“Digital Sequence Information” and non-monetary benefit-sharing

Chris Lyal

NATURAL HISTORY MUSEUM

Scope of my discussion – study area

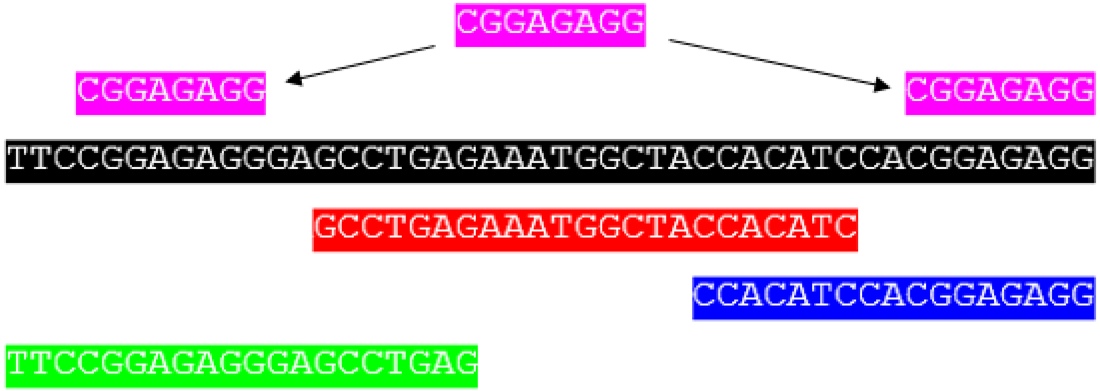
- Focus on information generated and used by taxonomy
  - Particularly in a non-commercial context
- This sector generates huge volumes of dsi

Global Taxonomy Initiative

Scope of my discussion – “DSI”

All of this is a personal understanding only!

- Sequence reads
- Sequence assembly (putting the reads together)



CGGAGAGG

CGGAGAGG CGGAGAGG

TTCCGGAGAGGGAGCCTGAGAAATGGCTACCACATCCACGGAGAGG

GCCTGAGAAATGGCTACCACATC

CCACATCCACGGAGAGG

TTCCGGAGAGGGAGCCTGAG

Scope of my discussion – “DSI”

- Sequences from DNA Barcodes to genomes
- Including RNA, “Junk” DNA etc.

Scope of my discussion – “DSI”

Gene functionality and structure

- More rarely used in taxonomy (other than bacteria)
- Considerable importance for conservation & sustainable use

McKenna *et al. Genome Biology* (2016) 17:227
DOI 10.1186/s13059-016-1088-8

Genome Biology

RESEARCH **Open Access**

CrossMark

Genome of the Asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle–plant interface

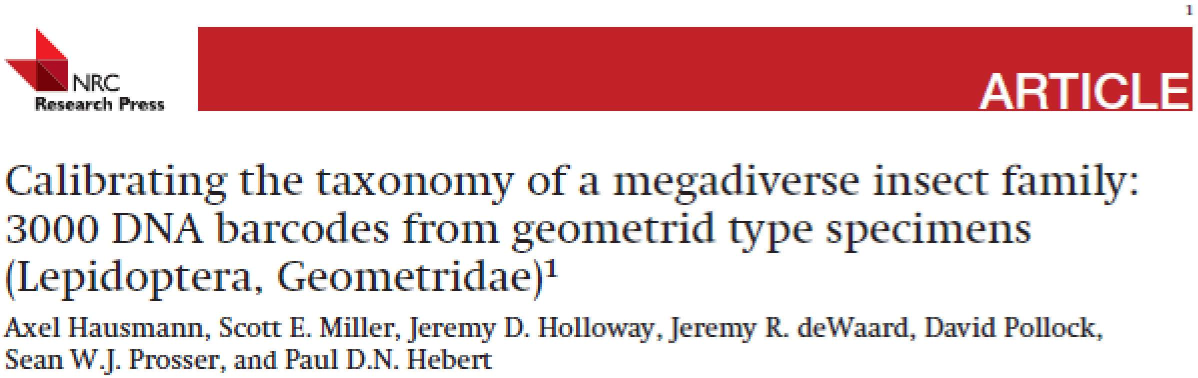
Where do taxonomists obtain sequence information?

- Generated during research
  - from recent GR accessed with PIC & MAT
  - from older (pre 1993 / 1962) specimens in collections
- In-house databases
  - Developed from earlier sequencing activities
- Public databases, particularly:
  - International Nucleotide Sequence Database Collaboration (INSDC)
    - DNA Data Bank of Japan (DDBJ)
    - European Bioinformatics Institute (EMBL-EBI)
    - National Center for Biotechnology Information (NCBII) (Genbank)
  - Barcode of Life Data System (BOLD)

What do taxonomists do with DSI?

Publication

- Research almost always intended for publication
- Standard condition of publication: data are made available
- So sequences placed on BOLD / INSCD etc.
  - These built on principle of open access



1

NRC Research Press

ARTICLE

Calibrating the taxonomy of a megadiverse insect family: 3000 DNA barcodes from geometrid type specimens (Lepidoptera, Geometridae)[1]

Axel Hausmann, Scott E. Miller, Jeremy D. Holloway, Jeremy R. deWaard, David Pollock, Sean W.J. Prosser, and Paul D.N. Hebert

What do taxonomists do with DSI?

Exchange Information

- Data formats used to exchange information about specimens / samples / sequences
- Elements now available to manage permit data

ACCESS & BENEFIT-SHARING

ABS

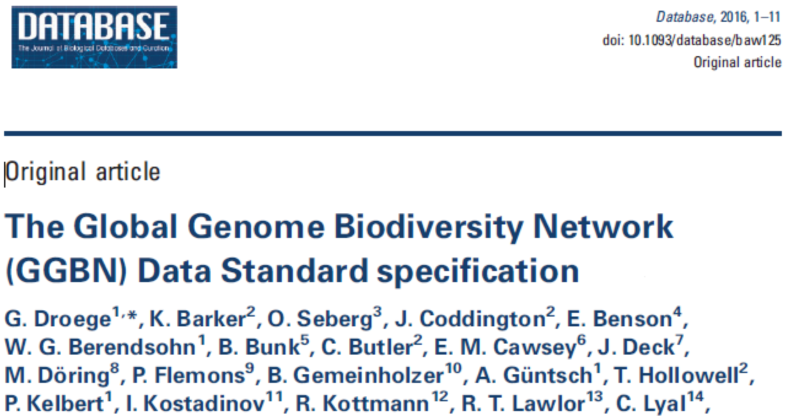
CODE OF CONDUCT & BEST PRACTICES

CETAF

CONSORTIUM OF EUROPEAN TAXONOMIC FACILITIES

Exploring and Documenting Diversity in Nature

includes "MATERIAL TRANSFER AGREEMENTS" TEMPLATES



DATABASE

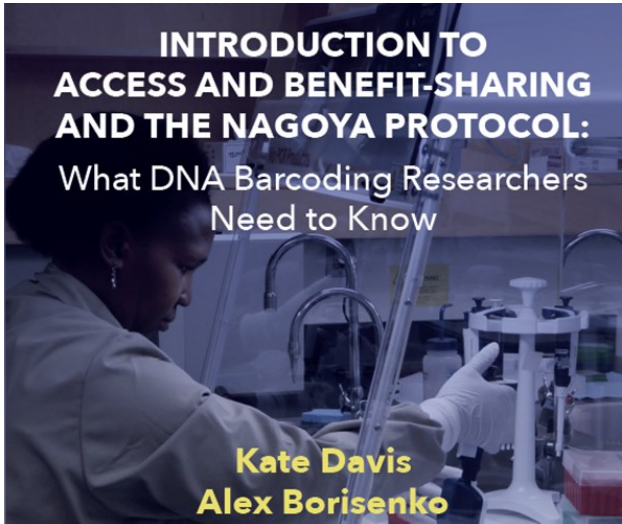
The Journal of Biological Databases and Curation

Database, 2016, 1–11
doi: 10.1093/database/baw125
Original article

Original article

The Global Genome Biodiversity Network (GGBN) Data Standard specification

G. Droege[1,*], K. Barker[2], O. Seberg[3], J. Coddington[2], E. Benson[4], W. G. Berendsohn[1], B. Bunk[5], C. Butler[2], E. M. Cawsey[6], J. Deck[7], M. Döring[8], P. Flemons[9], B. Gemeinholzer[10], A. Güntsch[1], T. Hollowell[2], P. Kelbert[1], I. Kostadinov[11], R. Kottmann[12], R. T. Lawlor[13], C. Lyal[14],



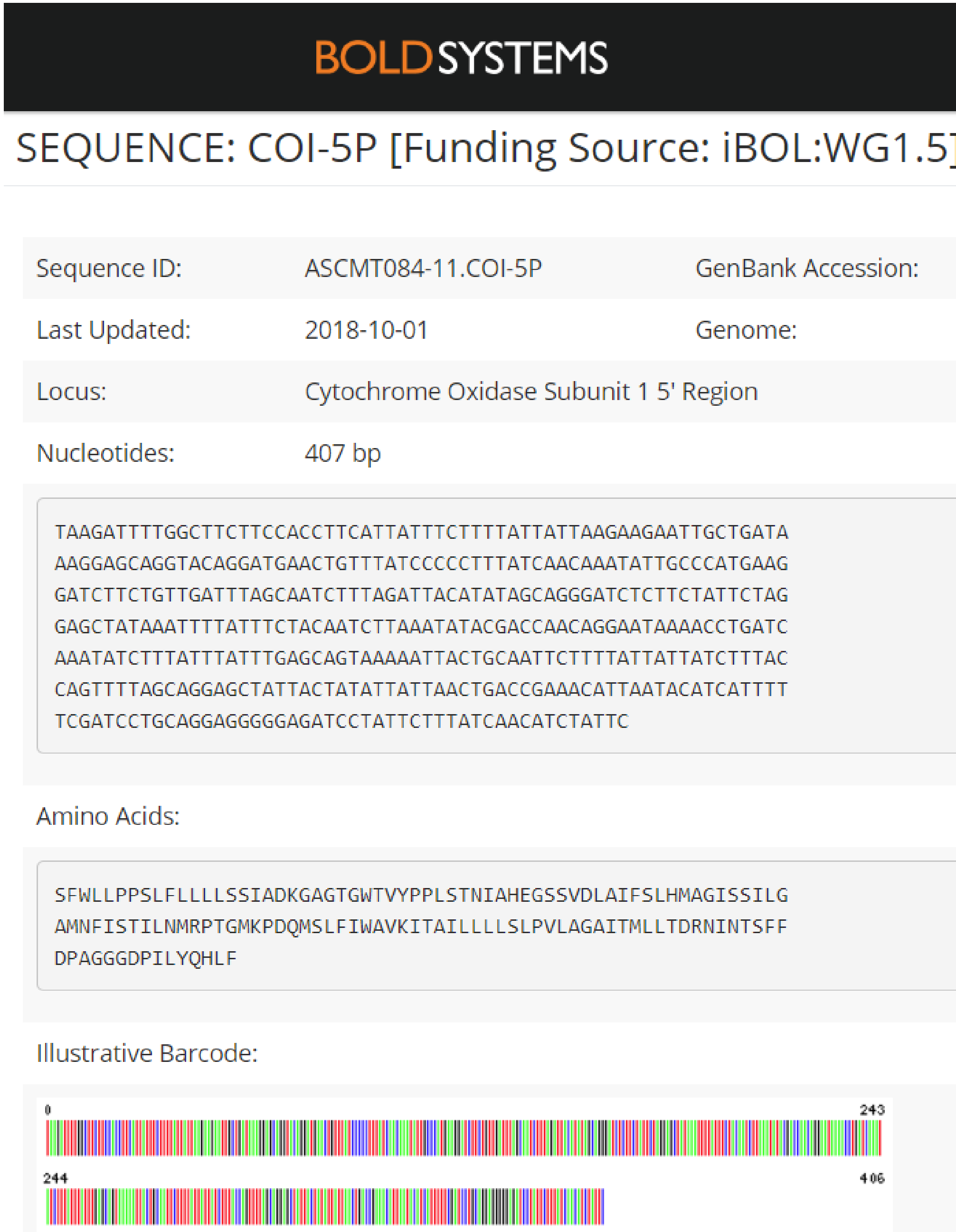
INTRODUCTION TO ACCESS AND BENEFIT-SHARING AND THE NAGOYA PROTOCOL:

What DNA Barcoding Researchers Need to Know

Kate Davis
Alex Borisenko

How do we use DSI?

- Digital Sequence Information
  - Basic step: Comparison of those sequences with others



SEQUENCE: COI-5P [Funding Source: iBOL:WG1.5

Sequence ID:	ASCMT084-11.COI-5P	GenBank Accession:
Last Updated:	2018-10-01	Genome:
Locus:	Cytochrome Oxidase Subunit 1 5' Region	
Nucleotides:	407 bp	

```
TAAGATTTTGGCTTCTTCCACCTTCATTATTTCTTTTATTATTAAGAAGAATTGCTGATA
AAGGAGCAGGTACAGGATGAACTGTTTATCCCCCTTTATCAACAAATATTGCCCATGAAG
GATCTTCTGTTGATTTAGCAATCTTTAGATTACATATAGCAGGGATCTCTTCTATTCTAG
GAGCTATAAATTTTATTTCTACAATCTTAAATATACGACCAACAGGAATAAAACCTGATC
AAATATCTTTATTTATTTGAGCAGTAAAAATTACTGCAATTCTTTTATTATTATCTTTAC
CAGTTTTAGCAGGAGCTATTACTATATTATTAACTGACCGAAACATTAATACATCATTTT
TCGATCCTGCAGGAGGGGGAGATCCTATTCTTTATCAACATCTATTC
```

Amino Acids:

```
SFWLLPPSLFLLLLSSIADKGAGTGWTVYPPLSTNIAHEGSSVDLAIFSLHMAGISSILG
AMNFISTILNMRPTGMKPDQMSLFIWAVKITAILLLLSLPVLAGAITMLLTDRNINTSFF
DPAGGGDPILYQHLF
```

Illustrative Barcode:

What do taxonomists do with DSI?

- Identification
  - Increasing use of ‘DNA barcodes’ - COI gene is ‘standard’ for many animals
  - Run ‘BLAST’ search on BOLD
    - finds regions of similarity between biological sequences
    - Potentially could read all sequences in the database
    - A match suggests an identification
    - Increasingly used for pests, invasive species etc
  - CBD (GTI) through Japan Fund has funded training Barcode use for relevant personnel from a number Parties

Global
Taxonomy
Initiative

DSI and CBD implementation

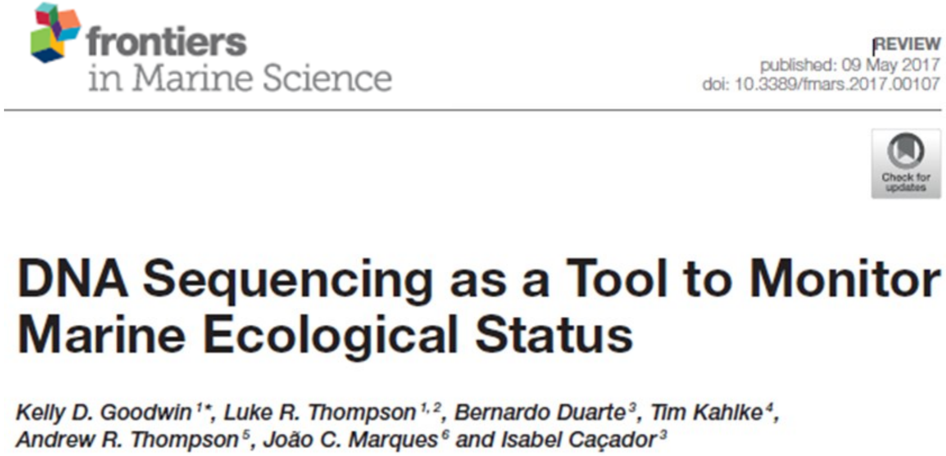
- Rapid species identification
  - Use of DNA barcode libraries in particular facilitates this.
- Particularly important for samples where rapid response is needed
  - Invasive species
  - Quarantine
  - Pest detection

Global Taxonomy Initiative

What do taxonomists (and others) do with DSI? – e-DNA

- Check for presence or absence of endangered or invasive species
- Detect unknown species
- Assess overall diversity of an environment

- Increasingly important tool for environmental management



frontiers in Marine Science

REVIEW
published: 09 May 2017
doi: 10.3389/fmars.2017.00107

DNA Sequencing as a Tool to Monitor Marine Ecological Status

Kelly D. Goodwin[1], Luke R. Thompson[1,2], Bernardo Duarte[3], Tim Kahlke[4], Andrew R. Thompson[5], João C. Marques[6] and Isabel Caçador[3]*

Environmental Science & Technology

Sampling of Environmental DNA

What do taxonomists do with DSI?

- Phylogenetic analysis
  - Use multiple genes (different genes evolve at different rates)
  - Use genomes where possible
  - Typically from many countries, collected over many years
  - Analysis may include many species, increasingly in hundreds
  - With standardisation of sequencing methodology, downloaded sequences increasingly useful

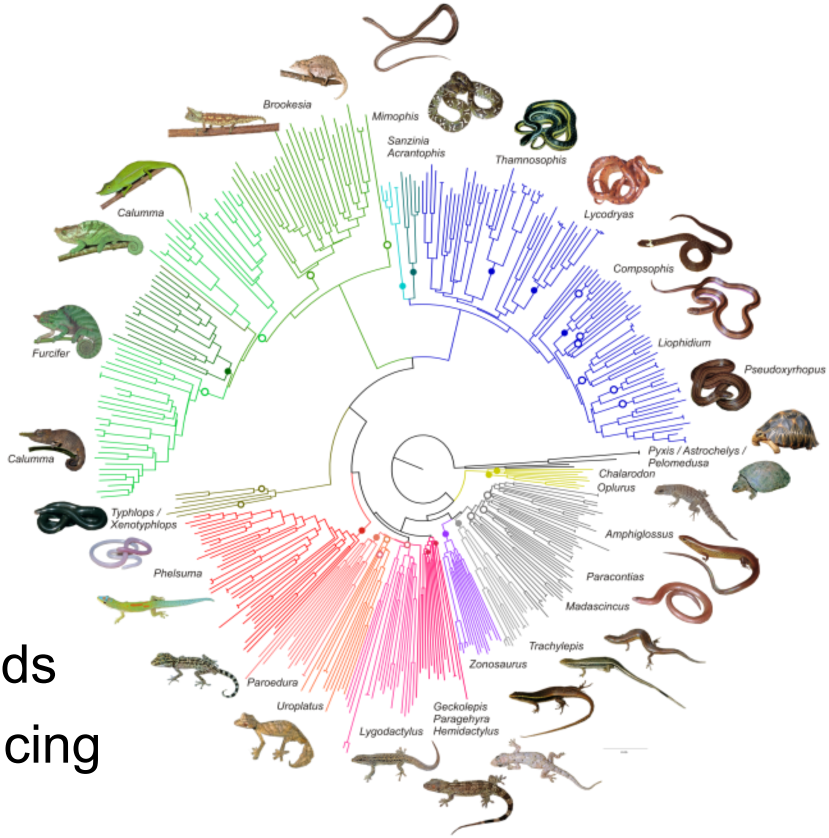
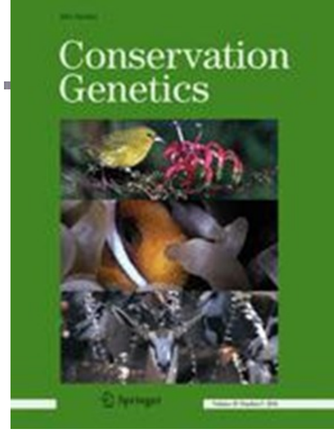


Image from: Nagy ZT, Sonet G, Glaw F, Vences M (2012) First Large-Scale DNA Barcoding Assessment of Reptiles in the Biodiversity Hotspot of Madagascar, Based on Newly Designed COI Primers. PLoS ONE 7(3): e34506.

DSI and CBD implementation

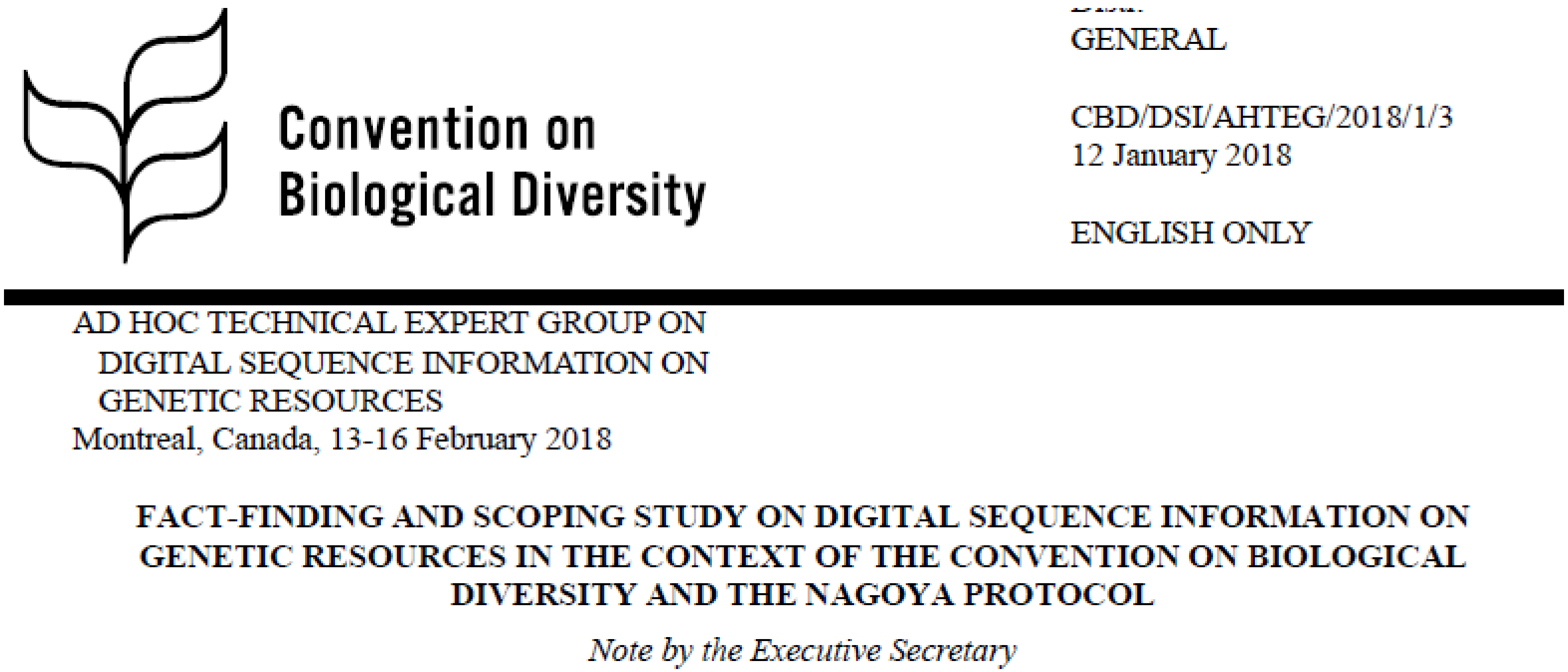


More than 1000 scientific papers were published 1/1/2016 – 1/7/2017

- Phylogenetic diversity across landscapes to assist large scale conservation planning
- Genetic structure of populations / species for effective conservation management of genetic diversity
- Use for identification when other methods inapplicable (e.g. invasive species; detection of protected species; species in trade).
- Understanding migrations and effect on conservation measures
- Effects of harvest rates on genetic diversity
- Changes in genetic diversity over time and potential causal factors
- Targeted sampling of *ex situ* collections to support *in situ* action
- Pollinator conservation
- Role of genes in plant resilience to environmental change.

DSI and CBD implementation

- Wide range of support for implementation identified in Report to CBD on DSI



Convention on
Biological Diversity

GENERAL

CBD/DSI/AHTEG/2018/1/3
12 January 2018

ENGLISH ONLY

AD HOC TECHNICAL EXPERT GROUP ON
DIGITAL SEQUENCE INFORMATION ON
GENETIC RESOURCES
Montreal, Canada, 13-16 February 2018

FACT-FINDING AND SCOPING STUDY ON DIGITAL SEQUENCE INFORMATION ON GENETIC RESOURCES IN THE CONTEXT OF THE CONVENTION ON BIOLOGICAL DIVERSITY AND THE NAGOYA PROTOCOL

Note by the Executive Secretary

DSI and CBD implementation

- No country holds sequence data for all of its biota
  - or of species likely to be intercepted by quarantine (Alien Species, pests etc.)
- Parties can only obtain the widest range of sequence data for supporting CBD implementation through global databases.

DSI and CBD implementation

Article 17. Exchange of Information

1. *The Contracting Parties shall facilitate the exchange of information, from all publicly available sources, relevant to the conservation and sustainable use of biological diversity, taking into account the special needs of developing countries.*
2. *Such exchange of information shall include exchange of results of technical, scientific and socio-economic research, …. It shall also, where feasible, include repatriation of information.*

DSI and CBD implementation

CBD COP repeatedly called for greater access to information of many types, including genetic information.

- Global Taxonomy Initiative Program of Work (COP VI/8) asks for a Global Taxonomic Information System
- Also Strategic Plan for Biodiversity 2011-2020 (XIII/31), Aichi goals C (Target 13) and E (Target 19), etc.

DSI and CBD implementation

Non-monetary benefits supporting conservation and sustainable use of biological diversity - Information sharing

- Sharing of research and development results;
- Admittance to ex situ facilities of genetic resources and to databases;
- Transfer of knowledge and technology under fair and most favourable terms;
- Access to scientific information;
- Research directed towards priority needs.

Application of PIC and MAT to DSI?

Could a system handle bilateral agreements?

- International Nucleotide Sequence Database Collaboration (INSDC)
  - Share data daily
  - hold quadrillions of nucleotides of DNA sequences
    - (1 quadrillion = 1,000,000,000,000,000)
  - 209,656,636 sequences in October 2018
  - collected from over 300,000 organisms
  - EMBL-EBI search engine: ca.12.6m user jobs every month
  - scientists at over 5m sites use EMBL-EBI websites every month;
  - more than 27m requests made every weekday to EMBL-EBI
- BOLD currently holds 1.7m public records of COI gene

Application of PIC and MAT to DSI?

- Very large number of transactions;
- Often very low incremental value of a single sequence to the research
- Risk that requiring separate PIC and MAT halts use
  - Reducing or eliminating non-monetary benefits
  - And acting directly against CBD implementation

Global Taxonomic Information System

Global Taxonomic Information System

- Still a need to increase capacity to make use of these databases and their content
  - Including adapting workflows

However:

- Public databases providing open access to dsi and other data types are in place and used across the globe
- Information pipelines to improve delivery from there are being improved and often linked to analytical tools
- This is a means of delivering information & sharing non-monetary benefits globally as part of the Global Taxonomic Information System