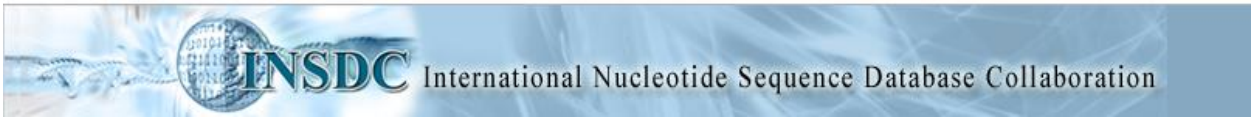

June 1 2019

Dr. Cristiana Paşca Palmer
Executive Secretary
Convention on Biological Diversity
United Nations Environment Programme
413 Saint-Jacques Street, Suite 800
Montréal, Québec, Canada H2Y 1N9

Dear Dr. Paşca Palmer:

Please find the response to the Secretariat's 5 February 2019 Notification No. 2019-012 regarding submission of views and information pursuant to paragraphs 9 of decision XIIII/20 on digital sequence information on genetic resources from the International Nucleotide Sequence Database Collaboration (INSDC) We appreciate the opportunity to contribute.

Thank you for your consideration.

Sincerely,

Ilene Mizrachi on behalf of the INSDC

National Center for Biotechnology Information
National Library of Medicine
National Institutes of Health, USA

Response from International Nucleotide Sequence Database Collaboration (INSDC) to CDB call for views and information on Digital Sequence Information on Genetic Resources

Ref.: SCBD/NPU/DC/VN/KG/RKi/87804, https://www.cbd.int/doc/notifications/2019/ntf-2019-012-abs-en.pdf

Key points

- INSDC databases provide the long-established and broadly adopted data infrastructure for the open sharing of sequences
- Open access to data is a central element of INSDC operations and lies under the governance of governments, institutions and scientific peers
- Submission to INSDC databases is an essential scientific process that adds value and drives knowledge generation from the entire corpus of data
- Use of INSDC databases is embedded in the scientific process, both for data submission and data access
- INSDC provides the effective free and open data infrastructure required to enable biodiversity-related Access and Benefit Sharing

Terminology and Scope

The International Nucleotide Sequence Database Collaboration (INSDC; http://www.insdc.org/) has for four decades captured, curated and preserved the world's output of DNA and RNA sequence data, including the sequences themselves (ordered strings of letters that represent the order of the nucleotides on chromosomes) and annotations thereon (such as indications of genes and their functions).

These data are variously referred to as "Digital Sequence Information; DSI" and "Genetic Sequence Data; GSD". These are data and do not constitute genetic material; rather they are data that describe genetic material.

Access to GSD and benefit-sharing

INSDC is a partnership between three major institutions, the National institute of Genetics' DNA Data Bank of Japan (DDBJ), the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) and the US National Institutes of Health's National Library of Medicine, National Center for Biotechnology Information (NCBI), who provide globally comprehensive coverage of open access data assured through data exchange systems and their supporting technical data standards. The open access policy (see below for details) is dictated by governments, funding sources, institutional governance and the INSDC International Advisory Committee, for the use of scientists worldwide. INSDC databases do not set these policies and could not make changes to suit any restrictions on open access that might be imposed by the Convention or the Nagoya Protocol.

The process of data submission into an INSDC database is one in which the data provider structures and describes the new data set in a standard manner, such that it can be integrated into the entire corpus of INSDC content. Data providers are encouraged to provide contextual metadata which describes when and where the biological sample was obtained and where it is stored, if applicable (for example, a museum voucher). Importantly, it is through this integrated corpus of data that the fullest value can be

extracted; few scientific interpretations rely on a single sequence and most rely on the aggregate of many different sequences considered together.

Submission to INSDC databases is broadly accepted by the scientific community as an essential part of the process of carrying out scientific work. Mandatory submission to INSDC is required prior to, or at the time of, publication in the academic literature in the vast majority of life science journals covering the life sciences. Equally, use of the data within INSDC databases is an accepted, and often daily, norm for life scientists. INSDC has thus long been adopted by the community as the system for managing sequence data.

INSDC is supported mainly by the host governments to promote and support open science primarily for human genetic research, food security and public health. Biological research, including biodiversity research, also benefits from this investment. While sequences that are CBD-relevant (in terms of legal scope) are substantial, but a minor part of the database (relative to human genome sequences or model organisms), these sequences play an essential enabling role in meeting the first two objectives of the CBD, the conservation of biological diversity and the sustainable use of its components.

Open Access in INSDC

Services provided by INSDC databases are free of charge to all users, both those submitting data and those searching and retrieving data.

Data presented in INSDC are open access and available freely across a host of online services, such as web sites, search tools, programmatic interfaces and FTP sites. It is only with open access that data can be fully integrated to support interpretation by original providers and the broader scientific community.

The INSDC partner databases place no conditions on the use and redistribution of data and records remain within the ownership of the data submitters; INSDC database are the hosts of the data. INSDC databases Terms of Use state that the databases provide data openly, but note that data owners or third parties may assert conditions on the use or application of sequences relating to rights such as intellectual property and access and benefit sharing (see https://www.ddbj.nig.ac.jp/policies-e.html#credit, https://www.ebi.ac.uk/about/terms-of-use, https://www.ncbi.nlm.nih.gov/home/about/policies/#data)

The INSDC partners believe that open access data are a key component of any successful Access and Benefit Sharing system. Without open access and data integration, the value (and ultimate benefit) from the data will be significantly reduced. Regulation of GSD would increase barriers to innovation and discovery which would affect attaining the goals of CBD as well as impacting food safety and health worldwide.

Illustrative Examples - Benefits of Open Genetic Sequence Data

Genomic sequence analysis is being used in the public health arena. In late 2017, there was an outbreak of Salmonella in the US. The genomes of bacteria isolated from clinical samples were sequenced and analyzed. Genomes were clustered to determine which are most closely related and potentially derived from a single source. The sequences from this outbreak clustered with a Salmonella genome isolated from a papaya which originated in Mexico (https://www.cdc.gov/salmonella/kiambu-07-17/). Open sequence data allowed public health scientists to determine the cause of the outbreak and prevent additional illness.

Open genetic sequence data enabled the discovery of fish sold in markets and restaurants may be mislabeled. DNA sequence can distinguish different species that may look very similar to the untrained eye. The DNA sequence of a gene from restaurant and fish market samples were compared with fish gene sequences in INSDC to determine the species of the sample. The species indicated was not always the species of fish that was detected by DNA sequencing. For example: (1) critically endangered shark species were being sold as different types of fish (https://www.nature.com/articles/s41598-018-38270-3) (2) less expensive fish were sold as more expensive ones (https://onlinelibrary.wiley.com/doi/full/10.1111/cobi.12888). This fraud would not have been uncovered without open GSD.

Across INSDC, we enable access and share the added value of integrated data to all. We believe that keeping GSD access open would enable more access and benefit sharing, not less.

Signatories

Ilene Karsch-Mizrachi (NCBI)
Guy Cochrane (EMBL-EBI)
Yasukazu Nakamura (DDBJ)
on behalf of the INSDC.