



DSI and the conservation and sustainable use of GRFA – an overview

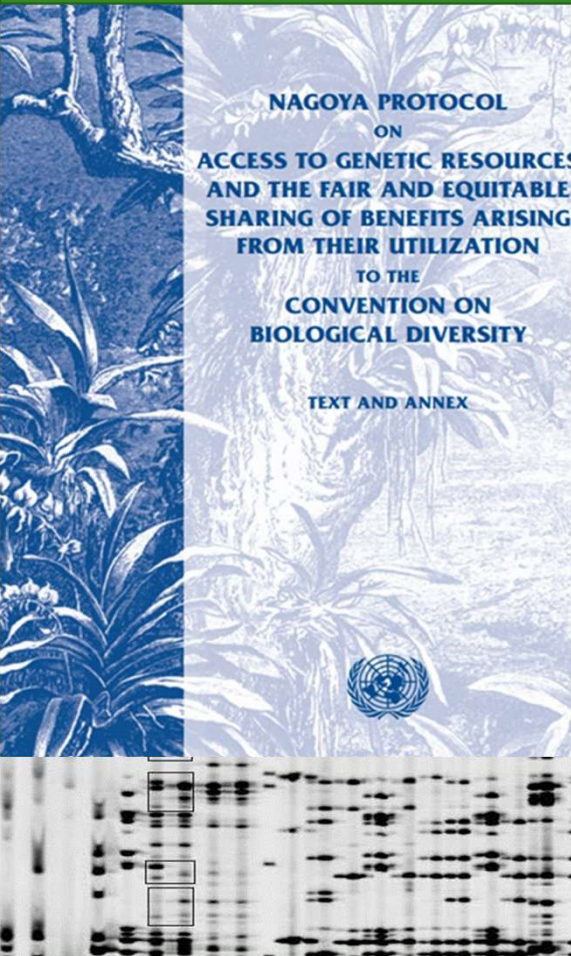
Digital sequence information and genetic resources for food and agriculture workshop

FAO, CBD, ITPGRFA, CABI and the CGIAR Genebank Initiative

14-15 November 2022

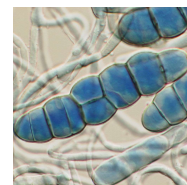
David Smith, Matthew J. Ryan and Alan G. Buddie, CAB International, United Kingdom

Study on DSI in the context of GRFA



Overview

1. Introduction – what is DSI and study methodology
2. Relevance of DSI
3. Generation and storage of DSI
4. The state of management of DSI
5. Access to and use of DSI
6. The regulatory environment





DSI Study methodology

- CBD outputs and commissioned reports
- Literature survey CAB Abstracts database 10.9 million literature records
- Comparison of outputs with PubMed and the DSI Network WiLDSI data portal (IPK Gatersleben, and Leibniz-Institute DSMZ, Germany)
- Survey of CABI member countries
- Country National Focal Points, National Authorities (only 26 returns from 157 contacts)

Definition and scope of DSI - Ad Hoc Technical Expert Group (AHTEG)

	Information related to a genetic resource			
	Genetic and biochemical information			Associated information
Group reference	Group 1	Group 2	Group 3	
High-level description of each group	DNA and RNA	Group 1 + proteins + epigenetic modifications	Group 2 + metabolites and other macromolecules	
Examples of granular subject matter	<ul style="list-style-type: none"> • Nucleic acid sequences; • Associated data; • Non-coding NA sequences; • Genotyping, microsatellite analysis, SNPs, etc.); • Structural annotation 	<ul style="list-style-type: none"> • Amino acid sequences; • gene expression; • Functional annotation; • Epigenetic modifications • Molecular structures of proteins; • Molecular interaction networks. 	<ul style="list-style-type: none"> • biochemical composition; • Macromolecules (other than DNA, RNA and proteins); • Cellular metabolites (molecular structures). 	<ul style="list-style-type: none"> • Traditional knowledge on GR • biotic and abiotic factors in the environment or associated with the organism • Other information associated with GR or its utilization.

- Term coined in the context of the CBD
- Debated in several fora:
 - United Nations Convention on the Law of the Sea
 - World Intellectual Property Organization
 - World Health Organisation
- Countries see the issue differently
- Current recommendations are not unanimously accepted.

Methods for analysing DSI and their citations in Food and Agriculture

Genome Sequencing (AHTEG Group 1 DSI)

Method	Search hits
De Novo Sequencing	9,877
Whole-Genome Resequencing	8,469
Transcriptome Sequencing, gene expression	32,945
Epigenetics	10,992
Targeted Resequencing	2,558
Genotyping by Sequencing (GBS)	16,213
Environmental DNA Sequencing	483
Genome editing e.g. CRISPR	5118

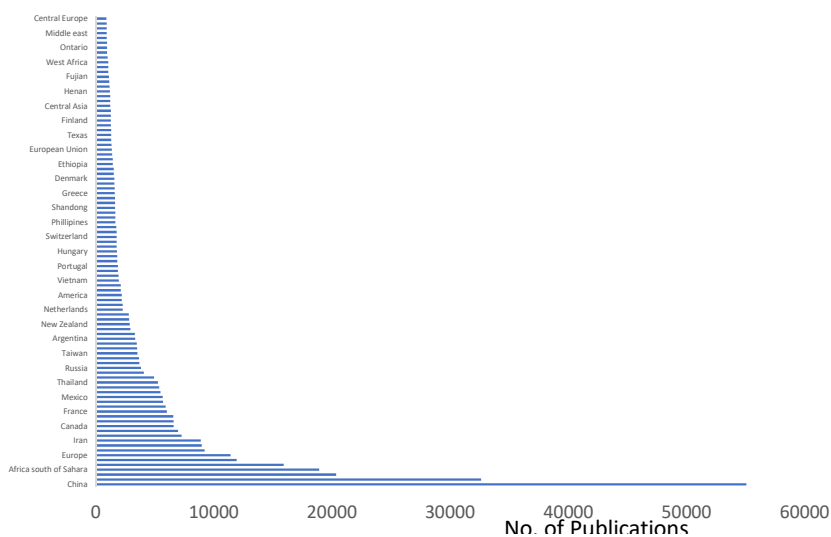
Proteins and epigenetic modifications (AHTEG Group 2 additions)

Method	Search Hits
Mass Spectrometry, gas- or liquid chromatography	13,887
Enzyme-linked immunosorbent assay - ELISA	24,332
Gel electrophoresis, Separation of DNA, RNA, protein	23,546
Chromatography	19,665
Protein microarrays	98

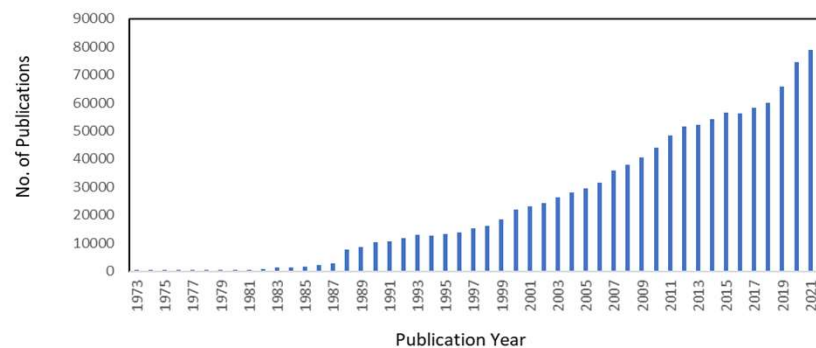
Metabolomic information (AHTEG Group 3 additions)

Method	Search hits
Matrix-assisted laser desorption/ionization coupled to time-of-flight mass spectrometry	2,813
Nuclear magnetic resonance (NMR) spectroscopy	3,130
High-resolution mass spectrometry (HRMS)	265

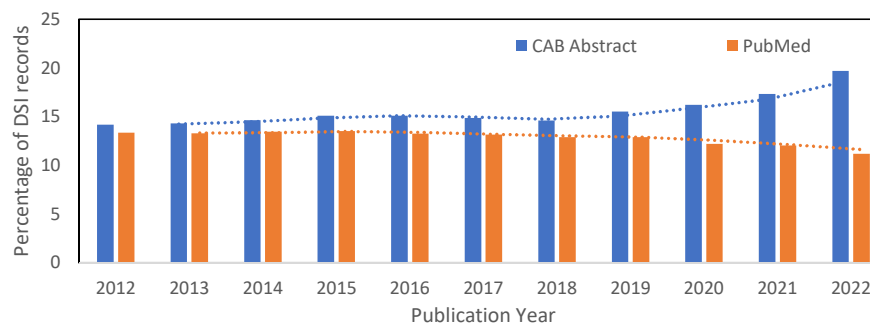
Number of publications in CAB Abstracts database from 1973 to 2022 that address DSI



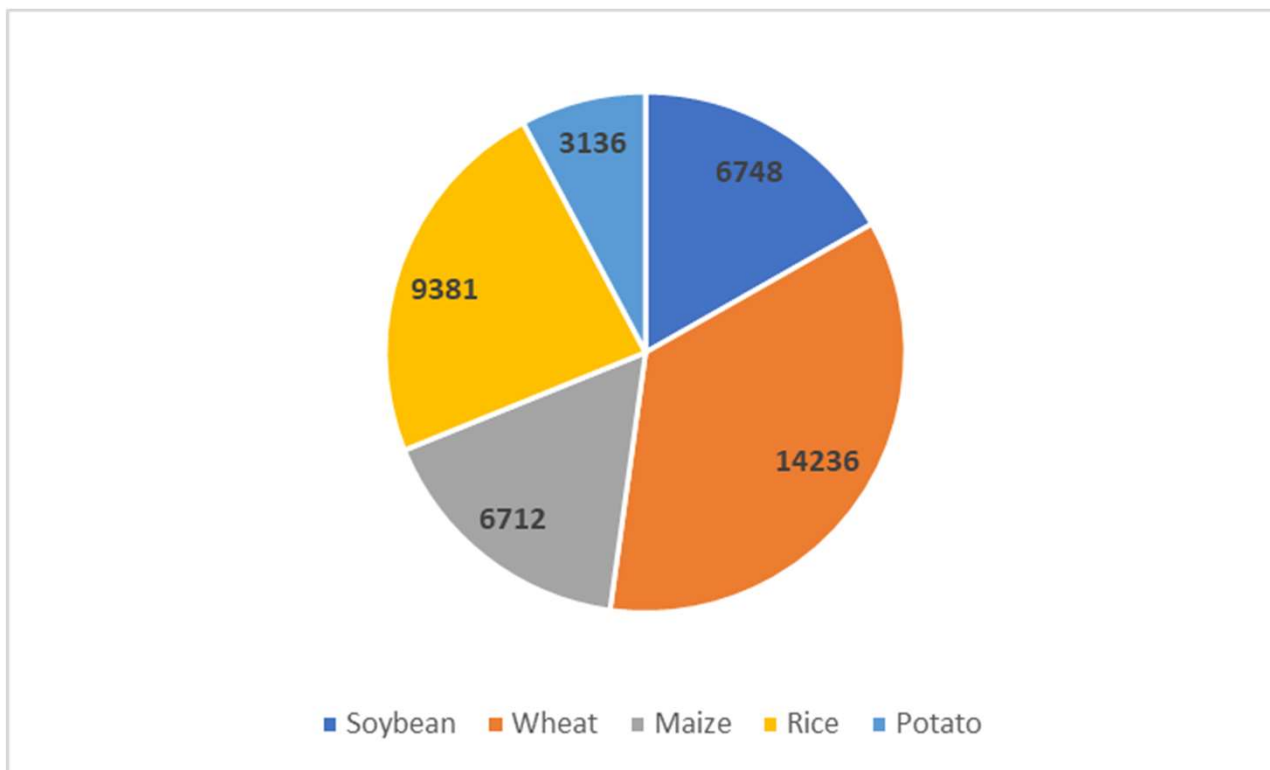
Distribution of publications related to DSI over different geographical location



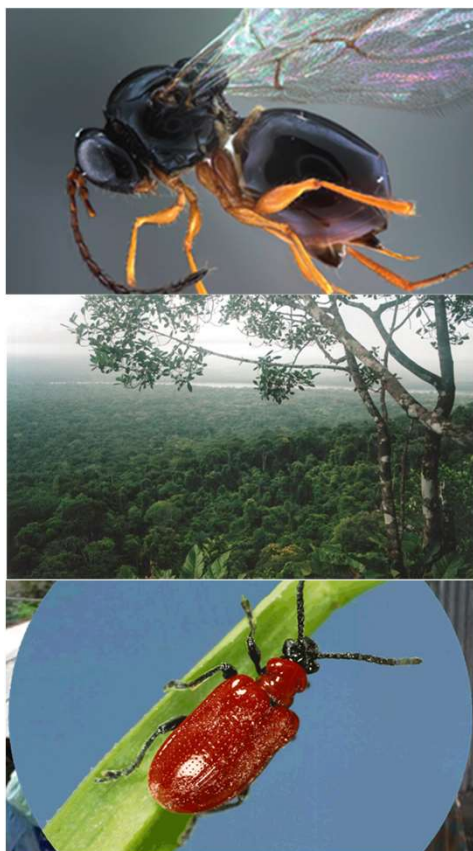
CABI and PubMed compared



The number of literature records for Soybean, Wheat, Maize, Rice and Potato referencing DSI in the CAB Abstracts database



A deeper look into the CABI Literature survey



Animal Genetic Resources

- a better understanding of an organism's traits for use in breeding programmes
- discovery and development of new livestock breeds

Forest Genetic Resources

- contributing to the assembly of breeding populations
- selecting genetic material for storage or micropropagation
- judicious control of pest infestation
- monitoring changes due to climate change

Insect genetic resources

- bio-surveillance of forest and agricultural pests
- the online data base BOLD has allowed the automatic identification of species, based on their DNA barcodes

A deeper look into the CABI Literature survey



Microbial genetic resources

- barcoding and whole genome sequencing are required to detect and identify the organisms yet to be cultured (99%)
- express properties through chemically synthesized genomes
- metabolomics detects and quantifies 1000s of compounds
- enables exploration of microbial communities that surround plants, animals and the environment
- studies of impact of agricultural practices

Plant genetic resources

- used to improve crop yields, resistance to disease, manage threats to biodiversity such as invasive species and mitigation of climate change
- identify and select for heritable alleles that control crop traits independent of underlying genotype
- enhanced the ability to identify plant specialized metabolic pathways to improve abiotic and biotic stress resistance

A deeper look into the CABI Literature survey

- Number of hits for the top crops in the FAO regions

FAO Regions	No. of hits in DSI pool	Top crops [common name] (no. of hits)
Africa	26,981	Wheat (967); Maize (743); Rice (313); Soybean (266); Potato (245)
Asia and the Pacific	128,783	Rice (8841); Wheat (5945); Soybean (2906); Maize (1840); Potato (925)
Europe	53,811	Wheat (2179); Maize (1193); Potato (730)
Latin America and the Caribbean	28,438	Soybean (1403); Maize (1005); Wheat (791); Potato (389)
Middle and 'Near' East	22,166	Wheat (1792); Rice (227); Potato (217)
North America	45,946	Wheat (2562); Soybean (2173); Maize (1931); Potato (630)

Relevance of DSI - Benefits to be shared from DSI studies

Activity	Benefits to be shared
Identification of biodiversity and its conservation and characterization	Contributes to biodiversity inventories and allows monitoring of biodiversity change and conservation
Diagnosis and identification of pests and diseases	Enabling appropriate management of threats to improve yields
Rapid identification of newly introduced (invasive) alien species	Early warning to facilitate containment and management; reducing losses
Studies to assess impact of land use and climate change on biodiversity and ecosystem services	Monitoring of soil health that can result in improved agricultural production and mitigation of climate change
Developing microbial solutions to improve health and nutrition security	identification of properties to improve yields and reduction in losses; monetary benefits arise from products developed
Developing biological control agents (BCAs)	Management of invasive species, reduction of crop losses and minimization of unnecessary pesticide use
Increasing and improving access to agricultural and environmental scientific knowledge	Sequencing DNA, RNA, proteins and metabolome of species adds to the knowledge base for innovation and discovery

Storage of DSI in Public databases

The main resources for storing and distributing sequence data are:

>205BACT

CTAGCACTTGTTCTTCCCTAACACAGAGT
TTTACGACCCGAAAGCCTTCATCACTCACG

CGGCGTTGCTCCGTCAGACTTTCGTCCATT
GCGGAAGATTCCCTACTGCTGCCTCCCGTA

GGAGTCTGGGCCGTGTCTCAGTCCCAGTGT
GGCCGATCACCTCTCAGGTCGGCTACGCA

TCGTTGCCTTGGTGAGCCGTTACCTCACCA
ACTAGCTAATGCGACGCGGGTCCATCCATA

AGTGACAGCCGAAGCCGCCTTTCAATTTTCG
AACCATGCAGTTCAAAATGTTATCCGGTAT

TAGCCCCGGTTTCCCGGAGTTATCCCAGTC
TTATGGGCAGGTTACCCACGTGTTACTCAC

CCGTCCGCCGCTAACTTCTTGAGAGCAAGC
TCTCAATCCATTTCGCTCGACTTGCATGTAT

TAGGCACGCCGCCAGCGTTCATCCTGAGCC
ATGA

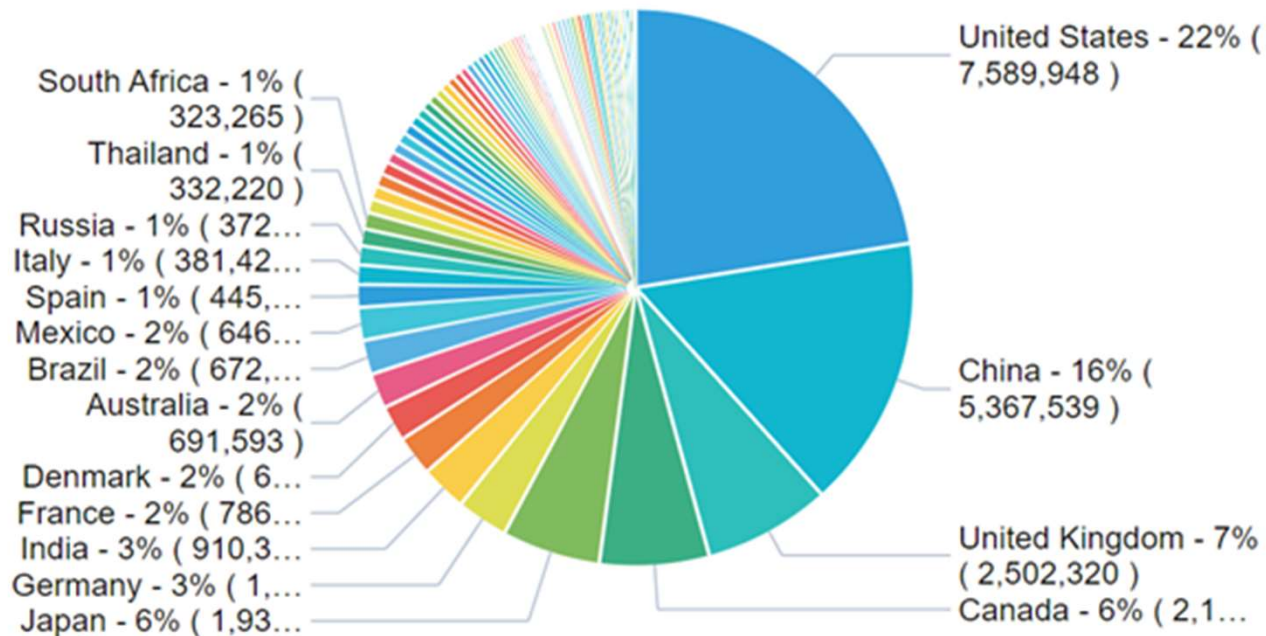
- NIH-NCBI – National Institute of Health, National Center for Biotechnology Information
- EMBL-European Nucleotide Archive (ENA)
- DNA Database of Japan (DDBJ)
- As of August, 2021, there were >15.3 trillion sequences stored in these open access databases
- The data in these databases originate from GR from all parts of the world, demonstrated by the WiLDSI report
- Not all sample metadata identifies the country of origin of the genetic resources from which the DSI was derived
- Sequences only become informative after comparison with known reference data

A snapshot from growth of selected GenBank Divisions (nucleotide base-pairs) 2020-2021 (ex-Sayers, 2022)

GenBank Division	Number of bp of sequence in GenBank by Division (at 08/2021)	Growth in sequence databases since previous year
Plants	350,590,744,188	30.12%
Phages	935,884,237	19.59%
Viruses	39,351,597,469	575.68%
Bacteria	130,518,385,589	32.07%
Primates	15,165,437,356	72.97%
Rodents	23,336,550,435	93.02%
Other mammals	28,568,850,588	37.06%
Other vertebrates	85,320,979,451	34.22%
Invertebrates	108,680,334,593	450%
Total - Above sets plus 12 other GenBank divisions	15,309,209,714,374	54.79%

^[1] <https://www.ncbi.nlm.nih.gov/genbank/samplerecord/>

The number of sequences countries provide and the percentage of the total number - WiLDSI Portal see <https://wildsi.ipk-gatersleben.de/apex/wildsi/r/wildsi/12>



Access and use of DSI from CABI Centres and in country partners

	Bahamas	Brazil	China	Ghana	India	Kenya	Malaysia	Pakistan	Trinidad & Tobago	UK	Zambia
Does the country use DSI?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Does the country generate DSI	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
In-country	-	Y	Y	N	Y	-	Y	Y	N	Y	Y/N
By out-sourcing	Y	Y	-	Y	N	Y	N	Y	Y	Y	Y
Are there constraints to generation or access of DSI?	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y
Sequencing Infrastructure	Y	Y	N	Y	N	Y	N	N	Y	Y	Y
Internet access	Y	N	Y	Y	-	Y	Y	N	-	N	Y
Trained staff	-	Y	-	Y	Y	Y	Y	N	Y	Y	Y
Financial	-	Y	-	Y	Y	Y	Y	N	Y	Y	Y
Other IT infrastructure	Y	Y	Y	Y	-	Y	Y	N	Y	Y	Y
Data analysis	Y	Y	Y	Y	-	Y	-	N	Y	Y	Y
Are data easily accessible?	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y

Bahamas and Trinidad & Tobago are covered by the same CABI Centre; Y: Yes; N: No; ; -: not stated view



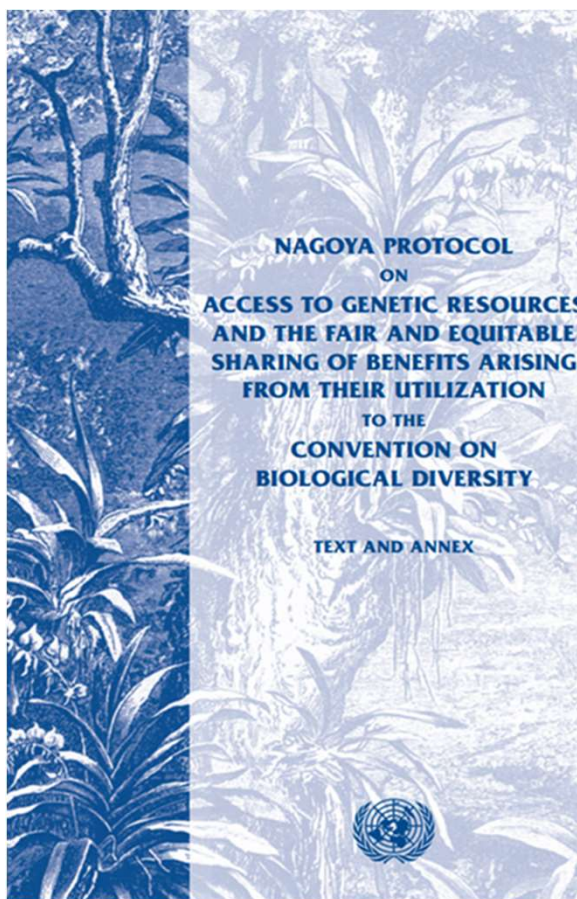
Addressing the imbalance

Coordinated and targeted capacity-building activities can be funded through multilateral or in bilateral arrangements through activities such as:

- a) On-site and/or virtual courses;
- b) Exchange of information, experiences and lessons learned;
- c) Joint scientific research, technology transfer, scientific visits, partnerships and collaborations;
- d) Support for development of scientific infrastructure, including through regional approaches (e.g., CGIAR centres);
- e) Intercultural dialogue through face-to-face meetings for indigenous peoples and local communities including dialogue between scientists and traditional knowledge holders;
- f) Integration in academic curricula;
- g) Integration in regional and international development agenda

The regulatory environment

– some countries have already taken action



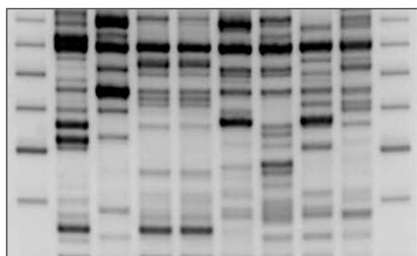
- A survey of ABS NFPs resulted in 26 Country returns from the 157 contacted, these were: **Zimbabwe**; Vietnam; Switzerland; Ethiopia; Croatia; Guinea-Bissau; Bosnia and Herzegovina; Argentina; Estonia; **Brazil**; Finland; Panama; Slovakia; Czech Republic; **Antigua and Barbuda**; Chile; Sri Lanka; The Netherlands; UK; New Zealand; Montenegro; Colombia; Malta; United States of America; **Guyana**
- 10 countries have ABS measures for genetic resources
 - 4 have measures for DSI
 - 4 others are intending to put them in place
- 8 responders agreed that publication of DSI was considered sharing appropriate benefits but 2 strongly disagreed
- Benefit-sharing arrangements are in place for countries with ABS measures
- The AHTEG fact-finding study (2020) stated that no country had reported receiving direct monetary benefits from the use of DSI to date



Options for a common approach to DSI use and benefit sharing

- Retaining the *status quo* allows individual countries to decide – complicated landscape
- Integrating DSI fully into the CBD and NP for benefit sharing - requires tracing of DSI and PIC and MAT potentially with multiple countries
- Benefit sharing being triggered by milestones along the value chain - standard MAT, license or terms and conditions.
- Payment or contribution to a multilateral fund which avoids tracing the source of the DSI - Access would not be regulated
- Enhanced technical and scientific cooperation with no PIC or MAT - a multilateral mechanism would be put in place and benefit sharing would not be linked to DSI data
- Currently, too little information to carry out a cost benefit analysis
- Little available evidence for significant monetary benefits

Discussion and conclusions



- The definition of DSI remains controversial and not agreed
- DSI is highly and increasingly relevant for research and development in all sectors of GRFA
- DSI is used similarly across all sectors, even if generated from BFA or other GR
- Open access to DSI is required to be able to characterise new sequence data – generation of data is not utilisation
- Deposit of data in publications is required by Scientific Journals; it is deposited freely by scientists
- A common approach should not hinder research and innovation
- Options still require further information and analysis
- Countries have already implemented their own approaches



The key messages from this study were:

- Many applications of DSI are highly relevant to GRFA, including from DSI that is not derived from GRFA
- DSI contributes significantly to innovation and discovery in GRFA - 12% of the CAB Abstracts records (1.18 million publications) cite DSI - dominating publications in 2021
- Making DSI available through public databases does not mean that DSI is accessible for everyone in the same way
 - serious obstacles for many countries:
 - urgent need to address these deficits
- The regulatory environment should be reviewed and revised with a view to facilitate access to DSI and the fair and equitable sharing of benefits arising from its use

Acknowledgements

Several individuals have helped review and ensure accuracy and adequate coverage of the content. These include the following:

Amber H. Scholz, DSI Network, WILDSI Project, Deputy to the Director, Leibniz-Institut DSMZ German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany;

Christopher Lyal, Scientific Associate, Natural History Museum, London, UK;

Dan Leskien, Senior Liaison Officer, Office of Climate Change, Biodiversity and Environment (OCBC), Commission on Genetic Resources for Food and Agriculture (CGRFA) FAO, Rome, Italy;

Devin Arbuthnott, Policy Advisor, Agriculture and Agri-Food Canada (AAFC), Ottawa, Canada;

Emmanuel Hala Kwon-Ndung, African BioGenome Project (AfricaBP), Professor of Plant Breeding and Genetics, Federal University of Lafia (FULafia), Nasarawa State, Nigeria;

Guy Cochrane, Data Coordination and Archiving Team Leader, Head of European Nucleotide Archive, European Bioinformatics Institute, European Molecular Biology Laboratory, Cambridge, UK;

Irene Maria Hoffmann, OCBC, CGRFA, FAO, Rome, Italy;

Justin Eze Ideozu, Co-Chair, Ethics Legal and Social Issues Subcommittee, African BioGenome Project (AfricaBP), Senior Scientist Pharmacogenomics, Genetic Research Center, North Chicago, USA;

Manuela da Silva, General Manager of Fiocruz COVID-19 Biobank, Fiocruz, Brazil;

Peter Mason, Research Scientist, Agriculture and Agri-Food Canada (AAFC), Ottawa, Canada;

Sally Mueni Katee, Chair, Ethics, Legal and Social Issues Subcommittee, Africa BioGenome Project (AfricaBP), ABS Legal Specialist/Officer, Livestock Genetics Program, International Livestock Research Institute, Nairobi, Kenya;

ThankGod Echezona Ebenezer, Founder and Co-Chair, African BioGenome Project (AfricaBP), Bioinformatician, EMBL's European Bioinformatics Institute, Cambridgeshire, UK.

CABI Centre contributions specifically coordinating, seeking and compiling on country feedback:

Brazil Centre: coordinated and compiled by Yelitza Colmenarez, Country Director Brazil.

China Centre: Feng Zhang coordinated the survey, Min Wan, Hongmei Li, and Jinping Zhang conducted the interviews with Chinese experts. Xin translated the feedbacks and compiled relevant literatures.

Ghana Centre: coordinated and compiled by Victor Clottey, Regional Representative, West Africa.

India Centre: coordinated and compiled by Gopi Ramasamy, Regional Director South Asia

Kenya Centre: coordinated and compiled by Joseph Mulema.

Malaysia Centre: coordinated and compiled by Sathis Sri Thanarajoo.

Pakistan Centre: coordinated Babar Bajwa, Senior Regional Director, Asia and compiled by Yusuf Zafar.

Trinidad and Tobago Centre: contribution for the Bahamas and the Caribbean was coordinated by Naitram (Bob) Ramnanan.

Zambia Centre: coordinated and compiled by Noah Phiri.

<https://www.cabi.org/what-we-do/cabi-centres/>



CABI is an international intergovernmental organisation, and we gratefully acknowledge the core financial support from our member countries (and lead agencies) including:



Ministry of Agriculture and
Rural Affairs,
People's Republic of China



Agriculture and
Agri-Food Canada



Ministry of Foreign Affairs of the
Netherlands



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Agency for Development
and Cooperation SDC