Laboratory Standard Operating Procedure

Feasability of Bad-Good Genotypes Screening using NIRS

High-Throughput Phenotyping Protocols (HTPP), WP3

Montpellier, France, April 2020

Denis CORNET, Centre de Coopération International en Recherche Agronomique pour le Développement (CIRAD), Montpellier, France

Pierre DARDENNE, Consultant, Leignon, Belgium (Validator)

Fabrice DAVRIEUX, CIRAD, Montpellier, France (Validator)

Emmanuel Oladeji ALAMU, International Institute of Tropical Agriculture (IITA), Lusaka, Zambia (Validator)

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

To be cited as:

Denis CORNET, Pierre DARDENNE, Fabrice DAVRIEUX, Emmanuel Oladeji ALAMU (2021). *Feasability of Bad-Good Genotypes Screening using NIRS. High-Throughput Phenotyping Protocols (HTPP), WP3.* Montpellier, France: RTBfoods Laboratory Standard Operating Procedure, 8 p. https://doi.org/10.18167/agritrop/00671

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).

RTBfoods

WP3: High-Throughput Phenotyping Protocols (HTPP)

SOP: Feasability of Bad-Good Genotypes Screening using NIRS

Date: 04/2020	**Release: 1**

Written by : Denis CORNET

Collaborators : Emmanuel ALAMU, Thomas zum FELDE, Karima MEGHAR

For information on this Protocol, please contact:

- Denis CORNET, denis.cornet@cirad.fr
- Emmanuel Alamu, O.ALAMU@CGIAR.ORG

This document has been reviewed by:

Pierre DARDENNE (Consultant)	24/04/2020
Emmanuel Oladeji ALAMU (IITA)	26/04/2020
Fabrice DAVRIEUX (CIRAD)	16/12/2020

Final validation by:

Fabrice DAVRIEUX (CIRAD)	Date 26/12/2020

CONTENTS

Table of contents

ABSTRACT

The current protocol's main objective is to determine the feasibility to calibrate a qualitative classification model allowing the distinction between good and bad genotypes using NIRS spectra collected on different product states (i.e. raw intact organ, chopped, puree). This protocol focuses on pounded yam and boiled cassava, but if the proof of concept is achieved, it can be extended to RTB product profiles.

The main principle is based on the existing traditional knowledge of varieties of good and bad qualities. This knowledge allows us to choose genotypes from both categories and train a binary classification model to predict their belonging. To test feasibility as soon as possible, this preliminary protocol focuses on the already available database. The idea is to train and test a classification model predicting good or bad genotypes using a convolutional neural network (CNN) with a binary cross-entropy loss function. Indeed, this type of algorithms showed excellent results for similar studies. In order to avoid a confusion effect of storage length on tuber quality, the only tuber with similar physiological age (i.e. meaning same storage length) should be kept in each database.

Key Words: Overall product quality, varietal selection, near infrared spectroscopy, convolutional neural network, binary classification

1 SCOPE AND APPLICATION

Functional and textural properties are amongst the most important quality traits for Root, Tubers and Banana products in Africa. Despite their importance, high throughput methods to evaluate them are still lacking. Many attempts have been made to quantitatively relate NIRS measurement with texture attributes (e.g. cohesion, elasticity). But these attempts faced two major constraints: (i) the link between texture attributes and sensory evaluation is not straightforward, and (ii) the performance of the NIRS calibration models are quite low.

The main objective of the current protocol is to determine the feasibility to calibrate a qualitative classification model allowing the distinction between good and bad genotypes using NIRS spectra collected on different product state (i.e. raw intact organ, chopped, puree).

This protocol focuses on pounded yam and boiled cassava, but if the proof of concept is achieved, it can be extended to RTB product profiles.

2 PRINCIPLE

The main principle is based on the existing traditional knowledge of varieties of good and bad qualities. This knowledge allows us to choose genotypes from both categories and train a binary classification model to predict their belonging. In order to build a robust model 2 datasets are necessary, a calibration data set and a validation dataset including untrained genotypes.

In order to test feasibility as soon as possible, this preliminary protocol focus on the already available database. Most breeding programs already have spectra of good and bad genotypes collected at different steps (entire, mashed or grated and cooked/transformed). For each step, the idea is to train and test a classification model predicting good or bad genotypes. Relationship between spectra and good/bad categories will be investigated using a convolutional neural network (CNN) with a binary cross-entropy loss function. Indeed, this type of algorithms showed excellent results for similar studies.

3 PROCEDURE

3.1 Choice of genotype

For yams, the choice of good/bad genotypes could be made very easily based on species. Indeed, on average *Dioscorea alata* exhibits relatively poor quality while *D. rotundata* offers high-quality genotypes.

For cassava, a classification between good and bad genotypes should be done “manually”. To ascertain the classification, only extremely good and bad genotypes should be kept for this preliminary study.

Idealy, at least 5 genotypes of each class (i.e. good and bad) originating from 5 different origins (i.e. place and season) should be studied. The choice of genotype and origin to investigate should be discussed with WP1 and WP4. However, because this protocol rely on already available data, we may not have the choice.

3.2 Spectra database

As many spectral databases should be investigated. Spectra should be gathered based on the same spectrometer and identical sample preparation. Moreover, minimal metadata should be attached to spectra (i.e. year, site, trial, genotype name, repetition, measurement sequence number,and good/bad classification).

As long as possible, all spectra should be retained; i.e. no outlier removal and spectra repetition averaging. Because fresh products are heterogenous we advices around 10 scans by sample (i.e. genotype and origin).

3.3 Model calibration

Pretreatments, calibration and validation, will be carried out using python language with a Keras framework and a TensorFlow backend. For each sample, feature augmentation will be done by generating 12 new spectra using pretreatments based on different degrees of multi scatter correction, the Haar transform, Gaussian derivatives, Savitsky-Golay and standard normal variate. If necessary, a data augmentation will be achieved by generating noised samples for each class using a combination of translation and rotation of the original spectra. A convolutional neural network composed of three convolutional layers followed by two dense layers will be fitted to the calibration data. Binary cross entropy will be used as the loss function. In order to avoid overfitting, a dropout of 20% of features will be applied between layers. The model will be calibrated using five-fold cross validation.

Model performance will be estimated based on a confusion matrix and traditional sensitivity, specificity, precision, recall, F1 score, Kappa statistic and overall accuracy. Accuracy is a popular metric that refers to the ability of the model to predict the class label of new or unseen data correctly. In addition to this, sensitivity and specificity are also used to assess how well classifiers can recognize true examples as well as false examples. The Kappa statistic evaluates the pairwise agreement between two different observers, corrected for an expected chance agreement. A Kappa value of 0 indicates chance agreement, and 1 shows perfect agreement between the classifier and the ground truth (true classes). Precision can be seen as a measurement of exactness or quality, whereas recall is a measurement of completeness or quantity. The F1 score is the harmonic mean of precision and recall, where an F1 score reaches its best value at 1 (perfect precision and recall) and worst at 0.

3.4 Model validation

Ideally, the model will be validated against an external dataset (i.e. different year, site and/or genotypes). However, depending on data availability, train/test sets could also be done using constrained split, cross-validation or leave-one-out sampling. Critical points or Note on the procedure

Same physiological age

4 CRITICAL POINTS OR NOTE ON THE PROCEDURE

In order to avoid a confusion effect of storage length on tuber quality, the only tuber with similar physiological age (i.e. meaning same storage length) should be kept in each database.

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/