Calibration Report

Predicting Sweepotato Sensory Attributes Using DigiEye and Image Analysis as a Breeding Tool

High-Throughput Phenotyping Protocols (HTPP), WP3

Kampala, Uganda, 13 December 2022

Joyce NAKATUMBA-NABENDE, Makerere University, Kampala, Uganda
Ann Lisa NABIRYO, Makerere University, Kampala, Uganda
Claire BABIRYE, Makerere University, Kampala, Uganda
Jeremy Francis TUSUBIRA, Makerere University, Kampala, Uganda
Andrew KATUMBA, Makerere University, Kampala, Uganda
Sudi MURINDANYI, Makerere University, Kampala, Uganda
Henry MUTEGEKI, Makerere University, Kampala, Uganda
Judith NANTONGO, International Potato Center (CIP), Kampala, Uganda
Edwin SSERUNKUMA, CIP, Kampala, Uganda
Mariam NAKITTO, CIP, Kampala, Uganda
Reuben SSALI, CIP, Kampala, Uganda
Fabrice DAVRIEUX, Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), Montpellier, France (Validator)

This report has been written in the framework of RTBfoods project.

To be cited as:

Joyce NAKATUMBA-NABENDE, Ann Lisa NABIRYO, Claire BABIRYE, Jeremy Francis TUSUBIRA, Andrew KATUMBA, Sudi MURINDANYI, Henry MUTEGEKI, Judith NANTONGO, Edwin SSERUNKUMA, Mariam NAKITTO, Reuben SSALI, Fabrice DAVRIEUX, (2023). *Predicting Sweepotato Sensory Attributes Using DigiEye and Image Analysis as a Breeding Tool. High-Throughput Phenotyping Protocols (HTPP), WP3.* Kampala, Uganda: RTBfoods Calibration Report, 14 p. https://doi.org/10.18167/agritrop/00733

Ethics: The activities, which led to the production of this document, were assessed and approved by the CIRAD Ethics Committee (H2020 ethics self-assessment procedure). When relevant, samples were prepared according to good hygiene and manufacturing practices. When external participants were involved in an activity, they were priorly informed about the objective of the activity and explained that their participation was entirely voluntary, that they could stop the interview at any point and that their responses would be anonymous and securely stored by the research team for research purposes. Written consent (signature) was systematically sought from sensory panelists and from consumers participating in activities.

Acknowledgments: This work was supported by the RTBfoods project https://rtbfoods.cirad.fr, through a grant OPP1178942: Breeding RTB products for end user preferences (RTBfoods), to the French Agricultural Research Centre for International Development (CIRAD), Montpellier, France, by the Bill & Melinda Gates Foundation (BMGF).

This document has been reviewed by:	
Fabrice DAVRIEUX (CIRAD) Judith S NANTONGO (CIP)	13/12/2022 13/12/2022
Final validation by:	
Fabrice DAVRIEUX (CIRAD)	13/12/2022

CONTENTS

Table of Contents

ABSTRACT

Context: This scientific report concerns DigiEye calibrations of sweetpotato sensory traits. The images plus sensory data collected from cooked roots were used.

Place: Uganda

Date: 12/12/2022

Authors: J Joyce NAKATUMBA-NABENDE, Ann Lisa NABIRYO, Claire BABIRYE, Jeremy Francis TUSUBIRA, Andrew KATUMBA, Sudi MURINDANYI, Henry MUTEGEKI, Judith NANTONGO, Edwin SSERUNKUMA, Mariam NAKITTO, Reuben SSALI (2022)

Content:

The objective of the work was to develop, test and evaluate a color and mealiness classification model based on images of sweetpotato roots. A total of 3018 images were collected from 950 samples from October 2021 to November 2022. The captured image data samples were harvested from several sites, including Namulonge, Arua, Bulindi, Nassari, Serere, Rwebitaba, Iganga, Kabarole, Mbale, Mpigi, Busia, Kamuli, Hoima, Kabale and Kenya. Calibrations were done using reference data collected by a sensory panel. Up to twelve cooked roots per genotype were used for sensory evaluation of traits per session. Calibrations used various linear and non-linear models.

Using linear regression, high performances were observed of the calibration for orange color intensity (R2 = 0.92, Mean Squared Error (MSE) =0.58), suggesting that the model is sufficient for field application. For mealiness-by-hand and positive area, the best model has a Mean Absolute Error (MAE) of 2.16 and 9.01 respectively.

Key words: DigiEye, cooked sweetpotato, sensorial profiles, textural properties, calibrations, chemometrics.

1 DATA

1.1 DigiEye calibration protocol

Prior to the use of the DigiProduction system, steps must be taken to ensure that the system is appropriately launched and calibrated before any images can be captured. These steps taken to calibrate the DigiEye are described in detail in the RTBfoods Standard Operating Procedure for DigiEye calibration (Nakatumba et al 2022a). Images were captured by clicking on the **capture button** that captured the image and automatically saved the image in a predefined location on the computer according to the standard Operating procedure (Nakatumba et al 2022b).

1.2 Sample Preparation

The samples were prepared according to the steps provided in the RTBfoods-Standard Operating Procedure for sample preparation (Nakatumba et al 2022b, Nantongo et al 2022a). Briefly, three roots per genotype were used for imaging. Each sample root had a label with a QR code which was scanned for details about the sample. i.e., genotype and site. The samples and the respective labels were placed on a blue board in the DigiEye illumination cabinet to take images. The image was taken by clicking the **capture button** in the **Camera TABS**. For raw samples, 2 sets of images were obtained; the top after peel and the cross-section of the same roots, as shown in Figure 1. A total of 3018 images (1487 images of cooked samples and 1531 images of raw sample) from 950 roots were collected from October 2021 to November 2022 in different sites (Supplementary Tables 1 & 2). These represented a total of 405 unique sweetpotato genotypes (Supplementary Table 3). Some of the varieties appeared multiple times across varying harvests and sites. The captured image data samples were harvested from several sites, including Namulonge, Arua, Bulindi, Nassari, Serere, Rwebitaba, Iganga, Kabarole, Mbale, Mpigi, Busia, Kamuli, Hoima, Kabale and Kenya. The image data was stored and backed up on External hard drives and a secure online file hosting service (OneDrive), and file access was shared between the CIP-Uganda and Makerere University teams.

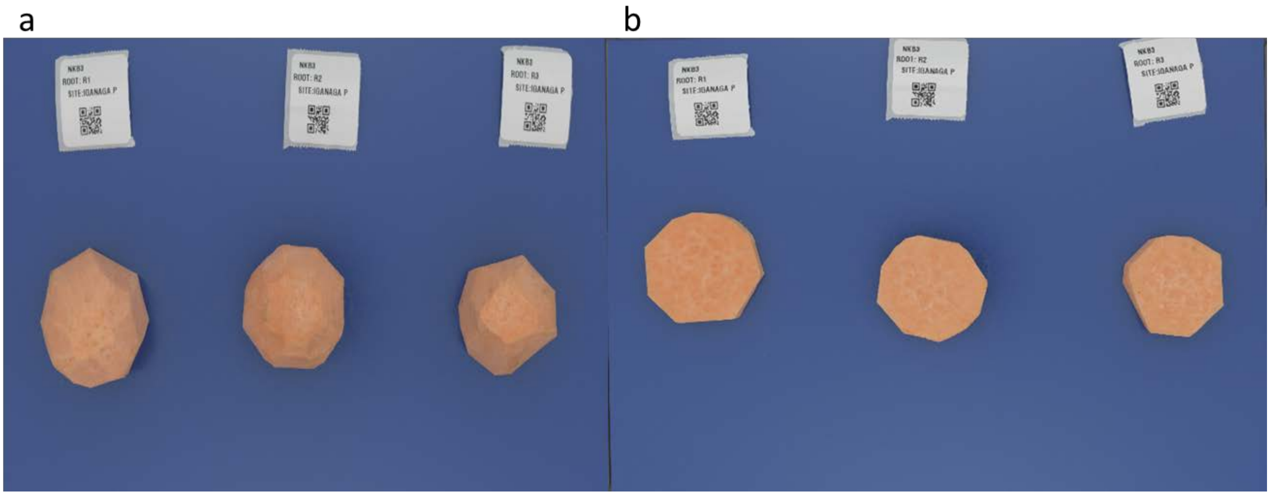

Figure 1: a) top of the root after peeling and b) cross-section for raw samples.

1.3 Sensory and texture parameters

Color and mealiness were assessed by the sensory panel while texture parameters were assessed using a texture analyser. The descriptive statistics of the sensory parameters assessed cooked sweetpotato roots are documented (Nantongo et al 2022b). Up to twelve cooked roots per genotype were used for sensory evaluation of traits. The protocol for descriptive sensory analysis established for sweetpotato that was used has been previously described (Nakitto 2020; Nakitto et al. 2022), where, up to 12 trained panelists consumed small cubes of each cooked sweetpotato genotype and rated the overall liking, color and mealiness of the samples on a 10-point hedonic scale ranging from 1 (dislike extremely) to 10 (like extremely), for each sensory trait per genotypes. In addition, the average peak positive area for the first and second compressions texture of each piece were analysed using a TA-XT texture analyzer (Stable Macro Systems, Godalming, UK) with 10 kg load cell, following a texture profile analysis (TPA) procedure.

2 RESULTS

1.4 Colour measurement model

For the classification task, experiments were run on different classification models, which include: The Decision Tree classifier and Random Forest classifier, which were evaluated using Precision, Recall, and F1-score. Precision which is used to show that out of those predicted as positive, this is how accurate the prediction was. Recall gives the ratio of the correctly predicted outcomes to all predictions. F1- score considers both the precision and recall by computing the average between those values. The Precision, Recall and F1-score lie between 0 and 1 and the higher the value the better the classification model (Table 1).

For the regression task, Regression experiments for orange color measurement were run with 6 models (Linear Regression, K-Nearest Neighbors (KNN) Regression, Decision Tree Regression, Support Vector Regression, Random Forest Regression, Lasso Regression), which were evaluated using R-squared score, Mean Squared Error. R-squared is used in machine learning to measure the goodness of fit or best-fit line of a regressor model whereas mean square error represents the error of the machine learning model created based on the given set of observations in the sample. The greater the value of R-squared, the better the regression model and the lesser the MSE, the better the regression model is.

For both tasks the input features for the model included the RGB features as the independent variables. The regression models performed better than the classification models (Table 1) given that we had an imbalance dataset across the different classes.

Table 1: Scores of the best models on both color measurement tasks (regression and classification) using standard evaluation metrics.

Regression Models	Linear Regressor	Random Forest Regressor	Classification Model	Random Forest Classifier
R2	0.92	0.87	Precision	0.73
			Recall	0.64
MSE	0.58	0.67	F1- Score	0.67

Since the regression models performed better than the classification models, the best regression model was later validated on a new batch. For all the samples in the testing set, the predicted values from the model lie within the range of the orange intensity value obtained from the ground truth data (Table 2). The range values were obtained from the scores on orange color intensity given by the human trained panelists on the sensory panel. From the comparisons drawn in Table 5, we deduced the regressor model is learning some relevant information from the features and thus it’s able to relate the extracted features to the mean orange intensity value which is the label or output for the regressor model.

Another validation exercise was conducted in situ on another batch from the 2022 harvest and a post-evaluation would be carried out by CIP-Uganda team after the ground truth data is obtained from the sensory panel.

Table 2: Comparison of samples from predicted values with the actual values and the ranges from the ground truth data

No	Variety	Actual Orange intensity value	Predicted Orange intensity value	Error rate (difference between the actual value and predicted value)	Range of orange intensity values from the ground truth data
1	C_P_UGP20170334-27 R3	0.40	0.88	-0.49	0 - 2
2	C_P_SILK OMUYAKA R3	0.31	1.40	-1.09	0 - 2
3	C_C_UGP20170335-6 R2	1.54	2.74	-1.19	0 - 4
4	C_P_UGP20170028-7 R1	3.91	4.88	-0.97	3 - 5
5	C_C_UGP20170015-26 R2	1.40	0.49	0.91	0 - 3
6	C_C_EJUMULA R1	8.18	7.47	0.71	7 - 9

1.5 Mealiness prediction models

For the experiments, the initial challenge was to determine an appropriate approach to model the mealiness problem. Although the mealiness-by-hand is classified discretely from 0-9, there was a challenge with determining the ground truth because each sample is independently scored by multiple human experts and the representation of the aggregate data was a mean value that cannot be used as a label for a classification experiment. After consultation with the experts at CIP, it was agreed that we should proceed and try different models while documenting and evaluating which models fit the data best. For the classification experiments, we considered the mode or median of the human expert scores as the ground truth, and the mean is used for the regression experiments. We have also considered using machine-measured attributes such as positive area 1R and positive area 2R as target labels for mealiness in the regression experiments.

Currently, three models have been trained for the regression task and evaluated using the Mean Absolute Error (MAE). The training dataset had 89 samples; for the Linear and Decision Tree regression models, the models were trained using a cross-validation approach (K-Fold) with 5 folds. The input features for the models are the CNN extracted features with 2 different targets; mealiness-by-hand (MBH) and mean Positive Area values. The third model was a neural network which was trained with 60% of the samples for training, 20% for validation, and 20% for testing. The results of the models are summarized in Table 3.

Table 3: Model scores for MAE and R2 for the 3 models when evaluated against mealiness-by-hand (MBH) and Positive Area as targets. (Mean cross-validation scores shown for linear Regression and Decision Tree models)

Model	MAE		R^2 score	
	MBH	Positive Area	MBH	Positive Area
Linear Regression	2.39	9.01	0.12	0.09
Decision Tree Regression	1.97	2.21	0.58	0.72
Neural Network	2.49	7.50	0.42	0.67

3 OPPORTUNITIES AND PLANS

Image-based analysis for the evaluation of different quality traits can be extended to other crops such as potatoes, and root crops like cassava, to inform the selection of samples but also aid in breeding decision-making such as the release of new varieties.

After the tool transfer to the CIP teams, there were discussions around extending the project to examine the possibility of developing a prediction tool to predict other sensory attributes using image analysis. The attributes that were discussed for possible exploration include; Browning, Fibrousness, Water absorption, and Firmness. For future improvement of the performance of the prediction models, we recommend that more data on sweetpotato varieties are collected and shared with the Makerere University team.

Plans are underway to deploy a web-based version of the sweetpotato sensory attribute prediction tool and integrate it into the existing SweetPotato Breed base.

4 REFERENCES

Nakitto M (2020) SOP for sensory evaluation on boiled sweetpotato. Biophysical characterization of quality traits, WP2.

Nakitto M, Johanningsmeier SD, et al. (2022) Sensory guided selection criteria for breeding consumer-preferred sweetpotatoes in Uganda. Food Quality and Preference 101, 104628.

Nantongo SJ, Serunkuma, Edwin, Burgos, Gabriela, Devrieux Fabrice, Meghar K, Ssali Reuben (2022) SOP for near infrared spectroscopy (NIRS) acquisition on sweetpotato roots and potato tubers. WP3.

Nantongo SJ, Serunkuma, Edwin, Burgos, Gabriela, Devrieux Fabrice, Meghar K, Ssali Reuben (2022b). NIRS Analyses of sensory traits in sweetpotato based on spectra collected from in raw-intact roots. RTBfoods Project Report, 12p.

Nakatumba J., Katumba A., et al., (2022a) Standard Operating Procedure for DigiEye Calibration, DigiEye and Image analysis as a breeding tool, WP3

Nakatumba J., Katumba A., et al., (2022b) Standard operating procedure for image capture in sweetpotato and potato, and sensory attribute prediction, WP3

5 APPENDICES

Supplementary Table 1: The number of DigiEye images that were captured in 2021 and 2022 by month

		JAN	FEB	MAR	APR	MAY	JUN	JUL	SEP	OCT	NOV	DEC	TOTAL
Number of images	2021				10				31	196	309	295	**841**
	2022	239	102	188		538	38	1072					**2177**
											Overall total		**3018**

JAN-JANUARY, FEB-FEBRUARY, MAR-MARCH, APR-APRIL, JUN-JUNE, JUL-JULY, SEP-SEPTEMBER, OCT-OCTOBER, NOV-NOVEMBER, DEC- DECEMBER

Supplementary Table 2: Sites from which the sweetpotato samples were harvested

No.	Site	Number of Images	No.	Site	Number of Images
1	Arua	48	9	Kenya	138
2	Buhanika	33	10	Kabarole	37
3	Bulindi	60	11	Kumi	21
4	Busia	32	12	Mbale	37
5	Hoima	32	13	Mpigi	37
6	Iganga	33	14	Namulonge	882
7	Kabale	38	15	Rwebitaba	1325
8	Kamuli	28	16	Serere	237

The images captured were from 405 unique sweetpotato genotypes harvested from the sites in table 2 above. These genotypes in the collected data are listed in table 3 below.

Supplementary Table 3: List of unique genotypes from which samples were collected for imaging and their corresponding number of occurrences in the dataset.

Unique varieties	OCCURRENCES	Unique varieties	OCCURRENCES	Unique varieties	OCCURRENCES
1.44	16	U0280_9	4	UGP20170015-26	4
20AKAB118	2	U0334_13	4	UGP20170023-11	4
20AKAB119	2	U0334_27	4	UGP20170023-17	4
20AKAB120	2	U0335_25	8	UGP20170023-2	6
20AKAB121	2	U0335_6	4	UGP20170023-20	4
20AKB118CT	1	U0341_20	8	UGP20170028-7	4
20KAB118	1	U0341_37	4	UGP20170334-13	4
ALAMURA	9	U0342_12	4	UGP20170334-27	4
AledaMan	4	U0342_14	8	UGP20170335-6	4
ARA209	12	U0342_17	4	UGP20170341-20	4
ARAKARAKA	8	U0342_21	8	UGP20170342-12	4
ARAKARAKA RED	8	U0342_32	7	UGP20170342-14	4

Unique varieties	OCCUR RENCES	Unique varieties	OCCURR ENCES	Unique varieties	OCCURREN CES
Beauregard	38	U0344_1	4	UGP20170342-21	4
Bela	4	U0344_16	4	UGP20170347-11	4
Bertran	5	U0344_17	4	UGP20170486-162	4
BF13CIP3	4	U0344_3	4	UGP20170490-982	2
BF59CIP1	2	U0347_1	4	UGP20170673-173	4
CACEAPE	4	U0347_11	4	UGP20170674-11	4
Cacearpe	2	U0348_18	4	UGP20170674-13	4
CARROT	6	U0348_19	4	UGP20170674-16	6
CEMSA	20	U0350_1	4	UGP20170679-18	4
CEMSA_74_228	12	U0350_10	4	UGP20170783-12	4
CIP1990	4	U0350_11	4	UGP20170793-120	4
CIP19906	8	U0350_16	12	UGP20170802-18	4
CIP199062.1	8	U0350_17	4	UGP20170893-11	4
D11	58	U0350_20	4	UGP20170893-20	4
D15	24	U0366_3	4	UGP20170902-54	4
D20	66	U0407_8	4	UGP20170903-101	4
D26	24	U0469_5	4	UGP20170907-31	4
Dadanyu	2	U0475_22	8	UGP20170910-18	4
DILLA	6	U0475_24	4	UGP201709334-49	4
D-Rex	4	U0476_16	4	UGP20170934-42	4
EJUMULA	73	U0486137	4	UGP20170943-39	4
GIHINGA	4	U0486152	4	UGP20170944-17	4
HUARMEY	4	U0489_2	4	UGP20170944-8	4
HUARMEYA	4	U0489_21	4	UGP20171152-1	4
Ininda	4	U0489_9	4	UGP20171155-17	4
IRENE	6	U0490_966	4	UKEREWE	4
JEWEL	4	U0490980	4	UMBRELLA	34
Kabode	4	U0491_15	4	WAGABOLI	12
Kadyaubw	4	U0493_30	4	U0678_19	8
KAWOGO	8	U0565_3	4	U0679_13	4
KBL648	12	U0654_8	4	U0679_18	8
Kemb36	2	U0667_12	4	U0707_2	4
Ken	4	U0667_17	4	U0710_1	4
Kenspt1	2	U0667_8	4	U0729_1	4
Kenspt2	4	U0672_28	8	U0752_1	4
Kenspt3	4	U0673_158	4	U0777_5	8
Kenspt4	4	U0673_173	4	U0782_13	4
KIEGEA	4	U0674_11	12	U0782_15	4
KIRIBWAMUKWE	4	U0674_13	4	U0782_18	4
KMI_61	9	U0674_15	16	U0782_27	8
KML942	4	U0674_16	4	U0783_12	4
KRE691	4	U0678_10	6	U0783_17	8
LOCAL	28	U0678_16	8	U0783_18	4
LUW1257	8	MPG1128	12	U0783_9	5
LUW1274	16	Mugamba	3	U0792_40	2

Unique varieties	OCCURRENCES	Unique varieties	OCCURRENCES	Unique varieties	OCCURRENCES
MAGABAL	4	MUGANDE	8	MAYAI	21
MAGABALI	8	MUWULU	38	MHistar	4
Mathuthu	4	Nakalbo	4	MLE179	4
NAROSPOT	28	MLE199	2	Nuti	2
NAROSPOT 1	40	NAS5-58	5	Nyumingr	4
NASP5-58	4	U0793_10	4	O_chingo	2
NASPOT 1	10	U0798_13	4	OKONYEDO	8
NASPOT 10	8	U0798_14	8	Olga	5
NASPOT 10 O	44	U0798_18	4	OMAMKAN 1	5
Naspot 11	50	U0798_19	4	OMAMKAN 3	8
NASPOT 5	6	U0799_13	8	OTADA	21
NASPOT 7	21	U0799_30	4	OTANDIBATA	4
Naspot 8	111	U0802_11	8	Pepris	4
NASPOT 9	4	U0802_7	4	POLISTA	4
NDEREERA BAANA	4	U0821_5	4	PZ06_077	4
NEW KAWOGO	28	U0836_6	8	PZ06_085	4
NIMIRA	4	U0846_8	4	RAK808	10
NK259L	4	U0855_1	4	RAK819	2
NK318L	16	U0893_11	4	RAK835	18
NKB105	44	U0893_13	4	RAK848	8
NKB135	8	U0893_20	8	RESISTO	21
NKB3	67	U0893-13	4	RESISTO_CIP	12
U0024_9	4	U0894_5	4	U0157_6	4
U0026_42	4	U0902_52	4	U0160_10	4
U0028_1	4	U0902_53	4	U0164_13	8
U0028_11	4	U0902_54	8	U0164_4	4
U0028_19	8	U0902_63	4	U0190_19	4
U0028_23	8	U0902_66	8	U0190_5	4
U0028_4	4	U0903_101	4	U0194_14	4
U0028_7	4	U0903_27	4	U0194_2	4
U0029_22	4	U0903_38	4	U0214_3	4
U0029_24	4	U0903_67	4	U0217_3	8
U0029_26	8	U0903_81	4	U0226_5	2
U0029-15	4	U0903_96	4	U0230_2	4
U0032_16	4	U0905_15	8	U0240_7	6
U0034_18	4	U0905_18	4	U0249-5	4
U0036_5	4	U0905_6	4	U0251_6	4
U0043_27	4	U0906_16	4	U0264_14	4
U0059_12	4	U0907_10	4	U0264_8	4
U0064_15	2	U0907_11	8	U0272_3	4
U0074_9	8	U0907_17	4	U0013_18	4
U0076_1	4	U0907_31	16	U0013_40	4
U0080_6	8	U0907_4	4	U0013_41	8
U0087_2	4	U0907_9	4	U0015_26	8

Unique varieties	OCCUR RENCES	Unique varieties	OCCURR ENCES	Unique varieties	OCCURREN CES
U0091_14	4	U0908_1	8	U0015_29	4
U0096_19	8	U0908_10	4	U0023_14	12
U0117_6	4	U0908_19	8	U0023_2	4
U0124_8	4	U0908_20	8	U0023_20	6
U0129_11	4	U0910_10	4	U0024_11	12
U0136_11	4	U0910_18	8	U0024_13	4
U0140_9	2	U0910_24	8	U0024_14	8
U0143_11	4	U0910_35	12	U0024_17	4
U0143_21	4	U0923_12	4	U0024_2	4
U0148_10	4	U0923_8	8	U0024_8	4
U0150_9	8	U0934_38	12	U0934_45	4
U0151_8	4	U0934_41	8	U0934_58	4
U0152_1	4	U1141_10	4	U0935_19	8
U0155_17	4	U1141_13	8	U0935_25	4
U0944_8	9	U1141_3	4	U0939_15	4
U0947_1	4	U1143_11	8	U0943_23	4
U0947_15	4	U1143_23	4	U0943_24	4
U1088_5	4	U1152_1	4	U0943_39	4
U1155_13	8	U1153_15	4	U0944_13	4
U1155_15	8	U1153_16	4	U0944_17	8
U1155_17	4	U1154_1	4	U0944_23	4
U486_137	4	U487_167	4	U490_985	4
U486_151	4	U487_171	4	U490_989	4
U486_152	4	U490_966	4	U673_173	4
U486_162	5	U490_982	4	U793_103	4
U846_143	4	U793_123	4		

Institute: Cirad – UMR QualiSud

Address: C/O Cathy Méjean, TA-B95/15 - 73 rue Jean-François Breton - 34398 Montpellier Cedex 5 - France

Tel: +33 4 67 61 44 31

Email: rtbfoodspmu@cirad.fr

Website: https://rtbfoods.cirad.fr/