IITA Transforming African Agriculture CGIAR	**Crop: Cowpea Function: Sampling and Genotyping**	**SOP #**	IITA-CP-SOP13
		Revision #	IITA-CP-SOP13-1
		Implementation Date	15/05/2022
Page #	1 of 14	**Last Reviewed/Update Date**	22/03/2022
SOP Owner	Breeder (Patrick O. Ongom)	**Approval Date**	10/05/2022

Standard Operating Procedure (SOP) for leaf sampling and genotyping

Authors & Contributors

Patrick Obia Ongom; P.Ongom@cgiar.org
Abou Togola; a.togola@cgiar.org
Christian Fatokun; c.fatokun@cgiar.org
Saba Baba Mohammed; s.mohammed@cgiar.org
Ousmane Boukar; o.boukar@cgiar.org

1. Introduction

Genotyping is the process of determining differences in the genetic make-up (genotype) of an individual by examining the individual's DNA sequence using biological assays and comparing it to another individual's sequence or a reference sequence. It reveals the alleles an individual has inherited from their parents. Effective deployment of molecular markers in breeding requires use of good quality DNA which begins with a good plant tissue sampling and preparations before going through the downstream processes of DNA extraction and eventual genotyping, irrespective of the genotyping methods used. In cowpea, genotyping is commonly done for various purposes including testing germplasm purity, authenticating hybridity of F1s, discovery of quantitative traits loci (QTL) and genomic-aided selections. The SOP describes the general processes involved in sampling

leaf tissues, processing the tissues and shipping where DNA will be extracted and genotyped, including handling of genotype data.

2. Purpose

The purpose of this document is to outline the roles, responsibilities, and steps involved in cowpea genotyping activities.

3. Scope

This document contains outlines of steps involved cowpea genotyping activities. It covers array events from preparing genotyping list, planting the seedlings, sampling leaf tissues, processing the samples and shipment, actual genotyping, genotype data processing and analysis, reporting and decision making.

4. Definition of terms

DNA, or deoxyribonucleic acid, is the hereditary material in almost all living organisms. It carries the genetic instructions used in growth, development, functioning, and reproduction. DNA is composed of two long strands forming a double helix, held together by base pairs.

Molecular markers are specific sequences of DNA that can be identified and used to distinguish between different genetic materials. These markers are valuable tools in genetics, genomics, and breeding programs because they provide a means to link specific traits or variations to particular regions of the genome. Molecular markers can be used for various purposes, including genetic mapping, marker-assisted selection, and the study of genetic diversity.

A single nucleotide polymorphism (SNP, pronounced "snip") is a variation at a single position in a DNA sequence among individuals. Each SNP represents a difference in a single DNA building block, called a nucleotide. For example, a SNP might change the DNA sequence from AAGCCTA to AAGCTTA

DNA, or deoxyribonucleic acid, is the hereditary material in almost all living organisms. It carries the genetic instructions used in growth, development, functioning, and reproduction. DNA is composed of two long strands forming a double helix, held together by base pairs.

Genotyping is the process of determining the genetic makeup of an individual by examining their DNA sequence and identifying specific genetic variations. Genotyping focuses on identifying variations at specific locations in the genome, such as single nucleotide polymorphisms (SNPs), insertions, deletions, and other types of genetic markers.

Sampling for genotyping involves collecting biological material from which DNA can be extracted and analyzed to determine genetic variations. Proper sampling is crucial to ensure the accuracy and reliability of genotyping results.

Hybridity testing is a quality control method used to determine whether a plant is a hybrid, meaning it is the offspring of two genetically distinct parents. In plant breeding, hybridity testing is particularly important for verifying the genetic purity of hybrid seeds, ensuring that they possess the desired combination of traits from both parent lines.

Quantitative Trait Locus (QTL) mapping is a statistical method used to identify the specific regions (loci) of the genome that are associated with variation in quantitative traits.

Genome-Wide Association Studies (GWAS) are a powerful tool used to identify genetic variants associated with complex traits and diseases in populations. Unlike traditional linkage studies that focus on familial relationships, GWAS analyzes genetic variation across the entire genome in a large group of unrelated individuals to find associations between specific genetic markers and traits.

5. Roles and Responsibilities

All staff involved in implementing breeding activities in the cowpea improvement program at IITA must use the genotyping SOP. No alteration should be made to the procedures unless approved exceptionally by the program leaders. The list of individuals responsible for each section of the SOP are listed below.

Crop Lead Breeder (CLB): Responsible for oversite and advise on genotyping materials and populations to be used. The CL could be a lead breeder, a principal investigator (PI), or a cluster lead.

Molecular Breeder: Responsible for designing and directing all genotyping activities. Molecular breeder should also identify relevant genotyping methods and services. He is also responsible for sending quality samples to service providers for genotyping.

Technicians: Responsible for implementing activities related to genotyping, especially establishing the seedling nursery, sampling, and processing tissues.

6. Procedure

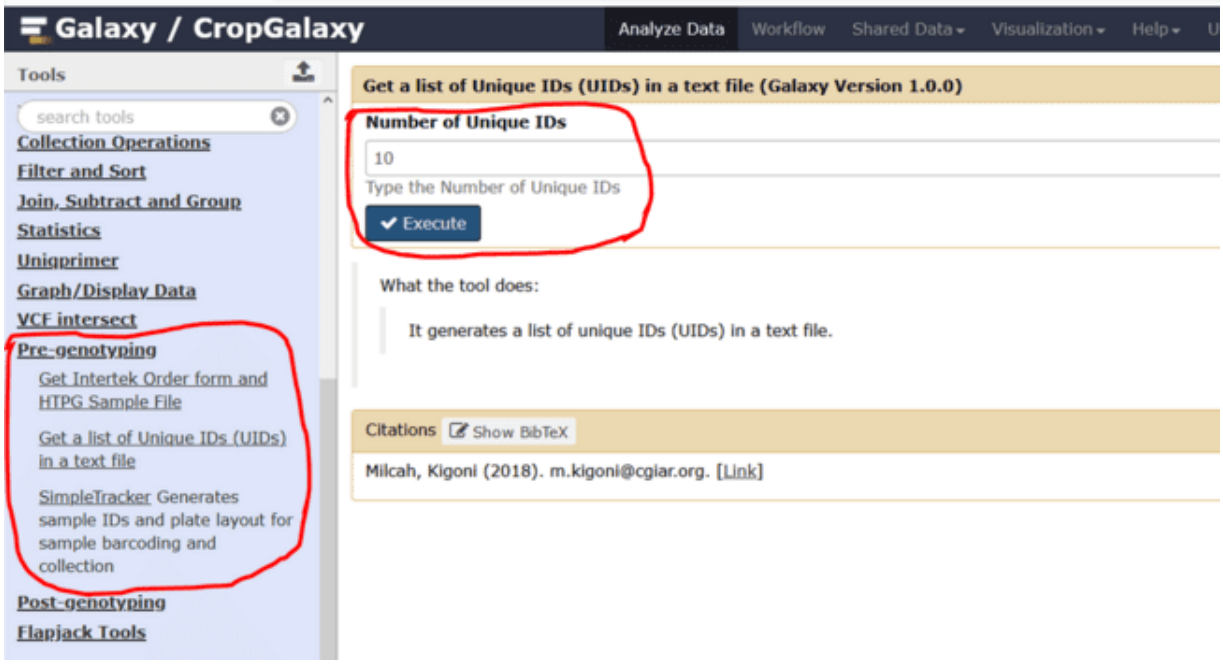
Genotyping is a key activity in cowpea breeding and must be guided by the steps outlined below:

Generating sample list

Genotyping for any purpose begins with defining an apprpriate population to be used from which a sample list is generted. In cowpea, genotyping is often done on the follwing types of populations i) Germplasm, parental lines or varieties to be fingerprinted for

future identy and for purity assessment ii) Populations of F1 progenies and their parents to be tested for hybridy iii) Baccross, F2, and Recombinanat Inbred Lines (RILs) to be used for QTL mapping and marker assisted selection (MAS) iv) Breeding lines for genomic selection.

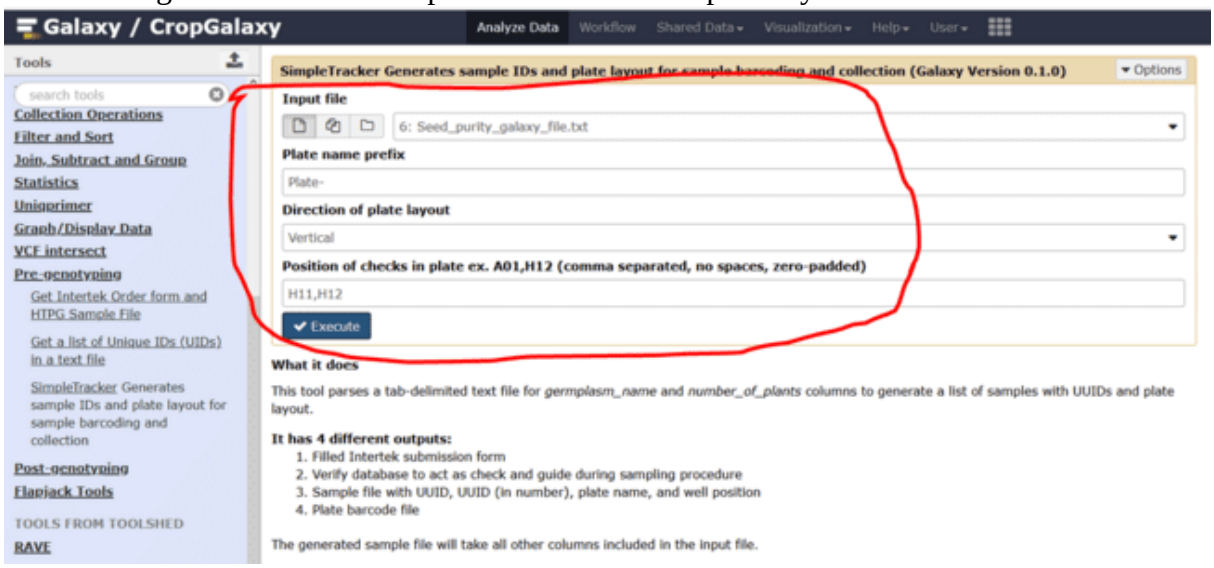
- For all these populations, the process begin by generting unique idenfiers (UIDs) for the samples to be genotyped.
- Unique IDs (UIDs) should be generated from the EiB galaxy website http://cropgalaxy.excellenceinbreeding.org/
- Utilize the pre-genotyping tool on this site to create UIDs. Three options are avaliable in this section:
  i) Intertek Order form: use this to get a template excle file for sending samples specifically to Intertek labs
  ii) Get a list of unique IDs: to generte random unique identifers for each sample. Enter the number of samples and click Execute.



Screenshot of cropgalaxy page showing pregenotyping tools and how to generate unique IDs.

iii) Sample tracker: Use this option to the generate IDs and plate layout for samples. This option requires uploading a tab-delimited text file with two columns; germplasm_name and number_of_plants. This tool parses a tab-

delimited text file for germplasm_name and number_of_plants columns to generate a list of samples with UUIDs and plate layout.



Screenshot of galaxy showing SampleTracker option

Sample ID (Subject ID) will be mentioned in the final report.
Plate ID: Unique identifier no longer than 21 characters long and excluding "," (comma).
Wells H11 and H12 should be left empty (contain no text).

Plate	IITA35_720_Plate-01											
	1	2	3	4	5	6	7	8	9	10	11	12
A	1940-1	573-1-1	P1	P2	P3	P4	P5	P6	P7	P8	P9	2130-1
B	573-1-1	P1	P2	P3	P4	P5	P6	P7	P8	1862-1	576-1-1	P1
C	P2	2791-1	573-1-1	P1	P2	2265-5	P1	P2	P3	P4	P5	2360-3
D	P1	P2	P3	P4	P5	P6	P7	2014-4	P1	P2	P3	P4
E	P5	P6	P7	P8	P9	1941-1	573-2-1	P1	P2	P3	P4	1960-5
F	P1	P2	P3	2354-4	573-1-1	P1	1734-1	P1	P2	P3	P4	P5
G	1983-4	P1	P2	P3	P4	P5	2052-1	573-2-1	P1	P2	P3	P4
H	2016-7	P1	P2	P3	P4	P5	2035-1	573-1-1	P1	P2		

Plate	IITA36_720_Plate-02											
	1	2	3	4	5	6	7	8	9	10	11	12
A	P3	P4	P5	P6	P7	P8	P9	P10	P11	1908-4	573-1-1	P1
B	P2	P3	P4	P5	2247-2	573-1-1	P1	P2	P3	P4	P5	1744-1
C	P1	P2	2235	P1	P2	P3	P4	P5	2438-1	P1	P2	P3
D	P4	P5	1699-4	573-1-1	P1	P2	P3	P4	P5	P6	P7	P8
E	P9	1852-2	573-1-1	P1	P2	P3	P4	P5	P6	P7	P8	P9
F	P10	P11	P12	1678-1	P1	P2	P3	1942-1	573-2-1	P1	P2	P3
G	1721-2	573-1-1	P1	P2	P3	P4	2553-4	573-2-1	P11	1885-1	573-2-1	P1
H	P2	2091-2	573-2-1	P1	2559-8	573-1-1	P1	P2	P3	P4		

Example plate layout from Sample Tracker tool of EiB CropGalaxy Website

Seedling establishement

- Treat seeds of materials chosen for geneotyping with approprite seed dressing chemical, normally, using Apron Plus at the rate of 10 g/4–5 kg of seeds (1 sachet), or Apron Star 42 WS at the rate of 10 g/8 kg of seeds/1 sache
- Plant seeds in pots pots of size 24 cm (height) x 25.4 cm (diameter), and fill three-quarter of the pot with sterilized topsoil, place the potd the crossing benches in the screenhouse
- Plant two samples per pot, each sampl planted with 2 seeds in a hill
- Water the pots well and after a week, apply NPK fertilizer at a rate of 0.5g per pot
- Inspect the seedlings for presence of white flies and spray accordingly

Leaf sampling

(a) Required materials and tools: The following are required for a good leaf sampling process:

- Plates and seals or mats. A 96-well plate, 1.2 ml AbGene Storage Plate (AB056) is required together with Sealing Mats (AB0674) (see leaf sampling protocol by Milcah Kigoni)
- Plate labels. The labels should be prepared prior to sampling and should have human readable plate ID, including the range of samples in the plate. Where possible, bar codes can be used.
- Plant labels or tags. It is important that the plants being sampled are tagged for traceability/tracking. Each plant being sampled should have its own tag/label, and the labels should be able to withstand harsh field conditions.
- Sampling tool: Broad leaf plants like cowpea, should be sampled using the single hole punchers, especially the ones that produces holes that is 6 mm in diameter
- Note: Unique sample identification (UIDs) can be automatically generated from the EiB galaxy website as descrbed in step 1 above.

(b) Cowpea Leaf sampling steps: The steps outlined below are adopted from the one used for broad leaf grasses.

1. First, Identify and tag/label the plants to be sampled.
2. Clean the single hole-puncher or the sampling tool with 50-80% ethanol. It is recommended that cleaning is done before each new sample is taken.
3. Scan or write down the plate ID

4. Scan or record the ID of the first plant to be sampled. Scanning is possible when use is being made of bar codes and Coordinate App (see leaf sampling protocol by Milcah Kigoni)
5. Select a young, healthy newly developed trifoliate leaf, punch and Collect 2 leaf discs per sample (per well) and not more!
6. Carefully transfer the leaf punches in one well using forceps, ensure the punches are at the bottom of the well.

Cowpea leaf being sampled and trasferred directly into a 96-well plate

7. Close the box with the sealing mat (or a temporary cling film) and place the plate in an ice box or cool place as you continue to sample more plates
8. Note:
    i. Control wells (H11 and H12 for HTPG/Intertek or G12 and H12 for DArT) should be left blank
    ii. Ensure the plate is in the right orientation before you start sampling, by noting the position of the letters and number
    iii. iMark all control wells with a marker pen before you start sampling. You can also mark wells with missing samples like this.

(c) Sampling leaf from the laboratory: Sometimes it can be difficult to sample leaf tissues in the field for reasons such as unfavorable weather conditions or insufficient logistics. Below, we have outlined the steps for leaf sampling in the laboratory:

1. Begin by preparing well labeled envelopes, showing plot number, germplasm ID, plate ID and well ID. The labels should match the tags on the actual plants to be sampled
2. Go to the field and select a young, healthy, tender leaf and place in the correct envelope. Note that, no punching is done in the field, but the entire leaf is plucked and placed in the envelope.
3. Organize and print the sample list with corresponding plate and well location
4. Ensure the ID on the envelope corresponds to the sample ID on the sample list
5. If all is in order, remove leaf from envelope and place on cleaned mat
6. With your single-hole puncher, collect the sample by gently but firmly, punching, twist and lift the sample
7. Transfer 2 leaf discs per well and not more for samples sent to Intertek-Agritech Labs
8. Close the box with the sealing mat (or a temporary cling film) and place the plate in an ice box or cool place as you continue to sample more plates
9. Note:
   i. Clean your sampling gadget with 50-80% Ethanol before you start sampling and between samples
   ii. Avoid sampling the midrib section
   iii. You can work faster by folding the leaf once to get 2 punches at a go but ensure you get 2 leaf discs

Sample processing and shipment

The steps below depict what happens to the samples after it has been collected in well labelled plates:

1. Dry the samples either by lyophilizing or using a dry oven at 40-50 ºC for 12-24 hours. Plates can be covered with seals during drying or one can use cling film
2. Seal the plates carefully and firmly. You can wrap the plates in a plastic bag(s) and secure the plastic bag with2rubberbands.Use Intertek recommended sealing mats (AB0674)
3. Complete filling the Intertek Order form and email a copy to the Selected Intertek-Agritech lab (see step 1 on “How to get the HTPG sample file and Intertek Order Form”)
4. Prepare a Commercial Invoice. A sample of the commercial invoice is provided below.

Company Logo

COMMERCIAL INVOICE

DATE OF EXPORTATION	IMPORT PERMIT NO.	EXPORT REFERENCES (ie order no, invoice no, etc)
	N/A	

SHIPPER / EXPORTER	CONSIGNEE
	Intertek-ScanBi Diagnostics AB Elevenborgs SV.2 SE-23053 Alnarp SWEDEN
COUNTRY OF EXPORT	IMPORTER *(DELIVER TO)*
	Intertek - ScanBi Diagnostics AB
COUNTRY OF ULTIMATE DESTINATION	Elevenborge SV.2
Sweden	SE-23053 Alnarp
COUNTRY OF ORIGIN OF GOODS	SWEDEN

Purpose: For analytical purposes only, not for breeding, will be destroyed

INTERNATIONAL AIRWAYBILL NUMBER	CURRENCY

NO. OF PKGS	TYPE OF PACKAGING	FULL DESCRIPTION OF GOODS	COMMODITY CODE	WEIGHT	QTY. (Envelopes)	UNIT OF MEASURE	TOTAL VALUE
							$ 3,00
							$ -
							$ -
							$ -
							$ -
SUB-TOTAL							$ 3,00

	TOTAL INVOICE VALUE $ 3,00

RETURN TO : SHIPPER/EXPORTER

I declare all the information contained in this invoice to be true and correct

Example of a Commercial Invoice

5. Print the following documents and ship them together with sample plates:
    a. Commercial invoice,
    b. Import permit: provided by Intertek labs, and
    c. Filled Intertek Order form. Also, indicate the purpose statement: E.g. "***For analytical purposes only, not for breeding, will be destroyed".***

6. Place the sample plates and a copy of the printed order form in a package in preparation for shipping

Dry cowpea samples ready for shipmenet

7. Attach/stick the printed import permit, commercial invoice and the statement of purpose on the surfaces of the shipping box/carton
8. Ship the samples to Intertek – Agritech Lab or DArT facility Australia. For now, our genotping services is provided by Intertek – Agritech Lab or DArT, this may change in future.
9. The shipping aderesses are:

INTERTEK SWEDEN LAB:
Elevenborgsvägen 2
230 53 Alnarp - Sweden
Phone: +46 40 69 28 001
agritech.sweden@intertek.com

INTERTEK AUSTRALIA LAB:
170 Greenhill Road
Parkside South Australia - 5063 Australia
Phone: +61 8 8301 1900
agritech.australia@intertek.com

INTERTEK INDIA LAB:
D-53, Phase-I, IDA, Jeedimetla
Hyderabad - Telangana 500 055 India
Phone: +91 40 2319 5257
agritech.india@intertek.com

Handling genotyping data

- The genotype data can be for different purposes, quality control, qtl mapping, marker-assisted breeding or genomic selection. In any either cases, genotyping report will return from the service provider with data presented in different formants: VCF format, intertek format (similar to hapmap format), and counts format. Details of the formats are often explained in a report file named "metadata". The metadata will also list the samples that failed and thus not reported in the data (It is important to not this).The intertek format is easy to manipulate in microsopft excel.
- Once data is receiced, it is impotant to allign the unique IDs with the actual sample names before any analysis can be done.

QC/QA of F1 hybridity:

1. Identify markers that are polymorphic between the parents of a cross and use the polymorphic markers to track parental alleles among the F1 progenies from that cross (See example below).

SUBJECT_ID	SAMPLE_ID	Type	#polymorphic markers	#Polymorphic loci heterozygous	%Polymorphic loci heterozygous	Hybridity
b68ac4d32e6f	**1940-1**	Parent	**7**			
8398b2d66218	**573-1-1**	Parent	**7**			
06c90b2918e1	P1	F1	7	7	100%	True F1 hybrid
d8971d812a02	P2	F1	7	NA	NA	Undetermined:Missing data
fe ae04fa381a	P3	F1	7	NA	NA	Undetermined:Missing data
e4423faf7df2	P4	F1	7	NA	NA	Undetermined:Missing data
bcbbe fb62fbf	P5	F1	7	7	100%	True F1 hybrid
b9f2470851ce	P6	F1	7	7	100%	True F1 hybrid
688abdaf3a85	P7	F1	7	7	100%	True F1 hybrid
a69bccb222cf	P8	F1	7	7	100%	True F1 hybrid
aba3d4f4211d	P9	F1	7	7	100%	True F1 hybrid
49fe 2ac9d472	2791-1	Parent	6			

Not a true F1

SNP	IT97K-573-1-1	Danila	F1-1	F1-2	F1-3	F1-4	F1-5
SNP1	C:C	T:T	C:T	C:T	C:T	C:C	C:T
SNP2	G:G	A:A	A:G	A:G	A:G	G:G	A:G
SNP3	A:A	G:G	A:G	A:G	A:G	A:A	A:G

Tracking parental alleles in the F1 progeny and declearly true hybridity of putative F1s

2. Determine the number of polymorphic markers that will detect a partiular F1 as being heterozygour (i.e. the F1 has alleles from both parents)
3. Compute hybdidty percentage (Ongom et al 2021).

$$Hybridity = \left(\frac{L_{het}}{Pm}\right) * 100$$

L_{het} =the number of polymorphic SNPs detecting an F_1 as heterozygous

Pm = number of polymorphic SNPS

4. Discard putative F1s with hybridity % age of less than 50%.

QTL mapping

- Two appraches are often used in our program: Traditional linkage mapping and GWAS.
- For traditional linkage analysis in bi-parental poulaion, use is being made of ICIM QTL IciMapping software http://www.isbreeding.net. The software mannual provides details of how to fomrat the genotype data.
    - Normally genotype data is obtained from progenies that are derived from a two parent cross.
    - Ensure that leaf samples are taken from all progenies and the two parents must also be sampled.
    - Data should be prepared in the right format acceptable by QTL IciMapping software

  - Once data is ready, the analysis undergo through varuaous stpes including: Marker data conversion and quality control, Binning of reduantand markers and creating the genetic map.The genetic map is then merged with trait data and a QTL mapping is conduted to identify QTL and flanking markers associated with the traits in question.
- For GWAS, samples are taken from special populations which does not require crossing. The populations can be a collection of diverse germplasm or breeding lines.
- Data should be in hapmap fomat for analysis using Tassel, GAPIT or rMVP R package.
- GWAS involves:
  - Chosing appropiate population,
  - Sampling leaf tissues from the population followed by DNA extraction and genotyping,
  - Data quality control he accuracy and reliability of genotyping data by filtering out low-quality samples and SNPs,
  - Collecting phenotypic data on the population,
  - Statistical abalysys that invoves use of association tests to identify SNPs that are significantly associated with the trait or disease.
  - Replication and validation: Replicate significant findings in independent cohorts to confirm the associations.
  - You can then annotate significant SNPs to identify the genes they are located in or near and predict their potential impact on gene function.

7. Forms/Templates to be used for monitoring and data collection

Contacts for support

Breeding support: Ousmane Boukar; O.Boukar@cgiar.org

Molecular breeding support: Patrick O. Ongom; P.Ongom@cgiar.org

Entomology: Abou Togola; a.togola@cgiar.org

Genetic support: Christian Fatokun; C.FATOKUN@CGIAR.ORG

Physiology support: Saba Baba Mohammed; s.mohammed@cgiar.org

8. Further reading

Ongom P.O, Fatokun C, Togola A, Salvo S, Oyebode OG, Ahmad MS, Jockson ID, Bala G and Boukar O (2021) Molecular Fingerprinting and Hybridity Authentication in Cowpea Using Single Nucleotide Polymorphism Based Kompetitive Allele-Specific PCR Assay. Front. Plant Sci. 12:734117. doi: 10.3389/fpls.2021.734117

EiB. How to generate UIDs using galaxy on Excellence in Breeding platform website. http://cropgalaxy.excellenceinbreeding.org/?tool_id=UIDs_generator&version=1.0.0&__identifer=qe1ge9awzzp