# We are IntechOpen,
# the world's leading publisher of
# Open Access books
# Built by scientists, for scientists

## 6,900
Open access books available

## 186,000
International authors and editors

## 200M
Downloads

Our authors are among the

## 154
Countries delivered to

## TOP 1%
most cited scientists

## 12.2%
Contributors from top 500 universities

**CLARIVATE ANALYTICS**
**BOOK CITATION INDEX**
**INDEXED**

**WEB OF SCIENCE™**

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

# Interested in publishing with us?
# Contact book.department@intechopen.com

# Domestication Syndrome in Cassava (*Manihot esculenta* Crantz): Assessing Morphological Traits and Differentially Expressed Genes Associated with Genetic Diversity of Storage Root

Luiz Joaquim Castelo Branco Carvalho,
James V. Anderson, Songbi Chen, Chikelu Mba and
Münevver Doğramaci

Additional information is available at the end of the chapter

## Abstract

Cassava (*Manihot esculenta* Crantz) provides a staple food source for millions of people in tropical and subtropical world regions. Brazil is the major center of diversification for species of the *Manihot*, and a center for domestication of the cultivated species originated from wild ancestral *M. esculenta subsp. flabellifolia.* Genetic breeding of cassava depends on landraces. Molecular phylogenetic technologies used to study genetic traits selected by mankind in crops, are likely to predict proposed "domestication syndrome." Phylogenetic trees use DNA sequences alignment to infer on gene historical events. A study on regulatory and structural complexity that dictates gene/protein function, will add non-sequence information to predict a more complete understanding of functional evolution. Transcriptional profile contains critical information on when and where a gene is manifested. These regulatory properties could explain functional genes diversity achieved within gene families across closely related species such as cassava and its ancestor. Microarray technologies measure transcriptional response of gene to a given environmental or genetic factor. Integration of genomic and transcriptomic data provides more detailed picture of molecular evolution. This chapter describes comprehensive study using the wild relative of cassava ancestor, recognition of natural morphological trait changes during domestication, and gene expression of cassava storage root.

**Keywords:** *Manihot, M. esculenta subsp flabellifolia, M. esculenta subsp esculenta*, cassava, domestication syndrome, genetic diversity, gene expression

## 1. Introduction

Although obtaining space for the intentional cultivation of edible plants often starts with the clearing of forests and modification of landscapes, it was the ability to domesticate plants that made agriculture possible in the first place. Domestication consists of a set of consecutive stages that begins with the original set of plant traits and evolves through the increase in selection frequency for desirable traits (the domestication traits). In the genus *Manihot*, the geographical occurrence and species relationships provided the first source of candidate species to be domesticated, which culminated in the emergence of cultivated plants adapted to both human needs and a cultivated environment that may be directly associated to the type of crop **(Table 1)**. A subset of traits that collectively form the morphological and physiological differences between the cultivated plant and their wild progenitors (the domestication syndrome) is specific for a plant adapted to human needs. For the case of cassava (*Manihot esculenta subsp. esculenta*), we selected a set of observable traits which suffered intense human-driven selection in relation to its ancestor (*M. esculenta subsp. flabellifolia*). These traits include early stages of environmental change resulting from the transition from forest (shade) to open-field cultivation. To determine what regulatory genes differ during the change in growth habit traits, such as vine-type (ancestor) to shrub-type (cultivated) growth, thickening of fibrous root (ancestor) into storage root (domesticated) and flowering set reduction as domestication progressed. The present chapter contemplate, the actual knowledge on the issue of cassava domestication report our current and forward studies on the evolutionary suite [1, 2] of genetic diversity in cassava landraces using genomic, transcriptomic, and proteomic technologies for cassava storage root, as the major domestication trait in cassava. Based on gene expression analysis, we identified a set of exploratory regulatory gene networks associated with diversity among minimum mankind artificial interference on the domestication of the major domestication trait (storage root formation) in cassava crop in relation to their ancestor originated in Brazilian Amazon (a major center for domestication of cassava). Finally, the pattern of exploratory regulatory gene networks linked to genotypes diversity was used to predict the early steps in the domestication process of

| Domestication index | Other crops | Cassava |
|---|---|---|
| Removal of the forest | Yes | Yes |
| Seed dispersion at campground | High/low | Refractive |
| Human migration | High/fair | Fair |
| Gardening crops | High/fair | Fair |
| Discarded undesirable genotype | None/low | Low |
| Storage root | None/high | High |
| Genetic diversity | Moderate/low | Moderate |
| Plasticity | Low/high | High |
| Genetic bottleneck | Moderate/low | High |

A set of hypothetical domestication index used to rank for their utilization by mankind and its effect on genetic processes.

**Table 1.** Crop domestication processes: a suite of procedures performed by humans to domesticate plants for food production.

the cassava ancestor as mankind first removed the forest (low-light shade environment) causing a clear open space (abundant light environment) with relevant landscape alterations that promoted alteration in the cassava ancestor plant as observed in its growth habit, storage root formation, and flowering set pattern.

## 2. Cassava wild relatives and the ancestral species

Mexico and Brazil are considered two relevant centers of diversification for *Manihot* species [3–5]. In Brazil, the geographical distribution for the genus *Manihot* is presented in **Figure 1**. Current systematic working models for botanical descriptions and classifications of species in the genus *Manihot* [6–9] contrasts in two major areas compared with an earlier classical monograph [9]. The early classical monograph suggests a total number of 98 species for the genus *Manihot*, which adopts a section classification system for grouping morphological closeness of species [9, 10] and proposes a compiled species concept to explain the origin of a single cultivated species (*M. esculenta subsp. esculenta*). The current model avoids the section classification system, reduces the number of species by using synonymies, uses an evolutionary approach, and permits a single ancestor species (*M. esculenta subsp. flabellifolia*) to explain the origin of the cultivated species (*M. esculenta subsp. esculenta*) of the genus *Manihot* [5–8]. In this chapter, we first contrast these two approaches by using a phylogenetic analysis of **ribosomal RNA internal transcribed spacer (ITS)** for 17 *Manihot* species from Brazil and Mexico to identify those most closely related to *M. esculenta subsp. esculenta* (cassava), which is recognized as the only cultivated species of the genus.
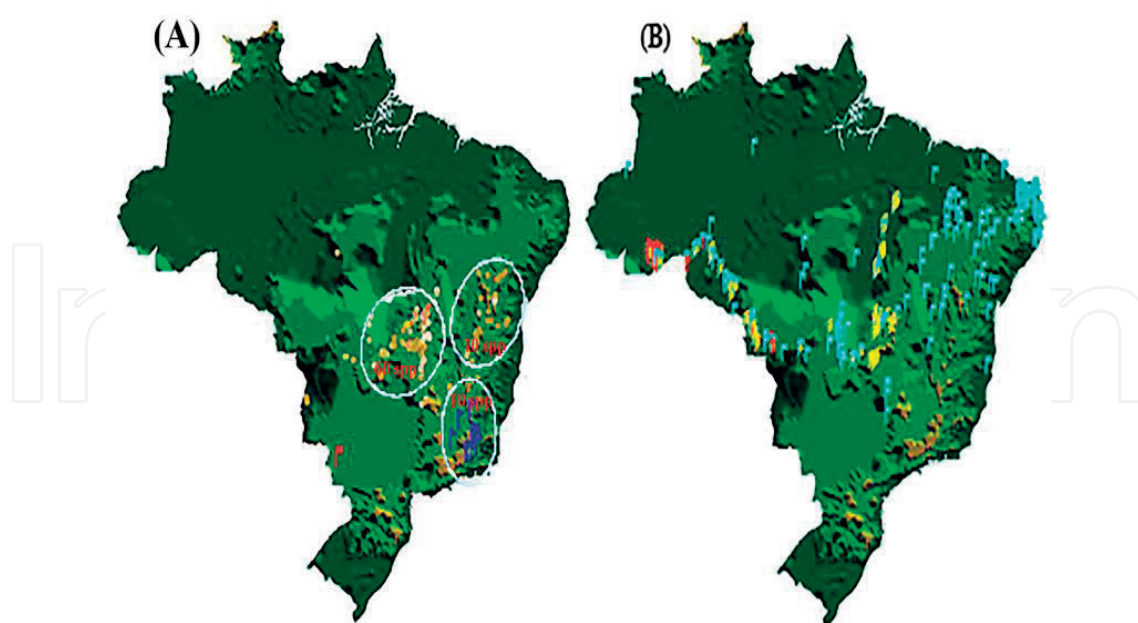


**Figure 1.** Occurrence and geographical distribution of *Manihot* genus in Brazil. (A) Refers to distribution of 60 species (larger circle) of the genus in the central part of Brazil, 10 species (upper small circle) of the genus in the east part, and 10 species (lower small circle) of the genus in southeast part of Brazil. (B) Refers to distribution of *M. esculenta subsp. flabellifolia* from Northwest Amazon to Northeast in the central part (yellow flag) of Brazil and *M. esculenta subsp. esculenta* from Northwest Amazon to Northeast of Brazil (blue flag).

Results indicate that species from Brazil, including *M. esculenta subsp. flabellifolia*, are phylogenetically related to the cultivated species *M. esculenta subsp. esculenta* rather than species from Mexico [3, 4]. These analyses indicated that cassava probably did not originate from Mexico. Therefore, it is not the result of a compiled species but instead possibly has a Brazilian single species ancestor originally named *M. esculenta subsp. flabellifolia* [5–7]. A gene pool analysis [8] for the cultivated species (*M. esculenta subsp. esculenta*) and its wild relative identify two gene pools involving 13 *Manihot* species in gene pool 2 (GP2) and 4 *Manihot* species in gene pool 1 (GP1), which has *M. esculenta subsp. flabellifolia* based on overall similarity of several morphological characteristics indicates *M. esculenta subsp flabellifolia* as the closest alias of cassava [8].

## 3. Domestication as evolutionary processes

Phylogenetic techniques used for determining the molecular evolution of a crop have relied predominantly on sequence information to model the evolutionary history that determines plant speciation and domestication. Phylogenetic trees are based on alignment of DNA or protein sequences, from which evolutionary distances between genes can be inferred. However, transcriptional behavior of a gene is poorly represented by DNA sequence data alone. A gene's transcriptional profile may contain critical functions, including when and where a gene is expressed, and the conditions under which gene expression is manifested. This chapter addresses questions on how function transcriptional profiles vary due to changes in the environment (light intensity) and due to genetic diversity of landraces and commercial breed varieties.

### 3.1. Molecular evolution of a crop species

Factors involved in regulation of expressed genes or gene sets could be crucial in explaining the key functional differences between related genes whose function, during selection (natural and artificial), cannot be distinguished from DNA sequence alone [11–19]. Attempts to predict expression patterns of genes using sequence information [20] have typically been limited by the complexity and diversity of factors influencing genes. Thus, sequence-based prediction of a gene's regulation remains a premature goal. However, transcriptomic approaches, for example, using microarray chip or RNA-Seq technology, allow for a direct, quantitative measurement of global transcriptional responses to a given environmental or genetic factor and are useful experimental sources for obtaining large-scale gene expression data [21, 22]. Genomic data sets spanning a wide selection of the cassava ancestor (*M. esculenta subsp. flabellifolia*), landraces from the Amazon, and breeding cultivars (cv.) are publicly available [23–33], providing a ready source of data for studying several aspects of gene transcription behavior. The integration of genome and transcriptome data [34–44] provides an increasingly detailed picture of molecular evolution by incorporating regulatory behavior into models of the evolution of gene expression and function. Here, we report steps toward understanding changes in gene expression to model gene evolutionary function using our cassava domestication syndrome hypothesis. Specifically, the cassava domestication syndrome hypothesis considers changes in traits such as plant growth habit, storage root formation and flowering from the cassava ancestor (*M. esculenta subsp. flabellifolia)* to becoming the cultivated species (*M. esculenta subsp. esculenta*). **Figure 2** illustrates these variables as observed from field trips to the Amazon and recorded images.
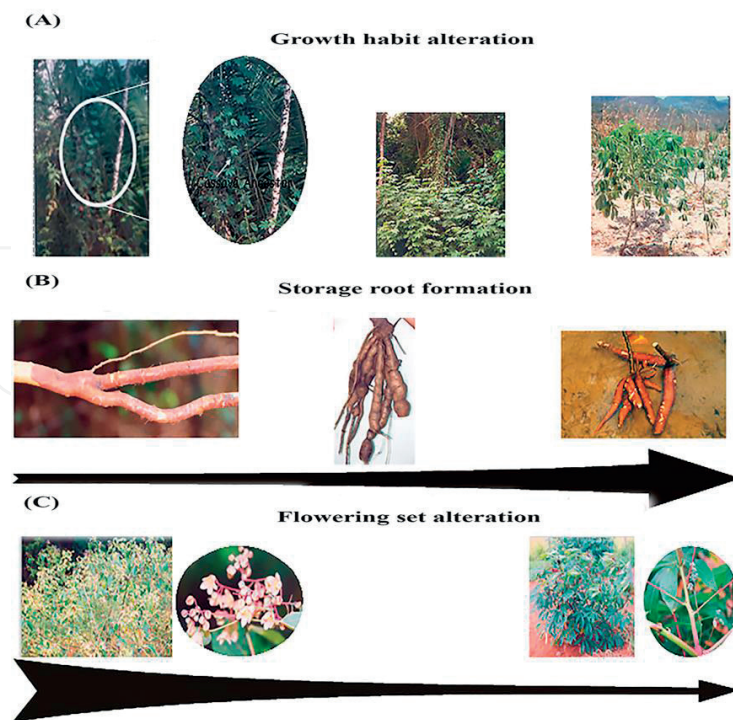
**Figure 2.** Diagram representing the early step of domestication of cassava crop. Field observed phenotype pattern differences between *M. esculenta subsp flabellifolia* (ancestor) in the forest (shade environment) and open field (full light environment, after removal of the forest) growth of the cultivated species *M. esculenta subsp esculenta*. (A) Growth habit change from vine type of growth (ancestor) to shrub type of growth (cultivated cassava). (B) Storage root formation from fibrous normal root (ancestor) in the forest (shade environment) and open field (C) flowering set pattern alteration between ancestor and cultivated cassava.

## 3.2. Differentially expressed genes

A cDNA microarray chip designed for Euphorbiaceae [24] was probed with total RNA extracted from storage root (31 samples total) of cassava with diverse storage root traits. This chapter documents a total of 569 genes which were identified as differentially expressed (*p-value* of 0.005) between storage root of *M. esculenta subsp. flabellifolia* (ancestor) and various *M. esculenta subsp. esculenta* landraces and cultivars. Hybridization intensity values were statistically analyzed to further identify groups among the differentially expressed genes (DEG). The complexity of the experimental design and analyses for screening groups of DEG was achieved using two statistical strategies [34, 35]. First, principal component analysis (PCA) was performed to observe the number of DEG among each group. The PCA results identified four groups among the DEG. Considering the cassava domestication syndrome hypothesis described above, the questions to address with the available data are (i) do these results occur due to differences in expression of genes per se or, (ii) in part, due to the selected genetic backgrounds in the experimental design? Therefore, the second approach used recursive partitioning to obtain tentative conclusions about the grouping patterns [34, 35], as shown in **Figure 3**.

## 3.3. Ontology and functional classification of differentially expressed genes (DEG)

Analysis of the DEG identified 22 distinct groups among gene ontologies and functional classifications. The groups (**Figure 4**) highlighted in yellow (i.e., "Protein with Binding Function
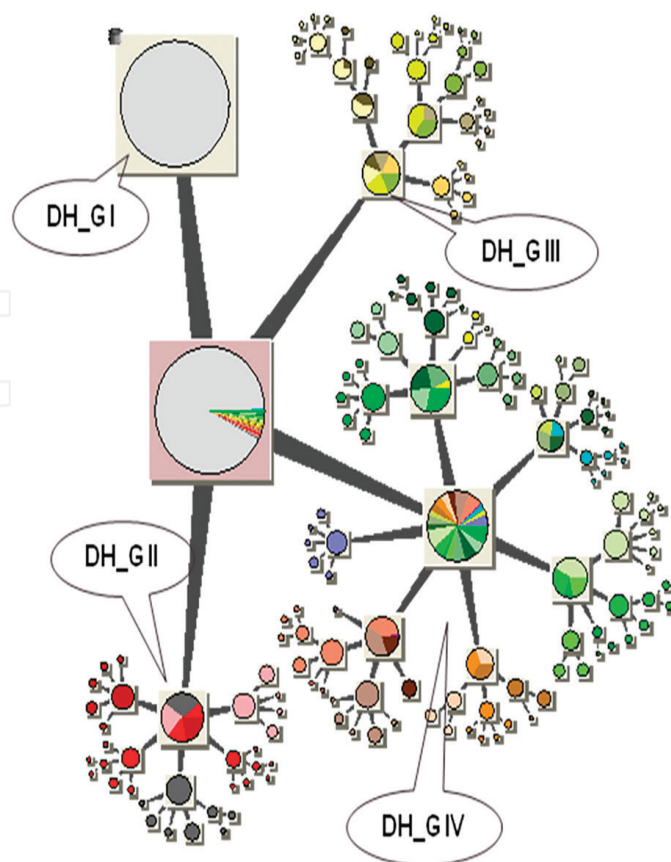
**Figure 3.** Partitioned pattern for a set of 569 differentially expressed genes (DEG) in storage root between cassava ancestor (*M. esculenta subsp flabellifolia*) and cultivated species (*M. esculenta subsp esculenta*). Group DH_G1 partitioning pattern showed no statistically significant further groups formation, while DH_G2, DH_G3 and DH_G4 formed variable numbers of subgroups and sub-sub groups of DEG.

or Cofactor Requirement," "Regulation of Metabolism and Protein Function," and "Cellular Transport, Transport Facilities and Transport Routes") were targeted to elucidate candidate genes involved in regulation of these key pathway networks.

### 3.4. Exploratory pathway networks and candidate regulatory genes

The program Pathway Studio [43] was used to conduct subanalysis (SNEA) to identify potential regulatory networks from transcriptome data obtained in this study and available databases, as previously described [36–42]. The results on statistics (shown in **Tables 2** and **3**) and visualization of gene networks (shown in **Figures 5** and **6**) took into consideration three types of molecular interaction mechanisms (expression target, protein binding, and protein modification).

These results indicate node operating gene/hub, edge genes which are regulated (activated or silenced), and their expression level—increased abundance (blue color) or decreased abundance (pink color) among genes visualized in the pathways; regulatory genes such as transcription factors and other gene products modulating functionality (protein binding and modification) were observed. The node/hub gene regulates the network and genes, while on the edge are
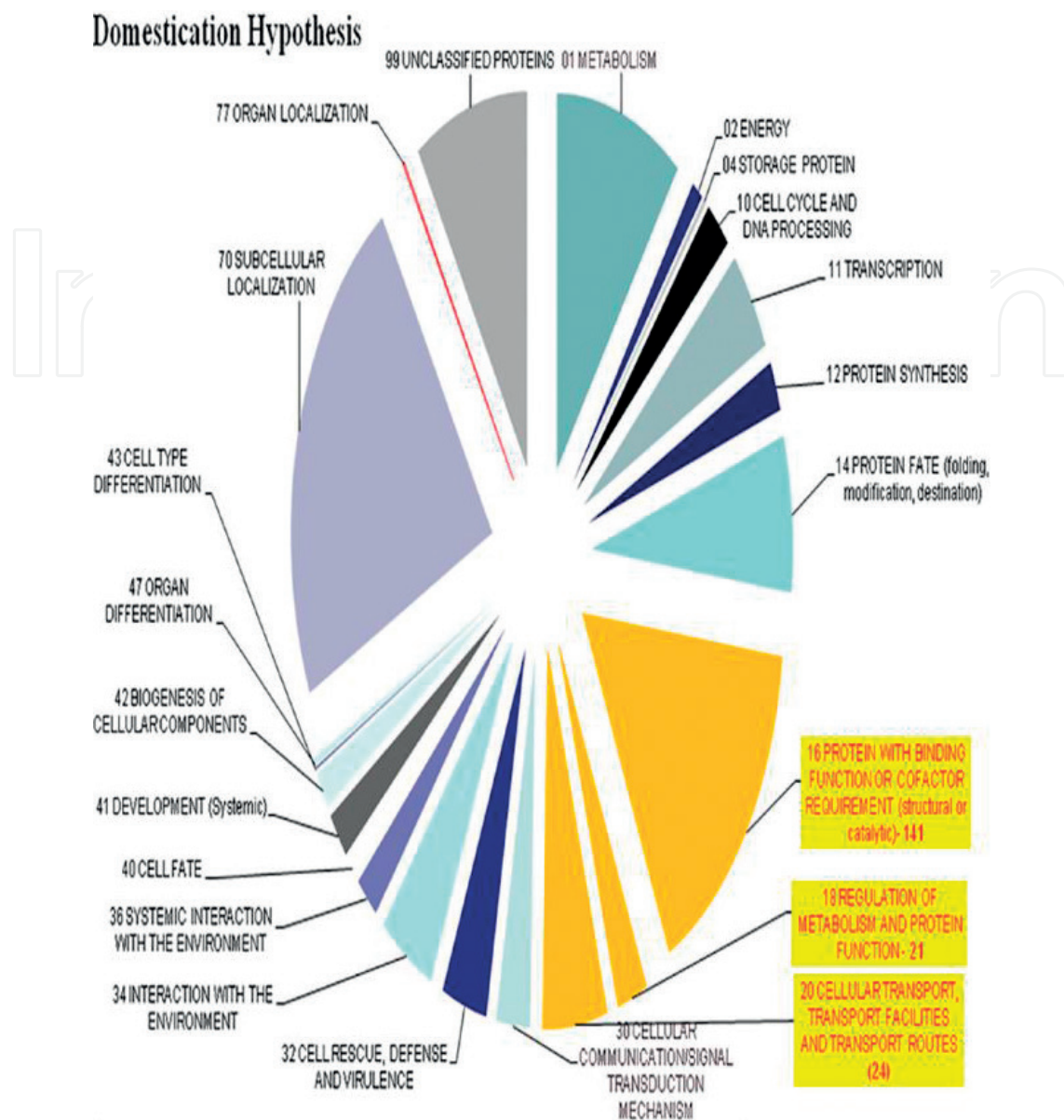
**Figure 4.** Ontology and functional classification of 569 differentially expressed genes (DEG) in cassava storage root using 31 hybridized RNA probes to the cDNA microarray chip considering a designed domestication syndrome hypothesis.

regulatory genes of a particular network. **Table 4** summarizes the list of nodes/hubs in the networks unique to each class of landraces based on comparisons to the cassava ancestor and the cv. IAC 12-829.

### 3.5. Identification of regulatory gene sets

Exploratory pathways network common among cassava ancestor (*M. esculenta subsp flabellifolia)*, cassava cultivars (*M. esculenta subsp esculenta))*, and cassava landraces (*M. esculenta subsp esculenta) were based on the analysis of DEG using* Pathway Studio software to conduct the Gene Set Enrichment Analyses (GSEA) algorithm and establish the level of significance (*p-value*) of

| Name | Gene set root | Entities | Neighbors | p-Value |
|---|---|---|---|---|
| **Comparison of FLA vs. a representative group from the World Collection (cv. MTAI16/1, cv. CM2177–2/1, cv. Reina1 and cv. Veronica1)** | | | | |
| Expression targets | Calmodulin | 8 | 7 | 0.0032 |
| | Ubiquitin | 7 | 6 | 0.0076 |
| | CRY1 | 9 | 8 | 0.0229 |
| | CTR1 | 7 | 6 | 0.0270 |
| | Proteasome | 9 | 8 | 0.0453 |
| Binding partners (protein interaction) | PHYB | 9 | 8 | 0.0004 |
| | PHYA | 7 | 6 | 0.0324 |
| | PRL1 | 6 | 5 | 0.0366 |
| | 14–3-3 | 13 | 12 | 0.0464 |
| **Comparison of FLA vs. cv. IAC12.829 (white-type storage root)** | | | | |
| Expression targets | Ubiquitin | 7 | 6 | 0.0074 |
| | Calmodulin | 8 | 7 | 0.0367 |
| | AXR3 | 6 | 5 | 0.0404 |
| Binding partners (protein interaction) | PHYB | 9 | 8 | 0.0063 |
| | SCF | 7 | 6 | 0.0125 |
| | WAVE1 | 6 | 5 | 0.0280 |
| | PRL1 | 6 | 5 | 0.0325 |
| | PHYA | 7 | 6 | 0.0386 |
| | TIR1 | 6 | 5 | 0.0391 |
| **Comparison of FLA vs. landrace Cas51 (pink-type storage root)** | | | | |
| Expression targets | ABI1 | 6 | 5 | 0.0027 |
| | Ubiquitin | 7 | 6 | 0.0192 |
| | CRY1 | 9 | 8 | 0.0324 |
| | FLC | 10 | 9 | 0.0452 |
| | Proteasome | 9 | 8 | 0.0477 |
| Binding partners (protein interaction) | PHYB | 9 | 8 | 0.0006 |
| | PHYA | 7 | 6 | 0.0147 |
| | SCF | 7 | 6 | 0.0307 |
| | TIR1 | 6 | 5 | 0.0318 |
| | COP1 | 6 | 5 | 0.0414 |
| **Comparison FLA vs. Landrace Cas36.1 (sugary-type storage root)** | | | | |
| Expression targets | Calmodulin | 8 | 7 | 0.0042 |
| | Ubiquitin | 7 | 6 | 0.0073 |
| | CTR1 | 7 | 6 | 0.0247 |

| Name | Gene set root | Entities | Neighbors | p-Value |
|---|---|---|---|---|
| Binding partners (protein interaction) | PHYB | 9 | 8 | 0.0008 |
| | PRL1 | 6 | 5 | 0.0349 |
| | SLY1 | 6 | 5 | 0.0429 |
| **Comparison of FLA vs. Landraces CAS36.7, Cas64, Cas56, and Cas62 (yellow-type storage root)** | | | | |
| Expression targets | Ubiquitin | 7 | 6 | 0.0043 |
| | CRY1 | 9 | 8 | 0.0230 |
| | Calmodulin | 8 | 7 | 0.0331 |
| | EIN2 | 13 | 12 | 0.0396 |
| Binding partners (protein interaction) | PHYB | 9 | 8 | 0.0006 |
| | SCF | 7 | 6 | 0.0369 |

Comparison between cassava ancestral *M. esculenta subsp. flabellifolia* (FLA) and genetic diversity of cultivated species *M. esculenta subsp. esculenta* (cassava).

**Table 2.** Sub network enrichment of classes of genes and their level of statistically significance (*p-value*).

| Common to all genotype | Landrace (pink) | Landrace (*sugary*) | Commercial cv. (white) | Landrace (yellow) | World core collection (white) |
|---|---|---|---|---|---|
| Ubiquitin(ET) | FLC(ET) | SLY1(PI) | PRL1(BP) | EIN2((ET) | 14–3-3(PI) |
| PHYB(PI) | TIR1(PI) | PRL1(PI) | WAVE1((PI) | SCF(PI) | |
| | ABI1(ET) | | | | |

Cassava classes of genotype include commercial cv. IAC 12–829 (white type), landrace Cas51 (pink color), landraces CAS36.7, Cas64, Cas56, and Cas62 (yellow), landrace Cas36.1 (sugary), and world core collection representatives cv. MTAI16/1, cv. CM2177–2/1, cv. Reina1, and cv. Veronica1 (white type). Abbreviation refers to gene name as in **Table 4**. Abbreviation (in parenthesis) accounts for molecular interaction mechanism as expression target (ET), protein interaction (PI), and protein binding (PB) regulatory gene function.

**Table 3.** Exploratory regulatory node/hub genes unique to different classes of cassava genotype compared to cassava ancestral (*M. esculenta subsp. flabellifolia*).

gene groups on the identification of functional gene according to the Sub-Network Enrichment Analysis (SNEA) for biological processes, cellular components, and molecular functions using based on annotation of cassava genes to *Arabidopsis* database. The gene sets were grouped per their function and selected for retrieving and visualizing regulatory networks or pathways they form. Together these results add new knowledge about the potential functionality of gene products previously unknown in cassava storage root, their potential roles in the domestication trait, as well as in the flowering set trait. As an example of these analyses, we propose a hypothetical hormonal gene regulatory model (**Figure 7**) to represent the effect of environmental light changes likely caused due to removal of ancestral cassava from the forest as illustrated in **Figure 2**. However, it is important to clarify that Gene Set Enrichment Analyses (GSEA) and Sub-Network Enrichment Analysis (SNEA) is based on annotation of cassava genes to *Arabidopsis*. Thus, for our proposed models, we are assuming that the cassava gene products are performing similar functions as their Arabidopsis homologues (all *Arabidopsis* annotations used in this chapter, and their known functions can be obtained from [44]).
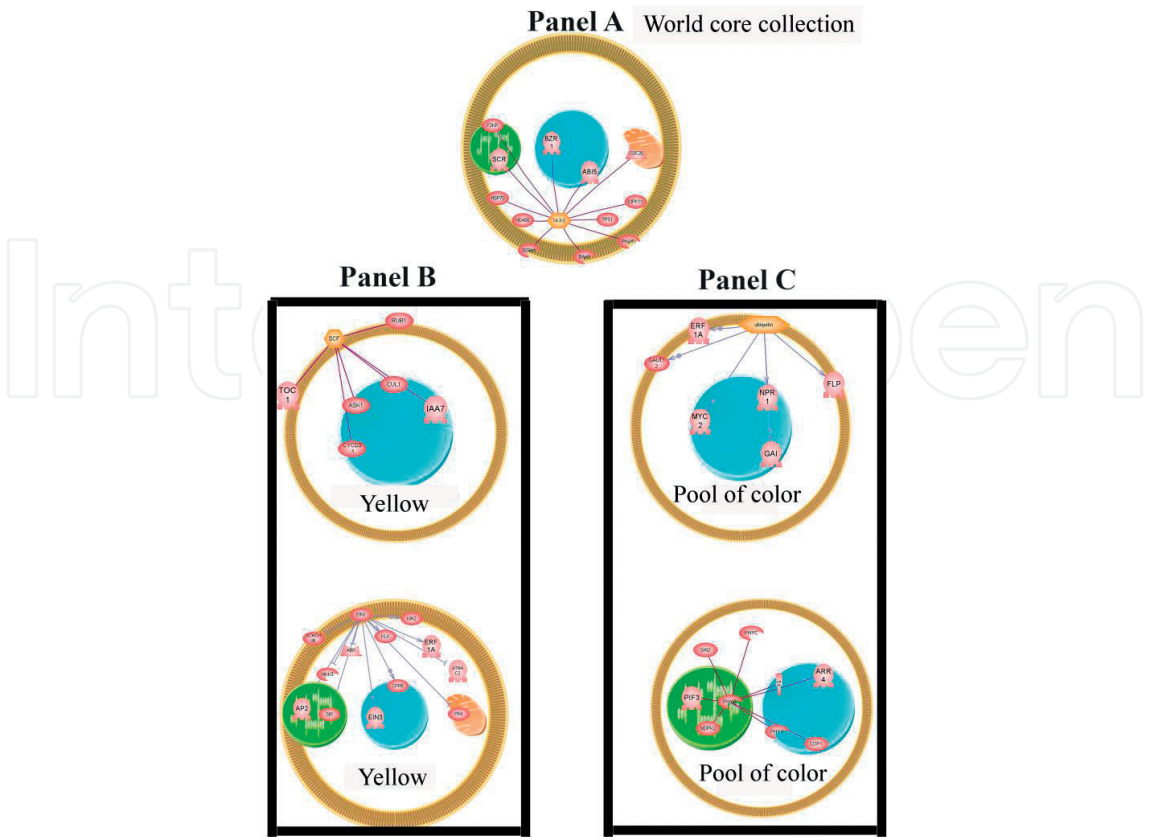
**Figure 5.** Diagram showing exploratory pathway networks for regulatory genes related to cassava ancestral (*M. esculenta subsp flabellifolia*) domestication storage root trait and its genetic diversity in the cultivated species (*M. esculenta subsp esculenta*). Panel A. Differentially expressed regulatory gene node/hub and regulated genes of ancestral and world core collection representatives. Panel B. Differentially expressed regulatory gene node/hub and regulated genes of ancestral and the yellow color storage root landraces. Panel C. Differentially expressed regulatory gene node/hub and regulated genes of ancestral and the pool of colors (white, yellow, and pink) storage root landraces genotypes. Blue and pink colors symbols are up and down regulated genes. Arrow (⟶) refers to positive regulation of the gene and block line (…⊣…) refers to silencing the gene.

The removal of ancestral cassava from a shaded forest environment would be expected to alter regulatory networks and pathways involved in light perception and signaling, as highlighted in **Figure 5** (Panel C) and **Figure 6** (Panel A) (the Brazilian collection). Further, altering light quantity and quality or selection for storage root traits during the domestication process also appears to have differentially impacted gibberellic acid (GA) signaling regulation of DELLAs, as indicated by **Figure 5** (Panel C) (GAI) and **Figure 6** (Panel B) (SLY1, RGA1, RGL2, GAI). These alterations are likely to have impacted known regulatory networks involving interactions between DELLAs and PIF3/PIF4 [39, 40]. Because light perception (lack of shade) would be expected to reduce the positive impact that GA signaling has on inhibiting DELLA, the function of DELLAs in reduced GA-induced elongation likely resulted in dwarfed and bushy phenotypes [41, 42]. In aboveground photosynthesizing tissues, shifts in expression of genes linked to the GA/DELLA regulatory pathway would also be expected to result in shifts between skotomorphogenesis and photomorphogenesis [39] and, potentially, reduced flowering as illustrated in **Figures 2** and **7**. However, in the underground storage root of cassava,
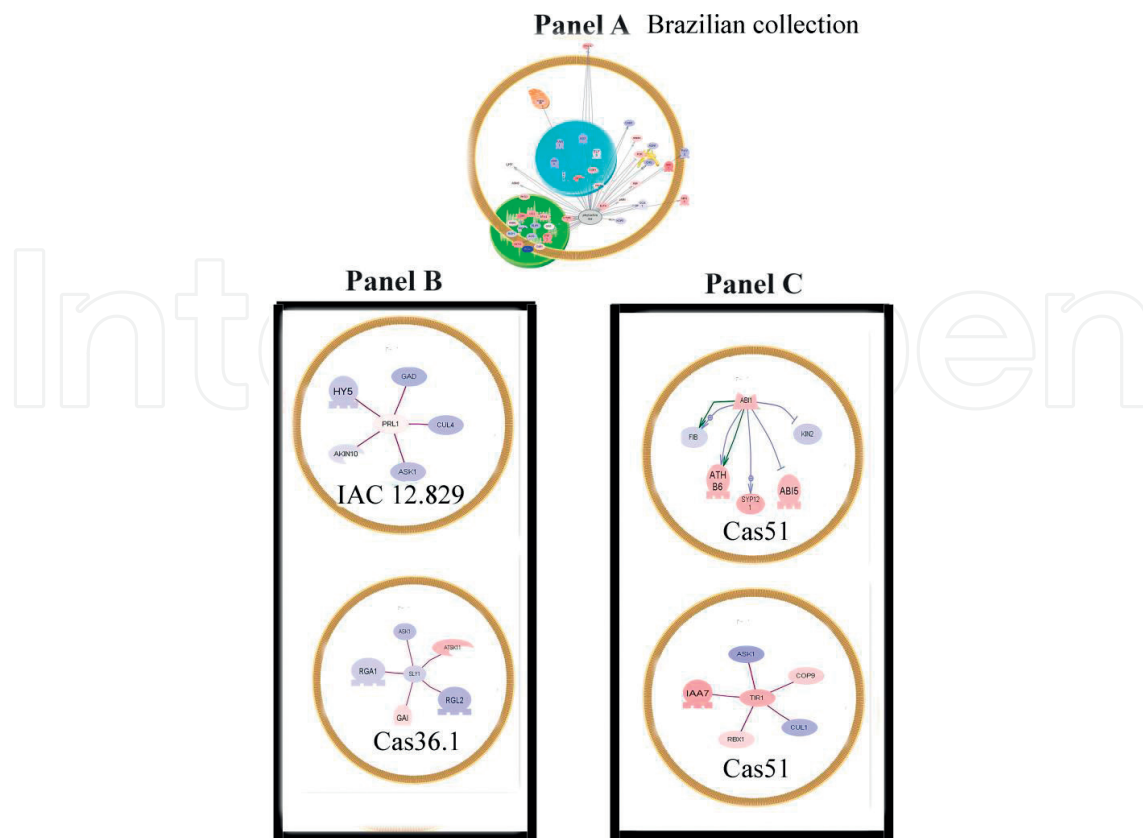
**Figure 6.** Diagram showing exploratory pathway networks for regulatory genes related to cassava ancestral (*M. esculenta subsp flabellifolia*) domestication storage root trait and its genetic diversity in the cultivated species (*M. esculenta subsp esculenta*). Panel A. Differentially expressed regulatory gene node and regulated genes of ancestral and Brazilian collection representatives. Panel B. Differentially expressed regulatory gene node/hub and regulated genes of ancestral and white color cassava (cv. IAC12.829 and sugary landrace Cas36.1). Panel C. Differentially expressed regulatory gene node/hub and regulated genes of ancestral and pink landrace Cas51. Blue color lines refers to more abundant expression in domesticated species, red color lines refers to less abundant expression in domesticated species, and purple color lines represents protein-protein interactions). Arrow (→) refers to positive regulation of the gene and block line (…⊣…) refers to silencing the gene.

altered regulation of DELLAs may have played some role in the shift observed from a fibrous type to a storage root type, as also illustrated in **Figures 2** and **7**. DELLAs have been reported to impact auxin signaling pathways, and, possibly, DELLA's impact on auxin via jasmonic acid (JA) regulation involving JAZ1 and MYC2, as reviewed by [42], could be involved in this process (**Figure 7**). Further, DELLA and SCARECROW (SCR, see **Figure 5 (Panel A— World Collection)** have known interactions that could also be involved in the domestication trait of cassava storage root formation by impacting GA/auxin/cytokinin cross-talk and signaling and root meristematic development and differentiation [44]. **Figure 5 (Panel A)** and **Figure 6 (Panel C)** highlighted abscisic acid (ABA) signaling as a potential component in the domestication process, which could also have some connection to DELLA and SCR proteins known in integration of GA/ABA cross-talk responses [45]. Finally, as presented in **Figure 5 (Panel B)**, **Figure 5 (Panel C)**, **Figure 6 (Panel B)**, and **Figure 6 (Panel B)**, regulatory pathways

| Gene abbreviation | Gene name | ATG number | Function (from www.arabidopsis.org) |
|---|---|---|---|
| 14-3-3 | 14-3-3 family protein | | 14-3-3 proteins are a family of conserved regulatory molecules that are expressed in all eukaryotic cells. 14–3-3 proteins have the ability to bind a multitude of functionally diverse signaling proteins, including kinases, phosphatases, and transmembrane receptors. More than 200 signaling proteins have been reported as 14–3-3 ligands |
| ABI1 | ABA INSENSITIVE 1 | AT4G26080 | Involved in abscisic acid (ABA) signal transduction |
| AP1 | APETALA1 | AT1G69120 | Floral homeotic gene encoding a MADS domain protein homologous to SRF transcription factors. Specifies floral meristem and sepal identity |
| AXR3 | AUXIN RESISTANT 3 | AT1G04250 | Transcription regulator acting as repressor of auxin-inducible gene expression |
| BRI1 | BRASSINOSTEROID INSENSITIVE 1 | AT4G39400 | Encodes a plasma membrane-localized leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. BRI1 appears to be involved in the autonomous pathway that regulates the transition to flowering, primarily through its effects on FLC expression levels |
| COP1 | CONSTITUTIVE PHOTOMORPHOGENIC 1 | AT2G32950 | Represses photomorphogenesis and induces skotomorphogenesis in the dark |
| CRY1 | CRYPTOCHROME 1 | AT4G08920 | A flavin-type blue-light photoreceptor with ATP-binding and autophosphorylation activity. Functions in perception of blue/green ratio of light |
| CTR1 | CONSTITUTIVE TRIPLE RESPONSE 1 | AT5G03730 | Homologous to the RAF family of serine/threonine protein kinases. Negative regulator in the ethylene signal transduction pathway |
| EIN2 | ETHYLENE INSENSITIVE 2 | AT5G03280 | Involved in ethylene signal transduction. Acts downstream of CTR1 |
| FLC | FLOWERING LOCUS C | AT5G10140 | Transcription factor that functions as a repressor of floral transition and contributes to temperature compensation of the circadian clock |
| LSD1 | LESION SIMULATING DISEASE 1 | AT4G20380 | LSD1 monitors a superoxide-dependent signal and negatively regulates a plant cell death pathway |
| PHYA | PHYTOCHROME A | AT1G09570 | Light-labile cytoplasmic red/far-red light photoreceptor involved in the regulation of photomorphogenesis |
| PHYB | PHYTOCHROME B | AT2G18790 | Red/far-red photoreceptor involved in the regulation of de-etiolation |
| PRL1 | PLEIOTROPIC REGULATORY LOCUS 1 | AT4G15900 | Mutations confer hypersensitivity to glucose and sucrose and augment sensitivity to cytokinin, ethylene, ABA, and auxin |
| SCF | Skp, Cullin, F-box containing complex | | A multi-protein E3 ubiquitin ligase complex catalyzing the ubiquitination of proteins destined for proteasomal degradation |
| SLY1 | SLEEPY1 | AT4G24210 | F-box protein that is involved in GA signaling. Component of E3 ubiquitin complex. Interacts with DELLA proteins |

| Gene abbreviation | Gene name | ATG number | Function (from www.arabidopsis.org) |
|---|---|---|---|
| TIR1 | TRANSPORT INHIBITOR RESPONSE 1 | AT3G62980 | Encodes an auxin receptor that mediates auxin-regulated transcription. It contains leucine-rich repeats and an F-box and interacts with ASK1, ASK2, and AtCUL1 to form SCF-TIR1, an SCF ubiquitin ligase complex |
| WAVE1 | WISKOTT-ALDRICH SYNDROME PROTEIN FAMILY VERPROLIN HOMOLOGOUS PROTEIN 1 | AT2G34150 | Encodes a member of the SCAR family. These proteins are part of a WAVE complex. The SCAR subunit activates the ARP2/ARP3 complex which in turn acts as a nucleator for actin filaments |

**Table 4.** List of functional node/hub genes with roles in domestication syndrome which play a role in flowering set trait and storage root formation trait in the cultivated plant of relevance for the domestication syndrome traits (storage root trait and flowering set) as reported in **Table 2**
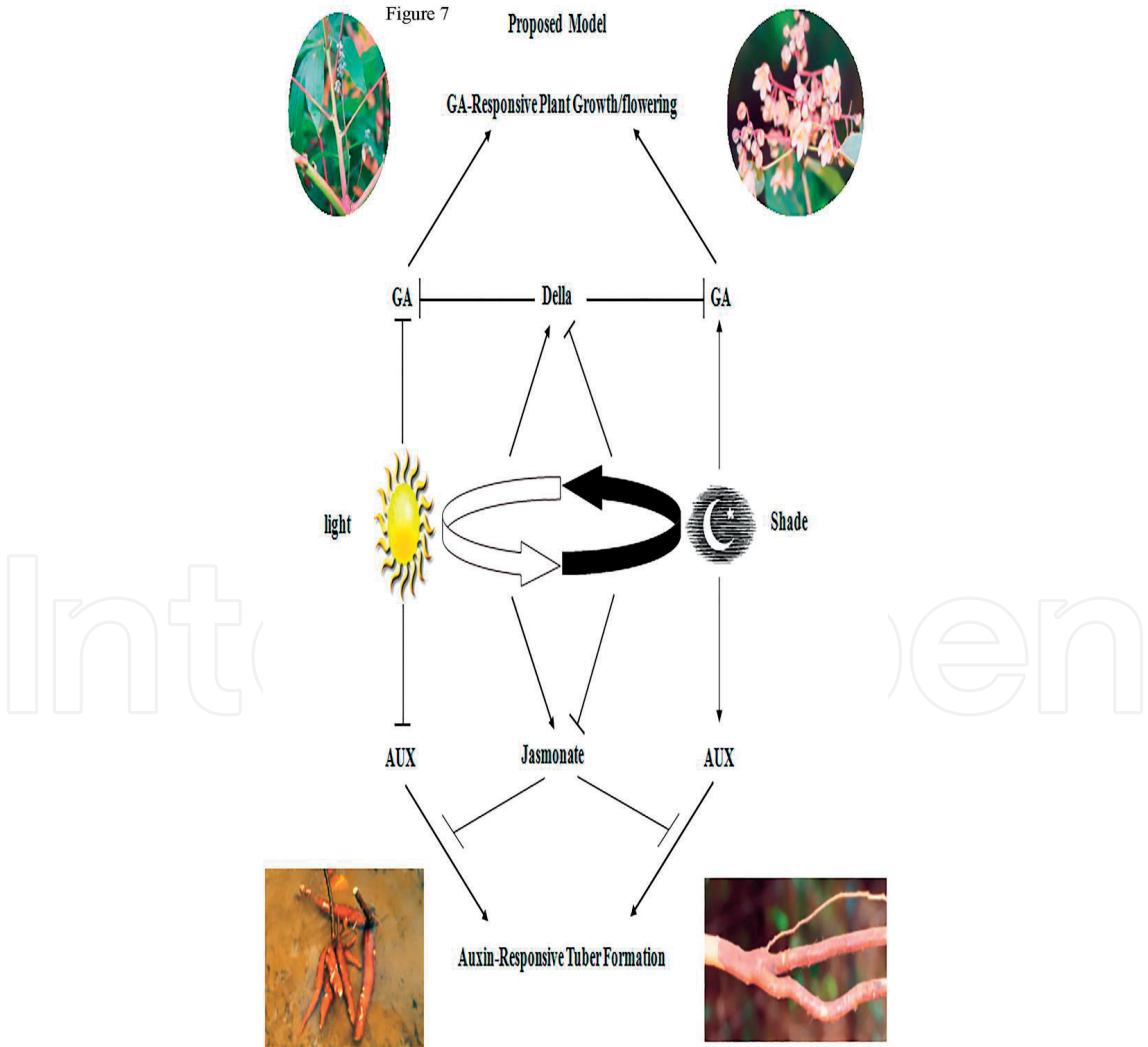


**Figure 7.** Diagram representing a proposed hormonal gene regulatory model to represent the environmental light alteration when human first removed the forest to cultivate cassava to harvest for edible food. The proposed model is based on results of differentially expressed genes of hormone signaling and biosynthesis pathways identified in storage roots of the cassava ancestor when compared to other domesticated cassava landraces and modern cultivars.

(with node/hubs SLY1, SCF, TIR1, and ubiquitin) also appear to function through proteasome degradation pathways. Collectively, these examples suggest that [39, 40] the domestication syndrome may have evolved through a combination of changes in [39, 40] environmental signaling factors and selection pressures that impacted phytohormone cross-talk, signaling, and protein regulation pathways during the removal of the cassava ancestor from the forest, which, in turn, lead to modern-day domesticated landraces and cultivars of cassava.

## 4. Synthesis and conclusions

This study highlighted some key factors influencing the fate of gene function in relation to cassava domestication syndrome traits including (1) landscape alterations resulting in sunlight exposure (alteration of light quality and quantity) in early stages of the domestication process of cassava crop due to removal of forest; (2) natural selective pressure, due to high light intensity under open-field cultivation leading to an edible cassava storage root; (3) artificial selective pressure (man selecting edible plant parts), due to harvest of plants showing storage root formation; and (4) artificial selective pressure (man involuntarily selecting plant traits such as plants with low flowering set in relation to the ancestor). Additionally, the data presented in the present study also allowed, for the first time, to propose a hormonal regulating model (**Figure 7**) on the involvement of GA/DELLA and Auxin/Jasmonate in cassava domestication traits. It appears that domestication of cassava originated with the removal of ancestral *M. esculenta subsp. flabellifolia* from the Amazon forest, which, through human trait preference and selection, evolved into current-day cultivated landraces and cultivars of *M. esculenta* due. As a result, ancestral cassava evolved from a vine, prolific flowering, and fibrous root phenotype into a domesticated bushy, reduced flowering, and tuberous storage root phenotype (**Figures 2** and **7**). Selection for specific storage root traits also resulted in domesticated cassava storage root with color diversity [46], storage root phenotypes, and carbon sequestration diversity with starch vs. sugary storage root phenotypes [47–49].

As our understanding of gene expression evolution improves, it should become possible to infer protein function into approaches focused on the use of proteomics technologies [50]. Ancestral protein functions can be estimated using this approach, and efforts to annotate current genes/proteins will benefit from knowledge of the behavior and factors influencing gene expression profiles. Ultimately, gene expression profiles should be equally integrated with structure and sequence to predict and assist in annotating protein function and evolution directly on the genome sequence of the ancestor and cultivated species.

Technological advances have aided our ability to rapidly and affordably obtain and compare transcriptomes from within and across plant species. As presented in this chapter, we compare the global transcriptomes from storage root of various landraces and cultivars of domesticated cassava and ancestral *M. esculenta subsp. flabellifolia* to identify differentially abundance of transcripts, in this case, at a very stringent level (p < 0.005). Modern technology also affords advances in bioinformatics approaches for analyzing these large transcriptomic data set using ever-evolving algorithms. In this chapter, we used GSEA and SNEA to identify

nodes/hubs and regulatory genes/proteins that provide a snapshot of potential regulatory networks/pathways that differ between ancestral and domesticated cassava storage root and are likely key pathways involved in the domestication syndrome. Based on our results, it appears that as ancestral cassava evolved into domesticated cassava several important regulatory pathways involved in light signaling and regulation, floral signaling and regulation, and hormone signaling and regulation (particularly GA and auxin) were altered. Many of these processes also appear to involve complexes involved in regulating protein degradation through Ubiquitination and Proteasomal trafficking.

## Acknowledgements

## Source of funds

## Author details

Luiz Joaquim Castelo Branco Carvalho[1*†], James V. Anderson[2†], Songbi Chen[3†], Chikelu Mba[4†] and Münevver Doğramaci[5†]

*Address all correspondence to: luiz.castelo@embrapa.br

1 Embrapa-Cenargen, Brasilia, DF, Brazil

2 USDA-ARS, Sunflower and Plant Biology Research Unit, Fargo, ND, USA

3 Tropical Crop Genetic Resources Institute, CATAS, Danzhou, China

4 FAO-The Plant Production and Protection Division, Rome, Italy

5 Department of Internal Medicine, Sanford School of Medicine, University of South Dakota, Vermillion, SD, USA

†All authors contributed equally to this work

# References

[1] Chapin FS III, Autumm K, Pungnaire F. Evolution of suites of traits in response to environmental stress. American Naturalist Supplement. 1993;**142**:S78-S92

[2] Isendahl C. The domestication and early events spread of manioc (*Manihot esculenta* Crantz): A brief synthesis. Latin American Antiquity. 2011;**22**(4):452-468. ISSN: 1045-6635 (Print), 2325-5080

[3] Schaal BA, Olsen KM, Carvalho LJCB. Evolution, domestication, and agrobiodiversity in the tropical crop cassava. In: Timothy J Otley; Nyree Zerega, and Hugh Cross, editors. Darwin's Harvest: New Approaches to New Origins, Evolution, and Conservation of Crops. 1 ed. New York: Columbia University Press; 2005. p. 269-284

[4] Schaal B, Carvalho LJCB, Prinzie T, Olsen K, Hernandez M, Cabral G, Moeller D. Phylogenetic relationships and genetic diversity in *Manihot* species. African Journal of Root and Tuber Crops. 1997;**2**:147-149

[5] Allem AC. The Origins and Taxonomy of Cassava. In: Hillocks RJ, Thresh JM, Bellotti AC, editors, Cassava: biology, production and utilization. CAB International; 2002. p 1-16. DOI: 10.1079/9780851995243.0000

[6] Allem AC. The origin of *Manihot esculenta* Crantz (Euphorbiaceae). Genetic Resources and Crop Evolution. 1994;**41**:133-150

[7] Allem AC. The closest wild relatives of cassava (*Manihot esculenta* Crantz). Euphytica. 1999;**107**:123-133. DOI: 10.1023/A:1026422229054

[8] Allem AC, Roa AC, Mendes RA, Salomão AN, Burle ML, Second G, Carvalho PCL, Cavalcanti J. The primary gene pool of cassava (*Manihot esculenta* Crantz). In: Carvalho LJCB, Thro AM and Vilarinhos AD, editors. Cassava Biotechnology - IV Intl Sci Meet-CBN, EMBRAPA-Genet Res Biotechn, Brazil: Brasilia-DF; 2000. p. 3-14

[9] Rogers DJ, Appan SG. *Manihot* and Manihotoides (Euphorbiaceae), a computer-assisted study. Flora Neotropica, monograph no 13. New York: Hafner Press; 1973; 272 pp

[10] Rogers DJ. Some botanical and ethnological considerations of *Manihot* esculenta. Economic Botany. 1965;**19**:369-377

[11] Olsen KM, Wendel JF. A bountiful harvest: genomic insights into crop domestication phenotypes. Annu Rev Plant Biol. 2013;**64**:47-70. DOI: 10.1146/annurev-arplant-050312-120048

[12] Olsen KM. Population history of *Manihot esculenta* (Euphorbiaceae) inferred from nuclear DNA sequences. Molecular Ecology. 2002;**11**:901-911

[13] Briana LG, Olsen KM. Genetic perspectives on crop domestication. Trends Plant Sci. 2010;**15**(9):529-537. DOI: 10.1016/j.tplants.2010.05.008

[14] Olsen KM, Schaal BA. Evidence on the origin of cassava: phylogeography of Manihot esculenta. Proceedings of the National Academy of Sciences of the USA. 1999;**96**:5586-5591

[15] Olsen K, Schaal B. New evidence on the origins of cassava: Phylogeography of *Manihot esculenta*. Proc. Natl. Acad. Sci. 1999;**96**:5586-5591

[16] Olsen K. Schaal B, Hernandez M. A survey of DNA sequence variation in cassava and other *Manihot* species. In Cassava Biotechnology: IV International Scientific Meeting – *CBN*, 129-134, eds. Carvalho LJCB, Thro AM, Vilarinhos AD. Brasília, DF, Brasil: Embrapa–CENARGEN; 2000. p. 129-134

[17] Olsen K, Schaal B. Microsatellite variation in cassava (Manihot esculenta, Euphorbiaceae) and its wild relatives: Further evidence for a southern Amazonian origin of domestication. Amer. J. Botany. 2001;**88**:131-142

[18] Olsen KM. Minisatellite variation in a single-copy nuclear gene: phylogenetic assessment of repeat length homoplasy and mutational mechanism. Molecular Biology and Evolution, 1999;**16**:1406-1409

[19] Schaal B, Olsen KM. Gene genealogies and variation within plant populations. Proc. Natl. Acad. Sci. 2000;**97**:7024-7029

[20] Beer MA, Tavazoie. Predicting Gene Expression from Sequence. Cell. 2004;**117**:185-198

[21] Gross BL, Olsen KM. Genetic perspectives on crop domestication. Trends in Plant Science. 2010;**15**:529-537. DOI: 10.1016/j.tplants.2010. 05.008

[22] Doebley JF, Gaut BS, Smith BD. The Molecular Genetics of Crop Domestication. Cell. 2006;**127**:1309-1321

[23] Wang W, Feng B, Xiao J, Xia Z, Zhou X, Li P, Zhang W, Wang Y, Møller BL, Zhang P, Luo MC, Xiao G, Liu J, Yang J, Chen S, Rabinowicz PD, Chen X, Zhang HB, Ceballos H, Lou Q, Zou M, Carvalho LJCB, Zeng C, Xia J, Sun S, Fu Y, Wang H, Lu C, Ruan M, Zhou S, Wu Z, Liu H, Kannangara RM, Jørgensen K, Neale RL, Bonde M, Heinz N, Zhu W, Wang S, Zhang Y, Pan K, Wen M, Ma PA, Li Z, Hu M, Liao W, Hu W, Zhang S, Pei J, Guo A, Guo J, Zhang J, Zhang Z, Ye J, Ou W, Ma Y, Liu X, Tallon LJ, Galens K, Ott S, Huang J, Xue J, An F, Yao Q, Lu X, Fregene M, López-Lavalle LA, Wu J, You FM, Chen M, Hu S, Wu G, Zhong S, Ling P, Chen Y, Wang Q, Liu G, Liu B, Li K, Peng M. Cassava genome from a wild ancestor to cultivated varieties. Nat Commun. 2014;**5**:5110-5119

[24] Lokko Y, Anderson JV, Rudd S, Raji A, HorvathD, Mikel MA, Kim R, Liu L, Hernandez A, Dixon AGO, Ingelbrecht IL. Characterization of an 18,166 EST dataset for cassava (Manihot esculenta Crantz) enriched for drought-responsive genes. Plant Cell Report. 2007;**26**:1605-1618

[25] http://www.phytozome.net/search.php?show=blast&targetType=genome&method= Org_Mesculenta

[26] http://cys.bios.niu.edu/plantcazyme/family.php?family=GH13&species=maes

[27] https://www.ncbi.nlm.nih.gov/nucest/?term=cassava%20carvalho

[28] https://www.ncbi.nlm.nih.gov/protein/?term=cassava+carvalho

[29] https://www.ncbi.nlm.nih.gov/genome/?term=cassava+carvalho

[30] http://www.cassavabase.org

[31] www.cenargen.embrapa/genome

[32] http://cassava.psc.riken.jp/ver.1/snp/index.pl?v=1

[33] http://www.R-project.org/

[34] Le S, Josse J, Husson F. Factominer: an R-package for multivariate analysis. Journal of statistical software, 2008;**25**(1):1-18

[35] Chavent M, Kuentz V, Saracco J. Orthogonal Rotation in PCAMIX. Advances in Classification and Data Analysis, 2011;**6**:131-146

[36] Doğramaci M, Horvath DP, Chao WS, Foley ME, Christoffers MJ, Anderson JV. Extended low temperature impacts dormancy status, flowering competence, and transcript profiles in crown buds of leafy spurge. Plant Mol. Biol. 2010;**73**:207-226. DOI: 10.1007/s11103-010-9621-8

[37] Doğramaci M, Foley ME, Chao WS, Christoffers MJ, Anderson JV. Induction of endodormancy in crown buds of leafy spurge (Euphorbia esula L.) implicates a role for ethylene and cross-talk between photoperiod and temperature. Plant Mol. Biol. 2013;**81**:577-593. DOI: 10.1007/s11103-013-0026-3

[38] Mazzella MA, Arana MV, Staneloni RJ, Perelman S, Rodriguez Batiller MJ, Muschietti J. Phytochrome control of the Arabidopsis tran¬scriptome anticipates seedling exposure to light. Plant Cell. 2005;**17**:2507-2516. DOI: 10.1105/tpc.105.034322

[39] Sun TP. Gibberellin metabolism, perception and signaling pathways in Arabidopsis. In: The Arabidopsis Book. American Society of Plant Biologists, Rockville, MD; 2008. doi/10.1199/tab.0103, http://www.aspb. org/publications/arabidopsis/

[40] Harberd NP, Belfield E, Yasumura Y. The angiosperm gibberellin-GID1-DELLA growth regulatory mechanism: How an "inhibitor of an inhibitor" enables flexible response to fluctuating environments. The Plant Cell. 2009;**21**:1328-1339

[41] Kazan K, Manners JM. JAZ repressors and the orchestration of phyhormone crosstalk. Trends in Plant Science, 2012;**17**(1):22-31

[42] Moubayidin L, Salvi E, Giustini L, Terpstra I, Heidstra R, Costantino P, Sabatini SA. SCARECROW-based regulatory circuit controls Arabidopsis thaliana meristem size from the root endodermis. Planta. 2016;**243**:1159-1168

[43] https://www.pathwaystudio.com/

[44] www.arabidopsis.org

[45] Golldack D, Li C, Mohan H, Probst N. Gibberellins and abscisic acid signal crosstalk: living and developing under unfavorable conditions. Plant Cell Rep. 2013;**32**:1007-1016

[46] Carvalho LJCB, Agustini MA, Anderson JV3, Vieira EA, de Souza CR, Schaal BA, Silva JP, Natural variation in expression of genes associated with carotenoid biosynthesis and accumulation in cassava (Manihot esculenta Crantz) storage root. BMC Plant Biol. 2016;**16**(1):133. DOI: 10.1186/s12870-016-0826-0

[47] Carvalho LJCB, de Souza CRB, Cascardo JMC, Bloch Junior C, Campos L. Identification and characterization of a novel cassava (Manihot esculenta Crantz) clone with high free sugar content and novel starch. Plant Mol Biol. 2004;**56**:643-659

[48] Carvalho LJCB, de Souza CRB, Cascardo JCM, Agostini MAV, Vieira EA Anderson JV, Lippolis J. Natural Genetic Variation in Cassava (Manihot esculenta Crantz) Landraces: A Tool for Gene Discovery. In: Shu QY, editor, Induced Plant Mutations in the Genomics Era. Food and Agriculture Organization of the United Nations, Rome; 2009. p. 313-316

[49] Carvalho LJCB, Schaal BA. Assessing genetic diversity in the cassava (Manihot esculenta Crantz) germplasm collection in Brazil using PCR-based markers. Euphytica. 2001;**120**:133-142

[50] An F, Chen T, Djabou MAS, Kaimian L, Qing XL, Carvalho LJCB, Keith T, Li BG, Chen S. Domestication syndrome is investigated by proteomic analysis between cultivated cassava (*Manihot esculenta* Crantz) and its wild relatives. PLoS One. 2016;**11**(3):e0152154. https://doi.org/10.1371/journal.pone.0152154