

## CGIAR Centers' use of Digital Object Identifiers (DOIs):

a submission to the Advisory Committee on the Global Information System

### Introduction

This submission has been prepared in response to a request from the Secretary of the International Treaty on Plant Genetic Resources (ITPGRFA), Dr Kent Nnadozie, to the eleven CGIAR Centers that signed Article 15 agreements with the Governing Body of the ITPGRFA. In his letter, the Secretary referred to a request from the Scientific Advisory Committee on the Global Information System (SAC-GLIS), to the Secretariat to:

“gather information from GLIS users, including CGIAR Centers and other institutions managing crop germplasm repositories, on the current application of DOIs to crop germplasm in the Multilateral System of Access and Benefit-Sharing for which DSI/GSD are available in compatible information systems”

He asked the Centers if they “could provide the Secretariat with information pertinent to the request of the SAC-GLIS”. In preparing this submission, the management team of the Policy Module of the CGIAR Genebank Platform conducted a survey among the eleven Article 15 Centers. All Centers responded to the questionnaire. The following sections provide a summary of their responses. It includes illustrative examples of Centers' practices using DOIs to date.

### Assignment of GLIS DOIs to PGRFA hosted by CGIAR Centers'

#### Current status

As of May 21, 2019, GLIS reported that 834,252 DOIs have been minted. Approximately 94% of those DOIs are for materials hosted by CGIAR Centers. CGIAR Centers made it a priority in 2018 to 'mint' DOIs for materials in the international collections that they host. The ITPGRFA Secretariat provided excellent technical support. Table 1 below provides an overall summary of the numbers of DOIs minted, per Center, as of February 25, 2019.

All of these DOIs have been assigned to materials that are accessioned in the concerned Center's Genebank, including landraces, wild relatives, and breeding and elite lines resulting from the Centers' breeding programmes when the decision has been made to conserve those materials long term and make them available through the Multilateral System.

Table 1: Numbers of DOIs minted, per Center, as of May 21, 2019

Center	PGRFA with GLIS DOIs
AfricaRice	21,300
Bioversity	1,598
CIAT	67,770
CIMMYT	204,646
CIP	17,982
ICARDA	156,901
ICRAF	12,999
ICRISAT	120,431
IITA	34,651
ILRI	18,639
IRRI	124,597
TOTAL	781,514

Datasource: GLIS, May 21, 2019

CGIAR Centers are not currently minting GLIS DOIs for breeding materials in a systematic manner (apart from those which are accessioned in the Genebank as just mentioned). That said, some Centers are experimenting with minting DOIs with some materials directly from their breeding programs. CIP will pilot minting DOIs for some improved materials being developed in a potato breeding program in Africa. Over the course of 2019, ICARDA breeders will work in collaboration with genebank managers to apply DOIs to 400 advanced breeding lines of faba bean and chickpea.

In addition, some Centers (AfricaRice, CIP) have organized seminars on the advantages of assigning GLIS DOIs for breeders' materials, and discussions are ongoing on how to best integrate DOIs in breeders' work. Some Centers (e.g., IITA) are making efforts to link the genebanks' databases (which have already integrated DOIs) with the breeders' databases.

## How the Centers use GLIS DOIs

### Use of GLIS DOIs in the SMTAs

Most Centers already include, or are in the process of putting systems in place to include, GLIS DOIs of germplasm in Annex 1 of the SMTA when transferring PGRFA samples from the genebanks to requestors. By the end of 2019, nine out of the eleven Centers which signed agreements under Article 15 of the ITPGRFA will be routinely providing GLIS DOIs with the SMTA. See image 1 below for an example of how Centers include this information.

In addition to including DOIs in SMTAs, CIP includes the accession DOI on each of its labels on seed packages and cryogenic vials and tubes containing germplasm for distribution (Image 2) and CIAT is modifying its software to also print DOIs on seed envelopes.

Image 1: Annex 1 of SMTAs sent by CIP (potato) and Bioversity (Musa), including DOIs for transferred materials:

**Annex 1**

**LIST OF MATERIALS PROVIDED**

This *Annex* contains a list of the **Material** provided under **this Agreement**, including the associated information referred to in Article 5b.

The following information is included for each **Material** listed: all available passport data and, subject to applicable law, any other associated, available, non-confidential descriptive information.

Crop: Potato Sample ID: doi:10.18730/90FJ  
Associated information: EE-2025  
Crop: Potato Sample ID: doi:10.18730/9KD\*  
Associated information: Unknown  
Crop: Potato Sample ID: doi:10.18730/9N09  
Associated information: Unknown  
Crop: Potato Sample ID: doi:10.18730/9N4D  
Associated information: Pana Piedra  
Crop: Potato Sample ID: doi:10.18730/9PW\*  
Associated information: Unknown  
Crop: Potato Sample ID: doi:10.18730/9Q0U  
Associated information: Camusa  
Crop: Potato Sample ID: doi:10.18730/D85Z  
Associated information: HHCH 4498  
Crop: Potato Sample ID: doi:10.18730/DADX  
Associated information: HHA 6450  
Crop: Potato Sample ID: doi:10.18730/DAJ\$  
Associated information: HHA 6496  
Crop: Potato Sample ID: doi:10.18730/DEP~  
Associated information: HHA 6656a  
Crop: Potato Sample ID: doi:10.18730/DEQ\$  
Associated information: HHA 6657  
Crop: Potato Sample ID: doi:10.18730/DHCY  
Associated information: OCH 15269  
Crop: Potato Sample ID: doi:10.18730/DHM1  
Associated information: OCH 12014  
Crop: Potato Sample ID: doi:10.18730/DPRH  
Associated information: OCHS 11805  
Crop: Potato Sample ID: doi:10.18730/DQC0  
Associated information: OCHS 11851  
Crop: Potato Sample ID: doi:10.18730/DRQ6  
Associated information: OCH 11917  
Crop: Potato Sample ID: doi:10.18730/DSZ9  
Associated information: OCH 12011  
Crop: Potato Sample ID: doi:10.18730/DT0A  
Associated information: OCH 12012  
Crop: Potato Sample ID: doi:10.18730/DT1B  
Associated information: OCH 12013  
Crop: Potato Sample ID: doi:10.18730/DT2C  
Associated information: OCH 12016  
Crop: Potato Sample ID: doi:10.18730/DT4E  
Associated information: OCH 12018  
Crop: Potato Sample ID: doi:10.18730/DX72  
Associated information: OCH 13166  
Crop: Potato Sample ID: doi:10.18730/EDG5  
Associated information: OCHS 14983  
Crop: Potato Sample ID: doi:10.18730/EDH6

---

*ANNEX 1*

---

**LIST OF MATERIALS PROVIDED**

This Annex contains a list of the **Material** provided under **this Agreement**, including the associated information referred to in Article 5b.

This information is either provided below or can be obtained at the following website URL:  
[https://www.crop-diversity.org/mgis/system/files/smta/order\\_646\\_7cc41da54e1e.pdf](https://www.crop-diversity.org/mgis/system/files/smta/order_646_7cc41da54e1e.pdf).

The following information is included for each **Material** listed: all available passport data and, subject to applicable law, any other associated, available, non-confidential descriptive information.

ITC Code / DOI	Accession Name	Biological Status	PGRFA under Development	Sp. / Group	Sp. / Subgroup	Acquisition date	Donor country	Origin
ITC0127 DOI: 10.1873 0/9JC7C	Kamaramasenge	traditional cultivar or landrace	no	AAB	subgr.Kamaramasenge	07 Feb 1987	BDI	ZZZ
ITC0277 DOI: 10.1873 0/9JNYV	Leite	traditional cultivar or landrace	no	AAA	subgr. Rio	16 Oct 1988	GLP	ZZZ
ITC1452 DOI: 10.1873 0/9MVAV	Huti(Shumbanyeelu)	traditional cultivar or landrace	no	AA	subgr. Mshare	01 Dec 2002		TZA

Image 2: Cryogenic vial and testing tube with CIP labels (including DOIs) for distribution



## Inclusion of GLIS DOIs in public databases with passport, characterization, evaluation and genomic sequence information

GLIS DOIs are displayed on the websites of all eleven Centers when users search for and order germplasm conserved in the genebanks<sup>1</sup>. GLIS DOIs are also included in Genesys (<https://www.genesys-pgr.org/welcome>) for all the CGIAR genebank accessions which have been assigned DOIs. These online public databases (i.e. the ones hosted by individual Centers' websites, and Genesys' database) include passport information and, when available, characterization and evaluation data, and images for accessions conserved in the genebanks. .

In addition, GLIS DOIs of Musa accessions maintained by Bioversity International are included in the databases of the Musa Germplasm Information System (MGIS, <https://www.crop-diversity.org/mgis/>).

CIMMYT has included DOIs of all wheat accessions maintained in CIMMYT genebank in the databases of Germinate, which will be publicly available shortly. An example (not yet published) of the Germinate Wheat Database is provided in Image 3 below. CIMMYT is also in the process of including DOIs of maize accessions (<https://ics.hutton.ac.uk/get-germinate/>).

Image 3: Screenshot of Germinate Webpages showing information (including DOI) for a wheat accession maintained by CIMMYT.

**Passport Data**

GID: 15750

Accession name: CMI 15905

Persistent unique identifier: 1515750-CIMMYT

Accession number: 15750

Institute code: ME002

Collecting institute address: Km. 45 Carretera Mexico-Veracruz, El Batán, Tlaxiaco, Edo. de Mexico, CP 86237

Genus: *Triticum*

Species: *turgidum*

Species authority: L.

Subtaxon: subsp. *dicoccum*

Subtaxon authority: ((Schrank) Thell.)

Acquisition date: 1984/10/1

Country of origin: Mexico

**Institution Data**

Id: 43

Name: International Maize and Wheat Improvement Center

Acronym: CIMMYT

Code: ME002

Country: Mexico

Contact: Wheat Germplasm Bank (<http://www.cimmyt.org>)

Phone: 52-55-5524-2004

Email: [CIMMYT.CMA@pgr.org](mailto:CIMMYT.CMA@pgr.org)

Address: Km. 45 Carretera Mexico-Veracruz, El Batán, Tlaxiaco, Edo. de Mexico, CP 86237

**Synonyms**

No data found.

**Pedigree**

No data found.

**Entity data**

No data found.

**Location Data**

Location: Mexico

Country of origin: Mexico

Latitude: 19.75

Longitude: 99.45

Elevation (m): 127.00

Location type: Collecting site

<sup>1</sup> The genebanks of CIAT, CIMMYT, CIP and IITA use GRIN-Global to manage genebank data. The genebanks of AfricaRice, ICRISAT, ILRI and IRRI are in the process of adopting GRIN Global.

**Additional Attributes**

Id	Name	Description	Data type	Value
158118	CID	Cross ID	int	155788
158119	SID	Selection ID	int	0
158120	ACCENAME	Accession Name: An MCPD term that describes either a registered or other designation given to the material received other than the donor accession number or collecting number	char	T.DICOCOCN P194820
634194	GLIS DOI	Digital object identifier assigned to germplasm through the Global Information System of the International Treaty on Plant Genetic Resources for Food and Agriculture that is maintained by the FAO. More information is available at <a href="https://ssl.fao.org/glis/">https://ssl.fao.org/glis/</a>	char	<a href="https://dx.doi.org/10.18730/CKNN1">https://dx.doi.org/10.18730/CKNN1</a>

**Annotations**

When logged in users can add annotations to this item which then become publicly available. Although there are only a few note classes defined here this can be customised based on the platform installation. You can add annotations below to test this feature. These notes are removed periodically from this test system.

No data found.

Add annotation

Select comment type

line annotation

Enter comment message

Submit

**External Links**

- [GRIN-Global CIMMYT Wheat Germplasm Bank Record](#)
- [Genesys Record for Global Germplasm Bank Accessions](#)
- [GLIS \(Global Information System\) Record](#)

In addition to passport information and evaluation data, these public databases (MGIS, IRIC and Germinate) provide phenotypic and genotypic data when they are available. For wheat genetic resources at ICARDA, the genotyping information is also included in Germinate along with CIMMYT data. CIMMYT and CIP have plans to include DOIs in public datasets they maintain at Dataverse (<https://dataverse.org/>). ICARDA is also planning to include the available genotyping data of its genetic resources in GIGWA.

Some Centers have started to include GLIS DOIs in online platforms that focus primarily on genomic sequence information: Bioversity has included accession DOIs in the Banana Genome Hub (<https://banana-genome-hub.southgreen.fr/>) when sequenced genomes come from Bioversity Musa genebank. An example is provided in Image 4 below.

IRRI is working towards extending the use of DOIs to the rice-specific platform SNP-Seek (<http://snp-seek.irri.org/>). AfricaRice is preparing a new website where genetic information about 3000 rice accessions maintained in the genebank (with GLIS DOIs) will be made publicly available, and linked to the GLIS.

Image 4: Screenshot of Banana Genome Hub Webpage showing information (including DOI) for a Musa accession maintained by Bioversity.

**Banana Genome Hub**  
A Next-Generation Information System for Musa genomics

Home Musa Genomes Search Genome Browsers Additional tools Download Help Germplasm Breeding Log in

Home » Musa Genomes

**Overview**

Full Name	Musa acuminata
Genus	Musa
Species	acuminata
Common Name	musa
Abbreviation	M. acuminata

**DH-Pahang** (doi:10.18730/9MZSP) is a doubled-haploid of the germplasm collection accession named Pahang (doi:10.18730/9K93B). Pahang was collected in Malaysia's Pahang province in the late 1940s. It is a wild accession that belongs to subspecies *Musa acuminata* ssp. *malaccensis*, whose genetic signature is commonly found in dessert and cooking bananas.

DH stands for Doubled Haploid which refers to the induced doubling of the chromosomes in a haploid cell (which has one set of chromosomes). In this case, the haploid cell was pollen. The doubled-haploid (DH Pahang) was produced through anther culture and spontaneous chromosomes doubling.

The *Musa acuminata* reference genome sequence results from collaboration between Genoscope and CIRAD (UMR AGAP) funded by ANR. The sequence was analyzed in collaboration with several teams within the **Global Musa Genomics Consortium (GMGC)** and was published as [version 1](#) in:

D'Hont A, Denoeud F, Aury JM., Baurens FC, Carreel F., Garsmeur O, Noel B, Boes S, Droc G, Rouard M, et al. (2012) The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature*, 10.1038/nature11241.

The reference sequence was then improved by a combination of methods and datasets, leading the release of the [version 2 of the assembly and gene annotation](#), which are described in the following publication:

Martin G, Baurens FC, Droc G, Rouard M, Cenci A, Kilian A, Hastie A, Dolezel J, Aury JM, Alberti A, Carreel F, D'Hont A (2016) Improvement of the banana "*Musa acuminata*" reference sequence using NGS data and semi-automated bioinformatic methods. *BMC Genomics*.

If you have any question, please don't hesitate to [contact us](#)

**LEGAL INFORMATION**  
Disclaimer  
Privacy policy

**FEEDBACK**  
Contact us  
How to cite us

It appears that GenBank, GigaDB, EMBL-EBI and other online platforms which are widely used for uploading big data in biology and genetics are not equipped to handle GLIS DOIs. They would likely need to adjust their databases to include this information for the GLIS and GLIS DOIs to facilitate the linking and layering of genomic sequence data associated with materials from CGIAR Genbanks (and any/all other sources of sequenced germplasm for that matter).

### Inclusion of GLIS DOIs in journal articles

Given that DOIs were minted only recently, in 2018, there has not been much time for scientists to include them in papers submitted to journals for publication. That said, scientists from a few Centers have already published peer reviewed journal articles which include GLIS DOIs of the germplasm involved in the work described in the articles. Two of those articles include genetic information (Seed Box 1 below). Additional papers including DOIs have been submitted to journals and are being reviewed. There are significant challenges – described in the next section -- to including DOIs in journal articles, and having them 'findable' by GLIS.



#### Box 1: Journal articles including DOIs.

Wang DR, Agosto-Perez FJ, Chebotarov D, Shi Y, Marchini J, Fitzgerald M, McNally KL, Alexandrov N, McCouch SR (2018) An imputation platform to enable efficient integration of rice genetic resources. *Nature Communications*, 7.

This article describes the development of a web-based service called the Rice Imputation Server, which allows leveraging haplotype information from more than 4500 genetically and geographically diverse rice accessions to impute rice datasets using the IMPUTE2 software. The information from the more than 4500 rice accessions comes from the Rice Diversity Project and the Three Thousand Genomes Project. The DOIs are listed in the Supplementary Table online at: [https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-018-05538-1/MediaObjects/41467\\_2018\\_5538\\_MOESM4\\_ESM.xlsx](https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-018-05538-1/MediaObjects/41467_2018_5538_MOESM4_ESM.xlsx)

Ellis D, Chavez O, Coombs J, Soto J, Gomez R, Douches D, Panta A, Silvestre R, Anglin NL (2018) Genetic identity in genebanks: application of the SolCAP 12K SNP array in fingerprinting and diversity analysis in the global in trust potato collection. *Genome*, 61:523-537.

This article describes the work and results of fingerprinting two hundred and fifty accessions from the cultivated potato collection at the CIP, using the Infinium 12K V2 Potato Array, and with the purposes to confirm genetic identity of the accessions and evaluate genetic diversity of the potato collection.

Whitehouse KJ, Owoborode OF, Adebayo OO, Oyatomi OA, Olaniyan AB, Abberton MT, Hay F. (2018). Further evidence that the Genebank Standards for drying orthodox seeds may not be optimal for subsequent seed longevity. *Biopreservation and Biobanking*, 16 (5).

The aim of the study described in the article was to test the effects of drying freshly harvested seed at a higher temperature (40°C–60°C) than the temperature recommended by the current genebank standards (5°C–20°C) on the longevity of the seed. Seeds of five cowpea accessions and four soya bean accessions were used in the tests.

## Challenges and opportunities

Challenges involved in minting DOIs for genebank accessions and breeding lines include identifying the best timing and manner. Centers have faced technical questions such as how many different DOIs should be assigned to an accession: Should each regeneration have its own DOI? Should the same material conserved under different conditions and forms be given different DOIs based on the location and type of conservation (seed, cryopreservation, DNA, herbarium)? In addition, some Centers have found practical problems of minor nature and able to be overcome, for example missing information required by GLIS to generate DOIs, such as date of acquisition.

As alluded to in the previous section, Centers (and likely all other DOI users) are encountering challenges referencing GLIS DOIs in journal articles. A major attraction of the DOI system is the potential to associate GLIS DOIs for PGRFA with DOIs for associated information, such as DOIs for publications and online datasets, thus bringing associated information within the scope of GLIS and ensuring that it does in fact remain associated with the PGRFA it describes. However, the technology to do so automatically is currently unable to retrieve DOIs that are embedded in the main body of a publication, including text, tables, figures and supplementary information. Currently, to enable GLIS to automatically discover the GLIS DOIs referred to in a publication, all those GLIS DOIs must be listed in the reference section. However, listing PGRFA in the reference section is not standard practice, so journal editors need to develop and implement appropriate new standards. CGIAR scientists are working closely with journal editors and with the ITPGRFA Secretariat to implement appropriate standards for listing GLIS DOIs in references; and in the longer term to seek an improved



solution. Of course, in the meantime, users can easily upload to GLIS links to associated information in any of the ways developed by the ITPGRFA Secretariat.

One challenge to linking germplasm accession DOIs and genetic sequence information arises from the fact that there is considerable genetic heterogeneity of material within accessions, particularly for outcrossing and inbred crops (less so with clonally propagated crops) or even for landraces of self-pollinated crops (composed of multilines). To address this complexity, it is necessary to purify single lines (via selfing and single seed descent) to develop materials for genomic sequencing. The purified line is not representative of the entire accession and likely needs to be handled as a unique accession with a specific purified genotype. The good practice to address this reality is to assign a new accession number and DOI for the purified line, and then associate the genetic sequence data with that line via the new accessions number and DOI. GLIS DOIs will forge a traceable connection between the purified line and the original accession. The sampling procedure for DNA extraction and the way genetic sequence information is reported (missing data imputed or not, etc.) are not necessarily straightforward and require considerable expertise and resources to address. Seed from that same purified accession should therefore be used for phenotyping for association analyses and could use the assigned DOI as a common reference.

Just as it is important for CGIAR Centers to follow this practice, it is equally important for recipients of materials from CGIAR Centers to assign new DOIs to purified lines that they sequence. If they do not, and they only refer to the DOI of the accession transferred to them by the Center, they will lose the value of a direct connection between sequence information and underlying genetic material (i.e. the purified line).

Some Centers have made efforts to raise awareness about DOIs and their value among different users beyond CGIAR. For example, when Bioversity released an updated version of the MGIS website, in March 2018, the website informed the visitors about the introduction of DOI in the passport data (<https://tinyurl.com/y9a3otpj>). Since February 2019, there is a dedicated webpage explaining DOIs (<https://tinyurl.com/y6ch3oa9>). This page is accessible from any page of the MGIS web site through a little clickable DOI logo visible at the bottom left of all web pages. All users of MGIS are encouraged to use the DOIs when citing germplasm from the MGIS database. Other Centers like ICRAF are considering this approach for the databases they maintain. All shipments of PGRFA from CIP include information for users about DOIs.

Once the value of DOIs is demonstrated and realized, and the communities working with genetic resources and genomic data integrate DOIs in their work, journal editors may be more willing to consider the possibility of facilitating the inclusion of DOIs within journal articles, in a way that makes the searches and linkages with the GLIS easier.

## Plans for further work promoting use of DOIs by CGIAR Centers' genebanks

Over the course of 2019, the eleven CGIAR Centers that have signed agreements with the Governing Body of the ITPGRFA under Article 15 will receive support from the CGIAR Genebank Platform to focus on full integration of DOIs into genebank activities. During 2019, each genebank will develop, document, and implement a protocol for the routine assignment and use of GLIS DOIs for genebank materials, to:

- Assign DOIs to new accessions, as a routine element of the standard operating procedures to process incoming samples;
- Determine whether and when to assign GLIS DOIs to parts of accessions such as individual harvests, seed lots or tissue samples, as appropriate for the genebank;
- Ensure that the list of PGRFA given in Annex 1 of outgoing SMTAs specifies GLIS DOIs for the material shipped
- If the genebank opts to identify material shipped out with an SMTA by individual harvests or seed lots, ensure that the SMTA and the shipping standard operating procedures accommodate minting of new GLIS DOIs where appropriate
- Ensure that all future publications and online databases and datasets on genebank accessions refer to the accessions by their GLIS DOI, preferably in a form that enables GLIS to discover those online resources automatically
- To the extent possible, identify online information resources (e.g. datasets, pages of databases, or publications) relating to existing genebank accessions with GLIS DOIs but not correctly referenced by GLIS DOIs, and associate GLIS DOIs and add them to GLIS
- Encourage the use of original GLIS DOIs (and minting of GLIS DOIs for distribution) beyond the genebank

Centers will collect best practices linked to implementing *Digital Object Identifiers for food crops - Descriptors and guidelines of the Global Information System*, which was developed and published by the ITPGRFA Secretariat in 2018, and will make suggestions for further revisions of that document.

The Centers will act collectively to request GLIS to approach journals to make the publication of DOIs mandatory in articles relating to genebank accessions.