BTI Discovery & Innovation in the Life Sciences

BOYCE THOMPSON INSTITUTE

IMPROVEMENT OF BANANA FOR SMALLHOLDER FARMERS IN THE GREAT LAKES REGION OF AFRICA

WP5 PROGRESS REPORT : 'HARNESSING DATA'

REPORTING PERIOD: APRIL 2019 – APRIL 2020

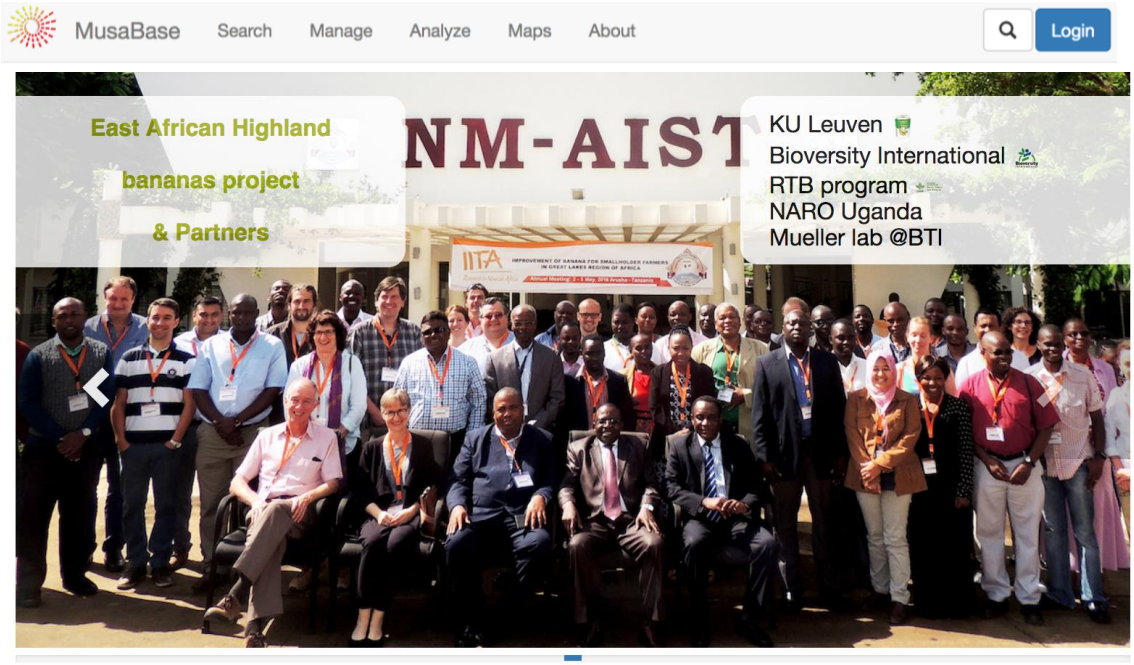

Table of Contents

1. Executive summary

BTI continued to work on the IITA funded "Improvement of Banana for Smallholder Farmers in the Great Lakes Region of Africa" with the objective of developing a banana breeding database that provides project partners and Musa researchers and breeders a virtual hub for information exchange, R&D collaboration and enhanced adoption of new hybrids.

In early 2020, Dr. Guillaume Bauchet left BTI to start a position as a plant breeder in a French breeding company. I would like to express my deep gratitude towards Dr. Bauchet for all he has done in this project during the past 6 years. He has been instrumental in moving the project forward. Dr. Christino Simoes, who started at the Mueller lab in mid-2019, and who has already travelled with Dr. Bauchet to the different breeding locations in Africa, will take over Dr. Bauchet's responsibilities in the project.

Due to the Covid-19 situation, some project related activities, such as the annual meeting 2020, have been postponed or cancelled. We will try to keep the impact of the pandemic on this project as small as possible, however, the circumstances are obviously difficult.

After the annual meeting in Mbarara in 2019, we focused on the following topics, which were a continuation of, or overlapping with, other activities previously pursued:

A- Recent tool development

B- Development and integration of a breeding tracking tool associated to musabase, deployment of field barcoding.

C- Trait ontology, integrate time and management components.

D- Inclusion of WP4 data.

E- Enhance germplasm curation of ITC material.

F- Germplasm collection morphological characterization image storage, as well as characterizations of pathogens (Sigatoka and Fusarium) was added to the website.

G- Workshops

2. Primary and intermediates outcomes

A- Musabase code developments: In the last year, we mainly focused on new trial analysis tools, in particular, heritability, stability, and general mixed models. Most of these are now completed and will be integrated in the site in the next few weeks.

B- Banana tracking tools development: We further developed the Musabase-end of btract http://btract.sgn.cornell.eduhttp://btract.sgn.cornell.edu/), which implements a system to track crossings in banana fields. However, btract supports very flexible data formats that can sometimes be problematic for re-import into the database, and a lot of work has been done to improve the quality of this data exchange. In the medium term, we would like to move the data exchange layer to BrAPI (https://brapi.org/), but this may be harder on the btract end. However, partnering with IITA Nairobi, the tool was further developed to store and retrieve btract data (https://musabase.org/breeders/odk) . An improved barcode labelling tool was released including specific labels for both office based and portable printer: Finally, two new portable zebra ZQ520 printers and label sets were purchased and delivered to Nairobi and Ibadan stations.

C- Trait ontology has been further refined with the help of the banana breeders. Currently the ontology is up to date, but we expect that further refinements will have to be made in the future.

D- Whole genome re-sequencing data was provided by university of North Carolina, matched to corresponding ITC codes and made available through a ftp site: ftp://ftp.musabase.org/musaWGS/ . It will also be integrated into the Jbrowse genome viewer tool.

E- Similarly, the marker storage system has been improved allowing storage of vcf files into musabase. The GBS data from Nyines 2018 was curated, linked with relevant pedigree and included in musabase using the new genotyping storage system https://musabase.org/breeders_toolbox/protocol/1 and under ftp://ftp.musabase.org/musaGBS/Nyines2018/. This new storage system allows complte information storage and further use in new tools (pedigree viewer). IITA Dar-es-salam research team (Dr Manoj Kaushal) recently requested to make available link to metagenomic study raw reads (fastq files) on fusarium, this is will be made available as soon as the archived links are publicly released.

F - The germplasm and corresponding image collections was further curated. The sigatoka and fusarium collections and characterizations, performed by the Viljoen lab, were added to the website. A new menu, 'Pathogens', was introduced to access this information. Currently, the data is linked simply through Excel files, but in the future will be loaded into the database.

G - Chris Simoes, a new member of the Mueller lab, travelled with Guillaume Bauchet to Uganda (Sendusu) and Tanzania (Arusha) to visit the respective breeding programs in the fall of 2019 and gave a number of workshops.

➔ **Table 1** gives the Framework and Results Tracker for WP5, as presented in the project document.

Table 1. Framework and result tracker WP5.

ID	Outcome/ Output	Indicator(s)	Data Sources (If Applicable)	Targets/ Milestones	Actual	Variance	Year 1	Year 2	Year
8	Primary Outcome 8 Create a banana Breeding database	Database online	clone cassavabase system with empty database	Database available online			-	-	-
8.1	Intermediate outcome 25 New database/website created based on Cassavabase software	Database online and accessible by breeders	clone cassavabase system with empty database	Database available online			-	-	-
8.1.1	Output 60 Banana database/website developed (Cornell univ.-IITA-NARO-EMBRAPA)	The website can be accessed by breeders	website access logs	Database available online	NARO and IITA as a partner compiled and submitted a list of Musa accessions available in the NARO fields for the construction of the Musabase by the Cornel University partner: https://musabase.org/	0%	na	na	na

ID	Outcome/ Output	Indicator(s)	Data Sources (If Applicable)	Targets/ Milestones	Actual	Variance	Year 1	Year 2	Year
8.2	Intermediate outcome 26 Adapt user interface/tools for the banana database	Adapted interfaces	feedback from banana breeders	-		na	na	Interfaces adapted based on banana breeder input	na
8.2.1	Output 61 Customized interfaces/tools developed for banana breeders (Cornell univ.-IITA-NARO-EMBRAPA)	new interfaces available	feedback from banana breeders	na	na	na	na	Interfaces adapted based on banana breeder input	na
8.3	Intermediate outcome 27 Database populated with historic data from NARO and IITA breeding programs	new database contents	Data provided by breeding programs	na	na	na	na	Database up to date with historic informatio from IITA and NARO	na
8.3.1	Output 62 Database populated with historic data from NARO and IITA breeding programs	new database contents	database contents	na	na	na	na	Database up to date with historic informatio from IITA and NARO	na

ID	Outcome/ Output	Indicator(s)	Data Sources (If Applicable)	Targets/ Milestones	Actual	Variance	Year 1	Year 2	Year
	(Cornell univ.- IITA-NARO)								
8.4	Intermediate outcome 28 Banana database integrated in workflows	Breeders use database to advance their breeding program	website access logs, database contents	na	na	na	Breeders use crossing/field design/phenotyping tools	Breeders use crossing/field design/phenotyping tools	Breeders use crossing/field design/phenoty tools
8.4.1	Output 63 Use banana database to manage crosses, trials, etc. (IITA-NARO-Cornell univ.)	database contents	website access logs, database contents	na	na	na	na	Breeders use crossing/field design/phenotyping tools	na
8.5	Intermediate outcome 29 Train breeders / IITA staff	pre- and post course surveys	Course attendance lists, surveys	held yearly	na	0%	held yearly	held yearly	held yearly
8.5.1	Output 64 Breeders trained (Cornell univ.- IITA-NARO-EMBRAPA)	Provided training	Course attendance lists, surveys	training course held	na	0%	training course held	training course held	training course

3. Results to date

In **Table 2**, we give an overview of the progress of activities to date for the outputs with a milestone in year 3. A more detailed description by primary outcome is given below.

Table 2. Progress for WP5 year 5 milestones.

ID	Outcome/ Output	Targets/ Milestones	Progress	Variance
8	Primary Outcome 8 Create a banana Breeding database	na	na	na
8.1	Intermediate outcome 25 New database/website created based on Cassavabase software	na	na	na
8.1.1	Output 60 Banana database/website developed (Cornell univ.-IITA-NARO-EMBRAPA)	Interfaces adapted to submit and store data collected with btract	na	na
8.2	Intermediate outcome 26 Adapt user interface/tools for the banana database	Interfaces adapted based on banana breeder input	na	na
8.2.1	Output 61 Customized interfaces/tools developed for banana breeders (Cornell univ.-IITA-NARO-EMBRAPA)	na	na	na

8.3	Intermediate outcome 27 Database populated with historic data from NARO and IITA breeding programs	Database up to date with historic information from IITA and NARO	na	na
8.3.1	Output 62 Database populated with historic data from NARO and IITA breeding programs (Cornell univ.-IITA-NARO)	Database up to date with historic information from IITA and NARO	**-1-** All current trials, field layouts and phenotypic data from IITA and WP4 are available. Data from the Genomic Selection: ftp://ftp.musabase.org/musaGBS/Nyines2018/ **-2-** NARO phenotypic data remain to be loaded. **-3-** MGIS germplasm data curation initiated with IITA is complete.	10%
8.4	Intermediate outcome 28 Banana database integrated in workflows	Breeders use crossing/field design/phenotyping tools	na	**na**
8.4.1	Output 63 Use banana database to manage crosses, trials, etc. (IITA-NARO-Cornell univ.)	Breeders use crossing/field design/phenotyping tools	**-1-** Banana trait ontology was established in collaboration with WP1, WP2, WP3 and WP4. http://musabase.org/tools/onto/ => Additional traits added to the current Ontology **-2-** Development of new nursery and crosses wish list to address the inclusion of the banana tracking tool ( )	0%

8.5	Intermediate outcome 29 Train breeders / IITA staff	held yearly	na	na
8.5.1	Output 64 Breeders trained (Cornell univ.-IITA-NARO-EMBRAPA)	training course held	Hosted Margaret Karanja btract developer at BTI in October 2018 (1 month). One annual training provided in West Africa (November 2018) 1 additional annual training to be provided in East Africa (August/September 2019).	0%

A. Several trial analysis tools – heritability, stability and general mixed model tool have been developed during the reporting period.

B. Ontology is up to date and post composing for traits has been added for te required time and/or physiological stage ontologies.

C. Code improvement to create a workflow between musabase and btract.

D. Integrate WP4 data remains to be completed.

E. ITC germplasm curation was updated.

F. Photo of morphological descriptors of ITC material and pathogens was uploaded.

G. All provided genomic resources (WGS and GBS) are available through musabase.org

4. Challenges Encountered

- WP4 integration ground work (trait mapping and data exchange protocol) were achieved but remain to be completed.
- The Btract export file format are very flexible and sometime problematic for uploads to musabase. In the future, the data exchange should be implemented in a BrAPI layer (https://brapi.org/).
- The Covid-19 situation has affected work around the globe. The Mueller lab has moved to working from home in mid-March; working in a completely web-based line of work, the situation has not been too detrimental for productivity in the group. However, if this situation persists, new and unforeseen problems may arise due to the lack of direct interaction between people. Events such as the annual meeting, which was cancelled for 2020, provided opportunities for deeper discussions and to socialize with the different teams, which is important for effective work and cohesion of a project.

5. Lesson learned

- Need to synchronise better btract and musabase code base to sustain long term development and not losing the interoperability between systems, preferentially implementing the data exchange layer with BrAPI.
- Unforeseen challenges such as the Covid-19 situation can arise and are difficult to prepare for, but we are making every effort to minimize its impact.

6. Workplan

- Pursue ODK integration in musabase (standardize btract input/output formats)
- Additional germplasm/experimental info to be integrated, finalize genome browser for display of resequenced genome data.
- Move pathogen data from Excel files to database.
- Provide additional training in East Africa later in 2020 or early 2021