envisioned as a growing 2x2 array with unique ASVs in rows and processed sample runs in columns.
4. Develop a service (Taxonomic Classifier) that runs monthly to update COI amplicons with current scientific taxonomic classification information based on an Ocean DNA team-provided curated reference library. One or many classification methods may be employed

based on conversation and feedback with the Ocean DNA team depending on precision and efficiency.

5. Create monthly updated versioned taxonomically-annotated occurrence records to enable reporting to biodiversity registries including NMNH's collection research system EMu and external repositories such as GBIF and OBIS.
6. Develop visualization tools to demonstrate proportions of unknown sequences, metrics of coverage and projections of completeness across both geography and taxonomy domains.
7. Create workflows using R scripts and R Markdown and or other public domain tools to develop extendable, repeatable, yet customizable workflows for other stakeholders and researchers.

B. Deliverables

The Consultant shall submit to the COTR a monthly written status report detailing progress made in each of the project areas outlined in the Specific Tasks of the Statement of Work. The consultant shall be paid on completion of each project area and once all deliverables have been completed per project area and submitted for acceptance and inspection. Such payment shall be inclusive of all direct labor, indirect costs (overhead and G&A expenses), other direct costs, required travel, materials, rental equipment, communication, reproduction, Consultants, etc. and profit for which the Consultant expects payment.

On acceptance of this contract, the consultant will provide to the COTR prior to a kick-off meeting a brief written work plan.

The Consultant shall complete all of the following deliverables:

1. Aggregate 2000 profiles of environmentally derived mixed taxonomy amplicon sequence variant (ASVs) datasets of COI across marine collections (eDNA, ARMS, fish guts, plankton tows, etc.) from demonstration sites (Moorea, Curacao, Ft Pierce, Chesapeake) and the Global ARMS Program, assembled and collated.
2. Develop minimum metadata standards for these derived libraries determined and documented.
3. Develop a COI ASV registry that can be easily augmented with newly generated datasets created.
4. Deliver a working Taxonomic classifier service.
5. Deliver visualization tools to demonstrate proportions of unknown sequences, metrics of coverage and projections of completeness developed based on taxonomy, time or space.
6. Deliver workflows using R scripts and R Markdown or other public domain tools created to develop extendable, repeatable, yet customizable workflows for other stakeholders and researchers.
7. Provide data and analyses for synthesis paper comparing approaches (ARMS, TOWS, GUTS), accumulation curves, alpha, beta and gamma, as well as projected known and unknown taxonomic coverage for publication.

Submittals

8. **Monthly Progress Reports.** Progress reports shall be provided once a month with the Ocean DNA program team, both in writing and via phone or in person. - Relevant progress reports shall be sent to relevant members of the Ocean DNA program team in writing two days in advance.
9. **Submit Mid-Project Report:** At the end of Year One, this deliverable will be completed by the Consultant. This will be completed no later than 13 months after the start of work.
10. **Submit Final Project Report:** At the end of Year Two, this deliverable will be completed by the Consultant. This will be completed no later than one month after all other tasks and deliverables are completed.

C. Performance Standards

All work shall be completed in accordance with SI National Museum of Natural History's Department of Invertebrate Zoology and Vertebrate Zoology collection and data management policies, procedures, and standards as indicated in the Smithsonian Directive 600, Collections Management and as indicated verbally to the Consultant. Taxonomic classifier can be run monthly using our constantly growing and improving NMNH curated reference library. Versioned registry can be used to report species occurrence records to biodiversity registries such as GBIF and OBIS. Workflows provide extendable, repeatable and customizable workflows for NMNH and other stakeholders.

Smithsonian Institution Furnished Materials

The Smithsonian will furnish all computational resources required for data management work, except those proposed by Consultant as part of work quote. All furnished materials shall be returned to Smithsonian upon completion of the contract.

Consultant Furnished Materials

All supplies and materials necessary to complete the work, are the responsibility of the Consultant, except those provided by the Smithsonian Institution.

Work Location

The Consultant shall perform the work at an appropriate on or off-site location.

Intent to Exercise Options

The Smithsonian Institution reserves the sole option to extend this order to engage the Consultant in providing similar services for two (2) 13-month periods. These option periods are subject to: 1) continuation of the need for the services, 2) acceptance and approval by the Contracting Officer's Technical Representative during the respective contract period, 3) availability of funds from which payment for contract purposes can be made, and 4) the contract price for services to be provided under the optional periods shall be as stated in Consultant's quote.

The opportunity to enter into an option extension is not automatic; however, may be determined in the best interest of the Smithsonian Institution. A written modification will be issued to exercise any options. In the event the Smithsonian exercises its right to extend the period of performance under this contract, all other terms and conditions hereunder shall remain unchanged.

Period of Performance:

Item 1-1 – Year One. The period of performance for this effort shall be from January 1, 2021 through January 31, 2022.

Item 1-1 – Year Two. The period of performance for this effort shall be from February 1, 2022 through March 1, 2023.

Payment Schedule

Multiple payments shall be made upon completion and acceptance of all work as required based upon the schedule below and following the receipt of proper invoices referencing this purchase order number:

1. Payment 1, Environmentally derived mixed taxonomy COI amplicon sequence variant (ASVs) datasets across marine collections (eDNA, ARMS, fish guts, plankton tows, etc.) from demonstration sites (Moorea, Curacao, Ft Pierce, Chesapeake) and the Global ARMS Program, assembled and collated. 25% of year one amount.
2. Payment 2, Minimum metadata standards for these derived libraries determined and documented, 25% of year one amount.
3. Payment 3, ASV registry created that can be augmented with newly generated datasets created, 25% of year one amount.
4. Payment 4, Strategy formulated and initial implementation for Taxonomic Classifier, visualization dashboard and beta testing, first year report submitted, 25% of year one amount.

Option I

5. Payment 5, Taxonomic classifier completed. 30% of year two amount.
6. Payment 6, Visualization tools to demonstrate proportions of unknown sequences, metrics of coverage and projections of completeness developed across taxonomic groups, sampling methods, time period or geographic regions, 30% of year two amount.
7. Payment 7, Workflows using R scripts R Markdown or other public tools created to develop extendable, repeatable, yet customizable workflows for other stakeholders and researchers, 20% of year two amount.
8. Payment 8, Data and analysis for synthesis paper comparing approaches (eDNA, ARMS, TOWS, GUTS), accumulation curves, alpha, beta and gamma, as well as projected known and unknown taxonomic coverage delivered for publication, final report submitted, 20% of year two amount.