

# CGIAR Centers' practices in the generation, deposit, sharing and use of DSI for research and breeding

Eleventh meeting of the Working Group for the Enhancement of the Functioning of the Multilateral System

April 16, 2024

Mathieu Rouard, Ruairaidh Sackville Hamilton, Michael Halewood, Claudio Chiarolla, Isabel López Noriega



# Outline

---

Types of genetic data that Centers generate and where they are stored and shared

---

Conditions under which the data are made available

---

How much DSI is used

---

CGIAR Centers use of DSI

---

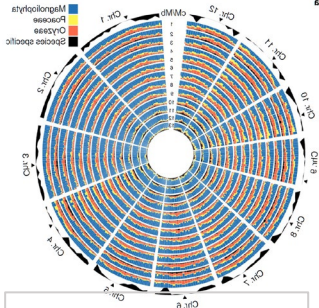
How DSI changes the use of genetic resources in ag R&D

---

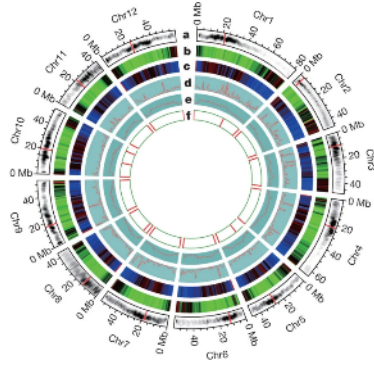
ABS policy implications

---

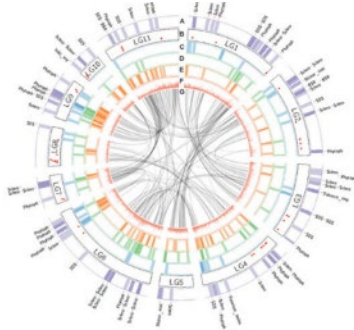
# CGIAR and reference genome sequences



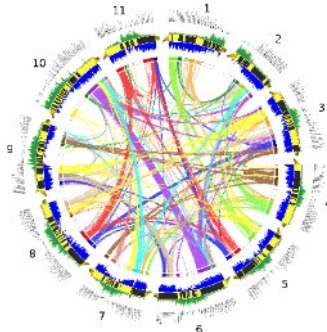
Rice  
(*Nature* 2018,  
2020)



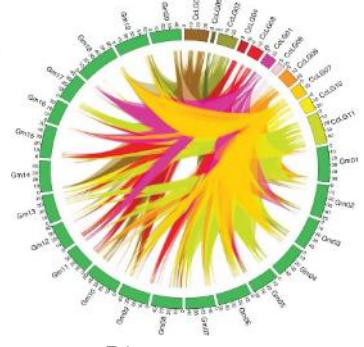
Potato  
(*Nature* 2011)



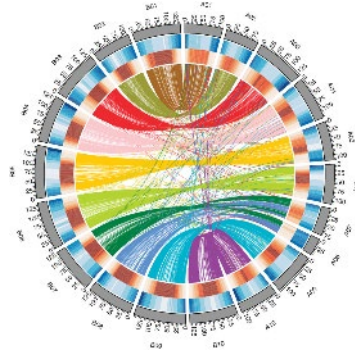
Adzuki bean  
(*PNAS* 2015)



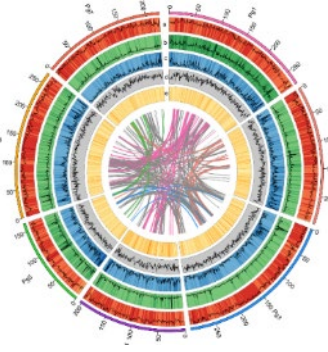
Banana  
(*Nature* 2012)



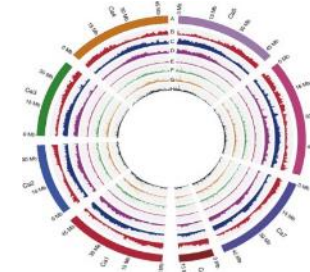
Pigeonpea  
(*Nature Biotech.* 2012)



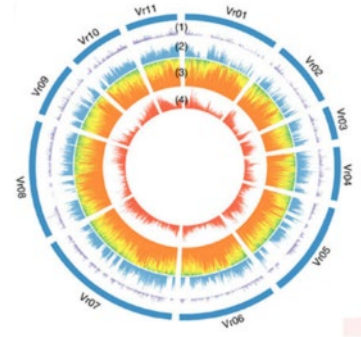
Groundnut  
(*Nature Genetics* 2016)



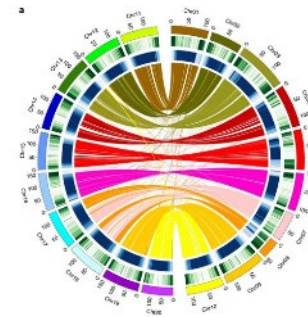
Pearl millet  
(*Nature Biotech.* 2017)



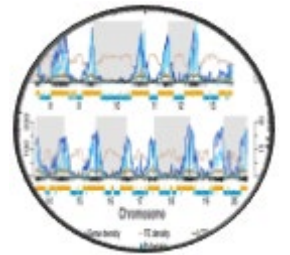
Chickpea  
(*Nature Biotech.* 2013)



Mung bean  
(*Nature Com.* 2014)



Cultivated  
Groundnut  
(*Nature  
Genetics* 2019)



Yam  
(*Nature Com.*  
2022)

# CGIAR genebanks and resequencing / genotyping

Crops	Resequencing (number of accessions)	NGS Genotyping (number of accessions)	Average % of resequencing and genotyping data currently available in public databases
<b>Major cereals</b> (maize, rice -Asian and African, wheat, barley, pearl millet)	3,676	110,829	81%
<b>Roots and tubers</b> (potato, cassava, sweet potato, yams, Andean roots and tubers - <i>Ullucus tuberosus</i> and <i>Oxalis tuberosa</i> ) and bananas	400	24,404	12%
<b>Minor cereals</b> (small millets, sorghum)	448	1,629	34%
<b>Pulses</b> (common bean, groundnut, chickpea, lentil, pea, cowpea, pigeonpea, faba bean, grasspea)	5,567	7,778	46%
<b>Forages</b> buffelgrass, napiergrass, rhodes grass, <i>Urochloa spp</i> , <i>Megathrysus maximus</i> , <i>Sesbania sesban</i> , <i>Neonotonia wightii</i> , Lablab, <i>Brizantha</i> , <i>Brachiaria</i> )		2,378	0%

# What can we conclude from latest statistics on genotyping

Center	Crop	High confidence/good quality	No data	% of good quality data	Notes
CIAT	Cassava	5703	254	95.73610878	
CIMMYT	Maize	27206	1381	95.16913282	
CIAT	Beans	35889	2047	94.60407001	
CIP	Sweet Potato	5770	503	91.98150805	
CIP	Potato	6243	1470	80.94126799	
ICARDA	Wheat	25886	18356	58.51001311	
CIP	<b>Andean Roots &amp; Tubers</b>	1239	1281	49.16666667	
Bioversity	<b>Banana</b>	833	872	48.8563	
CIMMYT	Wheat	36475	84338	30.19128736	
IITA	Cassava	448	2786	13.85281385	Result of DSI project?
ILRI	<b>Forages &amp; Fodder</b>	2282	16405	12.21169797	
IITA	<b>Yam</b>	261	2739	8.7	
ICARDA	Barley	2312	30175	7.116692831	Ok but pangenome from IPK
IRRI	Rice	6075	126577	4.579652022	Ok but pangenomes available
ICARDA	Chickpea	670	14715	4.354891128	Ok but 3000+ pangenome from ICRISAT (overlap?)
CIAT	<b>Forages</b>	618	20092	2.984065669	
AfricaRice	<b>Rice</b>	370	20311	1.789081766	
ICARDA	<b>Sorghum</b>	242	42627	0.564510485	

# Large-scale genotyping in CGIAR genebanks

- **To improve conservation of genetic resources**

- Diversity / core collections
- Taxonomic identification
- Gap analyses
- Duplicates
- Genetic integrity

- **To enhance use of genetic resources**

- identify useful alleles/haplotypes or candidate genes
- Identification of donors for specific traits



ARTICLE

OPEN  
<https://doi.org/10.1038/s41586-018-0063-9>

## Genomic variation in 3,010 diverse accessions of Asian cultivated rice

Wensheng Wang<sup>1,2\*</sup>, Ramil Maulon<sup>2,3†</sup>, Zhigang Hu<sup>3,3†</sup>, Dmytro Chebotarov<sup>2,3†</sup>, Shaoshuai Tai<sup>4,3†</sup>, Zhichao Wu<sup>5,3†</sup>, Min Li<sup>6,7,3†</sup>, Tianqing Zheng<sup>2,3†</sup>, Rowen Rommel Fuentes<sup>2,3†</sup>, Fan Zhang<sup>3,3†</sup>, Locedie Mansueto<sup>2,3†</sup>, Dario Copetti<sup>2,3†</sup>, Millicent Sanciango<sup>2</sup>, Kevin Christian Palis<sup>2</sup>, Jianlong Xu<sup>1,5,6</sup>, Chen Sun<sup>2</sup>, Binying Fu<sup>1,6</sup>, Hongliang Zhang<sup>2</sup>, Yongming Gao<sup>1,6</sup>, Xiqin Zhao<sup>2</sup>, Fei Shen<sup>2</sup>, Xiao Cai<sup>2</sup>, Hong Yu<sup>2</sup>, Zichao Li<sup>2</sup>, Miaolin Chen<sup>2</sup>, Jeffrey Detras<sup>2</sup>, Yongli Zhou<sup>1,6</sup>, Xinyuan Zhang<sup>2</sup>, Yue Zhao<sup>2</sup>, Dave Kudrna<sup>2</sup>, Chunshao Wang<sup>2</sup>, Rui Li<sup>2</sup>, Ben Ji<sup>2</sup>, Jinyuan Lu<sup>2</sup>, Xianchang He<sup>2</sup>, Zhaorong Dong<sup>2</sup>, Haisuo Xu<sup>4</sup>, Yanhong Li<sup>4</sup>, Miao Wang<sup>4</sup>, Jianxin Shi<sup>2</sup>, Jing Li<sup>2</sup>, Dabing Zhang<sup>2</sup>, Seunghee Lee<sup>8</sup>, Wushu Hu<sup>4</sup>, Alexander Poljakov<sup>9</sup>, Inna Dubchak<sup>10,11</sup>, Victor Jun Ular<sup>2</sup>, Frances Nikki Borja<sup>2</sup>, John Robert Mendoza<sup>12</sup>, Jaubar Ali<sup>2</sup>, Jing Li<sup>2</sup>, Qiang Gao<sup>2</sup>, Yongchao Ni<sup>4</sup>, Zhen Yue<sup>4</sup>, Ma. Elizabeth B. Naredo<sup>2</sup>, Jayson Talag<sup>2</sup>, Xueqiang Wang<sup>2</sup>, Jinjie Li<sup>2</sup>, Xiaodong Fang<sup>2</sup>, Ye Yin<sup>4</sup>, Jean-Christophe Glaszmann<sup>13,15</sup>, Jianwei Zhang<sup>8</sup>, Rayang Li<sup>1,15</sup>, Ruairadh Sackville Hamilton<sup>2</sup>, Rod A. Wing<sup>16,17</sup>, Jue Ruan<sup>18</sup>, Gengyun Zhang<sup>16,17</sup>, Chaochun Wu<sup>18,19</sup>, Nikolai Alexandrov<sup>20</sup>, Kenneth L. McNally<sup>21</sup>, Zhikang Li<sup>1,16,17</sup> & Hei Leung<sup>2</sup>

Article | [Open Access](#) | Published: 11 September 2020

## Diversity analysis of 80,000 wheat accessions reveals consequences and opportunities of selection footprints

Carolina Sansaloni , Jorge Franco, [...] Kevin Pixley

*Nature Communications* 11, Article number: 4572 (2020) | [Cite this article](#)

[nature](#) > [articles](#) > [article](#)

Article | [Open Access](#) | Published: 10 November 2021

## A chickpea genetic variation map based on the sequencing of 3,366 genomes

Rajeev K. Varshney , Manish Roorkiwal, Shuai Sun, Prasad Bajaj, Annapurna Chitkineni, Mahendar Thudi, Narendra P. Singh, Xiao Du, Hari D. Upadhyaya, Aamir W. Khan, Yue Wang, Vanika Garg, Guangyi Fan, Wallace A. Cowling, José Crossa, Laurent Gentsbittel, Kai Peter Voss-Fels, Vinod Kumar Valluri, Pallavi Sinha, Vikas K. Singh, Cécile Ben, Abhishek Rathore, Ramu Punna, Muneendra K. Singh, ... Xin Liu 

[+ Show authors](#)



# DSI storage and access by CGIAR

it of the Toronto Statement. doi:10.1038/461168a

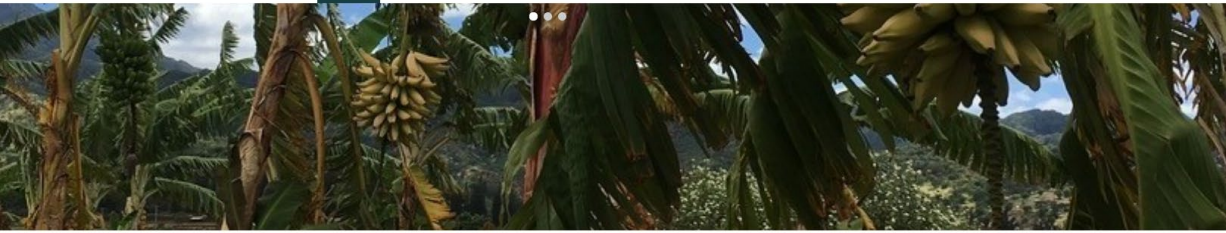
CicerSeq is a public repository global map of genome variation based on sequencing of 3,366 Cicer genomes representing 3,171 cultivated species accessions and 195 accessions from seven wild species. Information about genotype, passport data and different variants (called against CDC Frontier genome) including SNPs and various structural variations (SVs) reported in the article entitled "A global reference for chickpea genetic variation based on sequencing of 3,366 genomes" can be accessed here. This database provides information to facilitate genomics-informed decisions in chickpea improvement.

**Genotype calls**  
Query for SNPs

**Passport**  
Passport data

**JBrowse**  
Genome Browser

**Data**  
Data download



## Available Tools



### Synteny viewer

Retrieve all syntenic blocks and genes for your regions of interest



### GO enrichment

Perform GO enrichment on list of genes

TGACTCCAT  
TGACGCCAT  
TGACTCCCT  
TGACGCCCT

### Gigwa

Gigwa is a web-based tool for the exploration of large amounts of genotyping data.



### Crispor Tefor

CRISPOR is a program that helps design, evaluate and clone guide sequences the CRISPR/Cas9 system.

[See all tools](#)

## Genomes



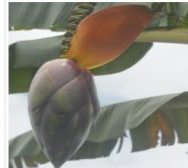
### Musa cavendish (Baxijiao)

Baxijiao (AAA, Cavendish), with a basic chromosome number of  $x=11$  ( $2n=3x=33$ ), belongs to the group of banana cultivars that the most common [read more](#)



### Musa troglodytarum (Utafun)

Musa troglodytarum (Fe'i banana) 'Utafun' (10.18730/9KVB3) is Fe'i banana collected from Papua New Gu [read more](#)



### Musa textilis (Abaca)

Musa textilis 'Abaca' (10.18730/9.998W) is an important industrial crop cultivated mainly in Philippines, Ecu [read more](#)



### Ensete ventricosum (Enset)

Ensete ventricosum, also known as Ethiopian banana or enset, is a perennial plant native to Ethiopia and parts of E Africa. It belo [read more](#)

# An example of a repository maintained by one CGIAR Center – Banana Genome Hub

## Users by Country



COUNTRY	USERS
Vietnam	1.7K
Philippines	944
China	493
Brazil	445
France	277
United States	255
Indonesia	219

## Visitors in last 3 months

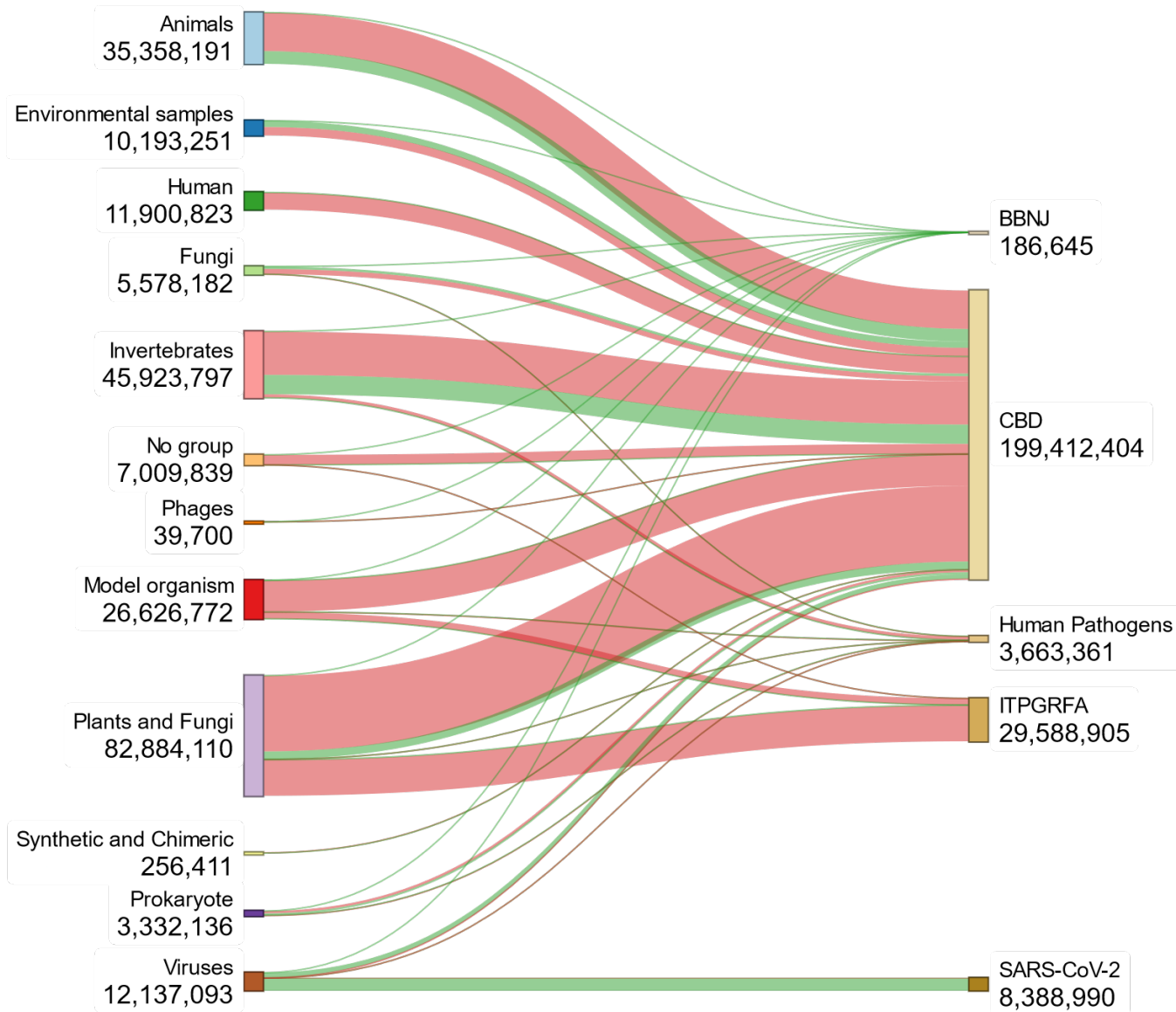
[View countries](#) →



# Conditions under which the data are made available

- Generally follow academic practices on data sharing:
  - Sequence data are deposited in INSDC databases
  - Access is open and free
  - Follows Toronto statement 2009 on prepublication data sharing
- Specialized portals maintained by CGIAR Centers usually don't require login, but do require acknowledgement / attribution and prohibit IPR claims over the data
- One CGIAR center requires login and click-wrap license agreement for non-exclusive non-transferable use of genetic data subject to conditions such as not claiming IPRs over it.

# How much DSI is out there from and for the agricultural sector and the crop sector in particular?



- Recent work by the DSI Scientific Network attempting to classify sequence data based on types of organisms and the scope of international conventions
- The number under ITPGRFA refers to sequences from Annex 1 species – around 10% of all sequences on INSDC
- **Limitations:**
  - Only INSDC – does not include genotyping data
  - Only Annex 1 species

# Approaches to investigate how much MLS material is being used by recipients to generate DSI

- Through DOIs included in publications and datasets
- Searches across multiple journals using keywords – name of particular genebanks, traits, sequences.
- Individual tracking of publications
- Automatic methods to be developed with AI?

# Building knowledge: Genome wide association studies

OPEN ACCESS Freely available online



## Genomewide Association Studies for 50 Agronomic Traits in Peanut Using the 'Reference Set' Comprising 300 Genotypes from 48 Countries of the Semi-Arid Tropics of the World

Manish K. Pandey<sup>1</sup>, Hari D. Upadhyaya<sup>1</sup>, Abhishek Rathore<sup>1</sup>, Vincent Vadez<sup>1</sup>, M. S. Sheshshayee<sup>2</sup>, Manda Sriswathi<sup>1</sup>, Mansee Govil<sup>1</sup>, Ashish Kumar<sup>1,3</sup>, M. V. C. Gowda<sup>4</sup>, Shivali Sharma<sup>1</sup>, Falalou Hamidou<sup>5</sup>, V. Anil Kumar<sup>1</sup>, Pawan Khera<sup>1</sup>, Ramesh S. Bhat<sup>6</sup>, Aamir W. Khan<sup>1</sup>, Sube Singh<sup>1</sup>, Hongjie Li<sup>7</sup>, Emmanuel Monyo<sup>8</sup>, H. L. Nadaf<sup>4</sup>, Ganapati Mukri<sup>4</sup>, Scott A. Jackson<sup>9</sup>, Baozhu Guo<sup>10</sup>, Xuanqiang Liang<sup>11</sup>, Rajeev K. Varshney<sup>1\*</sup>

[nature](#) > [scientific reports](#) > [articles](#) > [article](#)

Article | [Open access](#) | Published: 15 December 2020

## Genome-wide association studies in tropical maize germplasm reveal novel and known genomic regions for resistance to Northern corn leaf blight

[Zerka Rashid](#), [Mehrajuddin Sofi](#), [Sharanappa I. Harlapur](#), [Rajashekhhar M. Kachapur](#), [Zahoor Ahmed Dar](#), [Pradeep Kumar Singh](#), [Pervez Haider Zaidi](#), [Bindiganavile Sampath Vivek](#) & [Sudha Krishnan Nair](#)

[Scientific Reports](#) **10**, Article number: 21949 (2020) | [Cite this article](#)

3589 Accesses | 20 Citations | 1 Altmetric | [Metrics](#)

## Genome-wide association mapping for wheat blast resistance in CIMMYT's international screening nurseries evaluated in Bolivia and Bangladesh

[Philomin Juliana](#), [Xinyao He](#), [Muhammad R. Kabir](#), [Krishna K. Roy](#), [Md. Babul Anwar](#), [Felix Marza](#), [Jesse Poland](#), [Sandesh Shrestha](#), [Ravi P. Singh](#) & [Pawan K. Singh](#)

# Genomic data in breeding

## nature genetics

Perspective | Published: 30 August 2017

### Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery

John M Hickey, Tinashe Chiurugwi, Ian Mackay, Wayne Powell & Implementing Genomic Selection in CGIAR Breeding Programs Workshop Participants

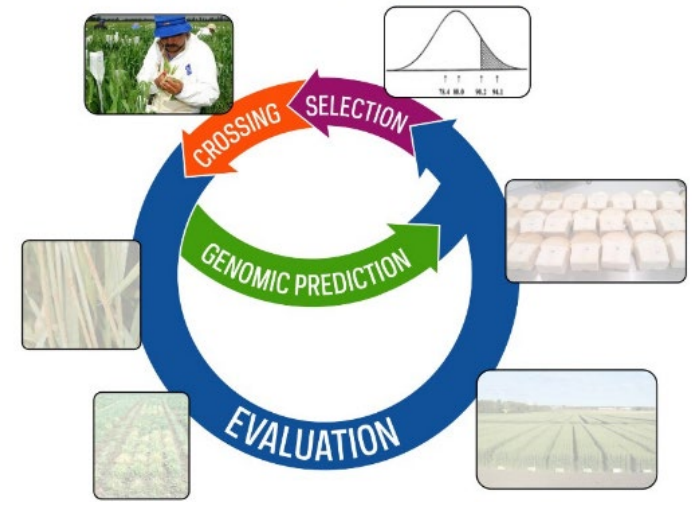
*Nature Genetics* 49, 1297–1303 (2017) | [Download Citation](#)

Fig. 1: Current genome-editing projects taking place at CGIAR centers.

Crop	Trait	Edited genes	Stage
Banana	Disease resistance (BXW, Fusarium wilt, BSV)	DMR6, BSV sequences	3,1
Cassava	Disease resistance (BB)	SWEET gene promoters	3
	Food safety (cyanide-free)	Linamarin synthase	3
	Quality (waxy starch)	GBSS1	3
Maize	Disease resistance (MLN)	C6 QTL	1
	Weed resistance (Striga)	Strigolactone	3
Potato	Disease resistance (PVY <sup>a</sup> , late blight)	eIF-4E, StDMR6-1, StCHL1	2
Rice	Disease resistance (BLB, RHB)	SWEET gene promoters, AGO4, STV11	4,3
	Food safety (low arsenic and cadmium)	OsNRAMP5, OsPT8, LS1, LS2	3
	Nitrogen remobilization, and methane emission reduction	Unpublished	3
	Insect resistance <sup>a</sup> (BPH)	BPH resistance alleles	2
Wheat	Disease resistance (rusts, mildew) <sup>a</sup>	Lr67 and others	3

Pixley, K.V. et al. Genome-edited crops for improved food security of smallholder farmers. *Nat Genet* 54, 364–367 (2022)

### (accelerating) The Breeding Cycle



# DSI-related capacity building



**Center of Excellence in Genomics & Systems Biology**

*Striving towards efficient breeding and research for better crop performance and human health*

## Plant Treaty Governing Body Resolution 16/2023

“The Governing Body ...

Welcomes and acknowledges the technical assistance ... undertaken by the CGIAR Centers in order to reduce the existing gap on capacity regarding generation, access to and use of DSI/GSD and recommends that this work be continued and strengthened”



Technical assistance to strengthen national agricultural research organizations' capacity to use digital sequence information

A submission from CGIAR



<https://hdl.handle.net/10568/131562>

# Training Workshop on Genotyping, Sequencing, and Principles on Data Analysis and Concept of Data Management



ICARDA's recent training workshop on genotyping and data management in Tunisia gathered participants from 10+ countries. They gained valuable insights into genotyping techniques, data management, and DSI applications in breeding & genebanks. Thanks to @CGIAR genebank initiative!

[Traduire le post](#)

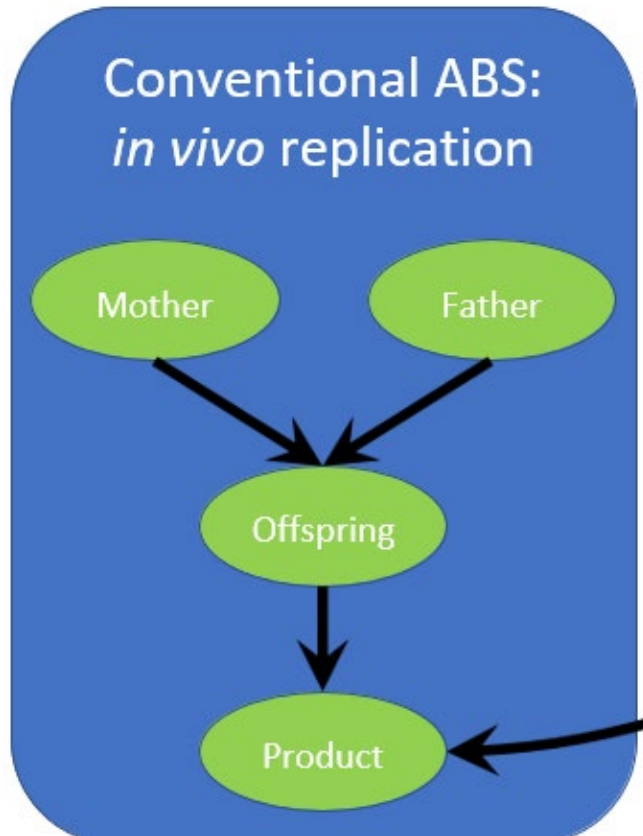


1:00 PM · 15 déc. 2023 · 1906 vues



1. Genotyping methods
2. BRI shared services & logistics
3. Genomic databases (incl. Gigwa)
4. Applications in genebanks
5. Applications in breeding programs
6. DSI and ITPGRFA policy

# Crop improvement – value creation without DSI



Incorporation of genetic material in final product as basis for fair and equitable benefit sharing

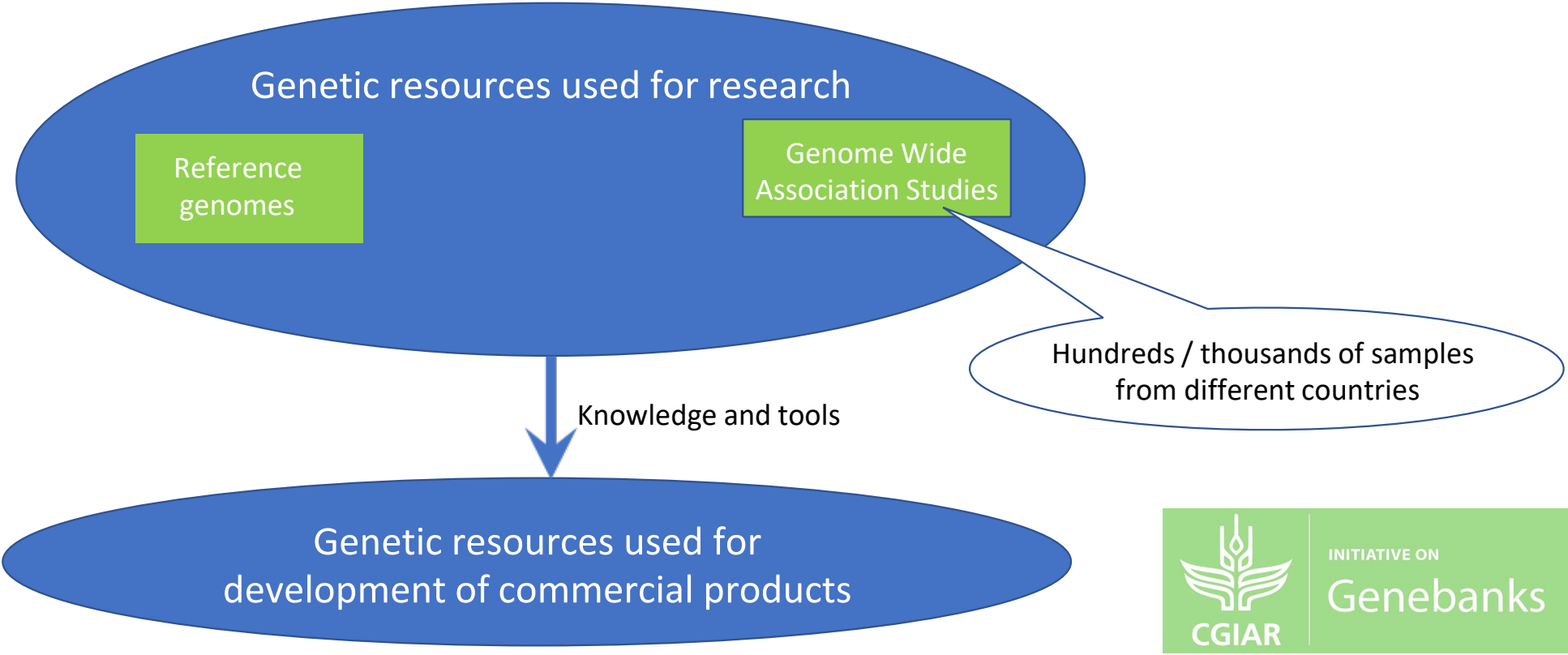


# DSI changes how we use genetic resources in agricultural R&D: benefit sharing implications

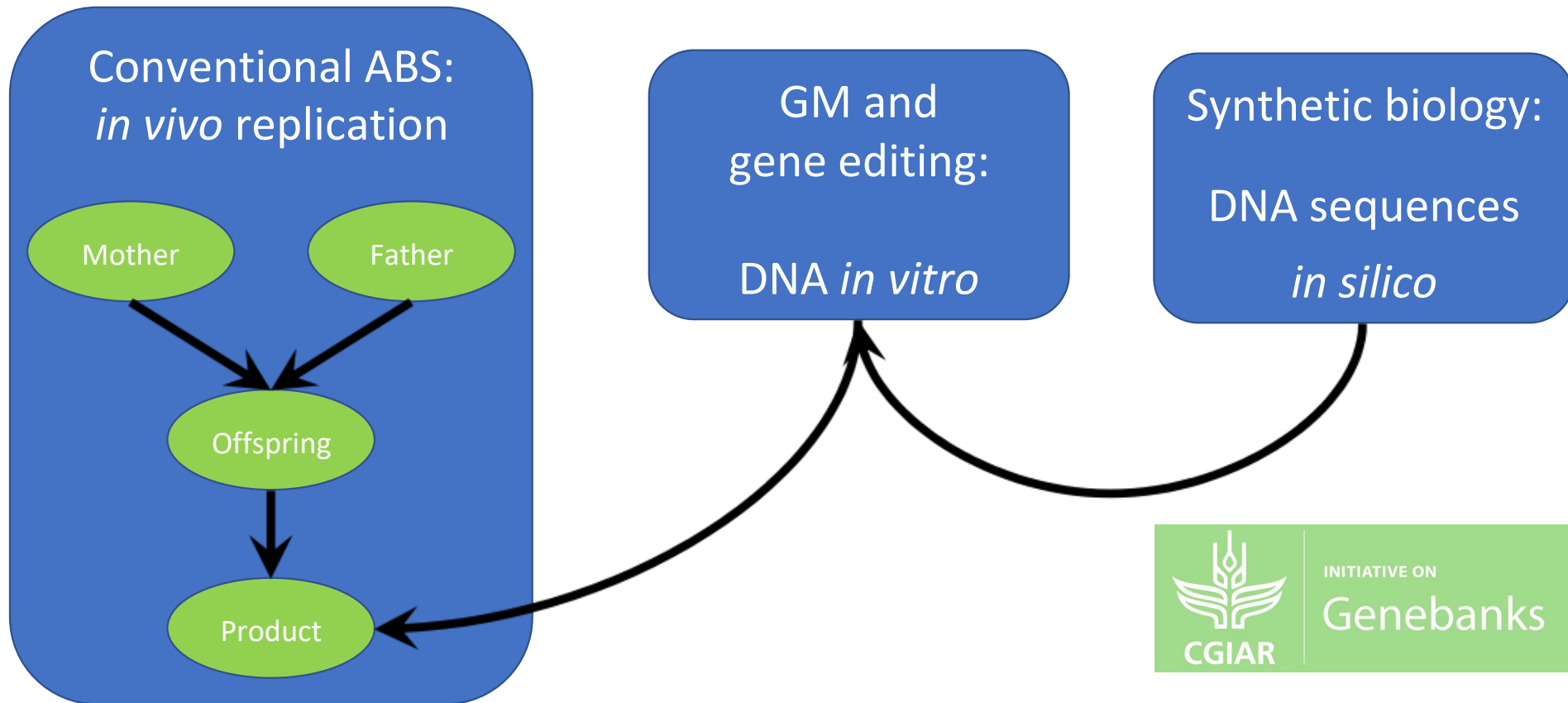
1. DSI delinks use of genetic resources for research and development
2. DSI-related technologies enable *in vitro* and *in silico* replication of units of heredity
3. DSI enables discovery of unrelated genetic resources containing the same sequences



# 1: DSI delinks use of genetic resources for research and development



## 2. DSI-related technologies enable *in vitro* and *in silico* replication of units of heredity



### 3. DSI enables discovery of unrelated genetic resources containing the same sequences

- Use materials from the MLS or other ABS regime in research phase to discover desirable sequence  
... then find/obtain desired sequence in an unregulated source and use it for product development



# ABS implications

Benefit sharing in the era of DSI can only be fair and equitable if it

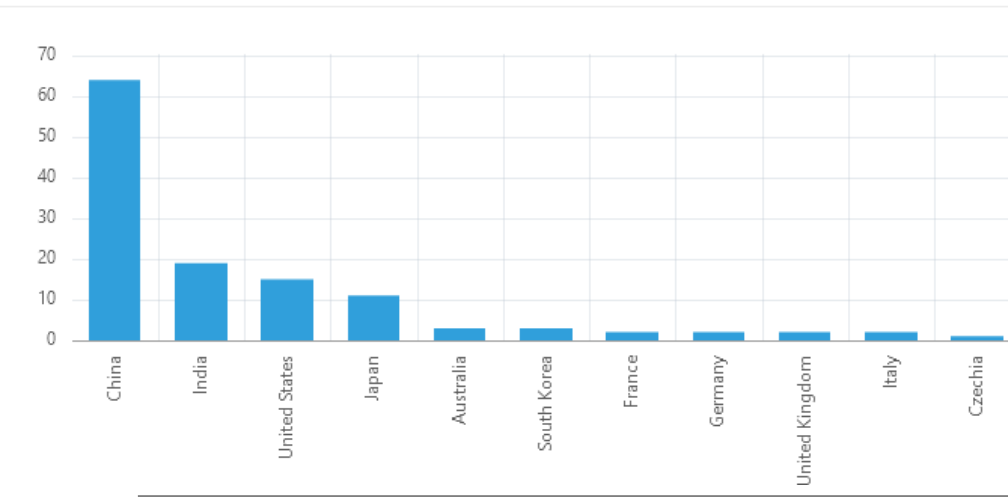
- delinks payment obligations from incorporation of DNA sequences into final products, and
- Recognizes, reflects and rewards the value contributed by DSI through upstream research

# *De facto* multilateral access and use of DSI

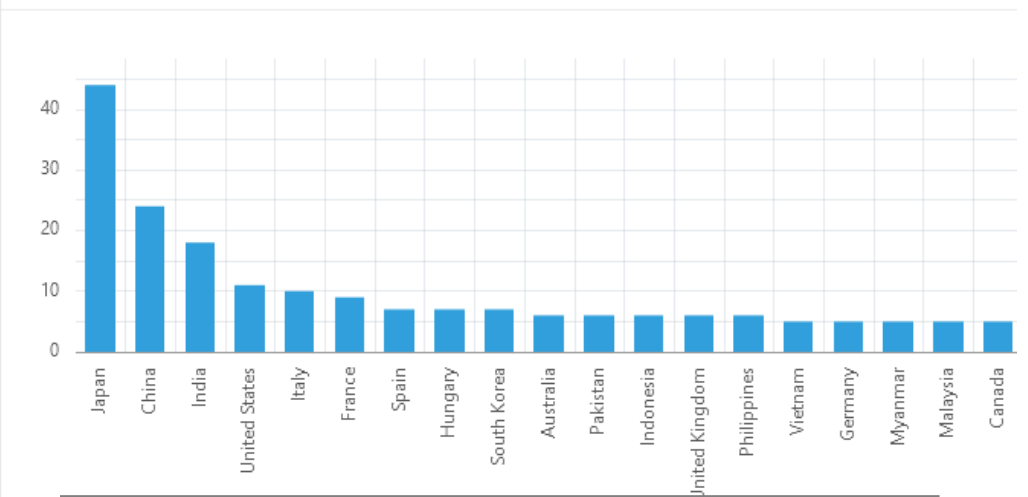
- Multilateral reality
  - Digital infrastructure facilitates pooling and availability of data from genetic resources world wide
  - Stimulating scientific break-through
  - Being used to create value at multiple levels
  - Ag R&D is now dependent on generation, availability and use of DSI.
- Reality check
  - Lots of non-monetary benefit sharing. Largely unmeasured
  - No monetary benefit sharing
    - Need a solution that raises funds, reflects value addition of DSI to ag R&D, shares those benefits, and supports continued availability of DSI

# Country-based view of the use of DSI of Annex 1 crops and forages

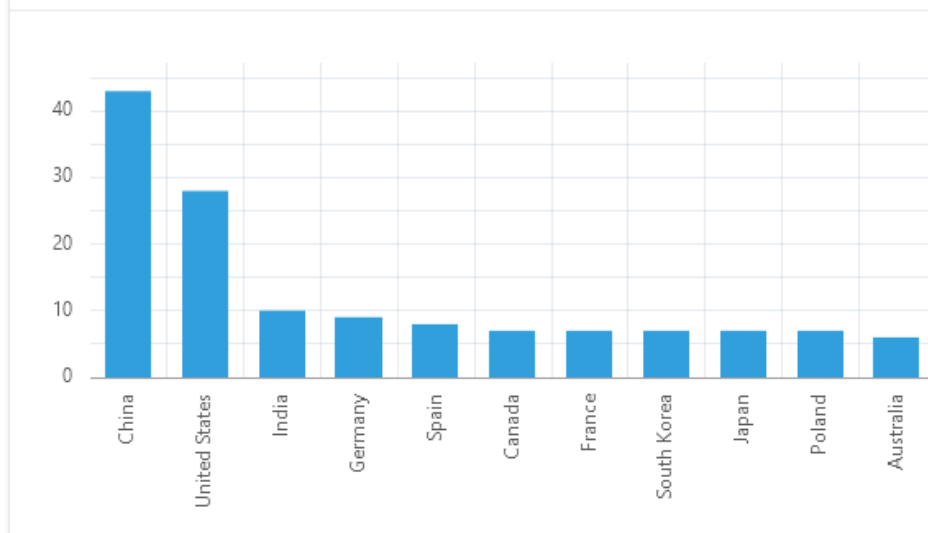
2.1 In-country use: How many authors from country x use DSI that originated in that country?



2.3 Out-of-country use: How many authors from outside of a country use that country's DSI?



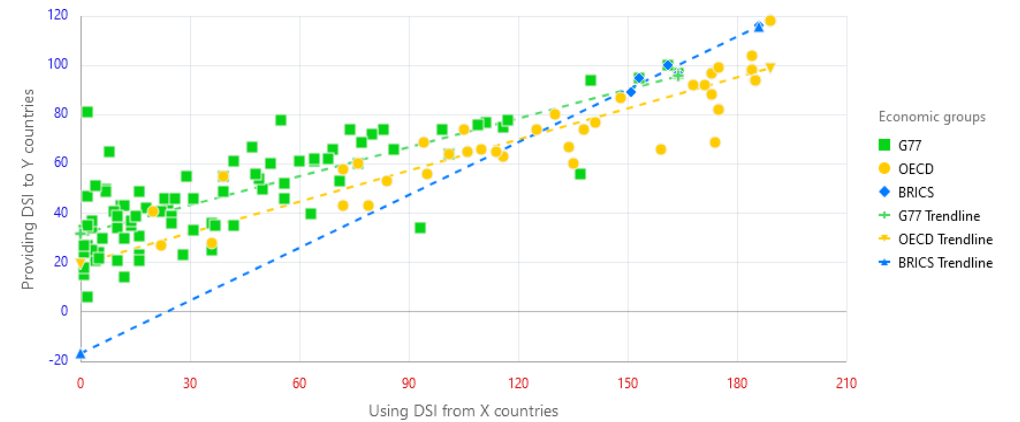
2.2 World use: How many authors from country x use DSI from other countries?



## Providing vs Using

6.1 What is the relationship between providing and using DSI?

The number of countries to which a country provides DSI is correlated to the number of countries from which it uses DSI, suggesting...



# What could delinked benefit-sharing look like?

Aggregate level payments on whole portfolios of products, whether they were developed using accessed DSI or not

Contributions from users based on %age additional fees for cloud-based services

Contributions from DSI service providers:  
reagents, sequencing, analysis, etc





Thank you.



Genebank  
Initiative