

# IARPA-RFI-14-07

**Posted Date: May 30, 2014**

**Response Date: July 14, 2014**

## Synopsis

The Intelligence Advanced Research Projects Activity (IARPA) seeks information from the community on the feasibility of using existing genomic, geographic, or other relevant databases to determine the geographic provenance (geospecificity) of a metagenomic sample. This request for information (RFI) is issued solely for information gathering and planning purposes; this RFI does not constitute a formal solicitation for proposals. The following sections of this RFI define the overall scope of the technical domain of interest, along with instructions for the preparation and submission of responses.

## Background & Scope

Metagenomics is the study of a community of organisms' genetic material recovered from environmental samples. We hypothesize that by correlating the genetic information within a metagenomic sample with information about species co-occurrence and range it may be possible to ascertain where, and perhaps when, a sample was collected. We are aware of a number of projects and databases across a spectrum of disciplines (genomics, biology, ecology, to name a few) that could be relevant to our objective. The purpose of this RFI is to learn more about the content and accuracy of the various relevant databases and their metadata, and to explore whether and how the community believes these databases could be used individually or together to determine the geospecificity of a metagenomic sample to a given degree of accuracy.

Species most relevant to our interests will be those expected to be present in metagenomic samples acquired via casual contact with the ambient environment, for example, those whose genetic residue would persist in air or soil. To achieve desired geospecificity, we expect to need to cross-correlate multiple sequences or genomic markers at the species and/or sub-species levels. We do not assume that all relevant databases will contain geographic metadata, so we are interested in bootstrapping methods that can associate sample species information with likely spatial range.

Examples of relevant databases include:

- Draft or complete genomes from genome databases, e.g., the DoE's Joint Genome Institute at [www.genomesonline.org](http://www.genomesonline.org) and others featured by the Nucleic Acid Research 2014 database issue at <http://nar.oxfordjournals.org/content/early/2013/12/06/nar.gkt1282.full>
- Species-specific sequences from species databases e.g., the International Barcode of Life at <http://ibol.org/>;
- 16S or 18S RNA datasets, e.g., the Earth Microbiome Project at <http://www.microbio.me/emp/>;

- Records of geo-referenced species occurrences, e.g., the Global Biodiversity Information Facility at [www.GBIF.org](http://www.GBIF.org);

Responses to this RFI may address any or all of the following questions:

- 1) Relying just on existing ecologic, biologic or other databases, public or private, is it possible to determine geospecificity of metagenomic sequences? If yes, why do you think so, and what would be your technical approach? If not, why not? What are the critical gaps in content and/or quality? How could these gaps be minimized or mitigated?
- 2) Which, if any, existing databases provide species-level identity with geographic and temporal metadata? Which existing databases provide species-level identity based on annotated sequences? What are the limitations of these datasets for geospecificity?
- 3) What ecological models currently exist that attempt to predict species or genome range on earth's land surfaces? How accurate are these models and what are their limitations?
- 4) What techniques or practices could dramatically improve the quality of existing datasets so that they would be more useful to studies of geospecificity? Which database elements require additional standardization or data? What standardization efforts are underway?
- 5) Citizen scientist projects worldwide, e.g., Wild Lives of Our Homes (<http://homes.yourwildlife.org/>), Earth Microbiome Project (<http://www.earthmicrobiome.org/>), Microbiology of the Built Environment (<http://microbe.net/>), and the American Gut Project (<http://americangut.org/>), are anticipated to expand. How is quality assured for these databases? Is quality sufficient for scientific purposes? Are there novel bioinformatics techniques that could compensate for some amount of inaccuracy in the data?

## **Preparation Instructions to Respondents**

IARPA requests that respondents submit ideas related to this topic for use by the Government in formulating a potential program. IARPA requests that submittals briefly and clearly describe the potential approach or concept, outline critical technical issues/obstacles, describe how the approach may address those issues/obstacles and comment on the expected performance and robustness of the proposed approach. If appropriate, respondents may also choose to provide a non-proprietary rough order of magnitude (ROM) regarding what such approaches might require in terms of funding and other resources for one or more years. This announcement contains all of the information required to submit a response. No additional forms, kits, or other materials are needed.

IARPA appreciates responses from all capable and qualified sources from within and outside of the U.S. Because IARPA is interested in an integrated approach, responses from teams with complementary areas of expertise are encouraged.

Responses have the following formatting requirements:

1. A one page cover sheet that identifies the title, organization(s), respondent's technical and administrative points of contact - including names, addresses, phone and fax numbers, and email addresses of all co-authors, and clearly indicating its association with **RFI-14-07**;

2. A substantive, focused, one-half page executive summary;
3. A description (limited to 5 pages in minimum 12 point Times New Roman font, appropriate for single-sided, single-spaced 8.5 by 11 inch paper, with 1-inch margins) of the technical challenges and suggested approach(es);
4. A list of citations (any significant claims or reports of success must be accompanied by citations, and reference material **MUST** be attached);
5. Optionally, a single overview briefing chart graphically depicting the key ideas.

## **Submission Instructions to Respondents**

Responses to this RFI are due no later than 4:00pm, Local Time, College Park, MD on 14 July 2014. All submissions must be electronically submitted to *dni-iarpa-RFI-14-07@iarpa.gov* as a PDF document. Inquiries to this RFI must be submitted to *dni-iarpa-RFI-14-07@iarpa.gov*. Do not send questions with proprietary content. No telephone inquiries will be accepted.

## **DISCLAIMERS AND IMPORTANT NOTES**

This is an RFI issued solely for information and new program planning purposes and does not constitute a solicitation. Respondents are advised that IARPA is under no obligation to acknowledge receipt of the information received, or provide feedback to respondents with respect to any information submitted under this RFI.

Responses to this notice are not offers and cannot be accepted by the Government to form a binding contract. Respondents are solely responsible for all expenses associated with responding to this RFI. IARPA will not provide reimbursement for costs incurred in responding to this RFI or reimbursement for travel. It is the respondent's responsibility to ensure that the submitted material has been approved for public release by the information owner.

The Government does not intend to award a contract on the basis of this RFI or to otherwise pay for the information solicited, nor is the Government obligated to issue a solicitation based on responses received. Neither proprietary nor classified concepts or information should be included in the submittal. Input on technical aspects of the responses may be solicited by IARPA from non-Government consultants/experts who are bound by appropriate non-disclosure requirements.

### **For information contact:**

Susan Alexander

Intelligence Advanced Research Projects Activity

*dni-iarpa-RFI-14-07@iarpa.gov*