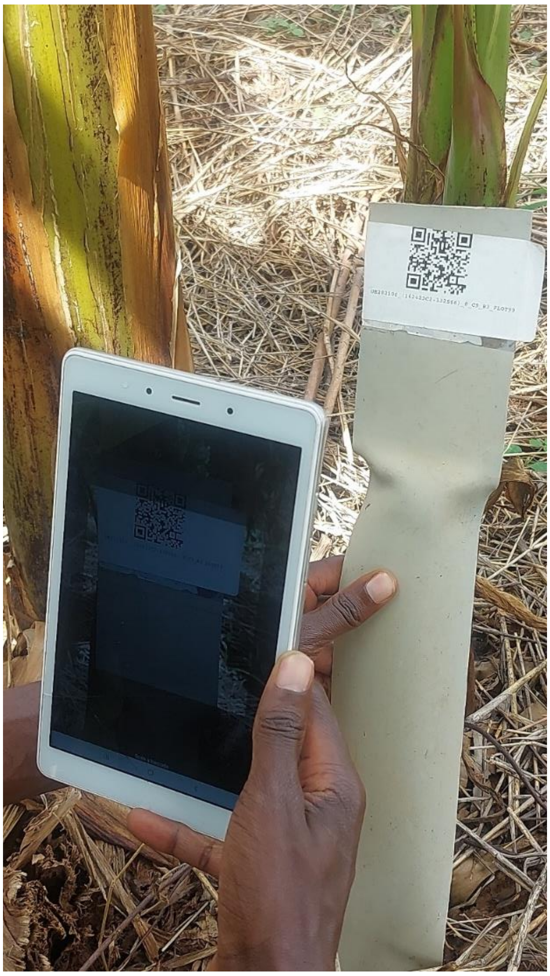
April 2025

SOP20:

Standard Operating Procedure (SOP) for Phenotypic Field data collection and management for hybrids evaluation



	Crop: Banana **Function:** Phenotypic Field data collection and management for hybrids evaluation	**SOP #**	20
		Revision #	
		Implementation Date	8th April 2025
Page #	1 of 8	**Last Reviewed/Update Date**	8th April 2025
SOP Owner	Stanley Bayo and Ronald Kivumbi	**Approval Date**	8th April 2025

Standard Operating Procedure (SOP) for Data Collection, Analysis and Statistical Design in Banana Breeding

Authors & Contributors

S. Bayo*, R. Kivumbi (R.Kivumbi@cgiar.org), R. Swennen (R. Swennen@cgiar.org), M. Batte (M.Batte@cgiar.org), M. Nyine (M.Nyine@cgiar.org), A. Brown (A. Brown@cgiar.org)

*Correspondence: s.bayo@cgiar.org

1. Introduction

Data ***management*** in breeding operations is of paramount importance. Decisions taken at every stage of the breeding cycle need to be accurate and fast but become complex when large amounts of data are generated. Inaccuracies with data is likely when established methods for data collection and management are not followed. Mistakes will slow the speed to boost genetic benefits for smallholder farmers. In order to guarantee that all required breeding data for bananas are accurately collected and made easily available, this document outlines standard operating procedures for data collection and management in banana breeding. This document will be updated when new knowledge is generated.

2. Purpose

This SOP provides detailed procedures for data collection, analysis, and statistical design for hybrid evaluations in the banana breeding program. It ensures data quality, consistency, and the application of appropriate statistical methodologies.

3. Scope

This procedure applies to all personnel involved in data collection and management in banana breeding trials, including data collectors, field technicians, data managers, and statistics consultants.

4. Responsibilities

4.1 Data Collectors

- Collect daily data from assigned trial
- Submit weekly data to the Data Manager
- Ensure proper labelling and handling of samples

4.2 Field Technicians

- Assist data collectors with the setup and organization of trials
- Maintain equipment and ensure tools are functioning properly

4.2 Data Manager

- Review submitted data for completeness and quality
- Conduct field validation if data quality is subpar
- Perform data analysis using statistical models
- Upload validated data to Musabase (https://www.musabase.org/)

4.3 Statistics Consultant

- Provides guidance on the choice of appropriate statistical models
- Assist in generating and interpreting statistical analyses

• *5. Data Collection Tools*

- Tablets: Equipped with the Field book app for data entry
- QR Codes: Fixed on every plant for easy identification and data tracking
- Tape Measure: For measuring plant height and circumference
- Weight Balance: For weighing bunches and fingers
- Meter Rule: For additional measurement accuracy

• *6. Data Collection Procedure*

6.1 Generation of trial layouts

- Trial layouts are generated in Musabase (https://www.musabase.org/)
- Barcodes for each genotype are generated and printed
- The barcodes are attached to the corresponding plants in the field
- Trial layout is exported to Field book
- All traits to be measured are entered in Fieldbook App

All traits to be measured are activated in field book before starting to collect data

6.2 Assignment of Trials

- Each data collector is assigned specific trials based on expertise and familiarity
- Assignments are communicated clearly, along with the location and specifics of the trial

6.3 Daily Data Collection

Data collectors must follow the schedule for data collection:

- Frequency: Daily
- Timing: Collect data in the morning (the coolest part of the day) to minimize stress on plants and ensure accurate measurements

6.4 Data Submission

- Data must be submitted weekly to the Data Manager
- Format: Data should be submitted via email or a shared drive in a predefined template

- Deadline: Submissions are due every Friday by 5 PM

6.5 Data Entry Guidelines

- Ensure all data is entered accurately into the Fieldbook App
- Double-check entries against physical measurements to prevent errors

• *7. Traits to be collected*

s/n	Trait list	Description
1	Bunch orientation 1-5| cycle 1 main plant| at harvesting| COMP:0000466	Observe the angle of the bunch with the pseudostem and choose the most similar option from the 5 schematic drawings of descriptor 6.4.6 in the reference material, 1 = hanging vertically; 2 = slightly angled; 3 = hanging at a 45-degree angle; 4 = horizontal; 5 = erect
2	Bunch orientation 1-5| cycle 2 main plant| at harvesting| COMP:0000473	
3	Bunch orientation 1-5| cycle 3 main plant| at harvesting| COMP:0001057	
4	Bunch weight in kg| cycle 1| at harvesting| COMP:0000453	Harvest the bunch by cutting the peduncle 10 cm above the first ridge above the most proximal hand and immediately below the most distal hand and weigh the bunch (including the rachis), using scales. In kg
5	Bunch weight in kg| cycle 2| at harvesting| COMP:0000455	
6	Bunch weight in kg| cycle 3| at harvesting| COMP:0000456	
8	Plant height in cm| cycle 1 main plant| at flowering| COMP:0000824	Measure the distance from the collar, or from the pseudostem base at the ground if the collar is not visible, to the intersection of the petioles of the two youngest leaves (leaf ranks 1 and 2), using a measuring pole or sliding ruler. In cm
9	Plant height in cm| cycle 1 tallest sucker| at harvesting| COMP:0000934	
10	Plant height in cm| cycle 2 main plant| at flowering| COMP:0000840	
11	Plant height in cm| cycle 2 tallest sucker| at harvesting| COMP:0000769	
12	Plant height in cm| cycle 3 main plant| at flowering| COMP:0000888	
13	Plant height in cm| cycle 3 tallest sucker| at harvesting| COMP:0000782	
14	Plant circumference at 100 cm from the collar| cycle 1| at flowering| COMP:0001091	Measure the circumference of the pseudostem of the plant at 1 m from the collar, or from the pseudostem base at the ground if the collar is not visible, using a tape measure. In cm
16	Plant circumference at 100 cm from the collar| cycle 2| at flowering| COMP:0001093	
18	Plant circumference at 100 cm from the collar| cycle 3| at flowering| COMP:0001095	
20	Number of suckers: count| cycle 1| "shooting"| COMP:0000088	Count number of all types of suckers of the mat.
21	Number of suckers: count| cycle 1| "harvesting"| COMP:0000087	

22	Number of suckers: count| cycle 2| "shooting"| COMP:0000090	
23	Number of suckers: count| cycle 2| "harvesting"| COMP:0000089	Count number of all types of suckers of the mat.
24	Number of suckers: count| cycle 3| "shooting"| COMP:0000092	
25	Number of suckers: count| cycle 3| "harvesting"| COMP:0000091	
26	Number of functional leaves: count| cycle 1| "shooting"| COMP:0000051	Count number of functional leaves (leaves that have 50% or more of their surface as green, healthy, photosynthetic tissue)
27	Number of functional leaves: count| cycle 1| "harvesting"| COMP:0000050	
28	Number of functional leaves: count| cycle 2| "shooting"| COMP:0000053	
29	Number of functional leaves: count| cycle 2| "harvesting"| COMP:0000052	
30	Number of functional leaves: count| cycle 3| "shooting"| COMP:0000055	
31	Number of functional leaves: count| cycle 3| "harvesting"| COMP:0000054	
32	Number of hands in bunch: count| cycle 1| COMP:0000093	Count number of hands in the bunch
33	Number of hands in bunch: count| cycle 2| COMP:0000094	
34	Number of hands in bunch: count| cycle 3| COMP:0000095	
35	Number of fingers in bunch: count| cycle 1| COMP:0000008	Count number of fingers in multiple hands
36	Number of fingers in bunch: count| cycle 2| COMP:0000009	
37	Number of fingers in bunch: count| cycle 3| COMP:0000010	
38	Finger weight in g| cycle 1| COMP:0000081	Weigh a finger, using scales. Associate the data with the Hand rank. Collect this data from six fingers in total - three fingers in the middle of the outer whorl from the second hand and from the second-most distal hand. In g
39	Finger weight in g| cycle 2| COMP:0000082	
40	Finger weight in g| cycle 3| COMP:0000083	
41	Finger external length in cm| cycle 1| COMP:0000096	Measure the length of a finger that is at the 'ready to eat' stage, along the external (dorsal) arc, excluding the pedicel and the fruit tip, using a tape measure Associate the data with the Hand rank and the Number of fingers measured. Collect this data from six fingers in total - three fingers in the middle of the outer whorl from the second hand and from the second-most distal hand. In cm
42	Finger external length in cm| cycle 2| COMP:0000097	
43	Finger external length in cm| cycle 3| COMP:0000098	
44	Finger circumference in cm| cycle 1| COMP:0000105	Measure the circumference of a finger at its widest point, using a tape measure and record the data in cm. Associate the data with the Hand rank. Collect this data from six fingers in total - three fingers in the middle of the outer whorl from the second hand and from the second-most distal hand.
45	Finger circumference in cm| cycle 2| COMP:0000106	
46	Finger circumference in cm| cycle 3| COMP:0000107	

47	Planting| date of plant| COMP:0000976	Record the date of planting following DD/MM/YYYY format
48	Flowering of the plant| cycle 1| date of| COMP:0000364	Record the date of flowering/shooting following DD/MM/YYYY format
49	Flowering of the plant| cycle 2| date of| COMP:0000365	
50	Flowering of the plant| cycle 3| date of| COMP:0000366	
51	Harvesting| cycle 1| date of| COMP:0000461	Record the date of bunch harvesting following DD/MM/YYYY format
52	Harvesting| cycle 2| date of| COMP:0000462	
53	Harvesting| cycle 3| date of| COMP:0000463	
54	Fruit parthenocarpy| cycle 1| COMP:0001045	Parthenocarpic (1) - ovary of all fruits on a bunch enlarged to form fruit filled with pulp mostly without seed; Not parthenocarpic (0) - ovary of some or all fruits remain small and lack pulp or are fully enlarged but have more seed than pulp
55	Fruit parthenocarpy| cycle 2| COMP:0001046	
56	Fruit parthenocarpy| cycle 3| COMP:0001047	
67	Youngest leaf spotted rank| cycle 1| at flowering| COMP:0000373	Record the rank (order) of the youngest leaf spotted (the first fully unfurled leaf or 50 % unfurled for pathology trials) with at least 10 discrete, mature, necrotic lesions or one large necrotic area with 10 light-coloured dry centres), counting the rank by starting with the youngest completely unrolled leaf as 1 and moving downwards, associating the data with the date of data collection
58	Youngest leaf spotted rank| cycle 1| at harvesting| COMP:0000379	
59	Youngest leaf spotted rank| cycle 2| at flowering| COMP:0000374	
60	Youngest leaf spotted rank| cycle 2| at harvesting| COMP:0000380	
61	Youngest leaf spotted rank| cycle 3| at flowering| COMP:0000375	
62	Youngest leaf spotted rank| cycle 3| at harvesting| COMP:0000381	
63	Fruit fill 1-2| cycle 1 main plant| at harvesting| COMP:0000465	Visual estimation of a fruit filling for Early Evaluation Trial stage only, 1 = filled; 2 = not filled
64	Fruit fill 1-2| cycle 2 main plant| at harvesting| COMP:0000472	
65	Raw pulp colour visual estimation 1-3| cycle 1| COMP:0001042	A quick field evaluation of the raw pulp to determine acceptability based on yellow colour using Mbwazirume (Matooke) as a standard control. Scale: 1 = White colour; 2 = Yellow colour slightly lighter than that of Mbwazirume; 3 = Deep yellow pulp colour same as that of Mbwazirume
66	Raw pulp colour visual estimation 1-3| cycle 2| COMP:0001043	
67	Raw pulp colour visual estimation 1-3| cycle 3| COMP:0001044	
68	Sap amount estimation 1-3| cycle 1| COMP:0001039	Quickly determine the amount of sap in the raw pulp on a mature bunch in the EET field. 1= a lot of sap; 2 = a little sap; 3 = no sap.
69	Sap amount estimation 1-3| cycle 2| COMP:0001040	
70	Sap amount estimation 1-3| cycle 3| COMP:0001041	

8. Statistical Design for Hybrid Evaluations

8.1 Early Evaluation Trial for Mchare Breeding

Design: Randomized Complete Block Design (RCBD)

Replications: 3 blocks

Copies per genotype: 6 copies per genotype

8.2 Layout

- Each block will consist of 2 randomized copies of each genotype
- Checks will be planted throughout the field after every two mats of the experimental unit to ensure uniformity and facilitate comparative evaluation

Example Layout:
Block 1: [Genotype A1] [Genotype A2] [Check] [Genotype B1] [Genotype B2] [Check]
Block 2: [Genotype A1] [Genotype B2] [Check] [Genotype A2] [Genotype B1] [Check]
Block 3: [Genotype B1] [Genotype B2] [Check] [Genotype A1] [Genotype A2] [Check]

8.3 Full Sib Family Evaluation

- Design: Randomized Complete Block Design (RCBD)
- Replications: Randomly replicated within blocks
- Family configuration: Each family will have 16 sibs planted together in a block
- Layout: Families should be distributed across multiple blocks to minimize environmental bias

Example Layout:
Block 1: [Family 1] [Family 2] [Family 3] [Family 4]
Block 2: [Family 3] [Family 1] [Family 4] [Family 2]
Block 3: [Family 2] [Family 4] [Family 1] [Family 3]

9. Statistical Design for Hybrid Evaluations

9.1 Initial Quality Check

- The Data Manager will review the submitted data for completeness, consistency, and accuracy
- Use a checklist to ensure all required traits have been recorded

9.2 Field Validation

- If the data quality is inadequate, the Data Manager will return to the field to cross-verify data with the physical plants
- Discuss any discrepancies with data collectors

9.3 Upload

- Upon confirming data quality, the Data Manager will upload validated data to Musabase
- Ensure the data is formatted correctly, and all necessary metadata is included

10. Data Analysis Procedure

10.1 Statistics Analysis Overview

- Data Managers and statistical consultants will perform the following analyses: Descriptive statistics (mean, median, standard deviation) for all traits, Analysis of variance (ANOVA) to evaluate differences among genotypes and families, Regression analysis to identify relationships between traits and environmental factors, Mixed models to account for both fixed and random effects, especially in multi-environment trials.

10.2 Statistical Models

Statistical models will be simplified and made accessible for easy use by team members. Key models to be used include:

- Fixed Effects Model: to analyse main effects of genotype, environment, and treatments
- Random Effects Model: to handle variability among blocks and random replicates
- Mixed Effects Model: to incorporate both fixed and random factors, allowing for better predictions and insights

10.3 Software for Analysis

- GenStat: Used for general statistical analysis and ANOVA
- R Studio: Utilized for advanced statistical modelling and graphical representation of data
- Code snippets will be provided for common analyses to ensure ease of use

10.4 Availability of Models

- All statistical models and analysis scripts will be documented and available on a shared drive
- Training sessions will be held quarterly to familiarize team members with the analysis tools and models

11. Data Flow Dendrogram

Data Collection

```
├── Data Collectors (Daily Data Collection)
│   ├── Assigned Trials
│   ├── Data Entry via Field book
│   └── Submit Data Weekly
                              ├── Data Manager (Data Quality Check)
                              │   ├── Review Data │    ├── Field Validation (if needed)
                              │   ├── Perform Data Analysis
                              │   └── Upload to Musabase
```

12. Record Keeping

- All collected data should be backed up electronically and stored securely.

- Maintain logs of daily activities and observations, including any issues encountered during data collection.

13. Review and Revision

This SOP will be reviewed annually and revised as necessary to incorporate improvements or changes in protocols.

14. Training and Compliance

- All personnel involved in data collection and management must undergo training on this SOP
- Compliance with this SOP is mandatory and will be monitored regularly

15. Conclusion

This SOP establishes a standardized framework for data collection, analysis, and statistical design in the banana breeding program. By ensuring data quality, consistency, and appropriate statistical methodologies, it supports reliable hybrid evaluations. The outlined procedures apply to all personnel involved in data management, fostering efficiency and accuracy across breeding trials. Adherence to this SOP will enhance the integrity of research outcomes, ultimately contributing to the development of improved banana hybrids.

Published by the International Institute of Tropical Agriculture (IITA)
in Ibadan, Nigeria

IITA is the leading research partner in Africa facilitating agricultural solutions to hunger and poverty in the tropics. It is a member of the CGIAR Consortium, a global research partnership that brings together organizations committed to research for sustainable development and a food-secure future.

International Address:
Suite 32
5th Floor, AMP House
Dingwall Road
Croydon
CR0 2LX, UK

Registered Office:
PMB 5320, Oyo Road
Ibadan, Oyo State

Headquarters
PMB 5320, Oyo Road, Idi-Oshe
Ibadan, Nigeria
Tel.: +1 201 6336094
+234 700 800 4482
Fax.: +44 (208) 711 3786 (via UK)

Follow our Social Media Platforms for regular updates on News, Training, Videos, Job openings

@IITA
@IITA_CGIAR
IITA_CGIAR
@IITA
IITA

www.iita.org