Short title : Cyanide genetic architecture in cassava

Title : **Genetic architecture and gene mapping of cyanide in cassava *(Manihot esculenta Crantz.*)**

Authors : Alex C Ogbonna[1,2], Luciano Rogerio Braatz de Andrade[3], Ismail Y. Rabbi[4], Lukas A. Mueller[1,2], Eder Jorge de Oliveira[3] and Guillaume J. Bauchet[2]

[1] Cornell University, Ithaca, NY, USA. [2] Boyce Thompson Institute for Plant Research, Ithaca, NY, USA. [3] Embrapa Mandioca e Fruticultura, Cruz das Almas, BA - Brazil.[4] International institute of Tropical Agriculture, Ibadan, Oyo state, Nigeria

One-sentence summary

Identification of an intracellular transporter gene and its allelic variation allow to point out cultivars with up to 30 percent decrease in cassava root cyanide content, toxic for human consumption.

Footnotes

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Eder Jorge de Oliveira (eder.oliveira@embrapa.br).

Abstract

Cassava is a root crop originating from South America and a major staple crop in the Tropics, including marginal environments. In this study, we focused on South American and African cassava germplasm and investigated the genetic architecture of Hydrogen Cyanide (HCN), a major component of tuber quality. HCN is a plant defense component against herbivory but also toxic for human consumption. We genotyped 3,354 landraces and modern breeding lines originating from 26 Brazilian states and 1,389 individuals were phenotypically characterized across multi-year trials for HCN. All plant material was subjected to high density genotyping using Genotyping-by-sequencing (GBS). We performed genome wide association mapping (GWAS) to characterize the genetic architecture and gene mapping of HCN. Field experiment revealed strong broad and narrow-sense trait heritability (0.82 and 0.41 respectively). Two major loci were identified, encoding for an ATPase and a MATE protein and contributing up to 7% and 30% of the cyanide concentration in roots, respectively. We developed diagnostic markers for

breeding applications, validated trait architecture consistency in African germplasm and investigated further evidence for domestication of sweet and bitter cassava. Fine genomic loci characterization indicate; (i) a major role played by vacuolar transporter in regulating HCN content, (ii) co-domestication of sweet and bitter cassava major alleles to be geographical zone dependant, and (ii) major loci allele for high cyanide cassava in *Manihot esculenta Crantz* seems to originate from its ancestor, *M. esculenta* ssp. *flabellifolia.* Taken together these findings expand insights on cyanide in cassava and its glycosylated derivatives in plants.

Introduction

Cassava (*Manihot esculenta Crantz.*) is a starchy root crop widely grown throughout the tropics (Southeast Asia, Latin America, the Caribbean and sub-Saharan Africa) for human and livestock consumption, and as feedstock for biofuels and other bio-based materials (Howeler, Lutaladio, and Thomas 2013). Mostly cultivated by low-income, smallholder farmers, cassava is a staple food crop for over 800 million people worldwide. Cassava is an efficient crop in marginal areas where poor soils and unpredictable rainfall dominates (FAO 2018). Cassava has developed defense mechanisms against herbivores and pathogens, including the biosynthesis of cyanogenic glucosides (CG) (Takos et al. 2011). However, some of the major challenges in cassava includes low tuber protein and carotenoid content as well as high content of linamarin and lotaustralin CG (K. Jørgensen et al. 2005; Blomstedt et al. 2012; Gleadow and Møller 2014). CG, characterized as α-hydroxynitriles, are secondary metabolites of plant products derived from amino acids (Gleadow and Møller 2014). Cyanogenesis is the release of toxic hydrogen cyanide in cassava upon tissue disruption. Its concentration is usually higher in young plants, when nitrogen is in ready supply, or when growth is constrained by nonoptimal growth conditions (Gleadow and Møller 2014).

Cultivars with cyanide content of < 100 mg kg^{-1} fresh weight (FW) are called 'sweet cassava' while cultivars with 100-500 mg kg^{-1} are 'bitter cassava' (Wheatley et al., 1993). In Brazil, cassava's center of diversity, the difference in preference of bitter and sweet cassava appears to be in its role in subsistence in regions they dominate. Areas where sweet cassava dominates, it is a component of a diet; whereas areas where bitter cassava dominates, it is the main carbohydrate source, generally complemented by a protein, such as a fish (Mühlen et al. 2019).

Cyanide in cassava is synthesized in the leaf and transported to the roots via the phloem (Jørgensen, Nour-Eldin, and Halkier 2015). The most abundant CG is linamarin (>85%), and total CG concentration varies according to the cultivar, environmental conditions, cultural practices and plant age (McMahon, White, and Sayre 1995). Degradation of linamarin is catalyzed by the enzyme linamarase, which is found in cassava tissues, including intact roots. The compartmentalization of linamarase in cell walls and linamarin in cell vacuoles prevents the accidental formation of free cyanide.

Disruption of tissues ensures that the enzyme comes into contact with its substrate, resulting in rapid production of free cyanide via an unstable cyanohydrin intermediary (Wheatley, Chuzel, and Zakhia 2003). Therefore, careful processing is required to remove Hydrogen Cyanide (HCN), especially in communities with poor nutritional status (Jørgensen et al. 2005; Blomstedt et al. 2012; Gleadow and Møller 2014). Incomplete processing could result in acute to chronic exposure to HCN (Leavesley et al. 2008). High dietary cyanogen consumption from insufficiently processed roots of bitter cassava combined with a protein-deficient diet leads to a neglected disease known as konzo (Kashala-Abotnes et al. 2019). Konzo is a distinct neurological disease characterized by an abrupt onset of an irreversible, non-progressive limbs paralysis (Tshala-Katumbay et al. 2001; Nzwalo and Cliff 2011; Kashala-Abotnes et al. 2019). Juice extraction, heating, fermentation, drying or a combination of these processing treatments aid in reducing the HCN concentration to safe levels (Wheatley, Chuzel, and Zakhia 2003). Gleadow and Moller (2014), reported efforts in cassava breeding programmes to actively select for varieties with lower levels of HCN. However some farmers favour cassava varieties with higher cyanide content as a source of resistance against herbivores and theft by humans (Lebot 2009). Modern breeding has not yet succeeded in developing cassava cultivars totally free of CG (Nweke, Lynam, and Spencer 2002; K. Jørgensen et al. 2005). Previous studies (Kizito et al. 2007; Whankaew et al. 2011) on cyanide using Quantitative Trait Locus (QTL) approach, could not provide a conclusive information on the genetic basis for cyanide variation in cassava, owing to the available genomic resources and narrow dataset background used so far.

In this study, we seek to (1) comprehensively understand the genetic architecture of HCN in cassava, (2) map gene(s) associated to CG variation, (3) develop a fast, cost effective molecular diagnostic toolkit for breeding purposes to increase selection efficiency, and (4) investigate evidences of HCN domestication.

Results

Large scaled analysis of Brazilian population for HCN content

Phenotypic distribution and variation for cyanide content was measured in a Brazilian population of 1,246 individuals using picrate titration method, on which a scale of 1 to 9 indicates the concentration of HCN content (1 and 9 representing extremes of low and high HCN, respectively) (M. G. Bradbury, Egan, and Howard Bradbury 1999). Based on a determined scale, the cyanide concentration varies from 2 to 9 with an average of 5.6 with individuals coming from across Brazilian states **(Figure 1A-B)**. About two-thirds of the total plots of 28,203 had missing values, with 9,139 plots having HCN observations (**Supplementary Table 1; Supplementary Table 2**). Broad-sense heritability (H^2) was calculated to 0.82 for cyanide content, similar to previous observations reported on several species (Barnett and Caviness 1968; Goodger, Ades, and Woodrow 2004; Gleadow and

Møller 2014). Using genotyping data previously recorded for this population (Ogbonna et al., in press), we observed a genotype variance (V_G) higher than genotype-by-year variance (V_{GxY}), with their ratio (V_{GxY}/V_G) showing an interaction of 0.29. HCN deregressed best linear unbiased prediction (BLUP) shows a very high correlation with non-deregressed BLUP with Pearson's correlation coefficient of 0.99, indicating a balanced replication of individuals among the studied population. See deregressed BLUPs (**Supplementary Table 3)**.

GWAS analysis revealed two SNPs associated with HCN accumulation

Single Nucleotide Polymorphisms (SNPs) calling using Tassel version 5, identified a total of 343,707 variants of which 30,279 were selected for phasing and imputation. After imputation, a total of 27,045 bi-allelic SNPs with an allelic correlation of 0.8 or above were kept for downstream analysis. The first three Principal Components (PCs) accounted for over 15.3% genetic variation (**Figure 1C-D; Supplementary Note1**).

To identify genetic correlation between HCN content and genotypic variation, mixed model Genome Wide Association Mapping (GWAS) was performed using GCTA software (Yang et al. 2011) with bonferroni correction as a test of significant SNPs. After the Bonferroni correction (-log10(0.05/27045) threshold, 5.733117), two significant peaks were identified on chromosomes 14 and 16 with 45 and 12 significant associated markers, respectively **(Figure 2A, Supplementary Table 4)**. Subsequent regional linkage disequilibrium (LD) analysis on chromosome 16 gives a 3.6 Mb interval and local LD analysis gives a 248 Kb interval (with r^2 threshold of > 0.8) on which 6 genes are annotated **(Table 1, Supplementary Figure 1A)**. The optimal strongest *p-value* indicates the SNP S16_773999 (p-value: 7.53E-22) located within the Manes.16G007900 gene. Manes.16G007900 is annotated as a Multidrug and Toxic Compound Extrusion or Multi-Antimicrobial Extrusion protein (MATE). MATE transporters are a universal gene family of membrane effluxers present in all life kingdoms. MATE transporters have been implicated directly or indirectly in mechanisms of detoxification of noxious compounds and able to transport CG (Darbani et al. 2016). Interestingly, S16_773999 SNP is predicted to induce a missense variant (A>G) in exon 4 **(Figure 2B - red star-gene model)**. This mutation causes an amino acid change from Threonine to Alanine, predicted as deleterious. A second MATE gene (Manes.16G008000) located at 22Kb from the candidate MATE gene (**Figure 2B - annotation panel**) also shows a high LD (pairwise correlation of 0.96; **Supplementary Figure 2A)**. The second MATE gene could be suggested as a paralog of our candidate gene following reported cassava genome duplication (Bredeson et al. 2016).

The second peak in chromosome 14 shows an association with log *p-values* of 1.08e-08, gives an interval of 615 Kb and the local LD analysis reduced it to 274 Kb on which 3 genes are located **(Table 1, Supplementary Figure 1B)**. The first candidate SNP indicates that S14_6050078 (*p-value* 1.08e-08) is located in Manes.14G074300, a gene coding for an integral membrane HPP family protein involved in nitrite transport activity (Maeda et al. 2014). In a recent study, Obata and colleagues (2020),

144 highlighted that linamarin, an abundant variant of CG in cassava, contains nitrogen and serves as a
145 nitrogen storage compound (Obata et al. 2020) as previously hypothesized (Siritunga and Sayre 2004).
146 This confirms previous observations that application of nitrate fertilizer to cassava plants increases
147 cyanide accumulation in the shoot apex (K. Jørgensen et al. 2005). The second candidate SNP indicates
148 that S14_6021712 (*p-value* 7.32E-08) is located in Manes.14G073900.1, coding for a plasma membrane
149 H(+)-ATPase. H+-ATPase mediated H+influx associated with Al-induced citrate efflux coupled with a
150 MATE co-transport system (Zhang et al. 2017). Wu et al. (2014) found that transgenic *Arabidopsis*
151 lines containing *Brassica oleracea* MATE gene had stronger citrate exudation coupled with a higher
152 H+efflux activity than wild-type plants (Wu et al. 2014).
153 As a validation step, we used a subset of 523 unique individuals (from the core Panel of 1,536
154 unique individuals [Ogbonna et al., in press]), with phenotypic and genotypic information to perform
155 GWAS. Results **(Figure 4-LA Unique; Supplementary Table 5)** revealed the same loci (as was
156 observed in the larger dataset of 1,246 individuals) associated with HCN variation in our initial GWAS
157 dataset, indicating that the core unique panel represents the overall genetic variation for HCN in the
158 Brazilian germplasm collection. However, it detected less significant loci (only 46%) than those
159 detected using a dataset of 1,246 individuals. This indicates that additional small effects QTL were
160 captured with the larger dataset conferring increased statistical power.
161 The alleles driving high cyanide at S16_773999 and S14_6050078 loci show dominance and
162 additive patterns, respectively (**Figure 2C-D**); homozygotes with alternate alleles for both loci show
163 higher cyanide content than heterozygotes, while homozygotes with reference alleles show lower
164 cyanide. This indicates that cyanogenic cassava can be alternate allele homozygous or heterozygous at
165 these loci, while acyanogenic cassava plants are more likely reference allele homozygous at these loci.
166 Joint allelic substitution effects at the associated loci for cyanide did not show any interaction between
167 the two loci as shown in **Supplementary Figure 1C**.
168
169 **Variance explained and evidence for Domestication in HCN reveals chromosome 16 as a good**
170 **candidate for KASP marker development**
171 To calculate narrow-sense heritability, the proportion of variance explained was calculated
172 using parametric mixed model multiple kernel approach (Akdemir and Jannink 2015). Single kernel
173 mixed model explained 0.41 of the marker based (narrow-sense heritability, h^2) proportion of the
174 variance for HCN across the genome. A multi kernel mixed model with the top significant SNPs in
175 chromosome 16 and 14 (S16_773999 and S14_5775892) as the first and second kernel with the rest of
176 the genome as the third kernel, explained 30%, 7% and 63% of the marker based variance respectively.
177 A three kernel mixed model to determine the variance explained by chromosome 14, 16 and the rest of
178 the genome, showed that the proportion of variance explained by the three kernels are 16%, 50% and
179 34% respectively. Chromosome 14 and 16 tag SNPs for the candidate SNPs explains 8% and 36%
180 proportions of variance, respectively; while the rest of the genome explains 56%. We found evidence

for local interactions within chromosome 16 which is most likely as a result of high LD around the region (**Supplementary Method1 for M&M**).

To validate the local interaction found in chromosome 16, we performed intra-chromosomal epistasis interaction using FaST-LMM (Lippert et al. 2011, 2013). Chromosome 16 revealed 242 significant interactions above the bonferroni corrected threshold (-log10(0.05/1131*(1131-1)/2); 1.6024), with three interactions clearly separated by 1 Mb between each pair of SNPs (**Supplementary Figure 1D; Table 2, Supplementary Table 9**). A biosynthetic gene cluster in cassava (Genome draft version Cassava4.1) was earlier identified by Andersen et al. (2000), which we identified to be on chromosome 12 in genome version 6.1 as shown in **supplementary figure 3A-B**. Inter-chromosomal epistasis interactions analysis involving about 400 million tests, did not reveal any significant interactions, neither for bonferroni or FDR threshold. Over 27 million tests had *p-values* less than 0.05 significant level (**Supplementary Method2 for M&M**).

Investigating evidence for domestication in HCN, we carried out differentiating loci analysis using cassava HapMap reference lines (Ramu et al. 2017) for cultivated *M. esculenta* and wild *M. esculenta* ssp. *flabellifolia* (**Supplementary Table 6**). We identified 294 biallelic ancestry-informative single-nucleotide markers that represent fixed or nearly fixed differences between cultivated and wild accessions (**Supplementary Figure 4)**. Interestingly, we observed high fixed loci (89) differentiating between the two groups in chromosome 16, of which over 54 of them are approximately 0.37 Mb away from the candidate MATE gene for HCN regulation (**Supplementary Figure 4).** Together these results indicate that (i) epistasis is observed within chromosome 16 around the main GWAS peak (**Supplementary Figure 1D**) and (ii) the identified epistatic region colocalizes with differentiating loci between *M. esculenta* and wild *M. esculenta* ssp. *flabellifolia* (**Supplementary Table 6**; **Supplementary Method3 for M&M**).

The Kompetitive Allele-Specific PCR assay (KASP) is a robust, high throughput and cost efficiency PCR based marker technology (He, Holme, and Anthony 2014; Neelam, Brown-Guedira, and Huang 2013). We used KASP to develop and validate diagnostic markers for HCN content, based on association peaks, local linkage disequilibrium and allelic effect. Candidate SNPs from the GWAS were subjected to KASP marker design (**Supplementary Table 7**) and assayed on Embrapa Breeding populations for a total of 576 individuals. The average percentage genotype score or call rate was 96.59% with a maximum of 97.92% and a minimum of 92.71% validated across population and validate allelic segregation for HCN content (**Supplementary Table 8**, **Supplementary Method4 for M&M**).

Phylogenetics and mutation predictions reveal altered function of MATE transporter

To identify homologues of the MATE transporter Manes.16G007900, a protein alignment and comparative phylogeny analysis were performed for genome-wide MATEs in cassava, sorghum and *Arabidopsis* using CLUSTAL OMEGA (Sievers et al. 2011). Results showed a close sequence homology between three additional MATE transporters in the cassava genome: Manes.16G007900,

Manes.17G038400, Manes.17G038300 and Manes.16G00800 with percentage identity of 91.09%, 78.05%, 68.59%, respectively. Highest interspecific homology analysis found SbMATE2 from sorghum (Sobic.001G012600; percentage identity of 67.84% [first isoform] and 71.00% [second isoform]) (Darbani et al. 2016) and AtMATE from *Arabidopsis* (AT3G21690; percentage identity of 72.80%) (Liu et al. 2009), characterized as vacuolar membrane transporters **(Figure 3A, Supplementary Note2 on Phylogeny Tree)**. Manes.16G007900 and Manes.16G00800 predicted topology of 12 transmembrane helices supporting their annotation (**Supplementary Figure 5A-B**), as previously reported for *Arabidopsis (Li et al. 2002)*, *sorghum (Darbani et al. 2016)* and blueberry (Chen et al. 2015) (**Supplementary Method5 for M&M**). Maximum likelihood tree using protein sequences from 241 HapMap individuals displayed distinct clade distribution of 64 homozygote individuals for SNP S16_773999 G:G allele (high Cyanide) in color red and 114 homozygotes for SNP S16_773999 A:A allele (low Cyanide) in green color. *M. esculenta* ssp. *flabellifolia* individuals (homozygote G:G) and other wild accessions *M.glaziovii and M.pruinosa* (homozygote A:A) clustered in distinct clades (**Figure 3B**).

The stability of a protein to denaturation is calculated by measuring changes in free energy, and the higher and more positive the change in the free energy is, the more stable the protein is against denaturation (Quan, Lv, and Zhang 2016). We mined 36 single point mutation predictions in GBS and whole genome resequencing data (Ramu et al. 2017) for Manes.16G007900 and Manes.16G008000 proteins. In the observed 36 single point mutations across the two proteins, this value ranges from 0.26 to -4.00 with an average of -1.57 (**Supplementary Figure 5C(1-4), Method6, Table 10**). The deleterious point mutations showed higher negative values in their structural change prediction. Mutations with sensitive stability changes can affect the motion and fluctuation of the target residues. All 36 point mutations except one (**Figure 2B, middle panel**), had a negative change in free energy, indicating loss of stability conferring fluctuations in the protein function (**Supplementary Method6 for M&M**).

Sweet and Bitter cassava Geographical Distribution

We represented the geographical distribution and HCN content of Brazilian germplasm, recently characterized (Ogbonna et al, in press) and presented a contrasted distribution (Figure 4A). Accessions with high cyanide contributing alleles are grouped mostly around the Amazonas and low cyanide contributing alleles are grouped in other areas of Brazil. Specifically, individuals with high cyanide are mostly found around the Amazonian rivers and the coastal areas while more variation in HCN content was observed in other regions of Brazil. The ancestry coefficient distribution for S16_773999, S14_5775892 and joint haplotypes (S16_773999 and S14_5775892) revealed 3 different ancestry coefficients for the candidate SNP S14_5775892 (Figure 4B) following an additive response (Figure 2C). Two different ancestry coefficients were observed for the candidate SNP S16_773999

(Figure 4C) following the complete dominant response observed (Figure 2D). Pseudohaplotype of candidate SNPs in chromosome 14 and 16 shows the distribution of 3 ancestry coefficients (Figure 4D), indicating low, intermediate and high cyanide ancestry coefficients (**Supplementary Method7 for M&M and Note3 for Discussion**).

Leveraging from open source data (available at cassavabase.org, see **Supplementary Method8 for M&M**), we explored the distribution of cyanide across sub-Saharan Africa datasets, including assayed individuals originating from 26 countries (**Supplementary Table 11, Supplementary Figure 6A**) and field trials carried out in different locations across Nigeria. This analysis indicated that Central and Southern Africa showed on average higher cyanide varieties compared to West Africa (**Supplementary Figure 6B**), while a trend of lowering cyanide was detected on landraces compared to improved varieties **(Supplementary Figure 6C)**.

Validating GWAS results in African and Joint Africa, Latin America population

Phenotypic distribution and variation for cyanide content was measured in African population of 636 individuals using the picrate titration method. Cyanide concentration varies from 1 to 9 with an average of 5.1 in the African population (**Supplementary Table 12**). H^2 and h^2 for cyanide content were 0.27 and 0.26 respectively, less than observed in Brazilian germplasm (**Supplementary Table 2**). The genotype variance (V_g) was higher than genotype-by-environment variance (G_{gxe}) with their ratio (V_{gxe} / V_g) showing a high interaction of 0.86. The estimated deregressed BLUPs ranged from 0.0009 to 2.5638 with an average of 0.5242 (**Supplementary Table 13**). After the Bonferroni correction (-log10(0.05/53547) threshold, 6.029765), two significant peaks were identified on chromosomes 14 and 16 respectively **(Figure 5 AF-panel; Supplementary Table 14)**. A third peak was observed in chromosome 11 but was not up to the significant threshold. GWAS dataset for HCN in African accessions showed peaks on chromosome 14 and 16 with SNP S14_6612442 and SNP S16_1298874 showing the highest *p*-values, congruent with the Brazilian GWAS dataset.

For Africa and Latin America combined analysis, phenotypic variation ranged between 1 to 9 with an average of 5.2 (**Supplementary Table 15**). H^2 and h^2 heritability for cyanide content in Africa and Brazilian combined analysis was 0.50 and 0.38 respectively. The genotype variance (V_g) was higher than genotype-by-environment variance (G_{gxe}) with a ratio (V_{gxe} / V_g) showing a lower interaction of 0.42 compared to that of African population alone (**Supplementary Table 2**). The estimated deregressed BLUPs (for the 1,875 individuals used in GWAS) ranged from 0.0027 to 4.2266 with an average of 1.2545 (**Supplementary Table 16**). After the Bonferroni correction, two significant peaks were identified on chromosomes 14 and 16 respectively, corresponding to the earlier reported candidate SNPs **(Figure 5 LA+AF-panel; Supplementary Table 17)**. A whole genome imputation of the African-Brazilian dataset using the hapmap as a reference panel for chromosome 16 **(Supplementary Figure 7A)**, further validate Manes.16G007900 and the associated SNP S16_773999, based on optimal

291 *p*-value (4.74E-22) (**Supplementary Table 18**; **Supplementary Method8 for M&M)**. See
292 distributions of phenotypes and deregressed BLUPs (**Supplementary Figure 8).**
293 We requested available open source RNA-sequencing dataset on the molecular identities for 11
294 cassava tissue/organ types using the TMEB204 (TME204) cassava variety to evaluate gene expression
295 (Wilson et al. 2017). Both Manes.16G007900 and Manes.16G008000 showed differential expression
296 between storage and fibrous root with p-values of 5.00E-05 and 0.00065, respectively (**Figure 6A-B**).
297 Manes.16G007900 is differentially expressed between fibrous root and leaf with FPKM values of
298 13.9219 and 89.5362 respectively, whereas Manes.16G008000 is not and shows low expression level
299 (**Figure 6A-B**). Selective sweeps detection using HapMap WGS between cassava progenitors, Latin
300 American and African accessions do not show sweeps overlapping with candidate and biosynthetic
301 regions (**Supplementary Figure 9&10)**.
302
303
304 Discussion
305
306 The potential of CG content in cassava varieties varies, even among the roots of the same plant
307 (Gleadow and Møller 2014). These variations are partly due to genetics, environmental conditions and
308 soil type (Bokanga et al. 1994; K. Jørgensen et al. 2005; Nzwalo and Cliff 2011). While germplasm
309 from Latin America shows higher genetic variance and heritability (Brazil, V_g=2.59, H^2=0.82, h^2=0.41;
310 Colombia, V_g=1.58, H^2=0.69), their African counterparts showed much less genetic variance and
311 heritability (V_g=0.21, H^2=0.27, h^2=0.26). Latin America/Brazil is the primary center of domestication
312 from where cassava was solely introduced in the XVI century into Africa, which could explain a
313 probable genetic bottleneck and the observed difference between the two populations (Bredeson et al.
314 2016). In addition, sweet and bitter cassava landraces are differentiated in Latin America but not in
315 Africa. This is attributed to post-introduction hybridization between sweet and bitter cassava and the
316 inconsistent transfer of ethnobotanical knowledge of use-category management to Africa (Bradbury et
317 al. 2013). Mislabeling of germplasm in Africa (Yabe, Iwata, and Jannink 2018) may also have
318 contributed to the observed difference. These differences were also observed with the distribution plots
319 for the individuals assayed in our analysis for HCN in Latin American (bimodal distribution) and
320 African (almost normally distributed) populations. The observed differences between broad and
321 narrow-sense heritability estimates, attributed to missing heritability, could be explained by the local
322 epistasis interactions involving few major genes explaining the variance for HCN in chromosome 16
323 (Akdemir and Jannink 2015), the large numbers of rare variants omitted through imputation (Yang et
324 al. 2015) and the use of only bi-allelic subsets of filtered SNPs, leaving behind multi-allelic loci that
325 may have explained additional variance.
326 Previous studies on the genetic architecture of HCN, found two QTL linked to loci SSRY105
327 and SSRY242 explaining 7 and 20% of the genetic variation in an S_1 population (Kizito et al. 2007).

Blasting the sequences of the loci, revealed SSRY105 location on chromosome 14 (57,582,253 bp) of cassava genome version 6.1 (http://phytozome.jgi.doe.gov) and was congruent with the region found on chromosome 14 associated with HCN variation in our current datasets. Whankaew et al. (2011), found five QTL (CN09R1, CN09R2, CN09L1, CN09L2, CN08R1) across two environments and years, but without any consistent QTL. Their corresponding locations on cassava genome version 6.1 were chromosomes 12 (CN09R1), 9 (CN09L1), 3 (CN09L2), and 4 (CN09R2). The sequence of SSRY242 and CN08R1 QTL could not be found to specify their locations in the genome. These studies could not provide a comprehensive information on the genetic basis for root cyanide variation in cassava given that; (1) cyanide content is affected by environment, (2) the use of population with distinct genetic background, (3) the stage of the field trials at which cyanide was assayed, and (4) the use of low marker density which limited the resolution and QTL detection power.

To provide comprehensive genetic architecture of HCN in cassava, we performed a GWAS using multi-year trials conducted in Brazil in 2016 through 2019 on individuals assayed for root cyanide using picrate titration method. Two major regions associated with HCN variation were identified in our dataset, the stronger one in chromosome 16 (within MATE efflux transporter coding region) and another in chromosome 14 (within an integral membrane HPP family protein and H+ ATPase coding regions). The validation of the genetic architecture of HCN in an African population, joint GWAS analysis between Africa and Latin America (Brazil) and whole genome imputation of the African-Brazilian dataset using HapMap as a reference for chromosome16 confirms results and shows that the genetic architecture of cyanide is conserved (based on our datasets). Homozygous reference alleles at identified loci showing lower cyanide content is in agreement with the finding that acyanogenic plants are homozygous recessive at one of the loci (Gleadow and Møller 2014). However, such homozygous cassava variety was yet to be identified given that they are recessive and difficult to discover because of cassava polyploid make-up (Jennings and Iglesias, 2002). Cyanide is maintained in cultivated cassava populations from Africa and Latin America via the selection of high and low cyanide phenotypes under different environmental and herbivore pressures, leading to a balanced selection. This phenomenon has been previously reported for cyanide in white clover (Corkhill L. 1942), *Sorghum bicolor* (Hansen et al. 2003) and *trifolium* (Kakes 1997) for cyanide. More recently, selective sweeps results between cultivated and cassava progenitors suggested that selection during domestication decreased CG content (Ramu et al. 2017).

Genome wide phylogenetic analysis of MATE genes in cassava, sorghum and arabidopsis suggested homology between our candidate gene and SbMATE2, a characterized vacuolar membrane transporter in sorghum for cyanogenic glucoside dhurrin (**Figure 3A)**. SbMATE2 functions in the accumulation of plant specialised metabolites such as flavonoids and alkaloids, and exports dhurrin and other hydroxynitrile glucosides, protecting against the self-toxic biochemical nature of chemical defence compounds. The transport of the pH-dependent unstable cyanogenic glucoside from its cytoplasmic site of production to the acidic vacuole likely contributes to reducing self-toxicity (Darbani

et al. 2016). Mechanistic studies on MATE transporters, such as sorghum SbMATE gene, strongly suggest that its transport cycle could be driven by proton and/or cation (H^+ or Na^+) gradients (Doshi et al. 2017). SbMATE shows high affinity for Na^+ & H^+, and H^+ constitute the main electrochemical driving force in plants, hence, it is likely that H^+ constitutes the main coupling ion for SbMATE. Darbani et al. (2016), reported that the biosynthetic gene cluster for dhurrin additionally includes a gene encoding a MATE transporter and glutathione S-transferase gene for dhurrin uptake in sorghum bicolor.

Our study identified a MATE transporter on chromosomes 16 and the Na^+ (from integral membrane HPP family protein) and a plasma membrane H^+-ATPase-coupled on chromosome 14, as involved in cyanide content regulation. In the cassava genome version 6.1, the cyanide biosynthesis gene cluster is located on chromosome 12 within 75 kb interval, including a couple of changes in orientation and gene arrangement (**Supplementary Figure 3B).** Interestingly, genome-wide epistasis study did not reveal interactions with other parts of the genome, including the biosynthesis gene cluster region on chromosome 12. This finding contrasts with sorghum, where cyanide biosynthesis and transport have been characterized within the same gene cluster (Darbani et al. 2016). This suggests a distinct evolutionary path for cyanide regulation in cassava than in sorghum. In view of this observation, we speculate that perhaps, cassava domestication targeted upstream or downstream genetic regulation steps of cyanide bio-synthesis. In cassava, CGs are synthesized in the shoot apex (Andersen et al. 2000) and then transported to the fibrous roots (Nartey 1968; Koch et al. 1992; K. Jørgensen et al. 2005). Jorgensen et al (2005), reported a reduction of cyanogenic content in leaves of RNAi transgenic cassava plants, but not in the roots, indicating a tissue-specific regulation of cyanide accumulation in roots. Candidate Manes.16G007900 (chromosome16) showed local epistasis interaction with a 1.36Mb region located at 772055 - 775833 bp downstream. Epistatic effects that arise from alleles in gametic disequilibrium, between closely located loci can contribute to long-term response since recombination disrupts allelic combinations that have specific epistatic effects and the detection of epistasis is a key factor for explaining the missing heritability (Akdemir, Jannink, and Isidro-Sánchez 2017; Santantonio, Jannink, and Sorrells 2019). This region spans over 54 biallelic ancestry-informative single-nucleotide markers fixed or nearly fixed between *M. esculenta* and *M. flabellifolia* (Ogbonna et al, in press), suggesting that domestication can impact metabolic content targeting transport regulation (Wang et al. 2019), as earlier reported in maize and rice (Sosso et al. 2015). In view of the above findings, we speculate that cassava domestication may have specifically targeted downstream genetic regulation steps of cyanide biosynthesis. This is supported by the fact that root size (starch storage) and cyanide content are the major traits of cassava domestication (Ramu et al. 2017). HCN is regulated in an oligogenic manner with two major loci explaining the variation across our datasets. To facilitate their use in breeding pipelines, SNPs tagging the major QTL loci were converted to robust, high-throughput and easy to use competitive allele-specific PCR (KASP) assays. The diagnostic markers for cyanide are available (**Supplementary Table 7**) to the global cassava improvement community through a commercial genotyping service provider under the High Throughput Genotyping Project

(https://excellenceinbreeding.org/htpg) via Intertek (https://www.intertek.com). We also observed that the closest homology observed for MATEs in cassava is in line with the results of the MATE protein alignment which displays the highest homology between MATE gene on chromosome 16 and chromosome 17 (Figure 3A). This is congruent with previously identified paleo tetraploidy in the cassava genome, where chromosomes 14 and 16 present partial conserved synteny with chromosome 6 and 17, respectively (Bredeson et al. 2016). We found the candidate gene to be paralog (68.59%) with Manes.16G008000 and homeolog (91.09%) with Manes.17G038400, indicating that our candidate had undergone double duplication events. This finding would need further investigation to clarify the potential fate of the observed tandem duplication (ie: subfunctionalization, neofunctionalization). MATE candidate gene topology prediction suggests that our candidate MATE protein shares a similar topology in the membrane as those observed in the MATE protein family and functions as an efflux carrier that mediates the extrusion of toxic substances (Brown, Paulsen, and Skurray 1999; Morita et al. 2000; Li et al. 2002). Further functional characterization of the putative cyanide transporters in cassava need to be performed.

Allele mining and mutation prediction (**Figure 2B)** on the HapMap dataset ensures that the current study captures the diversity of the HapMap panel. Moreover, DNA sequence analysis of Manes.16G007900 across HapMap individuals shows that *M. esculenta* ssp. *flabellifolia* individuals are preferentially homozygous G:G (high cyanide allele) for candidate SNP S16_773999, which is in line with its phenotypic characterization for cyanide content by Perrut-Lima and colleagues (Perrut-Lima, Mühlen, and Carvalho 2014)**.** Interestingly, for the same candidate SNP, *M. glaziovii* and *M. pruinosa* individuals gene sequences are all homozygous A:A (low cyanide alleles) and cluster separately from *M. esculenta subsp. flabellifolia* (**FIgure 3B**). However, sweeps on HapMap data groups (Latin American, African and progenitors) did not reveal selective sweeps associated with GWAS loci and biosynthesis clusters. Phenotypic spatial distribution analysis results for sweet and bitter cassava in Brazil, suggested clinal variation occurred along subregions gradient separating ancestral coefficients across ecoregions and agrees with the candidate marker response in the region regulating cyanide variation in cassava. This reflects the role environmental conditions and herbivore pressure had played on cyanide regulation and its synergy in maintaining balanced selection of cyanide traits in cassava **(see further discussion: Supplementary Note3)**.

Conclusion

In this study, we deciphered the genetic architecture of cyanide in cassava and mapped the genetic region in chromosome 16 and 14. The GWAS peak in chromosome 16 is strongly associated with the coding region of a MATE efflux protein, a transporter able to transport cyanogenic glucosides. In addition, the peaks on chromosome 14 is associated with the coding region of an integral membrane

HPP family protein involved in Nitrite Transport Activity and a plasma membrane H^+-ATPase mediated H^+ influx which potentially associated with MATE to participate in a cyanide glucosides cotransport system.

Haplotype defined from the region in chromosome 16 and 14 explained 36 and 8% of the total variance explained by the markers, while loci associated with the optimal p-values explained 30 and 7% variance respectively. Selected individuals carrying the alleles for high and low cyanide in chromosomes 16 and 14 were further validated by designing KASP markers for breeding applications. This approach also found the same regions explaining the variance in an African dataset for cyanide, a joint dataset for African and Latin American germplasm and a whole genome imputation of the African-Brazilian dataset for chromosome 16, validating the candidate SNP. Sweet and bitter cassava distribution have maintained pre-conquest distribution in Brazil, with breeding activities around Northern and Central regions creating a more balanced population with low, intermediate and high cyanide clones.

The broader impact of this study was to understand the genetic mechanism of cyanide regulation in cassava root and the identification of closely linked SNP markers to enhance efficiency and cost effectiveness through marker assisted selection. Further steps can include (1) deployment of diagnostic markers for breeding applications; (2) develop co-expression studies to further assess the source/sink relationship of cyanide metabolism in multi-environmental conditions on impact of low cyanide on pest and disease control in cassava. (3) breeding and introduction of low cyanide cassava varieties that are high yielding and disease resistant to regions often affected by agricultural and health related crisis such as konzo, especially in sub-Saharan Africa. Altogether the present study consolidates our understanding of the genetic control of HCN variation in cassava and provides new insights using genomics of diverse genetic background populations.

Material and Method

Plant material

A first dataset including a total of 1,389 accessions from the Cassava Germplasm Banks (CGB) of Brazilian Agricultural Research Corporation (Embrapa), located in Cruz das Almas, Bahia, Brazil were used for this study (**Figure 1B**). The region is tropical with an average annual temperature of 24.5ºC, relative humidity of 80%, and annual precipitation of 1,250 mm. The germplasm were collected from different cassava growing regions and ecosystems of Brazil, and consisted of land races and modern breeding lines (de Oliveira et al. 2014; Albuquerque et al. 2018).

A second data including 1,363 African accessions was obtained from the open source cassava breeding database, cassavabase.org. This dataset comprises plant material from the International Institute of Tropical Agriculture (IITA).

DNA extraction

DNA extraction was performed following protocol described in Albuquerque et al (Albuquerque et al. 2018) and Ogbonna et al (in press) on the Embrapa CGB collection. Briefly, from young leaves according to the CTAB protocol (cetyltrimethylammonium bromide) as described by Doyle and Doyle (1987). The DNA was diluted in TE buffer (10 mM Tris-HCl and 1 mM EDTA) to a final concentration of 60 ng/μL, and the quality was checked by digestion of 250 ng of genomic DNA from 10 random samples with the restriction enzyme *Eco*RI (New England Biolabs, Boston, MA).

Genotyping

Genotyping, imputation, filtering methods and parameters were performed as described in Ogbonna et al (in press). Briefly, Genotyping-By-Sequencing (Elshire et al. 2011) was conducted using the *Ape*KI restriction enzyme (Rabbi et al. 2014) and Illumina sequencing read lengths of 150 bp. Marker genotypes were called with the TASSEL GBS pipeline V5 (Glaubitz et al. 2014) using cassava reference genome version 6.1 available on Phytozome (http://phytozome.jgi.doe.gov). After filtering (mean depth values >5, missing data < 0.2 and minor allele frequency <0.01 per loci) and Imputation (AR^2>0.8) (Browning and Browning 2009), remaining markers were retained for downstream analysis.

Phenotyping

Brazilian dataset:

Phenotypic data were collected on 1,389 accessions over 4 trials in a single location with 3 replications each in 2016, 2017, 2018 and 2019. A total of 1,246 accessions had both phenotypic and genotypic information and were retained for further analysis. Cyanide content was measured using picrate titration methods (M. G. Bradbury, Egan, and Howard Bradbury 1999) as earlier described by Fukuda et al. (2010). Briefly, it involves a qualitative determination of HCN potential in cassava root, and given that HCN potential varies considerably in plants, we assayed 5 to 6 plants in a plot and 3 roots per plant. Cross sectional 1 cm^3 cut is made at mid-root position foreach root, between the peel and the center of the parenchyma. The cut root-cube and 5 drops of toluene (methylbenzene or phenyl methane) are added to a glass test tube respectively and tightly sealed with a stopper. To determine the qualitative score of HCN potential on a color scale of 1 to 9, a strip of whatman filter paper is dipped into a freshly prepared alkaline picrate mixture until saturation. The saturated filter paper is then placed above the cut root cube in the glass tube and tightly sealed for 10 to 12 hours before recording the color intensity (Maziya-Dixon, Dixon, and Adebowale 2007). See **Supplemental Table 1** for a HCN assay for Brazilian germplasm across four years.

African and Colombian datasets:

African phenotypic data was collected from a breeding database cassavabase (https://cassavabase.org) and include 18 locations, 23 years and 393 trials for a total of 8,244 accessions for a total of 33,523 observations from the Institute of Tropical Agriculture (IITA) (**Supplementary Figure 8C**). Colombian phenotypic data includes 41 locations, 11 years and 155 trials for a total of 13,111 observations from the Centro Internacional de Agricultura Tropical (CIAT). The Phenotyping protocol was performed using the same protocol as for the Brazilian dataset. A total of 636 unique accessions with phenotypic and genotypic information from 228 trials were retained for further analysis for the African dataset.

Statistical Analyses

Trials across years were combined and BLUPs were estimated for each clone from 9,138 observations on 1,389 genotypes for HCN. We used the lme4 (Bates et al. 2015) package in R (R. Core Team 2015) version 3.4.2 (2017-09-28) to fit a mixed linear model (MLM) following the method described in (Wolfe et al. 2016): $Y_{ijkl} = \mu + c_i + \beta_j + r_k + d_l + \epsilon_{ijkl}$, where $\epsilon_{ijkl} \sim N(0, \sigma^2_\epsilon)$ is the residual variance and assumed to be randomly distributed, Y_{ijkl} represents the phenotypic observations, μ is the grand mean, c_i is the random effects for clone with $c_i \sim N(0, \sigma^2_i)$, β_j is a random effect for clone nested in combination of year, r_k is a random effect for combination of year and rep and assumed to be normally distribution with $N(0, \sigma^2_k)$ and d_l is a fixed effect for years. Variance components from our mixed model were used to compute the broad-sense heritability according to the method described in (Holland, Nyquist, and Cervantes-Martínez 2010) Briefly, $H^2 = \frac{\sigma^2_c}{\sigma^2_c + \frac{\sigma^2_t}{\bar{t}} + \frac{\sigma^2_r}{\bar{r}} + \frac{\sigma^2_\epsilon}{\bar{p}}}$, where σ^2_c is the clone variance, σ^2_l is variance due to clone by year, σ^2_t is the variance due to years by replications and σ^2_ϵ is the variance due to error. While $\bar{t}$, $\bar{r}$ and $\bar{p}$ are the harmonic mean number of years, replications and plots in which the clone was observed, respectively. Given that the number of observations per clones varies across the four years dataset (replication varies from 1 to 9 with an average of 6), bias induced by precorrection and induced-heterogeneous residual variance (de Los Campos et al. 2013), estimated BLUPs (differentially shrunken to the mean) were deregressed using:

$$deregressedBLUP = \frac{BLUP}{1 - \frac{PEV}{\sigma^2_i}}$$

Where PEV is the prediction error variance for each clone and σ^2_i is the variance for the clonal component. **Supplementary Figure 8** shows the deregressed BLUPs distribution that was further used in the GWAS.

GWAS analysis:

We carried out a mixed model Genome Wide Association Mapping using GCTA software (Yang et al. 2011). Specifically, we used the mixed linear model based association analysis with the chromosome, on which the candidate SNP is located, excluded from calculating the genetic relationship matrix (GRM). The model is $y = a + bx + g^{-} + e$ where y is the deregressed-BLUP estimate, a is the mean term, b is the additive effect (fixed effect) of the candidate SNP to be tested for association, x is the SNP genotype indicator variable, g^{-} is the accumulated effect of all SNPs except those on the chromosome where the candidate SNP is located, and e is the residual. We plotted Manhattan with the Bonferroni threshold used as association tests of significant SNPs and compared the observed -log10(p-value) against the expected using the quantile-quantile plot. Local LD analysis will be performed on GWAS significant region based on r^2threshold of > 0.8 to identify candidate genes. GWAS was also performed on a unique set of 1,536 individuals (GU panel) from Ogbonna et al., (in press). This unique set was selected based on duplicate (identity-by-state) analysis on the total population of 3,354 individuals to ensure efficient germplasm and resource management at the Brazilian cassava program and to balance individual genetic contribution to population structure definition (Ogbonna et al., In press).

Candidate Gene Analysis

To investigate further GWAS candidate regions, we used the genomic resource from the cassava HapMap data (Ramu et al. 2017) to perform allele mining and predict genome-wide allelic mutation effect using SNPeff (Cingolani et al. 2012) and SIFT (Kumar, Henikoff, and Ng 2009).

Phylogenetic analysis of candidate gene sequence

We obtained MATE whole genome protein sequences from *Arabidopsis thaliana* (v.10), *Sorghum bicolor* (v3.11) and *Manihot esculenta* (v.6.1) genomes from phytozome (https://phytozome.jgi.doe.gov/pz/portal.html). The sequences were submitted to the transporter TP prediction tool (http://bioinfo3.noble.org/transporter/) for membrane domain identification and gene curation according to the TCDB guidelines (Saier et al. 2016). Sequences were aligned with CLUSTAL OMEGA (Sievers et al. 2011) and a phylogenetic analysis was performed using a Neighbour-joining tree without distance corrections (**Supplementary Sequence Data**, **Supplementary Multiple Alignment**). In addition, we generatedMATE candidate, Manes.16G007900, protein sequences from the cassava hapmap (Ramu et al. 2017). Briefly Manes16.G007900 annotated variants from HapMap II (ftp://ftp.cassavabase.org/HapMapII/) were used to generate coding sequences (CDS) and translated protein sequence for each 241 accessions in a fasta format. Subsequent alignment and maximum-likelihood phylogenetic trees were generated using MAFFT (Katoh and Standley 2013) and PhyML (Guindon et al. 2010) through the NGphylogeny portal (Lemoine et al. 2019).

Supplemental Data

The following supplemental materials are available.

Supplementary Figure 1. Manhattan plot and LD plots for chromosomes 16 and 14.
Supplementary Figure 2. Pearson correlation of top 5 significant SNPs.
Supplementary Figure 3. Schematic representation of the clustering of cyanogenic glucoside biosynthetic genes.
Supplementary Figure 4. Differentiating loci between cultivated and cassava progenitor.
Supplementary Figure 5. TMHMM posterior probability for transmembrane Protein and mutation prediction.
Supplementary Figure 6. Distribution of sweet and bitter cassava in Sub-Saharan Africa.
Supplementary Figure 7. Manhattan plot for whole-genome imputed chromosomes 16.
Supplementary Figure 8. Distribution of HCN assayed for Latin American and African.
Supplementary Figure 9. Selective sweeps between cassava Progenitors and Latin American.
Supplementary Figure 10. Selective sweeps between Latin American and African cassava.
Supplementary Figure 11. Genetic (cM) vs. Physical (bp) positions.
Supplementary Table 1. Raw HCN dataset from Latin America (Embrapa-Brazil).
Supplementary Table 2. Summary statistics, variance components and broad-sense heritability for HCN.
Supplementary Table 3. 1,389 BLUPs for Latin American (Embrapa-Brazil) dataset and the list of 1,246 BLUPs with genotype information used for GWAS.
Supplementary Table 4. Significant SNPs from Latin American dataset (Embrapa-Brazil).
Supplementary Table 5. Significant SNPs from GWAS on 523 Unique individuals.
Supplementary Table 6. Cultivated and Cassava Progenitor Differenting loci comparison; *M. esculenta*, *M. esculenta* ssp. *flabellifolia*.
Supplementary Table 7. Designed KASP Marker Sequences.
Supplementary Table 8. HCN kaspar ségrégation results.
Supplementary Table 9. 242 Significant epistasis interactions pair of SNPs higher than bonferroni correction threshold (2 way test result).
Supplementary Table 10. Single point mutation prediction for Manes.16G007900 and Manes.16G008000
Supplementary Table 11. List of Countries and regions in Sub-Saharan Africa with their average BLUP values.
Supplementary Table 12. Raw African dataset phenotypes.
Supplementary Table 13. African BLUPs used for GWAS analysis.
Supplementary Table 14. Significant SNPs from African Germplasm GWAS analysis.
Supplementary Table 15. Raw African (IITA) and Latin American (Embrapa) phenotypes.
Supplementary Table 16. 1882 Combined BLUPs for Africa (IITA) and Latin America (Embrapa) GWAS.
Supplementary Table 17. Significant SNPs from African and Brazil Germplasm.
Supplementary Table 18. Significant SNPs from Whole Genome Imputation of chromosome 16 GWAS using HapMapII and raw GBS dataset. 5000 SNP window was used.
Supplementary Sequence Dataset. Whole-genome sequence dataset for all MATE genes in Cassava, Arabidopsis and Sorghum.
Supplementary Multiple Sequence Alignment. Multiple sequence alignment for all MATE genes in Cassava, Arabidopsis and Sorghum.
Supplementary Method 1 Proportion of Variance Explained by Markers
Supplementary Method 2 Genome-wide Epistasis Interactions
Supplementary Method 3 Cultivated and Cassava Progenitor Differentiating Loci Analysis
Supplementary Method 4 KasparMarker Design and Assessment
Supplementary Method 5 Candidate gene protein Topology and Structure Prediction.
Supplementary Method 6 Single Point Mutation Prediction
Supplementary Method 7 Geographical Distribution of HCN

Supplementary Method 8 GWAS in African Population and Joint Africa, Latin America Analysis
Supplementary Note 1 Population Structure Analysis
Supplementary Note 2 Phylogenetic tree
Supplementary Note 3 Sweet and Bitter cassava Geographical Distribution

Availability of data and material

Genotyping (SNP) data used in this study were deposited on cassavabase.org hosted at "ftp://ftp.cassavabase.org/manuscripts/Ogbonna_et_al_2020/gwas_manuscript".

Acknowledgements

The authors appreciate Jean-Luc Jannink and Deniz Akdemir both from Cornell University, Ithaca, NY for their invaluable advice. The authors thank Kelly Robbins and Victoria Gomez for reviewing this manuscript. We are grateful to Embrapa Cruz das Almas breeding team for field experiment management. We thank the IITA cassava team and particularly Alfred Dixon, Peter Kulakow, Prasad Peteti, cassava breeders and data manager for making the historical African dataset publicly available. We thank the CIAT Breeding team and breeder Hernan Ceballos for making the historical Colombian dataset publicly available. This work was supported through Boyce Thompson Institute for plant science in collaboration with Embrapa Mandioca e Fruticultura of Brazil and partially supported by the NEXTGEN Cassava project, through a grant to Cornell University by the Bill & Melinda Gates Foundation and the UK Department for International Development.

Authors Contributions

Designed experiment: AO, EJO, GB; Performed experiment: AO, GB, LBRA. Project supervision: LM, GB, EJO; First draft of the manuscript: AO; IYR provided technical assistance. GB and EO agree to serve as authors responsible for contact and ensure communication. All authors reviewed and approved the manuscript.

668 **Tables**

669

670 **Table 1**. Summary of linkage disequilibrium analysis within the regions (chromosomes 14 and 16)
671 associated with HCN variation in cassava for Brazilian germplasm.

Chromosome	**GWAS-LD Intervals**	**High LD Intervals (r^2 > 0.8)**	**Main candidate gene**	**Annotation**
16	3.6 MB (409 genes)	248 KB (6 genes) 658264 to 800090 bp	Transporter	Multidrug and Toxic Compound Extrusion
14	615 KB (77 genes)	274 KB (3 genes) 5775892 - 6070331 bp	ATPase protein	plasma membrane H^+-ATPase

672

673
674
675 **Table 2.** Chromosome 16 interactions SNP Pairs for Separated (1 Mb apart) Epistasis Interactions in
676 Chromosome 16. SNP 1 and SNP 2 Showed Strong Significant Interactions. The table also contains
677 Single-Locus P-value for the interaction SNPs (SNP).
678
679

SNP 1		**SNP 2**		**Interactions**
SNP	**Single-Locus p Value**	**SNP**	**Single-Locus p Value**	**Interaction p Value**
S16_12540	1.00E-05	S16_1063230	4.31E-15	5.207858e-10
S16_1298874	2.01E-07	S16_12540	1.00E-05	2.969058e-09
S16_1298876	7.72E-08	S16_12540	1.00E-05	8.371752e-10

680
681
682
683
684
685
686

687 **Table 3.** Summary of genes within the regions associated with HCN variation in Brazilian cassava.

Chrom	SNP	Position, bp	Allele	P-value	Gene	Name	Function	Reference
16	S16_773999	773999	G/A	7.53E-22	Manes.16G007900	MATE efflux family protein.	Multidrug and Toxic Compound Extrusion	Darbani et al. 2016
16	S16_795990	795990	A/T	2.41E-10	Manes.16G008000	MATE efflux family protein.	Multidrug and Toxic Compound Extrusion	
16	S16_796041	796041	T/A	1.36E-20	Manes.16G008100			
16	S16_800090	800090	A/T	4.33E-21	Manes.16G008200	UPF0051 PROTEIN ABCI8, CHLOROPLASTIC-RELATED	The incorporation of iron and exogenous sulfur into a metallo-sulfur cluster.	
16	S16_698521	698521	A/G	2.08E-16	Manes.16G007000	F-type H+-transporting ATPase subunit gamma (ATPF1G, atpG) (ATP synthase).	The sector of a hydrogen-transporting ATP synthase complex in which the catalytic activity resides	
16	S16_658264	658264	T/C	2.34E-16	Manes.16G006300	ANKYRIN REPEAT FAMILY PROTEIN.		
14	S14_6050078	6050078	G/A	1.09E-08	Manes.14G074300	Integral membrane HPP family protein.	Involved in Nitrite Transport Activity.	Maeda et al. 2014
14	S14_5775892	5775892	G/T	1.63E-08	Manes.14G071000	K03355 - anaphase-promoting complex subunit 8 (APC8, CDC23).	Interacting selectively and non-covalently with any protein or protein complex.	
14	S14_6021712	6021712	A/T	7.32E-08	Manes.14G073900	H(+)-ATPase (The plasma membrane H+-ATPase)	Associated with the plasma membrane gradients coupled MATE co-transport system.	J. Zhang et al. 2017 and Wu et al. 2014

688

689

690

Figure Legends

Figure. 1. Geographic distribution and population structure of Latin American (Brazilian) germplasm. **(A)** Distribution of assayed HCN. HCN assayed phenotype score ranges from 2 to 9. **(B)** Distribution of germplasm based Brazilian states. A total of 1,821 cassava accessions have valid geographic information. The green dots show where the accessions come from while the size of the dots represent the number of clones from that state. The black numbers show how many accessions were sampled from each location. Population structure reveals the first three axes of the principal component analysis (PCA) explains about 15.3% of the variations in the population of 1246 individuals, 9,686 SNPs after filtering for Hardy–Weinberg equilibrium and LD of 0.01 and 0.2 respectively. **(C)** shows the first and second axis, while **(D)** shows the first and the third axis.

Figure 2. GWAS of HCN for LA germplasm. **(A)** Manhattan plot from a mixed linear model (MLM-LOCO) with the chromosome on which the candidate SNP is located, excluded from calculating the genetic relationship matrix (GRM). Bonferroni significance threshold is shown in red. A quantile-quantile plot is inserted to demonstrate the observed and expected -log10 of P-value for HCN. The red circle indicates the candidate SNP. **(B)** LocusZoomplot shows the HCN chromosome 16 associated region ($-\log_{10} P$ value) around the candidate gene. Top two lines show genomic coverage at the locus, with each vertical tick representing directly genotyped from GBS and HapMap SNPs. Each circle represents a SNP, with the color of the circle indicating the correlation between that SNP and the candidate SNP at the locus (purple). Light blue lines indicate estimated recombination rate (hot spots) in GBS. The middle panel shows 36 single point mutations (red are deleterious) between the region spanning ccMATE1 and ccMATE2. Bottom panel shows genes at each locus as annotated in the cassava genome version 6.1. The red and black rectangle indicate Manes.16G007900 and Manes.16G008000 gene respectively, with Pearson correlation coefficient of 0.96 (r^2) between both genes. The lower figure is the gene model with the position indicated of the associated SNP within the 4th exon. (**C** and **D**) Boxplot showing candidate SNP effect for HCN between each genotype class at the top markers, S14_6050078 and S16_773999, respectively.

Figure 3. Phylogeny analysis. **(A)** Protein sequences alignment of MATE genes in cassava, sorghum and Arabidopsis. Protein Alignment and comparative phylogeny show a close sequence homology between the GWAS candidate gene and SbMATE2 (Sobic.001G012600), a characterized vacuolar membrane MATE transporter in Sorghum, functions in the accumulation of plant specialized metabolites such as flavonoids and alkaloids. **(B)** Genomic sequence of Manes.16G007900 for the 242 HapMap accessions. Accessions highlighted in red are homozygous GG for SNP16_773999, identified as having high HCN content. Accession highlighted in green are homozygous AA for SNP16_773999, identified as low HCN content. Accessions in black are heterozygotes AG or GA for SNP16_773999.

Figure 4. Spatial distribution of ancestral coefficients for HCN candidate SNPs using 1,657 accessions. **(A)** Distribution of germplasm based on best linear unbiased prediction (BLUP) of HCN. Accessions with high cyanide contributing alleles are grouped around the Amazonas and low cyanide contributing alleles are grouped in other areas of Brazil. **(B)** Spatial distribution of allele frequency for HCN candidate loci in chromosome 14. **(C)** Spatial distribution of allele frequency for HCN candidate loci in chromosome 16. **(D)** Interactions of HCN candidate loci in chromosome 14 and 16.

Figure 5. Manhattan plot from a mixed linear model (MLM-LOCO) with the chromosome on which the candidate SNP is located, excluded from calculating the genetic relationship matrix (GRM). The

MLM-LOCO summarizes the genome-wide association results for HCN in **Latin American (LA, Brazilian)**, **African (AF)**, **joint Latin American + African (LA+AF)** and **Latin American unique (LA UNIQUE)** germplasms. Bonferroni significance threshold is shown in red. A quantile-quantile plot is inserted to demonstrate the observed and expected -log10 of P-value for HCN.

Figure 6. (A) Manes.16G007900 and (B) Manes.16G008000 issues/organs expression profiles, three month after planting (FPKM) of African cassava accession TMEB204 (*Manihot esculenta*) sampled for gene expression (Wilson et al. 2017). TMEB204, an African variety, was assayed for HCN in a 1997 field experiment carried out at IITA Mokwa location (Nigeria) and forms part of the individuals in our African dataset with average cyanide content of 5.67 (min=5, max=7). TMEB204 allelic profile for candidate SNP S16_773999 on chromosome 16 is heterozygous, indicating dominance of Manes.16G007900 high cyanide alleles.

Figure 1

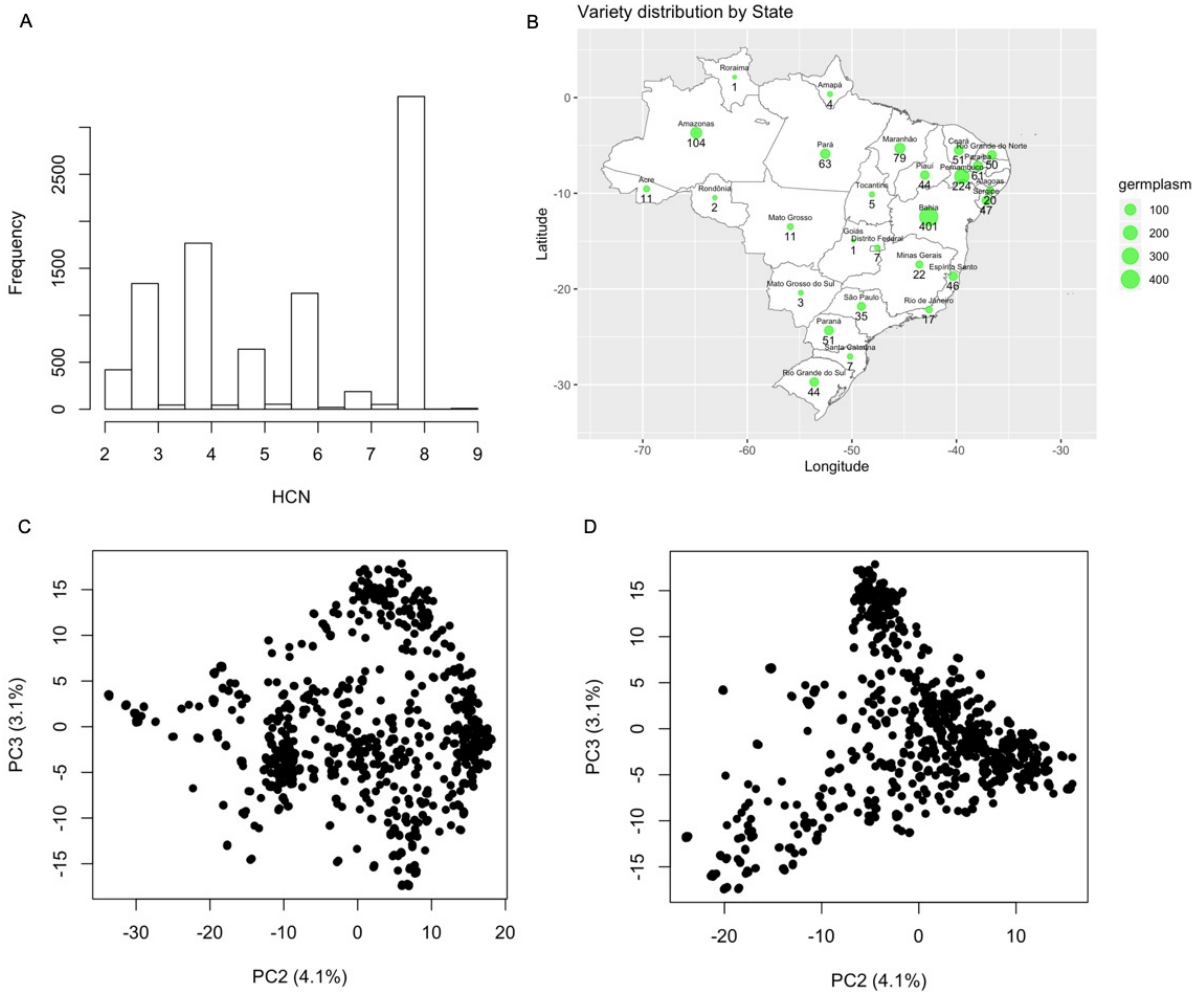


Figure. 1. Geographic distribution and population structure of Latin American (Brazilian) germplasm. **(A)** Distribution of assayed HCN. HCN assayed phenotype score ranges from 2 to 9. **(B)** Distribution of germplasm based Brazilian states. A total of 1,821 cassava accessions have valid geographic information. The green dots show where the accessions come from while the size of the dots represent the number of clones from that state. The black numbers show how many accessions were sampled from each location. Population structure reveals the first three axes of the principal component analysis (PCA) explains about 15.3% of the variations in the population of 1246 individuals, 9,686 SNPs after filtering for Hardy–Weinberg equilibrium and LD of 0.01 and 0.2 respectively. **(C)** shows the first and second axis, while **(D)** shows the first and the third axis.

Figure 2

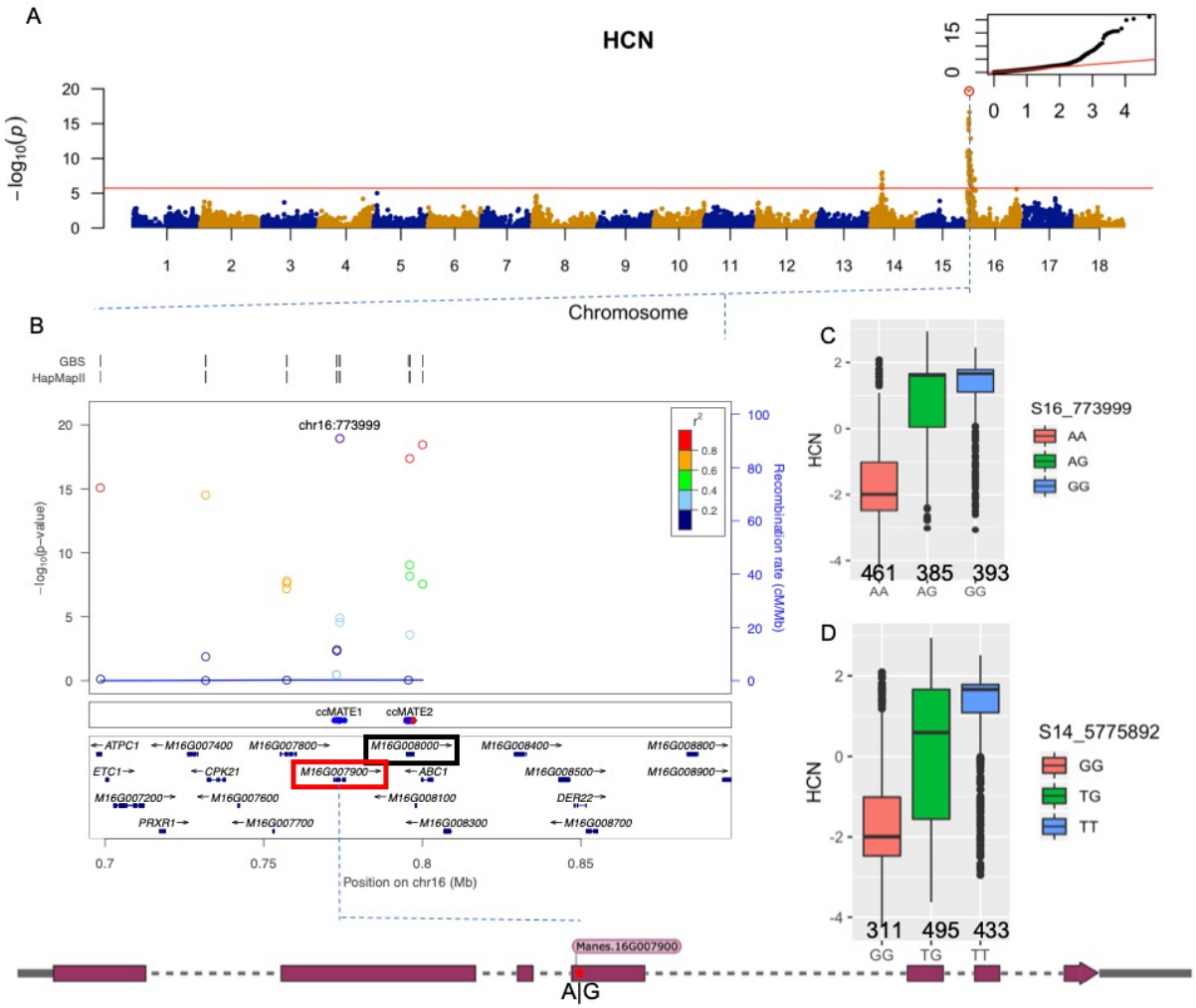
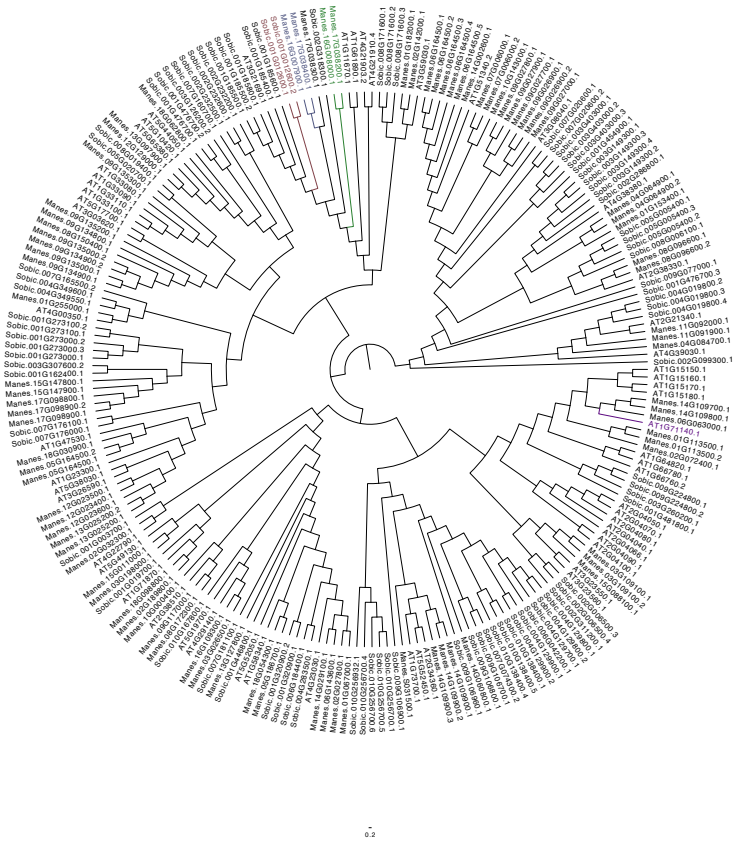


Figure 2. GWAS of HCN for LA germplasm. **(A)** Manhattan plot from a mixed linear model (MLM-LOCO) with the chromosome on which the candidate SNP is located, excluded from calculating the genetic relationship matrix (GRM). Bonferroni significance threshold is shown in red. A quantile-quantile plot is inserted to demonstrate the observed and expected -log10 of P-value for HCN. The red circle indicates the candidate SNP. **(B)** LocusZoomplot shows the HCN chromosome 16 associated region ($-\log_{10} P$ value) around the candidate gene. Top two lines show genomic coverage at the locus, with each vertical tick representing directly genotyped from GBS and HapMap SNPs. Each circle represents a SNP, with the color of the circle indicating the correlation between that SNP and the candidate SNP at the locus (purple). Light blue lines indicate estimated recombination rate (hot spots) in GBS. The middle panel shows 36 single point mutations (red are deleterious) between the region spanning ccMATE1 and ccMATE2. Bottom panel shows genes at each locus as annotated in the cassava genome version 6.1. The red and black rectangle indicate Manes.16G007900 and Manes.16G008000 gene respectively, with Pearson correlation coefficient of 0.96 (r^2) between both genes. The lower figure is the gene model with the position indicated of the associated SNP within the 4th exon. (**C** and **D**) Boxplot showing candidate SNP effect for HCN between each genotype class at the top markers, S14_6050078 and S16_773999, respectively.

Figure 3

A



B

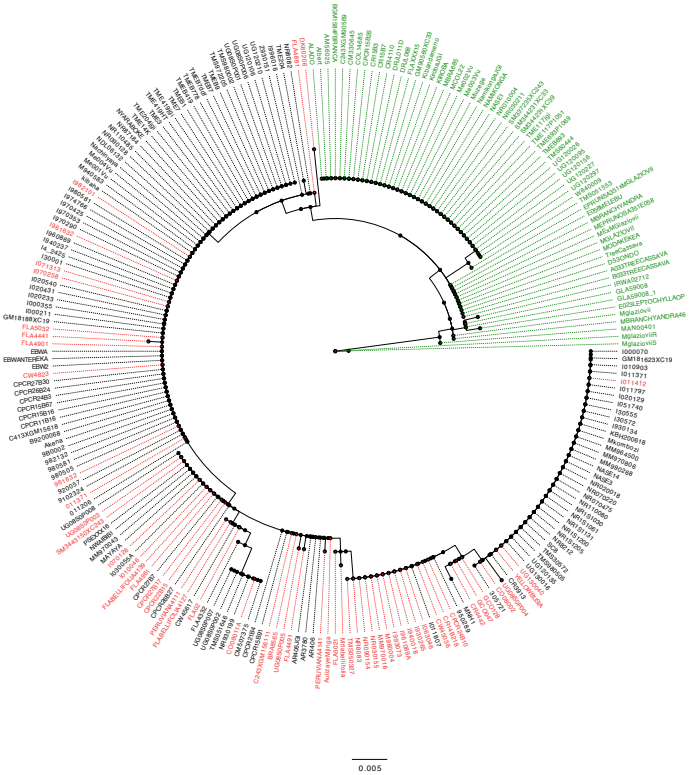


Figure 3. Phylogeny analysis. **(A)** Protein sequences alignment of MATE genes in cassava, sorghum and Arabidopsis. Protein Alignment and comparative phylogeny show a close sequence homology between the GWAS candidate gene and SbMATE2 (Sobic.001G012600), a characterized vacuolar membrane MATE transporter in Sorghum, functions in the accumulation of plant specialized metabolites such as flavonoids and alkaloids. **(B)** Protein sequence of Manes.16G007900 for the 242 HapMap accessions. Accessions highlighted in red are homozygous GG for SNP16_773999, identified as having high HCN content. Accession highlighted in green are homozygous AA for SNP16_773999, identified as low HCN content. Accessions in black are heterozygotes AG or GA for SNP16_773999.

Figure 4

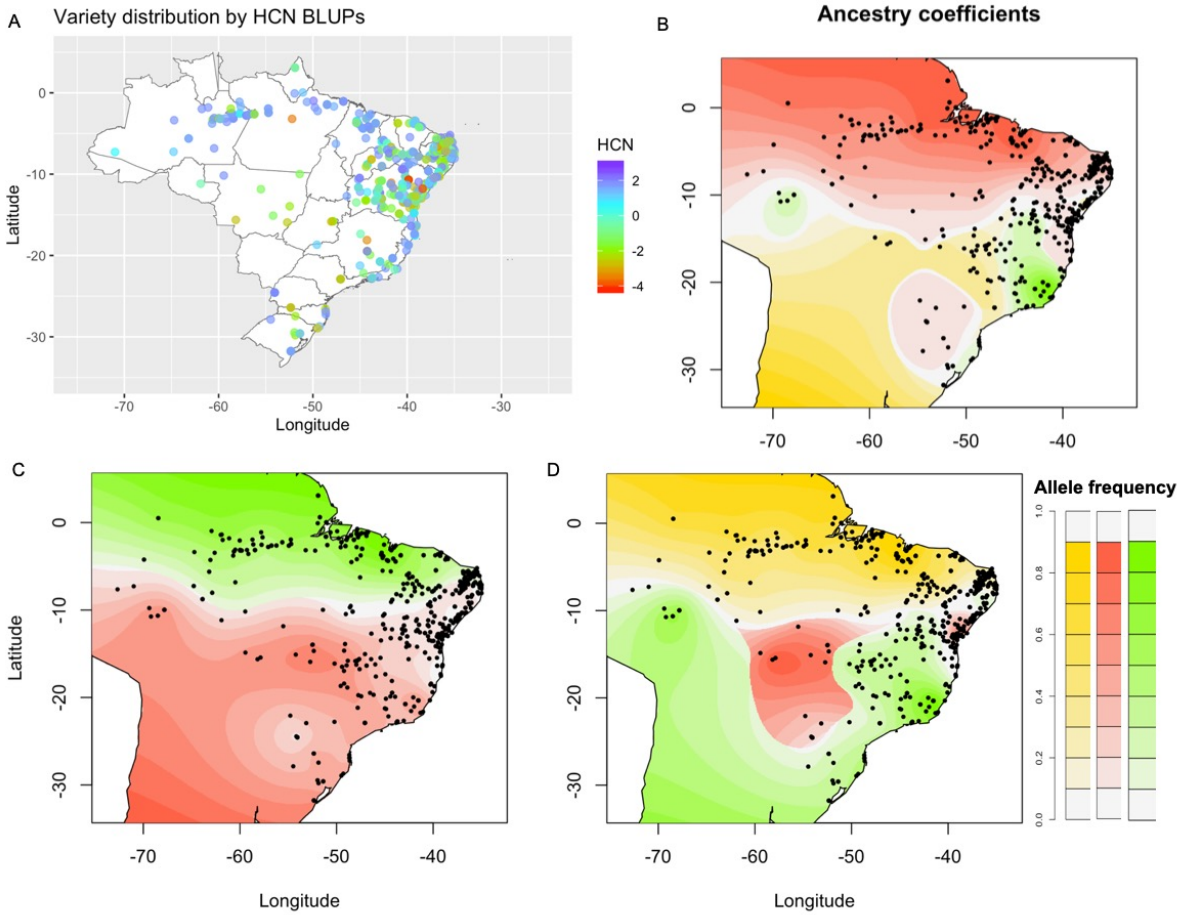


Figure 4. Spatial distribution of ancestral coefficients for HCN candidate SNPs using 1,657 accessions. **(A)** Distribution of germplasm based on best linear unbiased prediction (BLUP) of HCN. Accessions with high cyanide contributing alleles are grouped around the Amazonas and low cyanide contributing alleles are grouped in other areas of Brazil. **(B)** Spatial distribution of allele frequency for HCN candidate loci in chromosome 14. **(C)** Spatial distribution of allele frequency for HCN candidate loci in chromosome 16. **(D)** Interactions of HCN candidate loci in chromosome 14 and 16.

Figure 5

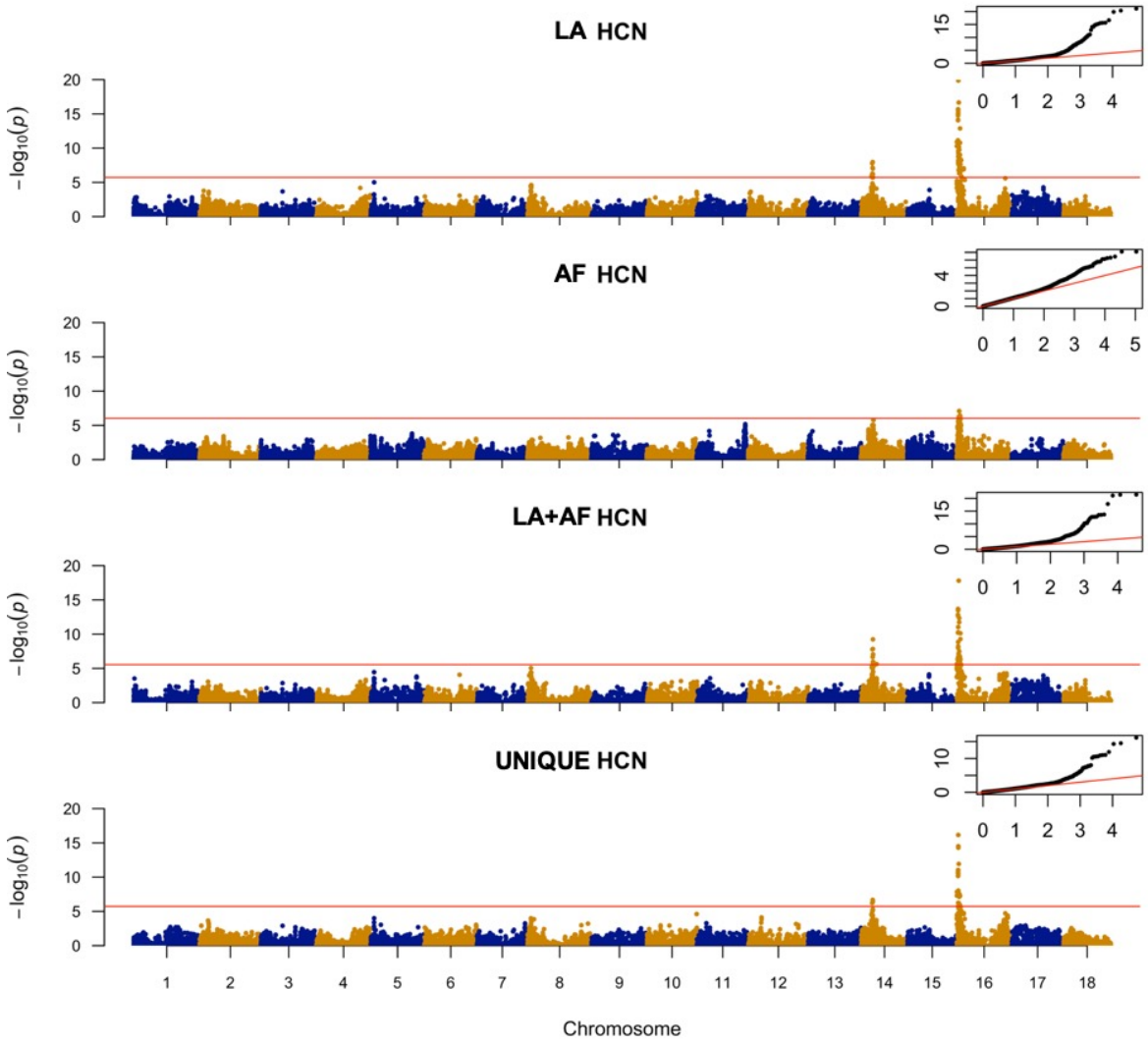
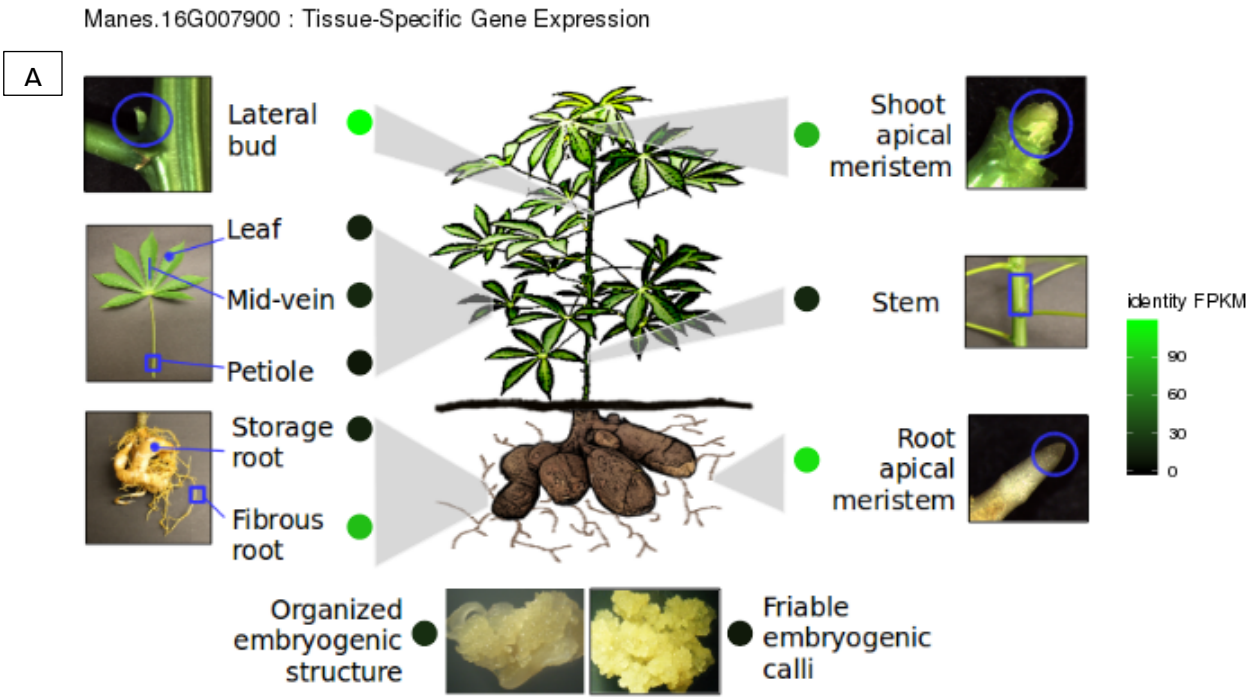


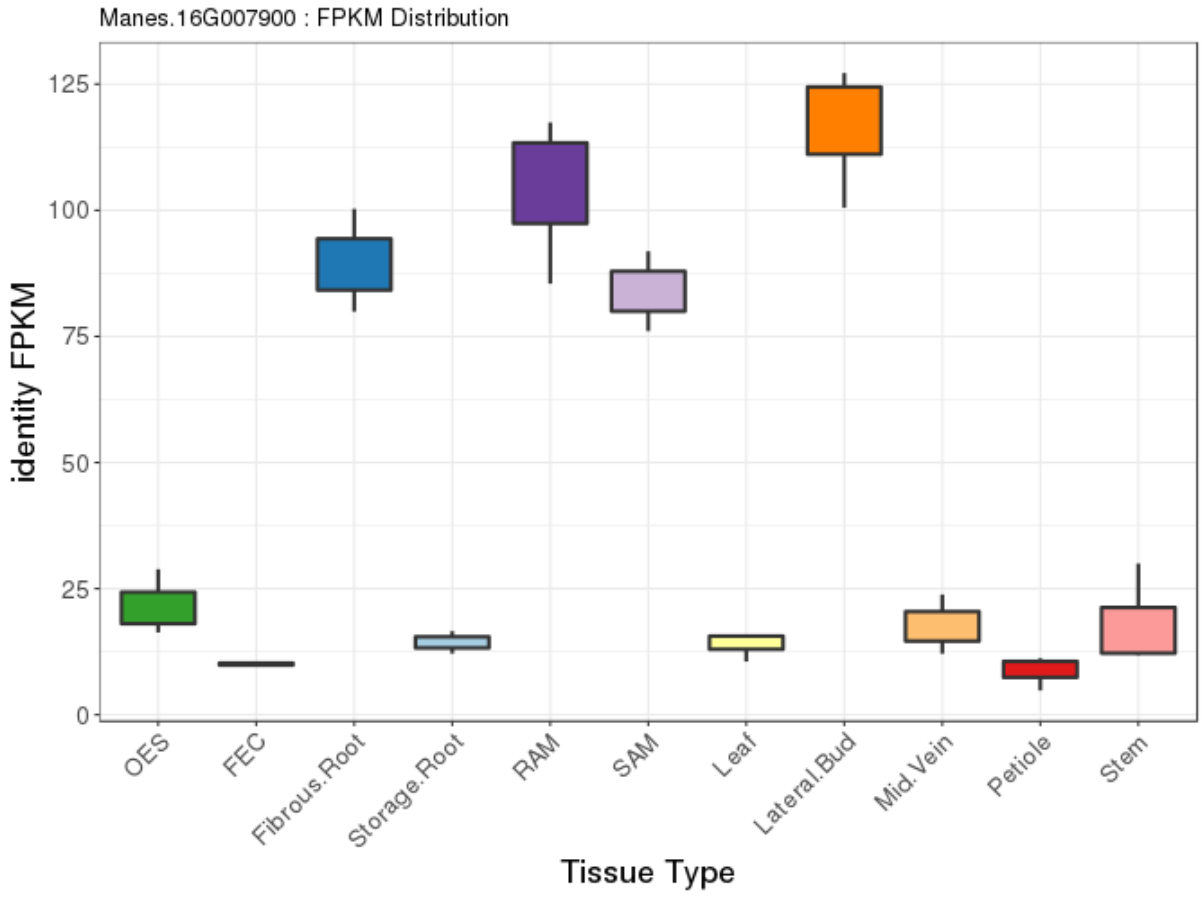
Figure 5. Manhattan plot from a mixed linear model (MLM-LOCO) with the chromosome on which the candidate SNP is located, excluded from calculating the genetic relationship matrix (GRM). The MLM-LOCO summarizes the genome-wide association results for HCN in **Latin American (LA, Brazilian)**, **African (AF)**, **joint Latin American + African (LA+AF)** and **Latin American unique (LA UNIQUE)** germplasms. Bonferroni significance threshold is shown in red. A quantile-quantile plot is inserted to demonstrate the observed and expected -log10 of P-value for HCN.

867

868 **Figure 6**

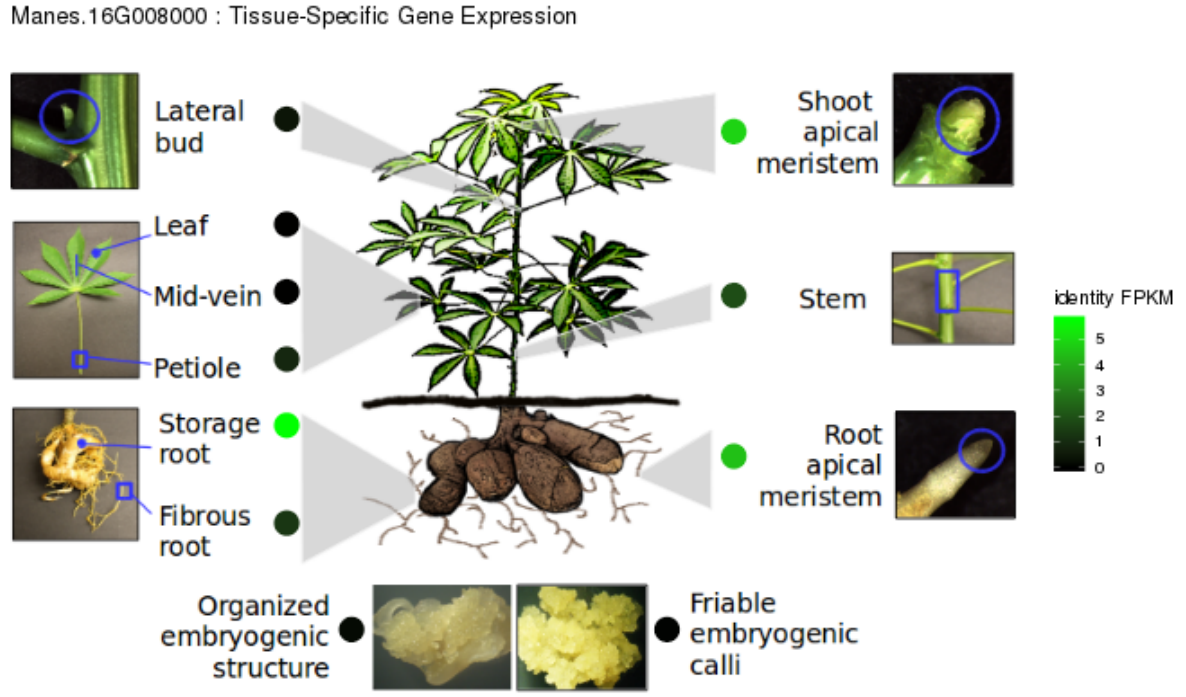


869

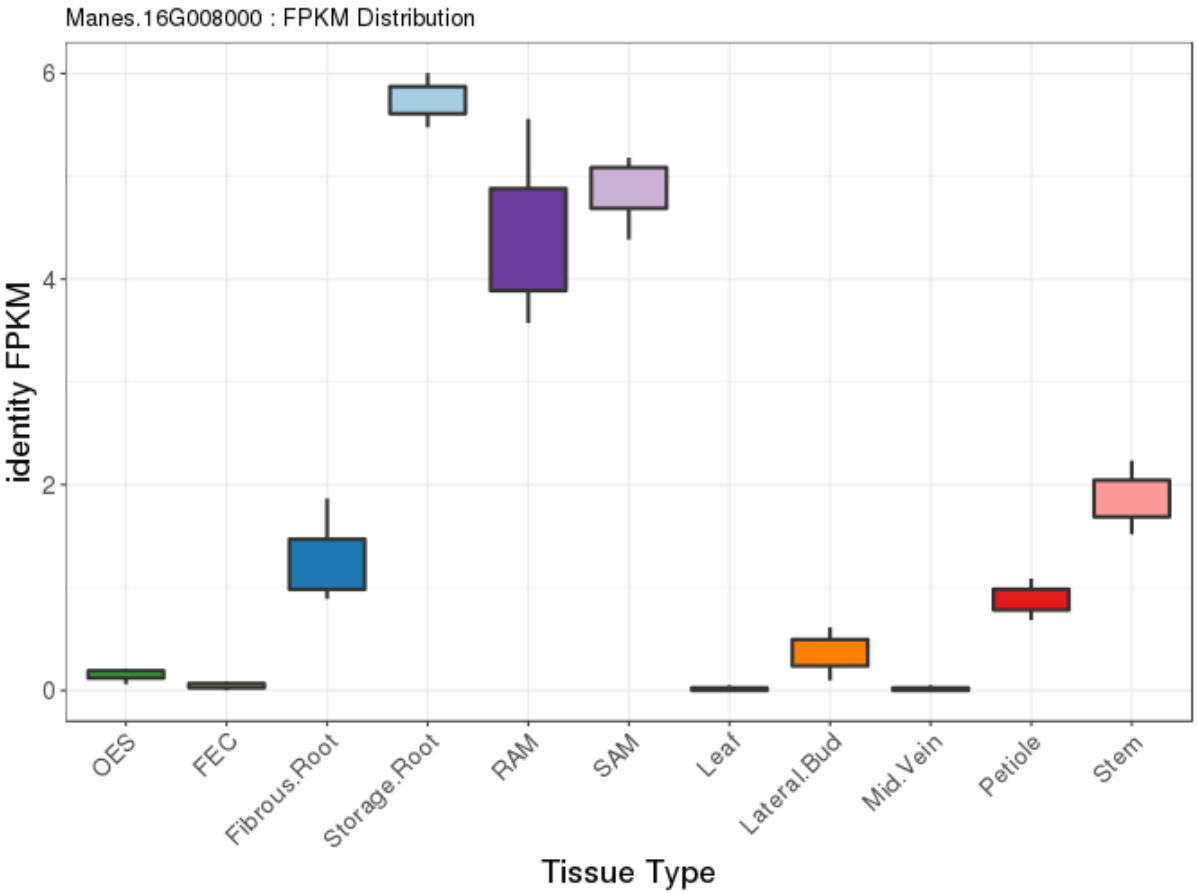


870

B



871



872
873
874
875

Figure 6. (A) Manes.16G007900 and **(B)** Manes.16G008000 issues/organs expression profiles, three month after planting (FPKM) of African cassava accession TMEB204 (*Manihot esculenta*) sampled for gene expression (Wilson et al. 2017). TMEB204, an African variety, was assayed for HCN in a 1997 field experiment carried out at IITA Mokwa location (Nigeria) and forms part of the individuals in our African dataset with average cyanide content of 5.67 (min=5, max=7). TMEB204 allelic profile for candidate SNP S16_773999 on chromosome 16 is heterozygous, indicating dominance of Manes.16G007900 high cyanide alleles.

915 Literature Cited

916 Akdemir, Deniz, and Jean-Luc Jannink. 2015. "Locally Epistatic Genomic Relationship Matrices for
917 Genomic Association and Prediction." *Genetics* 199 (3): 857–71.
918 Akdemir, Deniz, Jean-Luc Jannink, and Julio Isidro-Sánchez. 2017. "Locally Epistatic Models for
919 Genome-Wide Prediction and Association by Importance Sampling." *Genetics, Selection,*
920 *Evolution: GSE* 49 (1): 74.
921 Albuquerque, Hilçana Ylka Gonçalves de, Hilçana Ylka Gonçalves de Albuquerque, Catia Dias do
922 Carmo, Ana Carla Brito, and Eder Jorge de Oliveira. 2018. "Genetic Diversity of Manihot
923 Esculenta Crantz Germplasm Based on Single-Nucleotide Polymorphism Markers." *Annals of*
924 *Applied Biology*. https://doi.org/10.1111/aab.12460.
925 Andersen, Mette Dahl, Peter Kamp Busk, Ib Svendsen, and Birger Lindberg Møller. 2000.
926 "Cytochromes P-450 from Cassava (Manihot esculentaCrantz) Catalyzing the First Steps in the
927 Biosynthesis of the Cyanogenic Glucosides Linamarin and Lotaustralin." *Journal of Biological*
928 *Chemistry*. https://doi.org/10.1074/jbc.275.3.1966.
929 Barnett, R. D., and C. E. Caviness. 1968. "Inheritance of Hydrocyanic Acid Production in Two
930 Sorghum × Sudangrass Crosses1." *Crop Science*.
931 https://doi.org/10.2135/cropsci1968.0011183x000800010026x.
932 Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects
933 Models Usinglme4." *Journal of Statistical Software*. https://doi.org/10.18637/jss.v067.i01.
934 Blomstedt, Cecilia K., Roslyn M. Gleadow, Natalie O'Donnell, Peter Naur, Kenneth Jensen, Tomas
935 Laursen, Carl Erik Olsen, et al. 2012. "A Combined Biochemical Screen and TILLING
936 Approach Identifies Mutations in Sorghum Bicolor L. Moench Resulting in Acyanogenic Forage
937 Production." *Plant Biotechnology Journal* 10 (1): 54–66.
938 Bokanga, M., Indira J. Ekanayake, A. G. O. Dixon, and M. C. M. Porto. 1994. "GENOTYPE-
939 ENVIRONMENT INTERACTIONS FOR CYANOGENIC POTENTIAL IN CASSAVA." *Acta*
940 *Horticulturae*. https://doi.org/10.17660/actahortic.1994.375.11.
941 Bradbury, E. Jane, Anne Duputié, Marc Delêtre, Caroline Roullier, Alexandra Narváez-Trujillo,
942 Joseph A. Manu-Aduening, Eve Emshwiller, and Doyle McKey. 2013. "Geographic Differences
943 in Patterns of Genetic Differentiation among Bitter and Sweet Manioc (Manihot Esculenta
944 Subsp. Esculenta; Euphorbiaceae)." *American Journal of Botany* 100 (5): 857–66.
945 Bradbury, Meredith G., Sylvia V. Egan, and J. Howard Bradbury. 1999. "Picrate Paper Kits for
946 Determination of Total Cyanogens in Cassava Roots and All Forms of Cyanogens in Cassava
947 Products." *Journal of the Science of Food and Agriculture*. https://doi.org/3.0.co;2-
948 2">10.1002/(sici)1097-0010(19990315)79:4<593::aid-jsfa222>3.0.co;2-2.
949 Bredeson, Jessen V., Jessica B. Lyons, Simon E. Prochnik, G. Albert Wu, Cindy M. Ha, Eric
950 Edsinger-Gonzales, Jane Grimwood, et al. 2016. "Sequencing Wild and Cultivated Cassava and
951 Related Species Reveals Extensive Interspecific Hybridization and Genetic Diversity." *Nature*
952 *Biotechnology* 34 (5): 562–70.
953 Browning, Brian L., and Sharon R. Browning. 2009. "A Unified Approach to Genotype Imputation
954 and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals."
955 *American Journal of Human Genetics* 84 (2): 210–23.
956 Brown, M. H., I. T. Paulsen, and R. A. Skurray. 1999. "The Multidrug Efflux Protein NorM Is a
957 Prototype of a New Family of Transporters." *Molecular Microbiology*.
958 Chen, Li, Yushan Liu, Hongdi Liu, Limin Kang, Jinman Geng, Yuzhuo Gai, Yunlong Ding, Haiyue
959 Sun, and Yadong Li. 2015. "Identification and Expression Analysis of MATE Genes Involved in
960 Flavonoid Transport in Blueberry Plants." *PloS One* 10 (3): e0118578.
961 Cingolani, Pablo, Adrian Platts, Le Lily Wang, Melissa Coon, Tung Nguyen, Luan Wang, Susan J.
962 Land, Xiangyi Lu, and Douglas M. Ruden. 2012. "A Program for Annotating and Predicting the
963 Effects of Single Nucleotide Polymorphisms, SnpEff: SNPs in the Genome of Drosophila
964 Melanogaster Strain w1118; Iso-2; Iso-3." *Fly* 6 (2): 80–92.
965 Darbani, Behrooz, Mohammed Saddik Motawia, Carl Erik Olsen, Hussam H. Nour-Eldin, Birger
966 Lindberg Møller, and Fred Rook. 2016. "The Biosynthetic Gene Cluster for the Cyanogenic
967 Glucoside Dhurrin in Sorghum Bicolor Contains Its Co-Expressed Vacuolar MATE

968 Transporter." *Scientific Reports* 6 (November): 37079.
969 Doshi, Rupak, Aaron P. McGrath, Miguel Piñeros, Paul Szewczyk, Denisse M. Garza, Leon V.
970 Kochian, and Geoffrey Chang. 2017. "Functional Characterization and Discovery of Modulators
971 of SbMATE, the Agronomically Important Aluminium Tolerance Transporter from Sorghum
972 Bicolor." *Scientific Reports*. https://doi.org/10.1038/s41598-017-18146-8.
973 Elshire, Robert J., Jeffrey C. Glaubitz, Qi Sun, Jesse A. Poland, Ken Kawamoto, Edward S. Buckler,
974 and Sharon E. Mitchell. 2011. "A Robust, Simple Genotyping-by-Sequencing (GBS) Approach
975 for High Diversity Species." *PloS One* 6 (5): e19379.
976 Food and Agriculture Organization of the United Nations. 2018. *Save and Grow: Cassava: A Guide to*
977 *Sustainable Production Intensification*. Food & Agriculture Org.
978 Glaubitz, Jeffrey C., Terry M. Casstevens, Fei Lu, James Harriman, Robert J. Elshire, Qi Sun, and
979 Edward S. Buckler. 2014. "TASSEL-GBS: A High Capacity Genotyping by Sequencing
980 Analysis Pipeline." *PloS One* 9 (2): e90346.
981 Gleadow, Roslyn M., and Birger Lindberg Møller. 2014. "Cyanogenic Glycosides: Synthesis,
982 Physiology, and Phenotypic Plasticity." *Annual Review of Plant Biology* 65 (February): 155–85.
983 Goodger, J. Q. D., P. K. Ades, and I. E. Woodrow. 2004. "Cyanogenesis in Eucalyptus Polyanthemos
984 Seedlings: Heritability, Ontogeny and Effect of Soil Nitrogen." *Tree Physiology*.
985 https://doi.org/10.1093/treephys/24.6.681.
986 Guindon, Stéphane, Jean-François Dufayard, Vincent Lefort, Maria Anisimova, Wim Hordijk, and
987 Olivier Gascuel. 2010. "New Algorithms and Methods to Estimate Maximum-Likelihood
988 Phylogenies: Assessing the Performance of PhyML 3.0." *Systematic Biology* 59 (3): 307–21.
989 Hansen, Karina Sinding, Charlotte Kristensen, David Bruce Tattersall, Patrik Raymond Jones, Carl
990 Erik Olsen, Søren Bak, and Birger Lindberg Møller. 2003. "The in Vitro Substrate
991 Regiospecificity of Recombinant UGT85B1, the Cyanohydrin Glucosyltransferase from
992 Sorghum Bicolor." *Phytochemistry* 64 (1): 143–51.
993 He, Chunlin, John Holme, and Jeffrey Anthony. 2014. "SNP Genotyping: The KASP Assay."
994 *Methods in Molecular Biology*. https://doi.org/10.1007/978-1-4939-0446-4_7.
995 Holland, James B., Wyman E. Nyquist, and Cuauhtemoc T. Cervantes-Martínez. 2010. "Estimating
996 and Interpreting Heritability for Plant Breeding: An Update." *Plant Breeding Reviews*.
997 https://doi.org/10.1002/9780470650202.ch2.
998 Howeler, Reinhardt, Nebambi Lutaladio, and Graeme Thomas. 2013. "Save and Grow: Cassava. A
999 Guide to Sustainable Production intensificationProduire plus Avec Moins Ahorrar Para Crecer."
1000 FAO 633.6828 S266. FAO, Roma (Italia). http://www.sidalc.net/cgi-
1001 bin/wxis.exe/?IsisScript=orton.xis&method=post&formato=2&cantidad=1&expresion=mfn=100
1002 365.
1003 Jennings, D. L., and C. Iglesias. n.d. "Breeding for Crop Improvement." *Cassava: Biology,*
1004 *Production and Utilization*. https://doi.org/10.1079/9780851995243.0149.
1005 Jørgensen, Kirsten, Søren Bak, Peter Kamp Busk, Charlotte Sørensen, Carl Erik Olsen, Johanna
1006 Puonti-Kaerlas, and Birger Lindberg Møller. 2005. "Cassava Plants with a Depleted Cyanogenic
1007 Glucoside Content in Leaves and Tubers. Distribution of Cyanogenic Glucosides, Their Site of
1008 Synthesis and Transport, and Blockage of the Biosynthesis by RNA Interference Technology."
1009 *Plant Physiology* 139 (1): 363–74.
1010 Jørgensen, Morten Egevang, Hussam Hassan Nour-Eldin, and Barbara Ann Halkier. 2015. "Transport
1011 of Defense Compounds from Source to Sink: Lessons Learned from Glucosinolates." *Trends in*
1012 *Plant Science* 20 (8): 508–14.
1013 Kakes, P. 1997. "Difference between the Male and Female Components of Fitness Associated with
1014 the Gene Ac inTrifolium Repens." *Acta Botanica Neerlandica*.
1015 https://doi.org/10.1111/plb.1997.46.2.219.
1016 Kashala-Abotnes, Espérance, Daniel Okitundu, Dieudonne Mumba, Michael J. Boivin, Thorkild
1017 Tylleskär, and Desire Tshala-Katumbay. 2019. "Konzo: A Distinct Neurological Disease
1018 Associated with Food (cassava) Cyanogenic Poisoning." *Brain Research Bulletin* 145
1019 (February): 87–91.
1020 Katoh, Kazutaka, and Daron M. Standley. 2013. "MAFFT Multiple Sequence Alignment Software
1021 Version 7: Improvements in Performance and Usability." *Molecular Biology and Evolution* 30
1022 (4): 772–80.

Kizito, Elizabeth Balyejusa, Ann-Christin Rönnberg-Wästljung, Thomas Egwang, Urban Gullberg, Martin Fregene, and Anna Westerbergh. 2007. "Quantitative Trait Loci Controlling Cyanogenic Glucoside and Dry Matter Content in Cassava (Manihot Esculenta Crantz) Roots." *Hereditas*. https://doi.org/10.1111/j.2007.0018-0661.01975.x.

Koch, B., V. S. Nielsen, B. A. Halkier, C. E. Olsen, and B. L. Møller. 1992. "The Biosynthesis of Cyanogenic Glucosides in Seedlings of Cassava (Manihot Esculenta Crantz)." *Archives of Biochemistry and Biophysics* 292 (1): 141–50.

Kumar, Prateek, Steven Henikoff, and Pauline C. Ng. 2009. "Predicting the Effects of Coding Non-Synonymous Variants on Protein Function Using the SIFT Algorithm." *Nature Protocols* 4 (7): 1073–81.

Leavesley, Heather B., Li Li, Krishnan Prabhakaran, Joseph L. Borowitz, and Gary E. Isom. 2008. "Interaction of Cyanide and Nitric Oxide with Cytochrome c Oxidase: Implications for Acute Cyanide Toxicity." *Toxicological Sciences: An Official Journal of the Society of Toxicology* 101 (1): 101–11.

Lebot, Vincent. 2009. *Tropical Root and Tuber Crops: Cassava, Sweet Potato, Yams and Aroids*. CABI.

Lemoine, Frédéric, Damien Correia, Vincent Lefort, Olivia Doppelt-Azeroual, Fabien Mareuil, Sarah Cohen-Boulakia, and Olivier Gascuel. 2019. "NGPhylogeny.fr: New Generation Phylogenetic Services for Non-Specialists." *Nucleic Acids Research* 47 (W1): W260–65.

Li, Legong, Zengyong He, Girdhar K. Pandey, Tomofusa Tsuchiya, and Sheng Luan. 2002. "Functional Cloning and Characterization of a Plant Efflux Carrier for Multidrug and Heavy Metal Detoxification." *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.m108777200.

Lippert, Christoph, Jennifer Listgarten, Robert I. Davidson, Scott Baxter, Hoifung Poon, Carl M. Kadie, and David Heckerman. 2013. "An Exhaustive Epistatic SNP Association Analysis on Expanded Wellcome Trust Data." *Scientific Reports* 3 (January): 1099.

Lippert, Christoph, Jennifer Listgarten, Ying Liu, Carl M. Kadie, Robert I. Davidson, and David Heckerman. 2011. "FaST Linear Mixed Models for Genome-Wide Association Studies." *Nature Methods*. https://doi.org/10.1038/nmeth.1681.

Liu, Jiping, Jurandir V. Magalhaes, Jon Shaff, and Leon V. Kochian. 2009. "Aluminum-Activated Citrate and Malate Transporters from the MATE and ALMT Families Function Independently to Confer Arabidopsis Aluminum Tolerance." *The Plant Journal: For Cell and Molecular Biology* 57 (3): 389–99.

Los Campos, Gustavo de, John M. Hickey, Ricardo Pong-Wong, Hans D. Daetwyler, and Mario P. L. Calus. 2013. "Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding." *Genetics* 193 (2): 327–45.

Maeda, Shin-Ichi, Mineko Konishi, Shuichi Yanagisawa, and Tatsuo Omata. 2014. "Nitrite Transport Activity of a Novel HPP Family Protein Conserved in Cyanobacteria and Chloroplasts." *Plant & Cell Physiology* 55 (7): 1311–24.

Maziya-Dixon, Bussie, Alfred G. O. Dixon, and Abdul-Rasaq A. Adebowale. 2007. "Targeting Different End Uses of Cassava: Genotypic Variations for Cyanogenic Potentials and Pasting Properties." *International Journal of Food Science & Technology*. https://doi.org/10.1111/j.1365-2621.2006.01319.x.

McMahon, Jennifer M., Wanda L. B. White, and Richard T. Sayre. 1995. "REVIEW ARTICLE." *Journal of Experimental Botany*. https://doi.org/10.1093/jxb/46.7.731.

Morita, Y., A. Kataoka, S. Shiota, T. Mizushima, and T. Tsuchiya. 2000. "NorM of Vibrio Parahaemolyticus Is an Na(+)-Driven Multidrug Efflux Pump." *Journal of Bacteriology* 182 (23): 6694–97.

Mühlen, Gilda Santos, Alessandro Alves-Pereira, Cássia Regina Limonta Carvalho, André Braga Junqueira, Charles R. Clement, and Teresa Losada Valle. 2019. "Genetic Diversity and Population Structure Show Different Patterns of Diffusion for Bitter and Sweet Manioc in Brazil." *Genetic Resources and Crop Evolution*. https://doi.org/10.1007/s10722-019-00842-1.

Nartey, Frederick. 1968. "Studies on Cassava, Manihot Utilissima Pohl—I. Cyanogenesis: The Biosynthesis of Linamarin and Lotaustralin in Etiolated Seedlings." *Phytochemistry*. https://doi.org/10.1016/s0031-9422(00)85629-0.

Neelam, Kumari, Gina Brown-Guedira, and Li Huang. 2013. "Development and Validation of a Breeder-Friendly KASPar Marker for Wheat Leaf Rust Resistance Locus Lr21." *Molecular Breeding*. https://doi.org/10.1007/s11032-012-9773-0.

Nweke, Felix I., John K. Lynam, and Dunstan S. C. Spencer. 2002. *The Cassava Transformation: Africa's Best-Kept Secret*. Michigan State Univ Pr.

Nzwalo, Hipólito, and Julie Cliff. 2011. "Konzo: From Poverty, Cassava, and Cyanogen Intake to Toxico-Nutritional Neurological Disease." *PLoS Neglected Tropical Diseases* 5 (6): e1051.

Obata, Toshihiro, Patrick A. W. Klemens, Laise Rosado-Souza, Armin Schlereth, Andreas Gisel, Livia Stavolone, Wolfgang Zierer, et al. 2020. "Metabolic Profiles of Six African Cultivars of Cassava (Manihot Esculenta Crantz) Highlight Bottlenecks of Root Yield." *The Plant Journal: For Cell and Molecular Biology*, January. https://doi.org/10.1111/tpj.14693.

Oliveira, Eder Jorge de, Cláudia Fortes Ferreira, Vanderlei da Silva Santos, Onildo Nunes de Jesus, Gilmara Alvarenga Fachardo Oliveira, and Maiane Suzarte da Silva. 2014. "Potential of SNP Markers for the Characterization of Brazilian Cassava Germplasm." *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 127 (6): 1423–40.

Perrut-Lima, Poliana, Gilda S. Mühlen, and Cassia R. L. Carvalho. 2014. "Cyanogenic Glycoside Content of Manihot Esculenta Subsp. Flabellifolia in South-Central Rondônia, Brazil, in the Center of Domestication of M. Esculenta Subsp. Esculenta." *Genetic Resources and Crop Evolution*. https://doi.org/10.1007/s10722-014-0105-6.

Quan, Lijun, Qiang Lv, and Yang Zhang. 2016. "STRUM: Structure-Based Prediction of Protein Stability Changes upon Single-Point Mutation." *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btw361.

Rabbi, Ismail Y., Martha T. Hamblin, P. Lava Kumar, Melaku A. Gedil, Andrew S. Ikpan, Jean-Luc Jannink, and Peter A. Kulakow. 2014. "High-Resolution Mapping of Resistance to Cassava Mosaic Geminiviruses in Cassava Using Genotyping-by-Sequencing and Its Implications for Breeding." *Virus Research* 186 (June): 87–96.

Ramu, Punna, Williams Esuma, Robert Kawuki, Ismail Y. Rabbi, Chiedozie Egesi, Jessen V. Bredeson, Rebecca S. Bart, Janu Verma, Edward S. Buckler, and Fei Lu. 2017. "Cassava Haplotype Map Highlights Fixation of Deleterious Mutations during Clonal Propagation." *Nature Genetics*. https://doi.org/10.1038/ng.3845.

R. Core Team. 2015. *An Introduction to R*. Samurai Media Limited.

Saier, Milton H., Jr, Vamsee S. Reddy, Brian V. Tsu, Muhammad Saad Ahmed, Chun Li, and Gabriel Moreno-Hagelsieb. 2016. "The Transporter Classification Database (TCDB): Recent Advances." *Nucleic Acids Research* 44 (D1): D372–79.

Santantonio, Nicholas, Jean-Luc Jannink, and Mark Sorrells. 2019. "Prediction of Subgenome Additive and Interaction Effects in Allohexaploid Wheat." *G3* 9 (3): 685–98.

Sievers, Fabian, Andreas Wilm, David Dineen, Toby J. Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, et al. 2011. "Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega." *Molecular Systems Biology* 7 (October): 539.

Siritunga, Dimuth, and Richard Sayre. 2004. "Engineering Cyanogen Synthesis and Turnover in Cassava (Manihot Esculenta)." *Plant Molecular Biology* 56 (4): 661–69.

Sosso, Davide, Dangping Luo, Qin-Bao Li, Joelle Sasse, Jinliang Yang, Ghislaine Gendrot, Masaharu Suzuki, et al. 2015. "Seed Filling in Domesticated Maize and Rice Depends on SWEET-Mediated Hexose Transport." *Nature Genetics* 47 (12): 1489–93.

Takos, Adam M., Camilla Knudsen, Daniela Lai, Rubini Kannangara, Lisbeth Mikkelsen, Mohammed S. Motawia, Carl E. Olsen, et al. 2011. "Genomic Clustering of Cyanogenic Glucoside Biosynthetic Genes Aids Their Identification in Lotus Japonicus and Suggests the Repeated Evolution of This Chemical Defence Pathway." *The Plant Journal: For Cell and Molecular Biology* 68 (2): 273–86.

Tshala-Katumbay, Désiré, Jean-Pierre Banea-Mayambu, Théodore Kazadi-Kayembe, Raphaël Nunga-Matadi, Fidèle Bikangi-Nkiabungu, Eeg-Olofsson Karin Edebol, and Thorkild Tylleskär. 2001. "Neuroepidemiology of Konzo a Spastic Para-Tetraparesis of Acute Onset in a New Area of the Democratic Republic of Congo (English)." *African Journal of Neurological Sciences*. https://doi.org/10.4314/ajns.v20i1.7520.

Wang, Shouchuang, Saleh Alseekh, Alisdair R. Fernie, and Jie Luo. 2019. "The Structure and

Function of Major Plant Metabolite Modifications." *Molecular Plant* 12 (7): 899–919.

Whankaew, Sukhuman, Supannee Poopear, Supanath Kanjanawattanawong, Sithichoke Tangphatsornruang, Opas Boonseng, David A. Lightfoot, and Kanokporn Triwitayakorn. 2011. "A Genome Scan for Quantitative Trait Loci Affecting Cyanogenic Potential of Cassava Root in an Outbred Population." *BMC Genomics*. https://doi.org/10.1186/1471-2164-12-266.

Wheatley, C. C., G. Chuzel, and N. Zakhia. 2003. "CASSAVA | The Nature of the Tuber." *Encyclopedia of Food Sciences and Nutrition*. https://doi.org/10.1016/b0-12-227055-x/00181-4.

Wilson, Mark C., Andrew M. Mutka, Aaron W. Hummel, Jeffrey Berry, Raj Deepika Chauhan, Anupama Vijayaraghavan, Nigel J. Taylor, Daniel F. Voytas, Daniel H. Chitwood, and Rebecca S. Bart. 2017. "Gene Expression Atlas for the Food Security Crop Cassava." *The New Phytologist* 213 (4): 1632–41.

Wolfe, Marnin D., Ismail Y. Rabbi, Chiedozie Egesi, Martha Hamblin, Robert Kawuki, Peter Kulakow, Roberto Lozano, Dunia Pino Del Carpio, Punna Ramu, and Jean-Luc Jannink. 2016. "Genome-Wide Association and Prediction Reveals Genetic Architecture of Cassava Mosaic Disease Resistance and Prospects for Rapid Genetic Improvement." *The Plant Genome* 9 (2). https://doi.org/10.3835/plantgenome2015.11.0118.

Wu, Xinxin, Ren Li, Jin Shi, Jinfang Wang, Qianqian Sun, Haijun Zhang, Yanxia Xing, Yan Qi, Na Zhang, and Yang-Dong Guo. 2014. "Brassica Oleracea MATE Encodes a Citrate Transporter and Enhances Aluminum Tolerance in Arabidopsis Thaliana." *Plant & Cell Physiology* 55 (8): 1426–36.

Yabe, Shiori, Hiroyoshi Iwata, and Jean-Luc Jannink. 2018. "Impact of Mislabeling on Genomic Selection in Cassava Breeding." *Crop Science*. https://doi.org/10.2135/cropsci2017.07.0442.

Yang, Jian, Andrew Bakshi, Zhihong Zhu, Gibran Hemani, Anna A. E. Vinkhuyzen, Sang Hong Lee, Matthew R. Robinson, et al. 2015. "Genetic Variance Estimation with Imputed Variants Finds Negligible Missing Heritability for Human Height and Body Mass Index." *Nature Genetics* 47 (10): 1114–20.

Yang, Jian, S. Hong Lee, Michael E. Goddard, and Peter M. Visscher. 2011. "GCTA: A Tool for Genome-Wide Complex Trait Analysis." *American Journal of Human Genetics* 88 (1): 76–82.

Zhang, Jiarong, Jian Wei, Dongxu Li, Xiangying Kong, Zed Rengel, Limei Chen, Ye Yang, Xiuming Cui, and Qi Chen. 2017. "The Role of the Plasma Membrane H-ATPase in Plant Responses to Aluminum Toxicity." *Frontiers in Plant Science* 8 (October): 1757.

Koh Aoki,Ryosuke Sasaki,Miwa Ohnishi,Aya Anegawa,Yuko Sugiyama,Nozomu Sakurai,Tetsuro Mimura. 2010. "Metabolic changes by the overexpression of a putative vacuolar membrane transporter gene in Arabidopsis leaves and suspension cultured cell." 21ST INTERNATIONAL CONFERENCE ON ARABIDOPSIS RESEARCH. Publication:501737462.

McKey, D. and **S. Beckerman** (1993) "Chemical ecology, plant evolution and traditional manioc cultivation systems." In: Tropical Forests, People and Food: Biocultural interactions and applications to development, eds. C. M. Hladik, A. Hladik, O. F. Linares, H. Pagezy, A. Semple and M. Hadley. Man and the Biosphere Series, vol. 13, UNESCO and Carnforth: The Parthenon Publishing Group: Paris, pp. 83-112.

McKey, D.; Beckerman, S. Chemical ecology, plant evolution and traditional manioc cultivation systems. In *Tropical Forests, People and Food: Biocultural Interactions and Applications to Development*; Hladik, C.M., Hladick, A., Linares, O.F., Pagezy, H., Semple, A., Hadley, M., Eds.; Parthenon: Carnforth, UK, and UNESCO: Paris, France, 1993; pp. 83-112.

Indira P, Ramanujam T (1987) Distribution of hydrocyanic acid in high-cyanide and a low-cyanide variety of cassava in relation to the age of the plant. Indian J Agric Sci 57**:** 436–437

Corkhill L. 1942. Cyanogenesis in white clover (Trifolium repens L.) V. The inheritance of cyanogenesis. N.Z. J. Sci. Technol. B 23:178–93.

de Bruijn GH (1973) The cyanogenic character of cassava (Manihot esculenta). In B Nestel, R MacIntyre, eds, Chronic Cassava Toxicity: Proceedings of an Interdisciplinary Workshop, January 29–30, 1973, London. International Development Research Centre, Ottawa, pp 43–48.

Schappert PJ, Shore JS. 2000. Cyanogenesis in *Turnera ulmifolia* L. (Turneraceae): II. Developmental

1186 expression, heritability and cost of cyanogenesis. *Evol. Ecol. Res.* 2:337–52
1187 Fukuda, W.M.G., C.L. Guevara, R. Kawuki, and M.E. Ferguson. 2010. Selected morphological and
1188 agronomic descriptors for the characterization of cassava. International Institute of Tropical
1189 Agriculture (IITA), Ibadan, Nigeria. 19 pp.
1190 Fregene, M. and J. Puonti-Kaerlas (2002), "Cassava biotechnology", in: Hillocks, R.J., J.M. Thresh,
1191 and A.C. Bellotti (eds.), Cassava: Biology, Production and Utilization, CABI, Wallingford,
1192 United
1193 Kingdom, pp. 179-207, *http://dx.doi.org/10.1079/9780851995243.0179*.
1194 Lebot, V. (2009), Tropical Root and Tuber Crops: Cassava, Sweet Potato, Yams, and Aroids, Crop
1195 Production Science in Horticulture Series, Volume 17, CABI, Wallingford, United Kingdom.
1196 Yuan, Jingwei, et al. (2015) "Genetic parameters of feed efficiency traits in laying period of
1197 chickens." *Poultry Science* : pev122.
1198 Rubatsky, V.E., and M. Yamaguchi. (1997) World Vegetables: Principles, Production, and
1199 Nutritive Values, 2nd ed. Chapman & Hall, New York, NY.