**Exploring the chemotypes underlying important agronomic and consumer traits in cassava (*Manihot esculenta* Crantz)**

Margit Drapal1, Tatiana M. Ovalle Rivera2, Luis Augusto Becerra Lopez-Lavalle2 and Paul D. Fraser1\*

1 School of Biological Sciences, Royal Holloway, University of London, Egham Hill, Egham, Surrey, TW20 0EX, UK.

2 International Center for Tropical Agriculture (CIAT), Cali, Columbia.

Corresponding author: P.Fraser@rhul.ac.uk, Tel: +44 1784 443894

# Abstract

A broad diversity of phenotypes are available within the cassava germplasm collections. The phenotypes include improved nutritional, starch or culinary root quality as well as abiotic and biotic resistance properties. Some of these traits can be found naturally occurring in cassava landraces, whereas others are the result of targeted breeding efforts. For future breeding programmes it is important to know the underlying mechanisms of these desirable traits. Metabolomics can assist in the elucidation of these mechanisms by measuring the end products of the cellular processes conferring the traits of interest. The present study focused on the comparison of two or more variants of the same trait such as high and low culinary quality or resistance and susceptibility to thrips. Overall, eight different traits were assessed.

Results showed that amino acids and umami compounds were associated with superior culinary attributes and the phenylpropanoid superpathway plays an important role in pest resistance. Furthermore, the data highlighted a low chemodiversity in African cassavas and that the source-sink relation was still active at the harvest stage.

# Keywords

Cassava, characteristic traits, chemotypes, diversity, metabolite profiling, modern breeding approach

# Abbreviations

ANOVA, analysis of variance; CIAT, International Center for Tropical Agriculture; CMD, cassava mosaic virus; DW, dry weight; GC-MS, gas chromatography mass spectrometry; HCN, hydrogen cyanide; LC-MS, liquid chromatography mass spectrometry; PC, principle component; PCA, principle component analysis; QC, quality control; RSD, relative standard deviation; TCA, tricarboxylic acid; UPLC-DAD, ultra-performance liquid chromatography – diode array detector; UV/VIS, ultra violet/visible light

# Introduction

The tuberousroots of cassava (*Manihot* *esculenta* Crantz) provide a major source of calories for ~800 million people in the Americas, Asia and Africa (Liu *et al.*, 2014; Howeler *et al.*, 2013). Cassava plants can grow on degrades or marginal soils where other crops cannot survive and still provide feasible yields, even on small scales (Henry and Hershey, 2002). This presents cassava as an attractive staple crop for low-income, smallholder farmers (Howeler *et al.*, 2013). In South America, cassava progenitors have been cultivated since ~9000 years and farmers have realised the adaptation potential of cassava to a wide range of biotic and abiotic stresses (Piperno, 2011). Therefore, the International Center for Tropical Agriculture (CIAT) in Cali, Colombia generated a genepool for cassava landraces which perform better under challenging environments (Hershey and Jennings, 1992). Over the past decades, the genepool was significantly extended to include other cassava traits and modern breeding techniques such as marker-assisted breeding were incorporated to assist the international breeding efforts (Bredeson *et al.*, 2016; Hahn *et al.*, 1979; Wang *et al.*, 2014). The outcome of these efforts included cassava roots with improved nutritional values (e.g. high provitamin A content) and different starch properties (e.g. waxy phenotype) (Belalcazar *et al.*, 2016; Aiemnaka *et al.*, 2012). However, biotic stresses started to threaten the yield of these elite varieties, especially since the main propagation of cassava plants is through cuttings. Therefore, the focus of cassava breeding was divided into improving root quality and elucidating the mechanism of pest and virus resistant cassava plants (Hahn *et al.*, 1979; Alvarez *et al.*, 2009; Skovgård *et al.*, 1993; Omongo *et al.*, 2012).

The biological phenotype of a plant is a combination of cellular processes, which create the chemically distinct entity (chemotype) manifesting the phenotype. The chemotype can be measured through metabolomics to elucidate the metabolites or metabolic pathways responsible for the trait (Roessner and Bowne, 2009; Bino *et al.*, 2004). The distinct metabolite profile can then be used to facilitate and focus breeding efforts of complex crops such as the heterozygous cassava. The CGIAR Research Program on Roots, Tubers and Bananas ([www.rtb.cgiar.org](http://www.rtb.cgiar.org)) is promoting the metabolic characterisation of existing diversity panels which has resulted in established screening techniques (Drapal *et al.*, 2018; Perez-Fons *et al.*, 2019; Price *et al.*, 2020). These techniques were applied to a cassava diversity panel including different root phenotypes (e.g. amylose content, culinary quality) and abiotic and biotic stress phenotypes (Table 1). The biotic stresses included landraces known for the resistance or susceptibility to the herbivorous pests green mites (*Mononychellus tanajoa*), thrips (*Corynothrips stenopterus*) and whitefly (*Bemisia* *tabaci*) as well as to the virus-induced cassava mosaic disease (CMD). Metabolite data of the present study highlighted key metabolite pathways related to the characteristic traits such as the source-sink relation of monosaccharides affecting dry matter content, the high amino acid levels related to the high culinary quality and the different sub-pathways within the phenylpropanoid superpathway correlated to defensive traits.

Table 1: List of cassava varieties including country of collection and characteristic trait. Responses to biotic stresses are indicated as resistant (res.) or susceptible (susc.). CMD, cassava mosaic disease.

|  |  |  |  |
| --- | --- | --- | --- |
| Sample code | Country of collection | Characteristic trait | Response to green mites |
| BRA488 | Brazil | High cyanide content |  |
| COL113 | Columbia |  | res. |
| COL1505 | Columbia | Z01 and Z04 adaptation |  |
| COL1522 | Columbia | Good for bread baking (low amylose content) | res. |
| COL1684 | Columbia | Z03 adaptation | susc. |
| COL2017 | Columbia | High sugar content |  |
| COL2436 | Columbia | High carotene content/thrips susc. |  |
| COL912B | Columbia | Unknown |  |
| CUB25 | Cuba | Low amylose content |  |
| CUB74 | Cuba | High culinary quality/ Z06 adaptation/ Frog skin disease res. | res. |
| ECU72 | Ecuador | Whitefly res. /Bacteriosis susc. |  |
| GUA35 | Guatemala | Low sugar content/Frog skin disease susc. |  |
| PAN139 | Panama | Thrips res. |  |
| PAR36 | Paraguay | High amylose content |  |
| PER 496 | Peru | Low cyanide content |  |
| TME3 | Africa | CMD res. |  |
| NGA5\* | Africa | CMD moderate susc. |  |
| NGA11\* | Africa | Amenable to transformation, CMD susc. |  |
| VEN25 | Venezuela | Low culinary quality |  |
| VEN77 | Venezuela | Drought tolerant | susc. |

\*These varieties are also known as TMS30555 (NGA5) and TMS60444 (NGA11).

# Material and methods

## Plant material

Twenty cassava landraces (Table 1) were grown under CIAT’s standard field conditions. Six biological replicates were planted for each landrace. Leaf, stem and root tissue (approx. 200g) was collected from ten months old plants, immediately frozen in liquid nitrogen and lyophilized.

## Extraction and analysis of polar metabolites

Freeze-dried tissue was ground into a fine powder and a quality control (pool of all samples, QC) created. Samples and QCs were weighed (10±0.5mg) into plastic tubes and extracted with methanol/water/chloroform (Drapal *et al.*, 2019). After phase separation, aliquots of the polar phase were immediately used for LC-MS analysis or dried down for GC-MS analysis.

### LC-MS analysis of polar extracts

An aliquot of the polar phase (150µl) was filtered using syringe filters (0.45µm, nylon). The filtrate (100µl) was transferred to a glass insert and internal standard genistein (5µl of 0.2mg/ml stock solution) added. Analysis was performed with an Agilent 6560 Ion Mobility Q-TOF coupled to an Agilent 1290 Infinity II (Agilent Technologies, Inc.). All solvents used for LC-MS analysis were of LC-MS grade and comprised solvent A (water and 0.1% formic acid) and solvent B (acetonitrile and 0.1% formic acid). The samples were separated with an YMC-UltraHTPro C18 column (100 × 2mm i.d. 2μm) at 0.2ml/min. The solvent gradient started at 95% (A) for 0.5min, followed by a linear decrease to 75% (A) at 3min, 70% (A) at 6min, 0% (A) at 6.5min, which was held until 7.5min before return to initial conditions of 95% (A) at 9.5min. After the last step, the column was re-equilibrated for 1min. The eluate was split for simultaneously analysis by DAD (scan mode 200-600nm) and MS. Mass spectrum data of the eluting compounds was collected in negative centroid mode (100 to 1700m/z, 0.9spectra/second). The source settings included nozzle and capillary voltages at -500V and 4000V, nebulizer gas (nitrogen) at 35psi, dry gas at 5l/min and 325°C and sheath gas at 12l/min and 275°C. Calibration was performed during each run to a reference solution.

Molecular feature extraction was performed with Agilent Profinder (V10.0 SP1, Agilent Technologies, Inc.) with retention time tolerance 0.3min and mass tolerance 10ppm for peaks >400 counts. The resulting data table included molecular features which were a compilation of all adducts and isotopes detected. The data was corrected by corresponding internal standard and quality control. Quantification was performed relative to the internal standard. Metabolites were identified through comparison of UV/VIS spectrum, mass spectrum and retention time to an in-house library of authentic standards.

### GC-MS analysis of polar extracts

Aliquots of the polar phase (120µl) were dried down with internal standard d4-suiccinic acid (10µg). The dried extracts were stored at -20°C and derivatised (methoxymation and silylation) in batches before analysis (Drapal *et al.*, 2019). The analysis was performed as reported with an 7890A gas chromatography (GC) system coupled with a mass spectrometer (MS) 5795C MSD (Agilent Technologies, Inc.) in splitless mode. Metabolites were identified with AMDIS (V2.71) including an in-house library. Identification parameters included retention time, retention indices and mass spectrum. The resulting data was quantified relative to the internal standard and sample weight.

## Extraction and analysis of non-polar metabolites

The extraction and analysis of carotenoids and chlorophylls from cassava leaf tissue was performed as previously described (Nogueira *et al.*, 2013). After phase separation, an aliquot (350µl) of the non-polar chloroform phase was dried down and stored at -20°C until analysis. The dried aliquots were resuspended in ethyl acetate/acetonitrile (100µl, 1:9 (v/v)), centrifuged for 5 min at full speed and an aliquot (80µl) transferred to a glass insert. The injection volume was 3µl. Carotenoids and chlorophylls were identified by their specific retention time and UV/VIS spectrum (Fraser *et al.*, 2000). Quantification was performed with individual dose-response curves.

## Statistical analysis

Statistical analysis was performed with Simca P (13.0.3.0, Umetrics), XLSTAT (2017, Addinsoft), GraphPad Prism (8.3.0, ) and Metaboanalyst (Xia and Wishart, 2016). Statistical tests included Student’s *t*-test for pairwise comparisons and parametric ANOVA with false discovery rate correction for multiple comparisons. Graphs were produced with Simca P, GraphPad Prism and Metaboanalyst. Pathway displays were initially created in-house with BioSynLab (Royal Holloway University of London).

# Results

## LC-MS metabolite profiling of three tissues regarding phenotype

Polar extracts of leaf, stem and root tissue of 20 cassava landraces, grown under standard field conditions, were analysed by LC-MS as previously reported (Drapal *et al.*, 2018). Within the LC-MS metabolite profiling 118 molecular features were detected in all three tissues and an additional 300 molecular features solely in leaf tissue (Table A.2). The initial data analysis will focus on the molecular features detected in all three tissues to assess whether the chemotype of a trait is associated to a specific tissue type or can be observed in all plant organs. PCA analysis showed that the metabolite variation between the three tissues, leaf, stem and root, was explained by the first two principal components (PC) with 54.1% (Fig. 1A). Leaf samples separated along the x-axis (PC1) from stem and root sample, whereas stem and root separated along the y-axis (PC2). This suggests that the metabolite composition of stem and root is more similar compared to leaf. This was confirmed by RV coefficient, which only showed a significant correlation between stem and root samples (R=0.44 , *P*-value 0.033).

The combined score plot highlighted that the root tissue of GUA35 clustered with the stem tissue (Fig. 1A). In the individual score plots of roots and stem, GUA35 was located away from the main cluster including all other landraces (Fig. 1C, D). This trend could not be observed in the score plot of leaf tissue (Fig. 1B). After harvest, it became obvious that none of the GUA35 plants underwent proper tuber formation, resulting in thickened roots with brown epidermis-like structure throughout. No visual difference of the stem and leaf tissue of GUA35 could be observed compared to the other cassava landraces. Nevertheless, the LC-MS metabolite profiling indicated metabolic differences of the stem tissue of GUA35. As the cause of the extreme root phenotype and underlying metabolic extend throughout the plant could not be determined, GUA35 was excluded from any further analysis.

After removal of GUA35, differential expression between the tissues was performed and relative standard deviation (RSD) was calculated to depict the variation within the data set. The differential expression showed almost 90% of the detected molecular features were significantly different between the three tissues and coincides with the PCA analysis. The highest amounts for the majority of molecular features were detected in leaf tissue. PCA analysis of individual tissues highlighted prominent landraces COL2436 and BRA488 in leaf tissue and ECU72 in stem tissue and (Fig. 1B and C). The average biological variation of the diversity panel was ~84% for leaf and root tissue and 96% for stem tissue. The African varieties TME3, NGA5 and NGA11 had a RSD of ~68% for all three tissues.

## Targeted analysis of leaf, stem and root tissue

Targeted analysis identified 102 compounds in the polar extracts of leaf, stem and root tissue over the two platforms GC-MS and LC-MS (Table A.43). The compounds detected in all three tissues comprised 85 confirmed metabolites (identification level 1 and 2), 13 unidentified sugars (identification level 3) and four level 4 unknowns (Sumner *et al.*, 2007). Additionally, eight isoprenoids (carotenoids and chlorophylls) and 15 phenolic compounds were identified in leaf tissue by UPLC-DAD and LC-MS (Table A.4). This data will be used to (i) establish a correlation between the metabolites and phenotypic traits of roots such as dry matter and cyanide (HCN) content and (ii) assess the diversity panel for apparent chemotypes underlying phenotypic traits (e.g. zone adaptation and green mites resistance), including individual comparisons of opposed traits (e.g. high and low amylose content).

### Correlation between chemotypes and cyanide and dry matter content of roots

A comparison was performed between root dry matter content, the metabolites fructose, glucose and sucrose and the total sugar content and total content of TCA cycle intermediates measured by GC-MS. Only in root tissue, sucrose and total sugar content showed a significantly negative correlation to root dry matter content (R=0.3 *P-*value ~0.011; R=0.2 *P-*value ~0.047; Fig. 2C). Leaf and stem tissue showed a positive trend between the total sugar content and root dry matter content (Fig. 2A-B).

Correlation analysis between metabolites highlighted a positive correlation between intermediates of the TCA cycle and total sugar levels (R=0.16, *P*-value ~0.002) with all three tissues used (Fig. A. 1). Additionally, glucose and fructose showed a positive correlation to sucrose (R=0.31 *P*-value ~0.011; R=0.33 *P*-value ~0.008) in roots. Exceptions to this trend were PAR36 (high amylose content), CUB74 (high culinary quality) and PAN139 (thrips resistant). The root tissue of these three varieties contained glucose and fructose levels which corresponded to half or a third of the expected average (Fig. A. 1).

The cyanogenic glucosides detected in cassava diversity panel included linamarin and lotaustralin. Linamarin comprised 75-93% of the total cyanogenic glucoside content in stem and root; and 65-90% in leaf tissue with the exception of PAR36, which had a 40.8% linamarin and was the only variety with a higher lotaustralin content in leaf. The ratio between the two metabolites varied between 0.7 and 15- fold and did not correlate to overall amounts detected in the tissue. Nevertheless, linamarin and lotaustralin were significantly correlated (*P* <0.001) with a coefficient of 0.82. This was consistent across all three tissues analysed (Fig. A. 1). Furthermore, cyanogenic glucosides and HCN content showed a positive correlation in leaf and root tissue (R=0.3 *P-*value ~0.011; R=0.2 *P-*value ~0.047; Fig. 2D and F). The root tissue showed a better trend compared to the leaf tissue. The landrace BRA488, characterised with the trait high cyanide content, had the highest levels of linamarin in leaf and root (~680 and ~490µg/g DW). Contrary to this, the cyanogenic glucoside levels detected in stem tissue of BRA488 were ~30% lower than the average (~70µg/g DW).

### Metabolic changes as a result of zone adaptation

The four landraces representing a genepool adapted to edapho-climatic zones and drought grouped in a separate manner to each other across tissues (Fig. 3). In the leaf tissue comparison, VEN77 and COL1684 were the most similar and grouped with COL1505, whereas CUB74 was the landrace with the most variance in its metabolome. The heat map included 47 significantly different metabolites and did not include catechin gallates, procyanidins, chlorophylls and half of the carotenoids detected. COL1684 was distinguishable by high levels of feruloyl- and coumaroyl-malate and ferulic acid. VEN77 had high levels of catechin, xylose, trehalose, 4-hydroxy-proline and β-alanine. COL1505 had high levels of kaempferol and its glycosides, epicatechin, ribose, tryptophan, glycine, glutamic acid and precursors for glycerolipids. CUB74 showed the highest levels for half of the significantly different metabolites. These included glucose, sucrose, myo-inositol, mannose and other unknown sugars, the amino acids isoleucine, proline, valine, GABA, serine and threonine as well as galactosylglycerol.

The stem tissue comparison showed the least significant difference and included amino acids, sugars, intermediates of the TCA cycle, cyanogenic glucosides (COL1684 and COL1505) and epicatechin gallates (CUB74). CUB74 and VEN77 grouped together based on similarities of monosaccharides and sugar alcohols such as myo-inositol, ribose and galactinol and of amino acids GABA, valine and glutamine.

The most visual difference in root tissue was that the drought tolerant VEN77 showed the lowest levels for significantly different metabolites. CUB74 showed the highest levels of amino acids. COL1684 had the highest levels of malic and maleic acid and COL1505 had the highest levels of all other TCA cycle intermediates. In addition, COL1684 had the highest levels of phenolic compounds such as epicatechin gallates, procyanidin, flavonol glycosides and of aromatic amino acids, proline and valine.

### Assessing metabolite profiles of phenotypic cassava root traits

The cassava diversity panel included three root traits (culinary quality, cyanide and amylose content) which were represented by landraces with opposing traits. A direct comparison of the high and low content landraces highlighted metabolic differences related to the phenotypic trait. Only significant differences are described and summarised for each characteristic trait (Table 2).

The two cassava landraces with high (CUB74) and low (VEN25) culinary quality showed significant differences of the majority of amino acids detected in root tissue (Fig. A.2). The root tissue of CUB74 showed higher levels of amino acids with the exception of β-alanine and threonine. This chemotype of high amino acids levels was also detected in leaf and root tissue of CUB74. In a comparison to all other landraces, CUB74 had significantly higher (~2-times) levels of the amino acids glutamic acid, citrulline, ornithine, alanine, serine and norvaline. Other differences between the two culinary quality traits included lower levels of cyanogenic glucosides and four unidentified sugars and chlorogenic acid and higher levels of epicatechin gallate and procyanidin B. The comparison of sugar levels showed no definitive pattern with even amount of higher and lower levels. Within the diversity panel, CUB74 showed sugar levels in the lower range (~76mg/g DW). Additional differences in the stem tissue included higher levels of TCA cycle intermediates and lower levels of shikimic acid. The latter was also observed in leaf tissue as well as lower levels of xanthophyll epoxides.

As expected, roots of the two landraces classified with high (BRA488) and low (PER496) cyanide content had the highest (~490µg/g DW) and lowest (~62µg/g DW) linamarin content in the diversity panel. These amounts equated to more than double and less than half of the average linamarin content, respectively. The linamarin content in stem was ~2.5 times higher in BRA488 compared to PER496. No differences in cyanogenic glucosides could be observed in leaf tissue of the two landraces (Fig. A.3). Intermediates of the TCA cycle, epicatechin gallate and threonic acid were lower in BRA488 in all three tissues and over half of the sugars detected were lower in stem tissue of BRA488. The amino acids putrescine and threonine were detected in lower levels in leaf and stem tissue of BRA488. The phenolic compounds were detected in varying levels with no consistent trends between the three tissues.

The direct comparison of the landraces with low (CUB25) and high (PAR36) amylose content showed differences of photosynthesis related pigments in leaf tissue (Fig. A.4). Within the diversity panel, CUB25 had the highest content of chlorophylls and carotenoids in leaf (3.8 and 1.54µg/g DW), whereas PAR36 showed average levels of chlorophylls and carotenoids (~2.6 and ~1.4µg/g DW). Other differences in leaf included higher levels of the majority of phenolic compounds detected in CUB25. Chlorogenic acid and ferulic acid was also higher in stem and root tissue. In the root tissue, CUB25 showed higher levels of monosaccharides including glucose and fructose and of cyanogenic glucosides. A similar trend for sugars was observed in stem tissue. An additional comparison of CUB25 and PAR36 was performed with COL1522 (Fig. A.5). This landrace is known for its flour expansion properties during bread baking due to the lower amylose content. The dendrogram of the heatmap highlighted that CUB25 and PAR36 were more similar to each other than to COL1522. The metabolite levels related to this were higher levels of sugars, including glucose and fructose, in COL1522 and higher levels of amino acids glutamic acid, glutamine, valine, serine, glycine and tyrosine in the other two landraces.

Table 2. Metabolic difference detected between landraces with different physiological root traits. Landrace with a high content of the trait is indicated with an asterisk. Metabolic changes are listed by metabolite or chemical group, by their tissue location leaf (L), stem (S) and root (R) and by amounts higher (↑) or lower (↓).

|  |  |
| --- | --- |
| **Trait (landraces)** | **Significantly higher metabolites** |
| **Culinary quality**CUB74\*VEN25 | Amino acids (L,S,R)Cyanogenic glucosides (R) |
| **Cyanide content**BRA488\*PER496 | Cyanogenic glucosides (S,R)TCA cycle, threonic acid and epicatechin gallate (L,S,R)Sugars (S)Putrescine and threonine (L,S) |
| **Amylose content**PAR36\*CUB25 | Cyanogenic glucosides (R)Monosaccharides (S,R)Chlorogenic/ferulic acid (L,S,R)Chlorophylls, carotenoids (L)Amino acids of nitrogen metabolism (L,S,R) |

### Assessing metabolite profiles of phenotypic cassava leaf traits

The primary effect of biotic stresses on cassava plants can be detected in leaves and can lead to a detrimental impact on the tuberousroot formation. The present study included plants grown under standard field conditions and should not have been affected by any biotic stresses. Hence, the comparison of the landraces was focused on leaf metabolites, which could refer antixenosis properties and contribute to the resistance of the respective landrace. This was followed by a comparison of root tissues to assess the differences to the high culinary quality landrace CUB74. The metabolic differences detected between the landraces were summarised for each trait (Table 3).

The two cassava landraces with resistant (PAN139) and susceptible (COL2436) thrips properties showed distinct differences of chlorophyll and carotenoid levels, which were higher in PAN139 (Fig. A.6). The levels of cyanogenic glucosides were higher in COL2436 for both leaf and root tissue. The phenolic compounds in leaf tissue followed two trends. Kaempferol and derived compounds were higher in PAN139 and catechin and derived compounds (gallates and procyanidins) were higher in COL2436. Additionally, PAN139 had higher levels of ferulic acid and glycerolipid precursors and COL2436 had higher levels of quercetin, shikimic acid and chlorogenic acid. Fructose and intermediates of the TCA cylce were higher in COL2436 root tissue. Interestingly, the stem tissue only showed three significantly different metabolites including threonine, lotaustralin and chlorogenic acid.

Five landraces in the diversity panel were previously tested for their green mite resistance. The resistant landraces included COL113, COL1522 and CUB74 and the susceptible landraces COL1684 and VEN77. The hierarchical relationship of leaf metabolites showed that CUB74 was more similar to the susceptible landraces (Fig. A.7). Overall, 79 metabolites were significantly different between the five landraces. Chlorophyll and carotenoids were higher in COL113 and COL1522 as well as epicatechin gallates and kaempferol-pentoside. As described for culinary quality, CUB74 showed the highest levels of a majority of amino acids in both leaf and root tissue. The susceptible landraces COL1684 and VEN77 had higher levels of amino acids involved in the nitrogen metabolism, hydroxycinnamic acids and derivatives, catechin and procyanidin B3. Despite these similarities, the levels for the metabolite groups varied between the two landraces susceptible to green mites. COL1684 had the highest levels of hydroxycinnamic acids and VEN77 of catechin and procyanidin B3. The heatmap of root metabolites did not show any metabolic trends detected in leaf tissue.

A comprehensive comparison between eight landraces with herbivorous pest traits (green mites, thrips and whitefly) was performed and highlighted some metabolic similarities (Fig. A.8). The resistant green mite landraces COL113 and COL1522 were grouped with ECU72, a resistant whitefly landrace (Fig. 4). The two green mite resistant landraces show the highest levels for metabolites already described in the green mite trait comparison. ECU72 was distinguishable by higher levels of sucrose and precursors for glycolipids, whereas most other compounds showed average levels compared to the other seven landraces. The second hierarchical arm comprised the susceptible green mite landraces COL1684 and VEN77, the resistant green mite landrace CUB74 and the two landraces included for thrips traits. The thrips resistant PAN139 showed the most metabolite similarity to the two susceptible green mite landraces. Common metabolite levels included amino acids, intermediates of the TCA cycle and sugars. The thrips susceptible COL2436 had the highest levels of metabolites described for the thrips trait comparison. In the comparison of the root tissues, the green mite susceptible COL1684 and the thrips susceptible COL2436 were grouped with CUB74, known for high culinary quality and green mite resistance. These three landraces had the highest levels of metabolites described for CUB74 in the culinary quality comparison. All other resistant landraces were grouped with the low culinary quality landrace VEN25. Interestingly, COL113, COL1522 and ECU72 were grouped in one hierarchical arm, as observed for the leaf tissue. The heatmap highlighted that COL113 had levels of amino acids and TCA cycle intermediates similar to CUB74. In addition, COL113 had the highest levels of catechin, quercetin and their derivatives and rutin, which was the opposite of CUB74. In comparison, ECU72 and COL1522 showed the opposite trends in contents for the majority of metabolites.

The three African varieties TME3, NGA5 and NGA11 were the result of a CMD breeding program and are known for their resistance, moderate susceptibility and high susceptibility to the virus, respectively. As described for other traits, the hierarchical relationships between the varieties varied between the three tissues (Fig. A.9). The leaf tissue showed more similarity between TME3 and NGA11, the stem tissue between NGA5 and NGA11 and in root tissue between TME3 and NGA5. Yet, some metabolite trends could be observed between leaf and root tissue. NGA11 had higher levels of monosaccharides and NGA5 had higher levels of epicatechin gallate. Overall, 22 metabolites were significantly different between leaf tissue of the three varieties and included mainly primary metabolites. The exception were higher levels of kaempferol-rutinoside and –glucoside and coumaroyl-malate in TME3, epicatechin gallate and epigallocatechin gallate in NGA5. In the root tissue, only 17 metabolites showed significant difference and included amino acids, monosaccharides, linamarin and epicatechin gallate.

Table 3. Metabolic differences detected between landraces with resistant and susceptible phenotypes. Resistant landraces are indicated by an asterisk.

|  |  |
| --- | --- |
| **Trait (landraces)** | **Significantly higher metabolites** |
| **Thrips** |
| PAN139\*COL2436 | Chlorophylls and carotenoids (L)Kaempferol-rutinoside, ferulic acid (L)Cyanogenic glucosides (L,R)Catechin (L,R)Gallocatechin gallate, procyanidins (L) |
| **Green mites** |
| COL113\*, COL1522\*CUB74\*COL1684, VEN77 | Chlorophyll and carotenoids (L)Epicatechin gallates (L)Kaempferol-pentoside (L)Fructose, sucrose (L)Majority of amino acids (L)Chlorogenic acid (L)Nitrogen metabolism (L)Feruloyl- and coumaroyl-malate (L)Catechin, procyanidin B3 (L) |
| **Cassava mosaic virus (CMD)** |
| TME3\*NGA5NG11 | Kaempferol-sugars (L)Epicatechin gallate (L,R)Monosaccharides (L,R) |

# Discussion

Metabolite profiling is a useful tool to characterise potential breeding materials and elucidate the mechanisms underlying desired traits. The present study included 19 cassava landraces grown under standard field conditions, which presents an environment with reduced biotic and abiotic stress conditions. Hence, the metabolic information derived from this data set does not include induced metabolic responses to the stress, but rather the constitutive metabolite composition of the respective phenotypes.

## Metabolite composition of root phenotypes

Leaves produce carbohydrates throughout their life cycle and store them in the form of starch molecules. During the tuberous root development, the plant changes its signalling to transport carbohydrates from the leaf to the roots (Chaweewan and Taylor, 2015; Mitprasat *et al.*, 2011). The present study highlighted that the source-sink relation, observed by the positive correlation between leaf sugar root dry matter content, was still active at the harvest stage, contrary to previous findings (Luo and Huang, 2011). This was further corroborated by the positive correlation between the TCA cycle and sugar content in all three tissues. As expected, the correlation between sugar and dry matter content was negative in root tissue (Schreiber *et al.*, 2014).

A major proportion of cassava dry matter are carbohydrates, which are mainly present in the form of starch molecules (Montagnac *et al.*, 2009). Cassava starch has risen to the second most important source worldwide, due to its novel properties of low to no amylose content (Karlström *et al.*, 2016). The landraces with amylose traits showed different contents of glucose, fructose and dry matter, which was lowest in high amylose PAR36, followed by low amylose CUB25 and highest in low amylose COL1522. The levels of fructose and glucose is probably related to both the dry matter content of the roots and the mutations related to amylose synthesis. Cassava chromosome 2 contains the genetic code for four enzymes involved in starch biosynthesis including granule bound starch synthase (GBSS) responsible for amylose formation (Tappiban *et al.*, 2019). Overexpression and suppression efforts through breeding of these landraces could have led to additional mutations leading to an overall more or less efficient starch synthesis, which is reflected in the total sugar content present in root tissue.

Cyanogenic glucosides are present in all cassava plants and refer, depending on the content levels, a sweet or bitter phenotype. Previous studies elucidated the function of cyanogenic glucosides as herbivore deterrent and source of reduced nitrogen for protein synthesis (e.g. White *et al.*, 1998; Zidenga *et al.*, 2017). The linamarin levels detected in leaf tissue were much lower than expected (White *et al.*, 1998). This could be a result of several processes including depleting sink-source transport or high linamarase activity in roots upon root formation (Mitprasat *et al.*, 2011). All of these scenarios present a leaf chemotype unsuitable for elucidation of this particular root phenotype. The two landraces with high (BRA488) and low (PER496) cyanide content showed very few differences of root tissue including linamarin and lotaustralin. Another difference of BRA488 was higher sugar levels in root and leaf tissue, as previously described for “bitter” cassava (Araújo *et al.*, 2019).

The third root phenotype study included culinary quality. This particular trait is based on personal preference towards texture/mealiness, flavour and cooking properties of the cassava roots. All three properties are dependent on the composition and range of metabolites present in the tissue and how they behave during the food processing. Previous and current studies are focused on the assessment of bread mixes including cassava flour or culinary qualities affected by harvest stage (Jensen *et al.*, 2015; Dada *et al.*, 2017; Ngeve, 2003). The landrace CUB74 is commonly associated with good culinary quality, a product of its unique chemotype and an average dry matter content. The flavour of CUB74 should be conceived as savoury with average sweetness, sourness and low astringency (Izawa *et al.*, 2010). This is based on the higher than average levels of amino acids associated with umami flavour, low levels of phenolic compounds referring astringency and overall average levels of metabolites associated with sweet and sour flavours. This hypothesis is supported by the metabolite composition of the low culinary quality landrace VEN25, which contained above average cyanogenic glucosides, low amounts of umami compounds, average to high phenolic compounds, average sourness, and average to low amounts of metabolites referring sweetness.

## Zone adaptation

Cassava breeding programmes generated gene pools for six edapho-climatic zones which were selected for their good productivity within the respective environment (Hershey and Jennings, 1992). The present study included landraces adapted to lowland tropics, subtropics and drought conditions. Their constitutive metabolite profile placed the drought tolerant VEN77 in the middle of the leaf score plot (Fig. 1B), close to CUB74 and COL1684. This suggests that the general metabolism of these landraces does not include specialised or enhanced pathways prior to the abiotic stress of the edapho-climatic zones. On the contrary, COL1505 clustered towards the right side of the score plot with the African lines. In the direct comparison of the adapted landraces, COL1505 showed higher levels of kaempferol-glycosides, which are involved in auxin transport and herbivore defence (Zhang *et al.*, 2013; Peer and Murphy, 2007). This particular chemotype might have given COL1505 the advantage to remain productive in its genetic origin Savannah and adapt to the biotic and abiotic stresses of zone 1 and 4. So far, studies on zone adaptation have focused on the agronomic aspects of the plant and not yet considered the root quality of the best performing plants (Fraser *et al.*, 2012; Ogola and Mathews, 2011; Shan *et al.*, 2018). The metabolic comparison highlighted that VEN77 showed a similar profile to the landrace with low culinary quality VEN25, whereas COL1505 and COL1684 grouped with the high culinary quality CUB74.

## Metabolite profiles of cassava plants resistant and susceptible to biotic stresses

Biotic stresses are a great threat to cassava yield and food security leading to the search for sustainable solutions such as introduction of herbivore predators (Chen *et al.*, 2019; Onzo *et al.*, 2005). The most sought out solution against pests is the search for resistance and tolerance traits in cassava, through understanding the cellular plant processes during the herbivore attack (Mitchell *et al.*, 2016). A previous metabolomics study elucidated the defensive traits of ECU72 to whitefly infestation (Perez-Fons *et al.*, 2019). The trait included a reinforced cell walls deterring/reducing whitefly from feeding and a more effective phenylpropanoid superpathway during infection. The present study does not include any insect infestations. Hence, results from the present study present potential pathways or metabolites which could refer pest resistance and should be further investigated in infestation studies.

The comparison of thrips resistant and susceptible landraces highlighted different expression of the two phenylpropanoid pathways flavonoids and proanthocyanidins. Previous studies highlighted both compounds groups can have insect deterrent properties (Mierziak *et al.*, 2014; Barbehenn and Peter Constabel, 2011). Data from the present study indicates that kaempferol glycosides are more effective against thrips than proanthocyanidins. The data also showed higher levels of photosynthesis related isoprenoids and lower levels of total phenolic compounds including shikimic acid in PAN139, which indicates a different regulation of the two pathways in the resistant landrace. This might present the resistant PAN139 with superior chemical traits to reduce (i) the number of thrips attacks and resulting leaf damage and (ii) the negative effects of the thrips damage on plant growth (Mitchell *et al.*, 2016; Mierziak *et al.*, 2014).

The landraces COL113 and COL1522, resistant to green mites, had higher levels of photosynthesis related isoprenoids, which might give them the same fitness advantage as describe for PAN139. The total levels of phenolic compounds was similar between the green mites resistant and susceptible landrace. This emphasises the beneficial effects of certain phenolic compounds over others. The two green mite resistant landraces had higher levels of flavonols and flavan-3-ol derivatives. Both flavonoid subgroups are involved in the modulation of jasmonic acid, a well-known mediator of plant defence responses (Hong *et al.*, 2015; Peer and Murphy, 2007; Zhang *et al.*, 2017). The other green mite resistant landrace CUB74 was grouped with the two susceptible varieties suggesting that the resistance is an induced response rather than an antixenosis strategy.

The comparison between cassava plants with thrips, green mite and whitefly traits corresponded to many previous studies observing the influence of phenylpropanoids on insects (Simmonds, 2003; Barbehenn and Peter Constabel, 2011). The data from the present study suggested that resistance to thrips is based on different phenolic compounds than resistance to green mites and whitefly. However, the heatmap showed a different metabolite composition of the whitefly resistant ECU72, despite its close grouping with two green mite resistant landraces. The data also showed that ECU72 had the lowest levels of total phenolic compounds. This could be a result of the more effective phenylpropanoid superpathway and the reinforced cell wall, described previously by Perez-Fons *et al.* (2019). However, the resistance mechanism of ECU72 and two green mite resistant landraces had a detrimental effect on the metabolite composition of the root. All three landraces showed high levels of phenolic compounds in the root tissue, including flavonoid glycosides and gallates, and grouped with the low culinary quality landrace VEN25. This data highlighted the importance of investigating the resistance mechanism of the landrace CUB74 high culinary quality as it can maintain a good root quality despite the pest resistance. In addition it supports the need for robust molecular markers to facilitate efficient breeding without genetic drag of detrimental traits and the ability to pyramid favourable traits.

The other biotic stress affecting cassava productivity are viruses. The cassava mosaic geminivirus, causing CMD in cassava plants, is primarily transmitted through whiteflies causing yield loss of up to 100%. Furthermore, direct infection of leaf showed at best suppression of the virus-activated processes (Bi *et al.*, 2010; Lokko *et al.*, 2006). Based on current knowledge, all CMD resistance mechanisms have an associated molecular mechanism, which is not visible in the metabolite profile (Kuria *et al.*, 2017). This was supported by the comparison of CMD resistant (TME3), moderately susceptible (NGA5) and susceptible (NGA11) cassava plants showing less than 25% of the identified metabolites as significantly different. Additionally, the dendrogram of leaf tissue grouped the tolerant TME3 and susceptible NGA11 in one arm. Although TME3 showed clearly distinct and elevated levels of kaempferol glycosides, the functionality of this compound group in context of CMD resistance would need further investigation using metabolic analysis throughout the infection cycle (Peer and Murphy, 2007).

# Conclusion

This work has demonstrated the utility of metabolite profiling to assess key cassava landraces that have different agroecological adaption and confer key traits beneficial for quality and resilience to abiotic and biotic stresses. Using the approach key differentiating metabolites have been associated to traits. These data will help future pre-breeding strategies directed towards the identification of molecular/biochemical markers and pedigree breeding.

# Conflict of interest

The authors declare no conflict of interest.

# Acknowledgments

The authors would like to thank Chris Gerrish, Sarah Iftikhar and Kit Liew for their excellent technical assistance. This work was supported by a CGIAR Research Program on Roots, Tubers and Bananas (RTB) grant to PDF.

# Appendix A. Supplementary data

Table A.1. Library of metabolites detected in cassava diversity panel.

Table A.2. LC-MS metabolite profiling of leaf, stem and root tissues.

Table A.3. Identified compounds in leaf, stem and root tissue.

Table A.4. Additional compounds identified in leaf tissue.

Fig. A.1. Correlation between metabolites, cyanide and dry matter content.

Fig. A.2. Pathway display of culinary quality

Fig. A.3. Pathway display of cyanide content

Fig. A.4. Pathway display of amylose content

Fig. A.5. Heatmap amylose content II

Fig. A.6. Pathway display of thrips resistance

Fig. A.7. Heatmap green mite resistance

Fig. A.8. Heatmap of thrips, green mite and whitefly resistance

Fig. A.9. Heatmap CMD resistance

# Availability of data and materials

Processed data is available in the manuscript and appendices. Unprocessed data can be accessed at DOI: 10.17632/hm2hrmx3hg.2.

# References

**Aiemnaka, P., Wongkaew, A., Chanthaworn, J., et al.** (2012) Molecular Characterization of a Spontaneous Waxy Starch Mutation in Cassava. *Crop Sci.*, **52**, 2121–2130. Available at: https://doi.org/10.2135/cropsci2012.01.0058.

**Alvarez, E., Mejía, J.F., Llano, G.A., Loke, J.B., Calari, A., Duduk, B. and Bertaccini, A.** (2009) Characterization of a Phytoplasma Associated with Frogskin Disease in Cassava. *Plant Dis.*, **93**, 1139–1145. Available at: http://dx.doi.org/10.1094/PDIS-93-11-1139.

**Araújo, F. das C.B. de, Moura, E.F., Cunha, R.L., Farias Neto, J.T. de and Silva, R. de S.** (2019) Chemical root traits differentiate \_bitter\_ and \_sweet\_ cassava accessions from the Amazon . *Crop Breed. Appl. Biotechnol.* , **19**, 77–85.

**Barbehenn, R. V. and Peter Constabel, C.** (2011) Tannins in plant–herbivore interactions. *Phytochemistry*, **72**, 1551–1565. Available at: https://www.sciencedirect.com/science/article/pii/S0031942211000690?via%3Dihub#f0005 [Accessed April 17, 2020].

**Belalcazar, J., Dufour, D., Andersson, M.S., et al.** (2016) High-Throughput Phenotyping and Improvements in Breeding Cassava for Increased Carotenoids in the Roots. *Crop Sci.*, **56**, 2916–2925. Available at: http://dx.doi.org/10.2135/cropsci2015.11.0701.

**Bi, H., Aileni, M. and Zhang, P.** (2010) Evaluation of cassava varieties for cassava mosaic disease resistance jointly by agro-inoculation screening and molecular markers. *African J. Plant Sci.*, **4**, 330–338.

**Bino, R.J., Hall, R.D., Fiehn, O., et al.** (2004) Potential of metabolomics as a functional genomics tool. *Trends Plant Sci*, **9**, 418–425. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15337491.

**Bredeson, J. V, Lyons, J.B., Prochnik, S.E., et al.** (2016) Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat Biotechnol*, **34**, 562–570. Available at: https://www.ncbi.nlm.nih.gov/pubmed/27088722.

**Chaweewan, Y. and Taylor, N.** (2015) Anatomical Assessment of Root Formation and Tuberization in Cassava (Manihot esculenta Crantz). *Trop. Plant Biol.*, **8**, 1–8. Available at: https://doi.org/10.1007/s12042-014-9145-5.

**Chen, Qing, Liang, X., Wu, C., Gao, J., Chen, Qian and Zhang, Z.** (2019) Density threshold-based acaricide application for the two-spotted spider mite Tetranychus urticae on cassava: from laboratory to the field. *Pest Manag. Sci.*, **75**, 2634–2641. Available at: https://doi.org/10.1002/ps.5366.

**Dada, T.A., Barber, L.I., Ngoma, L. and Mwanza, M.** (2017) Formulation, sensory evaluation, proximate composition and storage stability of cassava strips produced from the composite flour of cassava and cowpea. *Food Sci. Nutr.*, **6**, 395–399. Available at: https://pubmed.ncbi.nlm.nih.gov/29564107.

**Drapal, M., Barros De Carvalho, E., Ovalle Rivera, T.M., Becerra Lopez-Lavalle, L.A. and Fraser, P.D.** (2019) Capturing Biochemical Diversity in Cassava (Manihot esculenta Crantz) through the Application of Metabolite Profiling. *J. Agric. Food Chem.*, **67**.

**Drapal, M., Carvalho, E. de, Ovalle Rivera, T.M., Becerra Lopez-Lavalle, L.A. and Fraser, P.D.** (2018) Capturing biochemical diversity in cassava (Manihot esculenta Crantz) through the application of metabolite profiling. *J Agric Food Chem*, **67**, 986–993. Available at: https://www.ncbi.nlm.nih.gov/pubmed/30557498.

**Fraser, J.A., Alves-Pereira, A., Junqueira, A.B., Peroni, N. and Clement, C.R.** (2012) Convergent Adaptations: Bitter Manioc Cultivation Systems in Fertile Anthropogenic Dark Earths and Floodplain Soils in Central Amazonia. *PLoS One*, **7**, e43636. Available at: https://doi.org/10.1371/journal.pone.0043636.

**Fraser, P.D., Pinto, M.E., Holloway, D.E. and Bramley, P.M.** (2000) Technical advance: application of high-performance liquid chromatography with photodiode array detection to the metabolic profiling of plant isoprenoids. *Plant J*, **24**, 551–558. Available at: https://www.ncbi.nlm.nih.gov/pubmed/11115136.

**Hahn, S.K., Terry, E.R., Leuschner, K., Akobundu, I.O., Okali, C. and Lal, R.** (1979) Cassava improvement in Africa. *F. Crop. Res.*, **2**, 193–226. Available at: http://www.sciencedirect.com/science/article/pii/0378429079900248.

**Henry, G. and Hershey, C.** (2002) Cassava in South America and the Caribbean. In R. J. Hillocks, J. M. Thresh, and A. C. Bellotti, eds. *Cassava: Biology; production and utilization*. Wallingford: CABI Publishing, pp. 17–40.

**Hershey, C.H. and Jennings, D.L.** (1992) Progress in breeding cassava for adaptation to stress. *Plant Breed. Abstr.*, **62**, 823–831.

**Hong, G., Wang, J., Hochstetter, D., Gao, Y., Xu, P. and Wang, Y.** (2015) Epigallocatechin-3-gallate functions as a physiological regulator by modulating the jasmonic acid pathway. *Physiol. Plant.*, **153**, 432–439. Available at: https://doi.org/10.1111/ppl.12256.

**Howeler, R., Lutaladio, N. and Thomas, G.** (2013) *Save and Grow: Cassava - A guide to sustainable production intensification*, Rome, Italy: Food and Agriculture Organization of the United Nations.

**Izawa, K., Amino, Y., Kohmura, M., Ueda, Y. and Kuroda, M.** (2010) 4.16 - Human–Environment Interactions – Taste. In *Comprehensive Natural Products II*. Oxford: Elsevier, pp. 631–671. Available at: https://www.sciencedirect.com/science/article/pii/B9780080453828001088.

**Jensen, S., Skibsted, L.H., Kidmose, U. and Thybo, A.K.** (2015) Addition of cassava flours in bread-making: Sensory and textural evaluation. *LWT - Food Sci. Technol.*, **60**, 292–299. Available at: http://www.sciencedirect.com/science/article/pii/S0023643814005453.

**Karlström, A., Calle, F., Salazar, S., Morante, N., Dufour, D. and Ceballos, H.** (2016) Biological Implications in Cassava for the Production of Amylose-Free Starch: Impact on Root Yield and Related Traits . *Front. Plant Sci.* , **7**, 604. Available at: https://www.frontiersin.org/article/10.3389/fpls.2016.00604.

**Kuria, P., Ilyas, M., Ateka, E., Miano, D., Onguso, J., Carrington, J.C. and Taylor, N.J.** (2017) Differential response of cassava genotypes to infection by cassava mosaic geminiviruses. *Virus Res.*, **227**, 69–81. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5130204/.

**Liu, Q., Liu, J., Zhang, P. and He, S.** (2014) Root and Tuber Crops. In N. K. B. T.-E. of A. and F. S. Van Alfen, ed. Oxford: Academic Press, pp. 46–61. Available at: http://www.sciencedirect.com/science/article/pii/B9780444525123001510.

**Lokko, Y., Dixon, A., Offei, S., Danquah, E. and Fregene, M.** (2006) Assessment of genetic diversity among African cassava Manihot esculenta Grantz accessions resistant to the cassava mosaic virus disease using SSR markers. *Genet. Resour. Crop Evol.*, **53**, 1441–1453. Available at: https://doi.org/10.1007/s10722-005-6841-x.

**Luo, X. and Huang, Q.** (2011) Relationships between Leaf and Stem Soluble Sugar Content and Tuberous Root Starch Accumulation in Cassava. *2011*, **3**. Available at: http://www.ccsenet.org/journal/index.php/jas/article/view/7514.

**Mierziak, J., Kostyn, K. and Kulma, A.** (2014) Flavonoids as Important Molecules of Plant Interactions with the Environment. *Molecules*, **19**, 16240–16265. Available at: http://www.mdpi.com/1420-3049/19/10/16240 [Accessed April 17, 2020].

**Mitchell, C., Brennan, R.M., Graham, J. and Karley, A.J.** (2016) Plant Defense against Herbivorous Pests: Exploiting Resistance and Tolerance Traits for Sustainable Crop Protection . *Front. Plant Sci.* , **7**, 1132. Available at: https://www.frontiersin.org/article/10.3389/fpls.2016.01132.

**Mitprasat, M., Roytrakul, S., Jiemsup, S., Boonseng, O. and Yokthongwattana, K.** (2011) Leaf proteomic analysis in cassava (Manihot esculenta, Crantz) during plant development, from planting of stem cutting to storage root formation. *Planta*, **233**, 1209–1221. Available at: https://doi.org/10.1007/s00425-011-1373-4.

**Montagnac, J.A., Davis, C.R. and Tanumihardjo, S.A.** (2009) Nutritional Value of Cassava for Use as a Staple Food and Recent Advances for Improvement. *Compr. Rev. Food Sci. Food Saf.*, **8**, 181–194. Available at: https://doi.org/10.1111/j.1541-4337.2009.00077.x.

**Ngeve, J.M.** (2003) Cassava root yields and culinary qualities as affected by harvest age and test environment. *J. Sci. Food Agric.*, **83**, 249–257. Available at: https://doi.org/10.1002/jsfa.1307.

**Nogueira, M., Mora, L., Enfissi, E.M., Bramley, P.M. and Fraser, P.D.** (2013) Subchromoplast sequestration of carotenoids affects regulatory mechanisms in tomato lines expressing different carotenoid gene combinations. *Plant Cell*, **25**, 4560–4579. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24249831.

**Ogola, J. and Mathews, C.** (2011) Adaptation of cassava (Manihot esculenta) to the dry environments of Limpopo, South Africa: growth, yield and yield components. *African J. Agric. Res.*, **6**, 6082–6088.

**Omongo, C.A., Kawuki, R., Bellotti, A.C., Alicai, T., Baguma, Y., Maruthi, M.N., Bua, A. and Colvin, J.** (2012) African Cassava Whitefly, Bemisia tabaci, Resistance in African and South American Cassava Genotypes. *J. Integr. Agric.*, **11**, 327–336. Available at: http://www.sciencedirect.com/science/article/pii/S2095311912600173.

**Onzo, A., Hanna, R., Negloh, K., Toko, M. and Sabelis, M.** (2005) Biological control of cassava green mite with exotic and indigenous phytoseiid predators—Effects of intraguild predation and supplementary food. *Biol. Control*, **33**, 143–152.

**Peer, W.A. and Murphy, A.S.** (2007) Flavonoids and auxin transport: modulators or regulators? *Trends Plant Sci.*, **12**, 556–563. Available at: https://doi.org/10.1016/j.tplants.2007.10.003.

**Perez-Fons, L., Bohorquez-Chaux, A., Irigoyen, M.L., Garceau, D.C., Morreel, K., Boerjan, W., Walling, L.L., Becerra Lopez-Lavalle, L.A. and Fraser, P.D.** (2019) A metabolomics characterisation of natural variation in the resistance of cassava to whitefly. *BMC Plant Biol.*, **19**, 518. Available at: https://doi.org/10.1186/s12870-019-2107-1.

**Piperno, D.R.** (2011) The Origins of Plant Cultivation and Domestication in the New World Tropics: Patterns, Process, and New Developments. *Curr. Anthropol.*, **52**, S453–S470. Available at: http://www.jstor.org/stable/10.1086/659998.

**Price, E.J., Drapal, M., Perez-Fons, L., et al.** (2020) Metabolite database for root, tuber, and banana crops to facilitate modern breeding in understudied crops. *Plant J.*, **101**, 1258–1268. Available at: https://doi.org/10.1111/tpj.14649.

**Roessner, U. and Bowne, J.** (2009) What is metabolomics all about? *Biotechniques*, **46**, 363–365. Available at: https://doi.org/10.2144/000113133.

**Schreiber, L., Nader-Nieto, A.C., Schönhals, E.M., Walkemeier, B. and Gebhardt, C.** (2014) SNPs in Genes Functional in Starch-Sugar Interconversion Associate with Natural Variation of Tuber Starch and Sugar Content of Potato (Solanum tuberosum L.). *G3 Genes|Genomes|Genetics*, **4**, 1797–1811. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4199688/.

**Shan, Z., Luo, X., Wei, M., Huang, T., Khan, A. and Zhu, Y.** (2018) Physiological and proteomic analysis on long-term drought resistance of cassava (Manihot esculenta Crantz). *Sci. Rep.*, **8**, 17982. Available at: https://doi.org/10.1038/s41598-018-35711-x.

**Simmonds, M.S.J.** (2003) Flavonoid–insect interactions: recent advances in our knowledge. *Phytochemistry*, **64**, 21–30. Available at: http://www.sciencedirect.com/science/article/pii/S0031942203002930.

**Skovgård, H., Tomkiewicz, J., Nachman, G. and Münster-Swendsen, M.** (1993) The effect of the cassava green mite Mononychellus tanajoa on the growth and yield of cassava Manihot esculenta in a seasonally dry area in Kenya. *Exp. Appl. Acarol.*, **17**, 41–58. Available at: https://doi.org/10.1007/BF00156943.

**Sumner, L.W., Amberg, A., Barrett, D., et al.** (2007) Proposed minimum reporting standards for chemical analysis. *Metabolomics*, **3**, 211–221. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3772505/pdf/nihms504189.pdf.

**Tappiban, P., Smith, D.R., Triwitayakorn, K. and Bao, J.** (2019) Recent understanding of starch biosynthesis in cassava for quality improvement: A review. *Trends Food Sci. Technol.*, **83**, 167–180. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0924224418305053 [Accessed April 14, 2020].

**Wang, W., Feng, B., Xiao, J., et al.** (2014) Cassava genome from a wild ancestor to cultivated varieties. *Nat Commun*, **5**, 5110. Available at: https://www.ncbi.nlm.nih.gov/pubmed/25300236.

**White, W.L.B., Arias-Garzon, D.I., McMahon, J.M. and Sayre, R.T.** (1998) Cyanogenesis in cassava. The role of hydroxynitrile lyase in root cyanide production. *Plant Physiol.*, **116**, 1219–1225. Available at: https://pubmed.ncbi.nlm.nih.gov/9536038.

**Xia, J. and Wishart, D.S.** (2016) Using MetaboAnalyst 3.0 for comprehensive metabolomics data analysis. In *Current Protocols in Bioinformatics*. John Wiley & Sons, Inc. Available at: http://dx.doi.org/10.1002/cpbi.11.

**Zhang, L., Zhang, F., Melotto, M., Yao, J. and He, S.Y.** (2017) Jasmonate signaling and manipulation by pathogens and insects. *J. Exp. Bot.*, **68**, 1371–1385. Available at: https://pubmed.ncbi.nlm.nih.gov/28069779.

**Zhang, Q., Zhao, X. and Qiu, H.** (2013) Flavones and Flavonols: Phytochemistry and Biochemistry BT - Natural Products: Phytochemistry, Botany and Metabolism of Alkaloids, Phenolics and Terpenes. In K. G. Ramawat and J.-M. Mérillon, eds. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1821–1847. Available at: https://doi.org/10.1007/978-3-642-22144-6\_60.

**Zidenga, T., Siritunga, D. and Sayre, R.T.** (2017) Cyanogen Metabolism in Cassava Roots: Impact on Protein Synthesis and Root Development . *Front. Plant Sci.* , **8**, 220. Available at: https://www.frontiersin.org/article/10.3389/fpls.2017.00220.

# Figure captions

Fig. 1. PCA analysis of untargeted analysis of leaf, stem and root of 20 cassava landraces. Untargeted data is shown as a combined plot of all three tissues (A) and as individual plots of leaf (B), stem (C) and root (D). Data displayed is based on features identified in polar extracts of all three tissues. Each variety is displayed as an average of six biological replicates.

Fig. 2. Correlation analysis between phenotypic measurements of the roots and chemotypic measurements of each plant tissue. Phenotypic measurements of root tissue included dry matter (A-C) and cyanide (HCN) content (D-F). Correlation trend is displayed by linear regression. Each variety is displayed as an average of six biological replicates.

Fig. 3. Heat maps of landraces with zone adaptation properties based on significantly different compounds. The data was displayed separately for leaf (A), stem (B) and root (C). Significance was established by ANOVA and included six biological replicates for each landrace.

Fig. 4. Dendrograms of leaf (A) and root (B) tissue of cassava landraces with herbivorous pest traits. Herbivorous pests include green mites (G), thrips (T) and whitefly (W). Resistant landraces are indicated by an asterisk. VEN25 was only included in root tissue comparison (B) as a reference for low culinary quality. Data includes only significantly different metabolites of six biological replicates per landrace. The associated heatmap can be found in supplementary data (Fig. A.8.).

# Figures



Fig. 1.



Fig. 2.



Fig. 3.



Fig. 4