Plant Physiology and Biochemistry 199 (2023) 107713

Contents lists available at ScienceDirect

Plant Physiology and Biochemistry

journal homepage: www.elsevier.com/locate/plaphy

Transcriptome and metabolome profiling identify factors potentially involved in pro-vitamin A accumulation in cassava landraces

Priscilla Olayide [a,b], Erik Alexandersson [a], Oren Tzfadia [c], Marit Lenman [a], Andreas Gisel [b,d], Livia Stavolone [b,e,*]

[a] Swedish University of Agricultural Sciences, Sundsvägen 10, SE-234 22, Lomma, Sweden
[b] International Institute of Tropical Agriculture, PMB 5320, Oyo Road, Ibadan, 200001, Oyo State, Nigeria
[c] Institute of Tropical Medicine, Kronenburgstraat 43/3, 2000, Antwerpen, Belgium
[d] Institute of Biomedical Technologies, CNR, Via Amendola 122/D, Bari, Italy
[e] Institute for Sustainable Plant Protection CNR, Via Amendola 122/D, Bari, Italy

ARTICLE INFO

Keywords:
Cassava
Carotenoids
Transcriptome
Metabolites
NCED
provitamin A
psy

ABSTRACT

Cassava (*Manihot esculenta* Crantz) is a predominant food security crop in several developing countries. Its storage roots, rich in carbohydrate, are deficient in essential micronutrients, including provitamin A carotenoids.

Increasing carotenoid content in cassava storage roots is important to reduce the incidence of vitamin A deficiency, a public health problem in sub-Saharan Africa. However, cassava improvement advances slowly, mainly due to limited information on the molecular factors influencing β-carotene accumulation in cassava.

To address this problem, we performed comparative transcriptomic and untargeted metabolic analyses of roots and leaves of eleven African cassava landraces ranging from white to deep yellow colour, to uncover regulators of carotenoid biosynthesis and accumulation with conserved function in yellow cassava roots.

Sequence analysis confirmed the presence of a mutation, known to influence β-carotene content, in *PSY* transcripts of deep yellow but not of pale yellow genotypes. We identified genes and metabolites with expression and accumulation levels significantly associated with β-carotene content. Particularly an increased activity of the abscisic acid catabolism pathway together with a reduced amount of L-carnitine, may be related to the carotenoid pathway flux, higher in yellow than in white storage roots. In fact, *NCED_3.1* was specifically expressed at a lower level in all yellow genotypes suggesting that it could be a potential target for increasing carotenoid accumulation in cassava.

These results expand the knowledge on metabolite compositions and molecular mechanisms influencing carotenoid biosynthesis and accumulation in cassava and provide novel information for biotechnological applications and genetic improvement of cassava with high nutritional values.

1. Introduction

Cassava (*Manihot esculenta* Crantz) is an important food and feed crop, particularly for subsistence of small-holder farmers in sub-Saharan Africa (SSA) where, due to the low standard of living, dietary diversification is quite hard to achieve. It is regarded as a food security crop due to its ability to grow and yield reasonable harvest even under poor soil conditions and during long periods of droughts (Orek et al., 2020; Imakumbili et al., 2021). While cassava has high starch content and serves as a good source of dietary carbohydrate, it is low in micronutrients, such as iron, zinc, and particularly provitamin A carotenoids (Ceballos et al., 2017). Vitamin A is essential for vision, epithelial cell synthesis and differentiation (Chiu et al., 2016). Therefore, a cassava-based diet increases the risk of vitamin A deficiency (VAD), a major public health problem in SSA resulting from an inadequate intake of dietary vitamin A (Sayre et al., 2011).

Carotenoids are a class of isoprenoid pigments found in plants and other photosynthetic organisms (Sandmann, 2021). In green plant tissues, carotenoids function in light harvesting in photosynthesis and are synthesized in tight coordination with chlorophyll (Stanley and Yuan,

* Corresponding author. International Institute of Tropical Agriculture, PMB 5320, Oyo Road, Ibadan, 200001, Oyo State, Nigeria.
E-mail addresses: priscilla.olayide@gmail.com (P. Olayide), erik.alexandersson@slu.se (E. Alexandersson), otzfadia@itg.be (O. Tzfadia), marit.lenman@slu.se (M. Lenman), a.gisel@cgiar.org, andreas.gisel@ba.itb.cnr.it (A. Gisel), l.stavolone@cgiar.org, livia.stavolone@cnr.it (L. Stavolone).

https://doi.org/10.1016/j.plaphy.2023.107713
Received 2 February 2023; Received in revised form 5 April 2023; Accepted 18 April 2023
Available online 24 April 2023

2019a). They have antioxidant properties, are synthesis precursors of strigolactones, abscisic acid (ABA) and other apocarotenoids involved in regulation of growth and development and for maintenance of normal health, also in humans (Kim et al., 2018). Provitamin A carotenoids, which include α-carotene, β-carotene and β-cryptoxanthin, are known for their characteristic bright colours and are converted into vitamin A after ingestion (D'Andrea and Rodriguez-Concepcion, 2019). Regrettably, VAD incidence is increasing globally, affecting mostly preschool children and pregnant women (Darnton-Hill, 2019) and, despite the efforts made to tackle VAD through dietary diversification, food fortification and vitamin supplementation, prevalence of VAD has not changed in SSA (Ceballos et al., 2017).

Most cassava varieties grown and consumed around the world have white storage roots, which are generally low in provitamin A, and their improvement through conventional breeding has been challenging due to the low flowering ability, heterozygous breeding material, and long cropping cycle (Guo et al., 2017). Also, several studies have reported a negative correlation between dry matter and carotenoid content in cassava (Esuma et al., 2012; Rabbi et al., 2017). This implies that genotypes with deeper yellow root colour are lower in dry matter content, a trait combination that could affect the acceptance of varieties lacking either of them. However, our previous study and others have revealed that carotenoid and dry matter contents behave as independent or only weakly correlated variables, suggesting that cassava germplasm can be improved for both traits in parallel (Sanchez et al., 2014; Olayide et al., 2020). To this end, a better understanding of the molecular mechanisms governing both carotenoid and starch metabolism in cassava, and the identification of key genes there involved, is essential to support biofortification of cassava. The discovery of novel and generally valid targets for cassava improvement becomes particularly important in the face of the climate change, which brings new biotic and abiotic challenges to be addressed in parallel with nutrients availability.

Phytoene biosynthesis is the first committed step in the carotenoid biosynthesis pathway (CBP). This reaction is catalysed by phytoene synthases (PSY), which is regarded as the limiting enzyme in tissues with low carotenoid composition (Hannoufa and Hossain, 2012; Nisar et al., 2015). *Psy* sequence comparison, in low and high carotenoid-accumulating cassava storage roots, revealed sequence variations and/or deletions, including a single nucleotide polymorphism in *MePSY2* resulting in a nonconservative amino acid exchange, which increases the carotenoid accumulation (Welsch et al., 2010).

Studies have been carried out to elucidate the carotenoid profiles of tuberous crops and gain insights into the expression of carotenoid related genes and their functions in these non-photosynthetic organs. In potato, a gene encoding PSY and another encoding β-carotene hydroxylase (BCH) are linked to carotenoids regulation, which is thought to be responsible for the large variation in colour intensity observed across cultivars (Morris et al., 2004). Comparative expression studies between cultivars of *Solanum tuberosum* L. and yellow potato *Solanum phureja,* at different stages of tuber development, revealed similar transcript levels of many carotenogenic genes, including those encoding 1-deoxy-D-xylose-5-phosphate reductoisomerase (DXS), ζ-carotene desaturase, lycopene-β-cyclase (LYCβ), BCH, and neoxanthin synthase (NXS) (Morris et al., 2004). Also, expression profile and abundance of carotenoid related genes, e.g. those encoding PSY, PDS, carotenoid isomerase (CRTISO), BCH, LYCβ and NXS, and of genes regulating plastid division, were found to correlate with carotenoid composition and total β-carotene accumulation among three Amazon cassava landraces (Carvalho et al., 2016). In a more recent study, we also found that genes encoding PSY2, lycopene-ε-cyclase LCYε, and CHYβ were generally overexpressed in the roots of yellow cassava landraces compared to white ones (Olayide et al., 2020).

The Orange (*OR*) gene has also been linked to carotenoid accumulation in several crops, including sweet potato and carrot (Kim et al., 2018; Shim et al., 2020). The OR cysteine-rich zinc finger domain has been found to interact with PSY and mediate increase in PSY protein levels and stability leading to higher carotene content (Zhou et al., 2015). The *OR* gene is also known to mediate high levels of carotene accumulation by regulating chromoplast differentiation to provide a sink for stable storage (Osorio, 2019).

Variations in carotenoid contents have been found associated to variations in ABA levels in coloured citrus and *Arabidopsis thaliana* mutants, and since ABA is formed by the cleavage of neoxanthin by 9-cis-epoxycarotenoid dioxygenase (NCED), also the regulation of the gene encoding this enzyme can influence carotenoid accumulation (Barrero et al., 2008; Liu et al., 2021).

While quite a number of putative transcriptional regulators of carotenoid biosynthesis and accumulation have been identified, only few of them have seemingly conserved function in the transcriptional control of carotenoid biosynthesis genes across tissue types or plant species, particularly in roots of model systems and of root crops (Stanley and Yuan, 2019b).

Carotenoid biosynthesis and regulation is complex, and increasing evidence indicates that there is more to carotenoid accumulation than expression and regulation of carotenoid-related genes and proteins.

Seeking for regulators of carotenoid biosynthesis and accumulation with conserved function in cassava roots, we performed comparative transcriptomic and untargeted metabolic analyses of leaves and roots of a panel of yellow and white African cassava landraces. A number of genes and metabolites with expression and accumulation levels significantly associated to β-carotene content were identified. Among them, *NCED_3.1* was specifically expressed at a lower level in all yellow genotypes suggesting that its regulation could generally influence β-carotene accumulation in cassava. Functional gene enrichment, metabolite profiling and data integration, performed to identify novel functions and processes indirectly related to carotenoid accumulation revealed conserved features and targets specifically linked to the level of β-carotene root content.

This study provides a basis for deeper understanding of the molecular mechanisms influencing carotenoid biosynthesis and accumulation in cassava.

Further studies are ongoing to confirm the function of the newly identified putative carotenoid regulators, which will be useful and innovative targets for biotechnological applications and breeding programs.

2. Materials and methods

2.1. Plant materials

Eleven cassava landraces of different root colour ranging from white to deep yellow were cultivated in open fields at the International Institute of Tropical Agriculture (IITA), Nigeria (N 7.490250, E 3.884143; Table 1). Planting materials consisted of stem cuttings (or stakes) from fresh portions of mature cassava stems. Field layout was in a randomised block design with two replications. Each genotype was allotted a plot

Table 1
Properties of cassava genotypes.

Genotype	Code	Root colour	Root starch (%)	Dry matter (%)	β-carotene (μg/g)	
					roots	leaves
Kaleso	A	White	**85.29** ± 1.17	**44.13** ± 1.87	**0.26** ± 0.06	**416.75** ± 106.30
TMEB419	R	White	**76.31** ± 1.13	**34.45** ± 2.15	**1.60** ± 0.14	-
TMS980505	P	White	**74.44** ± 1.07	**32.20** ± 1.39	**1.73** ± 0.57	-
IBA011231	C	Pale Yellow	**77.15** ± 5.31	**32.12** ± 3.08	**4.38** ± 1.2	**928.61** ± 86.55
IBA050128	B	Pale Yellow	**79.77** ± 4.87	**35.83** ± 1.83	**6.15** ± 0.98	**1340.35** ± 29.76
IBA083724	D	Yellow	**64.85** ± 2.34	**32.62** ± 1.68	**11.15** ± 3.10	**731.49** ± 38.62
IBA083565	G	Yellow	**52.72** ± 7.83	**20.75** ± 2.51	**12.37** ± 1.55	**1479.30** ± 47.58
IBA130799	O	Yellow	**69.29** ± 3.33	**18.48** ± 9.03	**12.37** ± 3.21	**1046.25** ±105.19
IBA102103	J	Yellow	**58.02 ± 4.37**	**26.37** ± 3.30	**13.98** ± 7.18	**917.58** ± 34.53
IBA011663	I	Yellow	**74.19** ± 4.23	**23.34** ± 1.90	**15.27** ± 0.65	**943.84** ± 72.54
IBA141092	L	Deep Yellow	**69.62** ± 2.77	**25.45** ± 1.13	**36.06** ± 0.78	**1441.76** ± 35.94

and adjacent plots were separated by 1 m alleys. Twenty plants were planted per genotype in each plot which contained four rows of plants with five plants in each row and a 1m spacing between each plant. Plants were selected randomly for sampling and the entire field was harvested at the end of the planting season. Source leaf and storage root samples were collected in three biological replicates from 240-days (8 months) old cassava plants, planted in the 2018 planting season. Individual plants were regarded as biological replicates, and three source leaves and one root were sampled from each plant. Storage roots were peeled and chopped into small cubes and mixed together for homogeneity. Source leaf and storage root samples were snap frozen in liquid nitrogen immediately after collection and stored at −80 °C.

2.2. β-carotene and starch content

Extraction and analysis of carotenoids and chlorophyll are as described in Olayide et al. (2020). Two additional genotypes TMS980505 and TME419 added to the study were analysed for β-carotene content by Spectrophotometry Multiskan GO (Thermo Scientific, Waltham, USA). 1 mL of ethyl acetate: ethanol: butylated hydroxytoluene (80:20:0.1, v/v/w) was added to 100 mg lyophilized cassava powder and placed in ultrasonic water bath (VWR, Radnor, Pennsylvania, USA) at 30 °C for 10 min followed by incubation at 60 °C for 60 min. The mixture of sample and solvent was centrifuged at 4 °C, 5400 rcf for 10 min and approximately 2 mL of supernatant was transferred to collection Eppendorf tubes, centrifuged at 4 °C, 5400 rcf for 10 min. β-carotene content was measured at 450 nm, path length 10 mm and expressed as ug/g dry weight.

Starch content of lyophilized root samples was determined using a Total Starch Assay kit by Megazyme (Bray, Ireland) using the same methods described in Olayide et al. (2020).

2.3. Untargeted analysis by GC- and LC-MS

Metabolic profiling by GC-MS and LC-MS was performed at the Swedish Metabolomics Center in Umeå, Sweden, as described previously by (Diamanti et al., 2019). A full detailed method description can be found in the **Supplementary data**. After identification and quantification, the metabolite data structure was assessed by Normalyzer (Chawade et al., 2014), and based on the results it was decided not to normalise metabolite abundance. Metabolite data were visualized using ClustVis, a web tool for visualising clustering of multivariate data (Metsalu and Vilo, 2015).

2.4. RNA extraction and purification

RNA from root and leaf tissue of 10 genotypes in 3 replicates (60 samples in total) was extracted from root and leaf samples by combining cetyl trimethylammonium bromide (CTAB)-extraction method and spin-column based purification (Carluccio et al., 2022). The purified RNA was resuspended in RNase-free water, and RNA quality was preliminarily assessed using a NANODROP 8000 spectrophotometer (Thermo Scientific, Waltham, MA USA). RNA samples with OD260/280 and OD260/230 values ranging between 1.9 and 2.2 were selected for further analysis. 28S and 18S RNA quality were then assessed by gel electrophoresis using 1.2% agarose gel with 1 × Tris/Borate/EDTA buffer (Sigma-Aldrich, St. Louis, MO, USA), at 80V for 40 min. Samples with good quality were selected for RNA-sequencing (RNA-seq).

2.5. Library construction and sequencing

RNA-seq was performed by SNP & SEQ Technology Platform in Uppsala, Sweden. Sequencing libraries were prepared from 75 ng total RNA, using the TruSeq stranded total RNA library preparation kit with RiboZero Gold treatment (cat# 20020598/9, Illumina Inc. San Diego, USA). The library preparation was performed according to the manufacturers' protocol (# 1000000040499). Sequencing was done in a NovaSeq S4 flowcell, paired-end 150bp read length, v1 sequencing chemistry. Paired-end sequencing data was deposited in ENA under the project accession PRJEB48562.

2.6. Quality control and mapping of Illumina reads

Read quality was checked using FastQC v0.11.3 (Andrews, 2010). Raw reads were mapped against the reference genome of Manihot esculenta v6.1 (Phytozome, https://phytozome-next.jgi.doe.gov/) using the spicing site sensitive mapping tool STAR_2.4.1c and STAR index including the corresponding GFF annotation file (Phytozome, https://phytozome-next.jgi.doe.gov/). STAR parameters were set to eliminate low quality mappings and including: i) only reeds aligning with fragments longer than 50% of the full read length (paired end), ii) only matching fragments with less than 10 mismatches and iii) only alignments with a score of 66% of the full length. The output SAM file was ordered by coordinates and converted into a BAM file using the PICARD tool MergeSamFiles.

2.7. Gene expression analysis

BAM files were used to reconstruct the transcripts and calculate the expression values for each gene using CuffLinks v2.2.1.' and normalisation and differential expressions were calculated using CuffDiff v2.2.1 (Trapnell et al., 2010). Only genes with FPKM >1 in at least one sample were considered significant and used for downstream analyses (Table S1). Gene clustering by expression profile was performed using EXPANDER software (Hait et al., 2019).

Functional enrichment analysis was performed using the gProfiler web tool (Raudvere et al., 2019) providing significant enriched Gene Ontology (GO, http://www.geneontology.org) terms for molecular function, biological process and cellular component. The PCAs were produced using the R (version 3.6.3) and the ggplot2 (version 3.3.5) packages.

2.8. Transcriptome assembly and cassava gene identification

The RNA-seq reads were assembled for each genotype merging the replicates using the *de novo* transcript assembler Trinity v2.10.0 (Grabherr et al., 2011).

Cassava genes involved in the different metabolic pathways mentioned below were extracted from Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/) and the RefSeq annotations converted into the IDs used in this study using blastp against the predicted protein sequences of the Manihot esculenta v6.1 annotation.

2.9. Integration of expression and metabolomic data

For multi-omics data integration, we used the MOFA (Multi-Omics Factor Analysis) software (Argelaguet et al., 2018). The input data used were expression data, filtered for genes which are differentially expressed between the white and at least one yellow genotype (2781 genes), and metabolomics data produces by LCMS (73 metabolites) and GCMS (59 metabolites. Both metabolite datasets were normalized with the SERRF (Systematical Error Removal using Random Forest) software (Fan et al., 2019). MOFA was used with the default parameters and number of factors equal to eight.

3. Results and discussion

3.1. Phenotypical and biochemical characterization of cassava landraces

The cassava landraces selected for this study vary in root colour, ranging from white to deep-yellow (Table 1). In accordance with their visible appearance, landraces with white storage roots have very low

β-carotenoid content while in yellow root landraces the carotenoid content varies according to the level of root yellowness (from pale-yellow to deep yellow, Table 1). Consistently with previous reports (Ceballos et al., 2017), we observed significantly higher ($p < 0.001$) β-carotene accumulation in leaves compared to roots (Table S2), regardless of the landrace root colour. In particular, leaf and root contents were not correlated as indicated by pale yellow genotypes accumulating similar or higher levels of β-carotene in leaf compared to yellow genotypes (Table 1). This suggests that carotenoid synthesis is not the limiting factor for carotenoids accumulation roots of white or pale-yellow genotypes. In roots, percent starch ranged from 52 to 85% of the dry weigth, with the highest percentage in the white-root genotype A, and the lowest in the yellow landrace G. However, some yellow genotypes had relatively high starch content, up to 77%. Percent dry matter in these landraces ranged from 21 to 44% with the highest in A and the lowest in G; and dry matter content varied substantially between yellow cultivars. These results confirm that several genotypes may deviate from the reported negative correlation between dry matter and carotenoid content (Ceballos et al., 2013; Sanchez et al., 2014), and such correlation cannot be generally assumed in cassava.

3.2. Global comparison of cassava landraces

To understand the molecular mechanisms influencing carotenoid biosynthesis or accumulation in cassava, we investigated molecular differences between field-grown white and yellow genotypes by comparative gene expression analysis. Paired-end RNA-seq was performed of leaves and roots samples collected eight months after planting in biological triplicates.

An average of 34.9 million paired-end reads were obtained per genotype and tissue, 91% of which mapped/matched to the reference genome. In total, 26,514 genes (out of the 33,033 predicted genes in the cassava genome V6.1) were detected in at least one sample and were used in subsequent analysis. Comparative gene expression profile analysis was used to investigate molecular differences between white and yellow landraces. The results showed that 14,136 genes were expressed (FPKM >1) in roots while 13,383 genes were expressed in leaves.

To assess similarities in global gene expression among genotypes, Principal Components Analysis (PCA) was performed on both leaf and root significantly expressed genes. In leaves, genotypes do not obviously group according to the colour of the storage root except for the white genotypes A, which stands out (Fig. 1). This indicates that regulation of the carotenoid pathway in leaves is not directly influencing carotenoids accumulation in roots. Instead, in roots a partial separation of white and yellow rooted genotypes is shown in PC1and PC2, where the three genotypes with the highest β-carotene content in roots (I, J and L; Table 1) clearly form a separate group (Fig. 1, top right). Furthermore, the projection of PC2-PC3 shows clear separation between yellow and white genotypes (Fig. 1, bottom right). Interestingly, the pale-yellow and yellow cultivars tend to fall in separate clusters in roots, albeit their relative positions suggest that there are clear similarities in the overall gene expression. Consistently with the PCA results, genotype pairwise comparisons revealed that higher number of differentially expressed genes (DEG) is observed between root sample with most different β-carotene content (Table S3).

3.3. Carotenoid-related gene expression

To examine differences among cassava genotypes, we compared the expression levels of genes known to be involved in the cassava carotenoid pathway (Fig. 2) and expressed in leaf and root tissues of white and yellow landraces. PCA biplots, based on the expression of only these selected genes, suggest that the difference in the expression profile of carotenoid-related genes does not consistently correspond to the yellow and white root phenotypes. This is particularly evident in leaves, despite the highly similar trend of expression among most of the analysed gene expression (vector arrows close together, Fig. 3A) and further suggests no direct correlation between carotenoids synthesized in leaves and those accumulated in roots. In roots, white genotypes P and R cluster together, albeit opposite to the other white genotype A (Fig. 3A). Yellow rooted genotypes are distributed all around the vector origins, showing that carotenoid-related genes can have positive or negative influence on different yellow genotypes (Fig. 3A). The expression of genes involved in β-carotene accumulation in roots, e.g. those encoding Deoxyxylulose 5-phosphate synthase (DXS), ζ-Carotene desaturase (ZDS), PDS, PSY_1.2, OR_3.0, NXSb. LYC, is positively correlated (positive side of the vector origin) with root carotenoids accumulation in the genotypes with highest β-carotene content (J, I and L). Consistently, the expression of

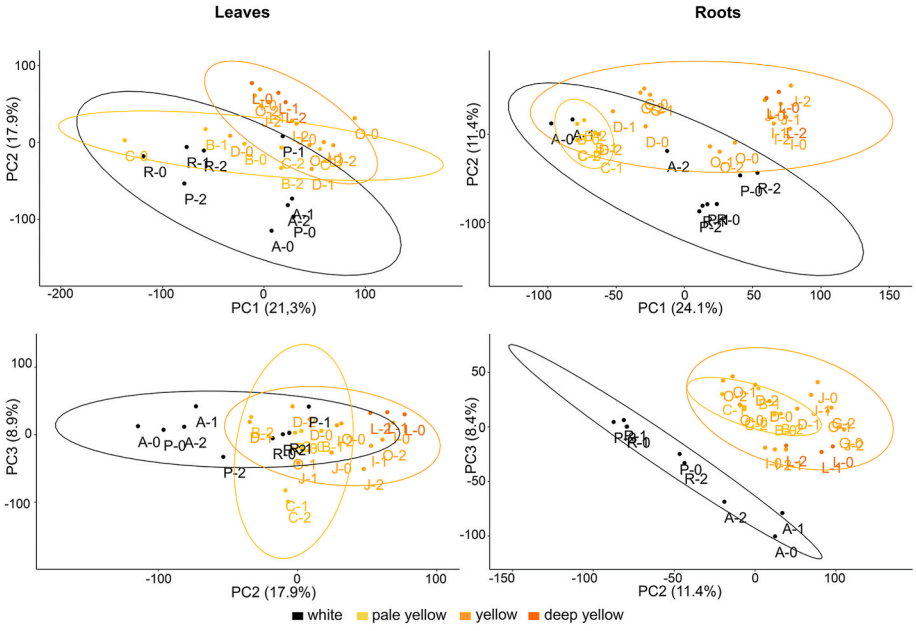

Fig. 1. Transcriptome relationships among cassava genotypes. Principal component analysis (PCA) of leaf (left panel) and root (right panel) genes (FPKM>1). The clustering of the replicates of the different genotypes and the separation between genotypes is visualized with PC1 vs PC2 and PC2 vs PC3.

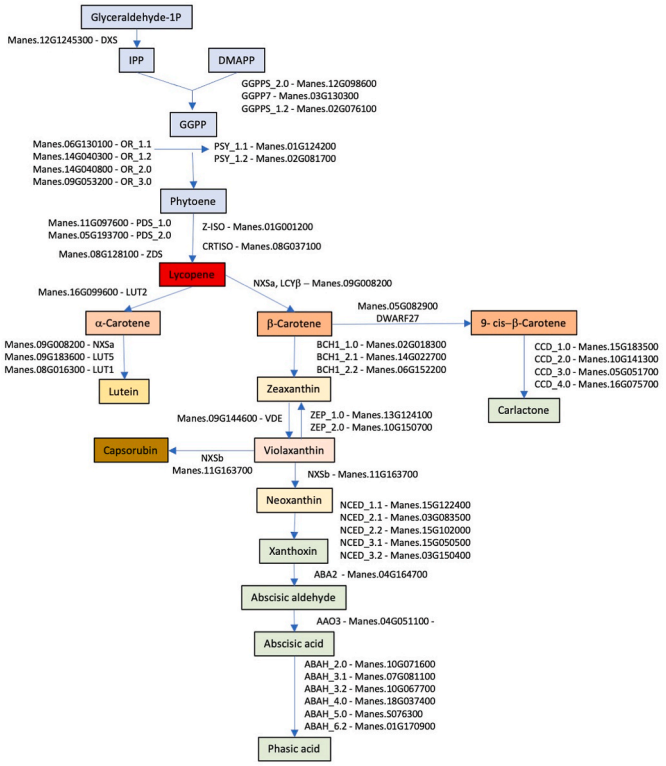

Fig. 2. Carotenoid biosynthesis pathway in cassava. Metabolites are visualized in coloured boxes and the predicted cassava genes catabolizing the corresponding step are reported between the boxes. Blue boxes are colourless metabolites needed for the β-carotene synthesis. Green boxes are colourless, possible catabolites of β-carotene. Red, orange, and yellow boxes indicate products responsible for the yellow colour of the cassava roots.

NCEDs, ZEP_1.0, CCDs, VDE, BCH1_2.2 encoding genes, among others, is negatively correlated in the same lines, but positively correlated with respect to the white genotypes P and R (Fig. 3A). We noticed that Carotenoid isomerase (CRTISO) and Geranylgeranyl pyrophosphate synthase (GGPS_2.0) encoding gene vectors were specifically pointing at (correlated with) genotype O and might be considered discriminant for carotenoid accumulation. However, other genes on the same side, such as those encoding BCH_2.1a and LUT5, which negatively influence β-carotene accumulation are also highly correlated with the yellow genotype O, and the same genes are inversely correlated to the yellow genotype B, placed opposite direction from O (Fig. 3A). Taken together this observation indicate the absence of a general relationship between carotenoid gene expression and the yellow cassava landraces analysed.

The level of gene expression was generally higher in leaf samples compared to root samples, and no obvious pattern of differential expression among genotypes was visible, both in leaves and in roots (Fig. 3B). Also, there was no general correspondance between carotenoid gene expression and root colour and/or β-carotene content in root samples. For instance, the expression of *PSY_1.1* was lower in deeper yellow landraces compared to white landraces P and R and pale yellow B and C (Fig. 3B). Based on the knowledge that, rather than expression level, a single nucleotide polymorphism in *PSY* of cassava and other species is often associated to coloured root phenotype (Welsch et al., 2010), in parallel, we performed a multisequence alignment of the assembled *PSY* transcripts of the analysed cassava genotypes and other selected species. In yellow landraces (D, O, I, J, L), this analysis revealed the presence of the aspartic acid, previously reported to be critical for β-carotene accumulation, replacing the alanine, still present in the three white genotypes (A, R, P; Fig. S1). However, an alanine was also present in the PSY sequence of the two pale yellow genotypes (B and C, Fig. S1), confirming that this mutation solely does not determine carotenoid

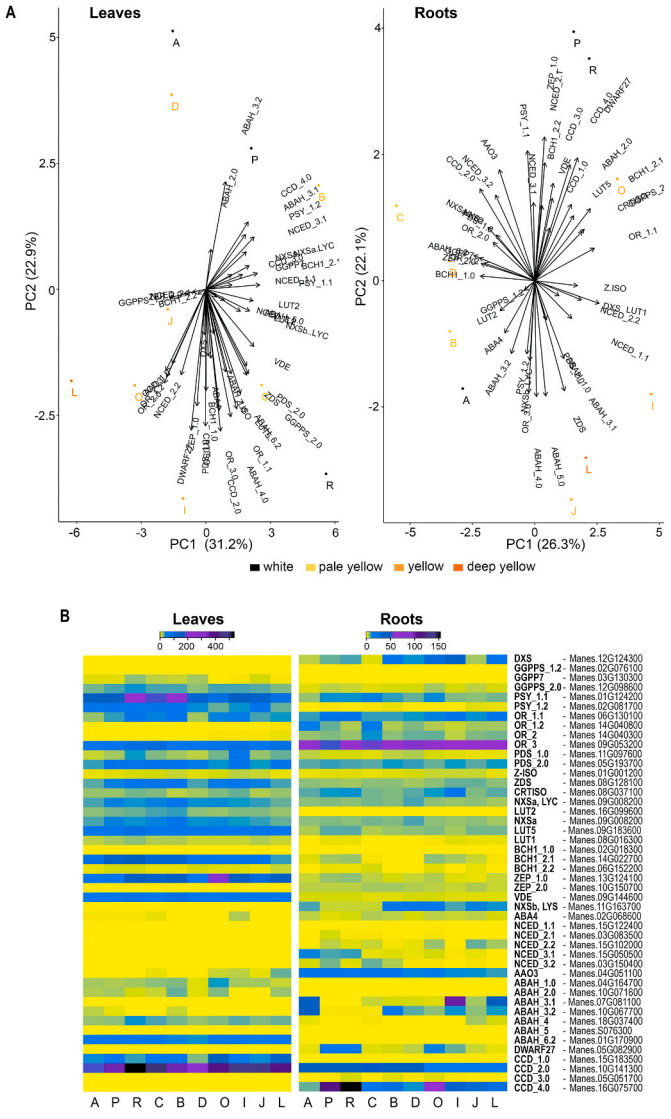

Fig. 3. Expression of carotenoid pathway genes in leaves and roots. **(A)** Principal components Biplot analysis of the significantly expressed genes in leaves (left panel) and roots (right panel). The separation between genotypes is visualized with PC1 vs PC2 **(B)** Heatmap of the significantly expressed genes in leaves and roots. Columns (genotypes) are ordered from left to right by increasing β -carotene content in roots. Rows (genes) are ordered from top to bottom according to the pathway sequence.

accumulation, which is therefore influenced by other factors as well.

As illustrated in Fig. 3B, there is no significant difference in *OR* gene expression among the analysed landraces that could be linked to higher accumulation of carotenoids in the yellow genotypes as previously observed in transgenic sweet potato (Kim et al., 2013). While *OR1.2* and *OR2* were generally more expressed in roots than leaves, all four homologs were found more expressed either in white or yellow landraces. Likewise, comparison of OR gene sequences between yellow and white cassava genotypes did not reveal any mutation (Sun et al., 2020) that could explain differential functions independent from the level of gene expression (Fig. S2). However, as higher OR proteins accumulation was previously detected in yellow genotypes compared to white genotypes, post-transcriptional regulation of OR genes rather than the expression level could influence a possible higher OR protein accumulation in yellow cassava (Jaramillo et al., 2022).

Genes encoding PDS, ZDS, and CRTISO, catalysing the dehydrogenation reactions that convert phytoene to lycopene (Nisar et al., 2015)

also showed similar or lower expression in yellow cultivars compared to white cultivars, except for landrace I where CRTISO-encoding gene was slightly higher expressed compared to all other genotypes (Fig. 3B). On the other hand, we observed higher expression of genes involved in the conversion of carotenoids to downstream products in the roots of white landraces. A reduced β-carotene cleavage in yellow landraces compared to white landraces could contribute to the higher β-carotene content. In particular, the expression of the carotenoid cleavage dioxygenases *CCD_2.0, CCD_4.0,* and *NCED_3*, and of the *DWARF27* gene, involved in strigolactone biosynthesis (Nisar et al., 2015), was higher in white genotypes (and in the two pale yellow lines for *CCD_2.0*) compared to the yellow landraces (Fig. 3B). This is consistent with previous results showing negative correlation between carotenoid and ABA accumulation mediated by lower expression of *NCED* and ABA catabolism genes in coloured fruits of orange and melon mutants (Chayut et al., 2021; Liu et al., 2021). In the metabolite analysis performed in this study, ABA levels were below the detection limit and we could not compare ABA accumulation in the landrace's roots. However, both the reduced expression of NCED and a GEM encoding gene (Manes.18G061700, Fig. 3B and Table S4) are involved in the ABA signalling pathway (Mauri et al., 2016), and the higher expression of *ABAH* genes (Fig. 3) in yellow genotypes compared to white landraces, suggest an increase in ABA catabolism, which warrants further investigations. Interestingly, the evidence that *NCED_3.1* was specifically expressed at a higher level in the three white genotypes (A, P, R) compared to all yellow genotypes (Fig. 3B) suggests that regulation of this gene could generally influence β-carotene accumulation in cassava. However, expression levels of *CCD_4.0* and *DWARF27,* are high also in the yellow genotype O and I compared to other yellow cultivars. Taken together these results confirm that accumulation of carotenoids can be linked to both the rate of biosynthesis and the rate of catabolism and degradation (Hannoufa and Hossain, 2012).

3.4. Starch metabolism-related gene expression

To investigate in more detail the correlation between dry matter and carotenoid content in cassava, we analysed the expression of the major genes supporting starch metabolism (Pfister and Zeeman, 2016). KEGG analysis demonstrated that starch biosynthesis has a common precursor metabolite with the carotenoid biosynthesis: glyceraldehyde-3-phosphate (GA3P). In the KEGG terpenoid backbone pathway (pathway mesc00900), GA3P initiates the MEP/DOXP pathway converting GA3P into Geranylgeranyl-PP, which is the precursor of the β-carotene synthesis in the carotenoid pathway (KEGG pathway mesc00609). On the other hand, in the KEGG glycolysis pathway (pathway mesc00010), GA3P is converted into α-D-Glucose-1P, which is a precursor of starch synthesis (KEGG pathway mesc00500). However, comparison of the expression profiles of genes involved in glycolysis in the roots of cassava landraces, did not reveal specific correlation patterns clearly differentiating low and high starch content genotypes (Fig. S3). Interestingly, the two white genotypes P and R, with intermediate starch content, share almost identical expression profiles, which, however, are different from the profile of the white high starch genotype A.

Furthermore, we observed that the expression of certain glycolysis-related genes was correlated with the β-carotene content of the cassava landraces. The chloroplastic fructose-bisphosphate aldolase (chlFBA1, EC:4.1.2.13, Manes.04G083800), which catalyses the conversion of GA3P to β-D-Fructose-1,6P_2, has highest expression in high β-carotene-accumulating genotypes (I, J, and L), suggesting that in these genotypes the glycolysis pathway might be disfavoured (Fig. S3). Additionally, the lower expression of the chloroplastic phosphoglucomutase (chlPGM, EC:5.4.2.2, Manes.10G046600), converting α-D-Glucose-6P into α-D-Glucose-1P at the last step before α-D-Glucose-1P synthesis, might hinder the production of the precursor of the starch pathway in the yellow genotypes O, I, J, and L compared to white and pale yellow genotypes. However, high expression of the cytoplasmic cytFBA8 (EC:5.4.2.2, Manes.07G000100) in the yellow genotypes O, I, J, L, might compensate the reduced chlFBA1 activity and/or support glycolysis in different cellular compartments (Anoman et al., 2016).

Furthermore, we noticed that the expression levels of certain genes involved in starch biosynthesis were correlated to the β-carotene content of the genotypes. In the high β-carotene genotypes (D, O, I, J, L), two genes were expressed at higher level than in the pale yellow and white genotypes: the 6-phosphogluconate dehydrogenase (EC:1.1.1.44, Manes.01G105600), involved in the synthesis of ribulose-5-phosphate needed for nucleic acid and fatty acid synthesis, and a pyrophosphatase (EC:3.6.1.1, Manes.S022600), probably providing phosphate to the ATP/ADP antiporter exchanging ADP and ATP between amyloplast and cytoplasm (Fig. S3).

3.5. Differential gene expression in storage roots

To gain more insight into the transcriptomic signature of the different landraces and to identify potential candidate genes involved in regulation of carotenoid accumulation, we selected genes not significantly differentially expressed (p-value >0.05) among the 3 white genotypes (A, P, R), but significantly differentially expressed (p-value ≤0.05, foldchange ± 1.5) between each of the three white and at least 5 of the 6 yellow genotypes (C, D, O, I, J, L). Based on these conditions, we identified 47 genes, most of which upregulated in the yellow landraces with expression levels positively related with the β-carotene content of the different genotypes (Fig. 4). Two major clusters were observed in the heatmap, the top one being characterised by genes mainly annotated in plant defence and development. Among them, Cystatin B, a cysteine proteinase inhibitor known for its role in drought tolerance (Martinez et al., 2012; Velasco-Arroyo et al., 2018), was upregulated in yellow genotypes compared to white genotypes. Two genes in this cluster and a third one in the other main cluster (Fig. 4, bottom) encode members of the RING/U-box protein superfamily mediating protein–protein interactions and involved in the protein ubiquitination pathway (Trujillo, 2018). Four genes were identified as potential regulators of

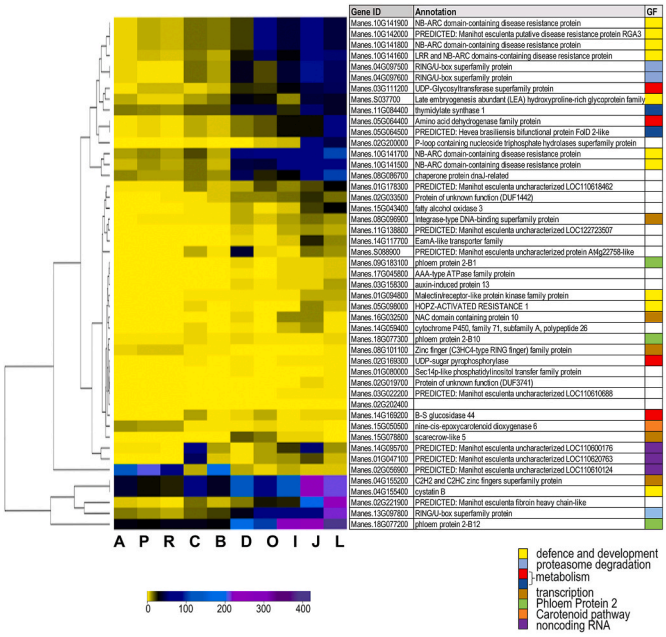

Fig. 4. Heatmap of genes differentially expressed between white and yellow genotypes. Columns (genotypes) are ordered from left to right by increasing β-carotene content in roots. Rows (genes) are clustered according to the expression profile. For each gene the gene ID, the gene description and the gene function (GF) according to the gene ontology are reported.

transcription. Three of them were upregulated in yellow landraces among which, one encoding a C2H2-type domain-containing protein (Manes.04G155200) is a member of the Zinc-finger protein family, known for its important roles in plant growth and development, and responses to biotic and abiotic stresses (Han et al., 2020). Another member of the Zinc finger family protein (Manes.08G101100), containing a cysteine-rich domain with potential role in the ubiquitination pathway (Trujillo, 2018), was slightly more expressed in the white genotypes compared to the yellow ones. Importantly, the expression of these RNA regulators is correlated to carotenoid accumulation in melon fruits (Chayut et al., 2021).

Several genes involved in secondary metabolism, among which a UDP-glycosyltransferase (Manes.03G111200) and two in the one-carbon metabolism (Manes.11G084400, Manes.05G064500) were upregulated in yellow genotypes (Fig. 4).

Interestingly, three genes (Manes.09G183100, Manes.18G077300, Manes.18G077200), annotated as phloem protein 2-B, were more expressed in yellow than in white landraces (Fig. 4). These proteins host the F-box domain and are involved in long-distance phloematic transport of various molecules including proteins, RNAs and photoassimilates (Dinant et al., 2003). A member of this family in Arabidopsis, AtPP2-B11, was reported to modulate ABA signalling (Cheng et al., 2017).

Two genes upregulated (Manes.14G095700 and Manes.01G047100) and one downregulated (Manes.02G056900) in yellow landraces, particularly in the pale yellow ones, were predicted as non-coding RNA (ncRNA) and are potentially involved in regulation of gene expression (Hajieghrari and Farrokhi, 2021). The only member of the carotenoid pathway identified in this study as significantly higher expressed in all white compared to all yellow genotypes, the carotenoid cleavage dioxygenases *NCED_3.1* (Manes.15G050500), appeared also in this gene expression analysis confirming its potential role in regulating ABA biosynthesis thus favouring accumulation of metabolites upstream in the carotenoid biosynthesis pathway (Fig. 2).

3.6. Functional enrichment analysis

We performed GO-based gene enrichment analysis to identify functions and processes potentially related to carotenoid biosynthesis and accumulation. For this analysis, we selected genes not significantly differentially expressed (p-value >0.05) among the three white genotypes (A, P, R) and compared the average expression value of these genes with each yellow genotype (C, B, D, O, G, J, I, L). This analysis produced eight lists of significantly (p-value <0.05) differentially expressed genes, one for each of the eight yellow genotype (Table S5). The obtained eight lists of genes were then analysed for GO enrichment (p-value <0.05) by three independent analyses including: (i) the full gene list, (ii) the subset of downregulated genes or (iii) the subset of upregulated gene (Fig. 5). Catalytic activities were generally overrepresented in yellow landraces. In particular, RNA polymerase, peptidase and hydrolase activities were upregulated in genotypes with high β-carotene content while genes involved in other activities (e.g. glycosyltransferase activity), were downregulated in pale genotypes (C and B) and in genotype G.

We found an overrepresentation of peptidase complex linked to overexpression of proteasome-related genes and downregulation of genes active at the plastidial level in the three genotypes with the highest β-carotene content. Consistently, in the same three genotypes and partly in the genotype G, we observed an enrichment of downregulated photosynthesis-related genes and overrepresentation of certain proteasomal catabolism terms (Fig. 5). Interestingly, genes active in the Golgi apparatus and involved in cell wall biogenesis and organisation were specifically downregulated in both genotypes C and G, with low and intermediate levels of β-carotene content, respectively.

To further investigate these findings, we used an additional approach of analysis performing a hierarchical clustering of significantly expressed genes (FPKM >3) according to expression profile similarity

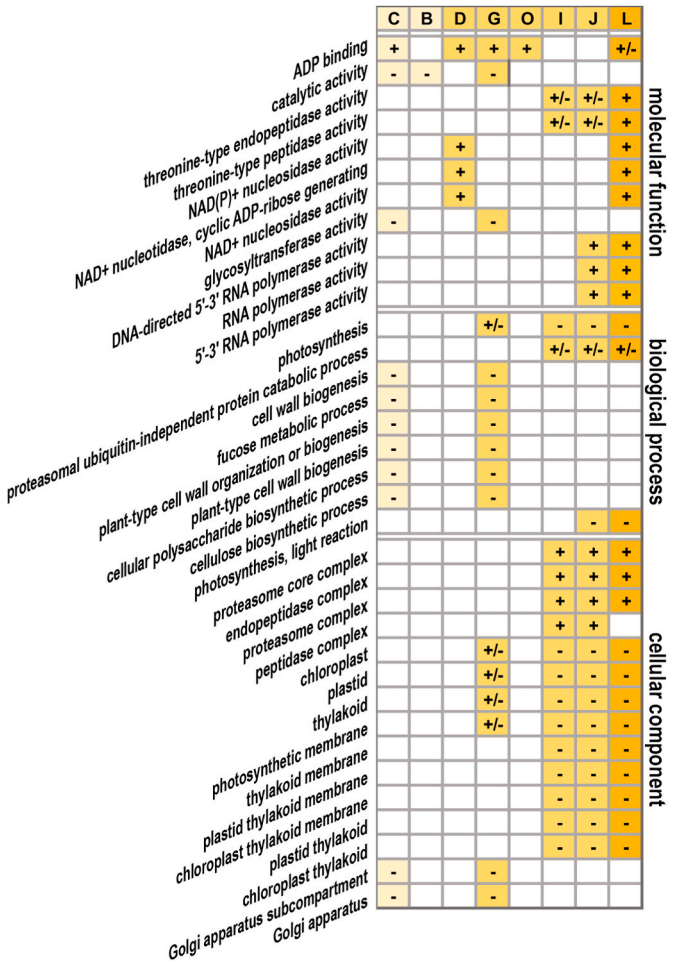

Fig. 5. Gene Ontology (GO) analysis of the differentially expressed genes between white and yellow genotypes. Each column shows significantly over-represented GO terms (rows) of each yellow genotype compared to the average expression value of white genotypes. The columns are ordered from left to right by increasing β-carotene content in roots with the corresponding colour code. The plus sign indicates GO terms of upregulated genes; the minus sign GO terms of downregulated genes; and the plus/minus sign GO terms of genes which are either up- or downregulated. The terms are grouped into the three domains molecular function, biological process, and cellular component.

within the 10 investigated genotypes (EXPANDER, Hait et al., 2019). This generated 32 gene expression clusters for root samples, among which we selected the three clusters better distinguishing yellow and white genotypes (Fig. 6, Table S6). Cluster 1 contained transcripts lowly expressed in the three genotypes with the highest β-carotene content (J, I, L) while cluster 2 contained genes upregulated only in white genotypes (A, P, R). Cluster 3, the second largest cluster of all as for number of genes, included genes with low expression in the white and pale yellow genotypes (A, P, R, C, B), and with higher expression level in the yellow genotypes. Interestingly, this cluster contains the same group of genotypes hosting the alanine to aspartic acid substitution in the *psy1* gene (Fig. S1). For more accurate analysis of the large Cluster 3 expression profile, we clustered only the 1583 genes there included and selected four out of the 14 obtained sub-clusters better distinguishing yellow and white genotypes (Fig. 6, Table S6). The genes in each cluster were further analysed using the GO enrichment analysis webtool gProfiler (Raudvere et al., 2019). The enriched GO terms associated to genes downregulated in the carotenoid-rich genotypes I, J and L, confirmed the overrepresentation of catalytic activities (Fig. 6, cluster 1), while genes with low expression in all yellow genotypes, were enriched for functions associated to plastidic cellular components and photosynthesis

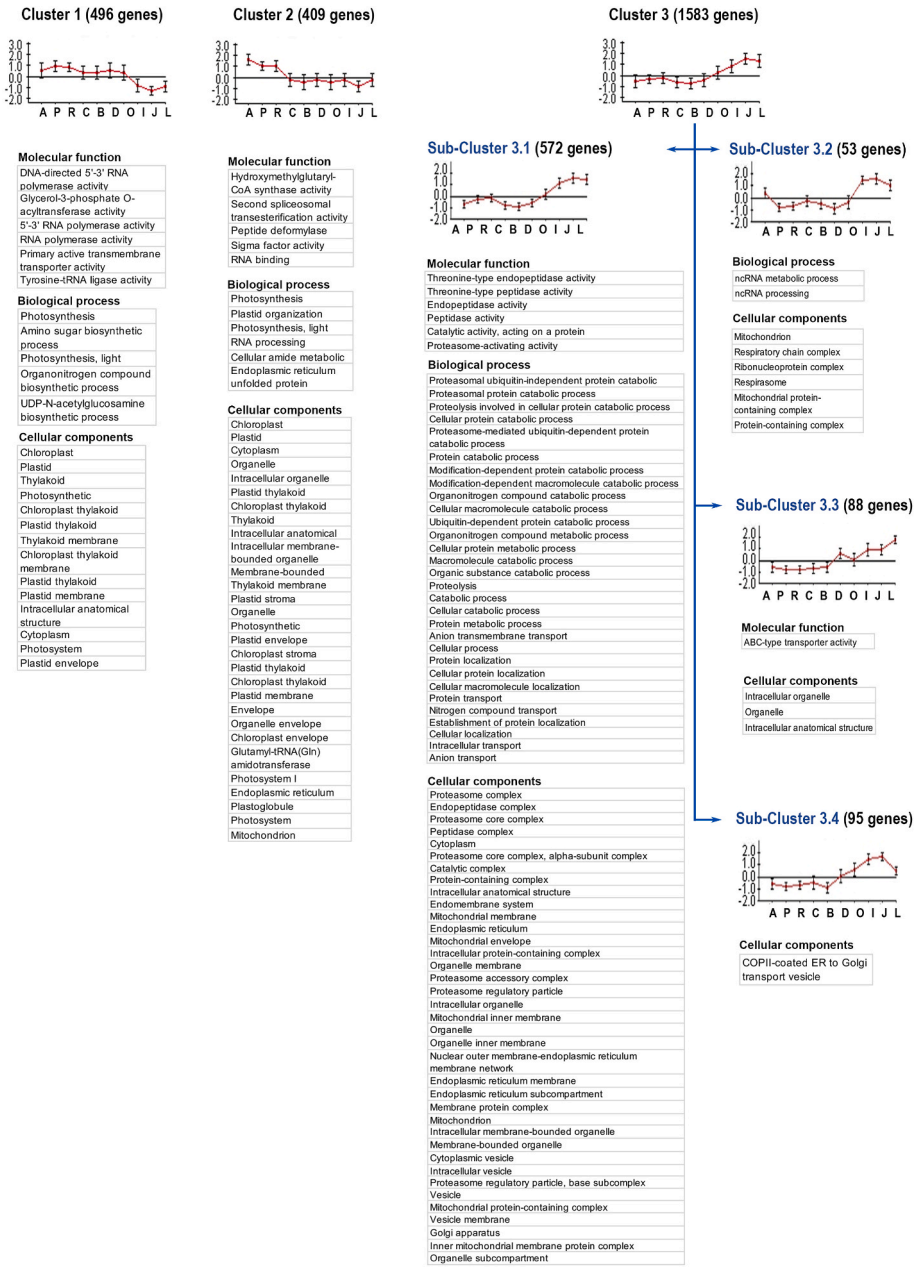

Fig. 6. Expander expression profile clusters with corresponding significant overrepresented GO annotations. Root genes significantly differentially expressed genes in roots were clustered according to the expression profile of the different genotypes. Cluster 1, 2 and 3 are primary clusters. Cluster 3.1, 3.2, 3.3 and 3.4 are secondary clusters obtained by clustering again Cluster 3 genes. GO terms below the cluster graphs are significantly over-represented term of the respective cluster gene pools.

(Fig. 6, cluster 1 and 2). Downregulation of genes involved in transcriptional maintenance of the photosynthetic apparatus has been previously reported for carotenoid-poor melon mutants, where it is correlated to the arrest of the carotenoid pathway mediated by *CRITSO* (Chayut et al., 2021), a gene that we find more expressed in the yellow line I (Fig. 3B). This regulation of expression in non-photosynthetic cassava yellow roots compared to white roots suggests that the lack of connection between carotenoid accumulation and photosynthesis expression machinery might be conserved and independent of the photosynthetic activity of the specific tissue. Indeed, it has been recently shown that stimulation of carotenoid biosynthesis in the absence of photosynthetic competence, triggers a major gene expression reprogramming, similar to the one observed in our comparison, to stimulate the differentiation of chromoplasts from chloroplasts (Llorente et al., 2020).

Genes low expressed in the yellow genotypes are associated to an overrepresentation of plastid membrane GO terms (Figs 5, 6). This observation is consistent with previous studies reporting that higher carotenoid accumulation is coupled to changes in the anatomical structure of plastids, resulting in increased carotenoid storage capacity (Paolillo Jr. et al., 2004; Yu et al., 2008; Guan et al., 2019). The differential representation of photosynthetic, light reactions, and plastid related functions between yellow and white genotypes implies that enhanced carotenoid accumulation might be triggered by changes in cell structure allowing for storage of more carotenoids. This also suggests further investigations of plastid-to-nucleus retrograde signalling that may somehow regulate CBP gene expression (Nott et al., 2006).

In the sub-cluster 3.1, the largest in gene number, genes are low expressed in white and pale-yellow genotypes, whereas increasing level of gene expression positively correlated to the root β-carotene content in the yellow genotypes (Fig. 6). In this sub-cluster, protein catabolic processes (biological process) were overrepresented together with

protein transport and localization while the enriched molecular functions were all related to catalytic activity of proteins. Consistently, overrepresented terms of the cellular components are peptidase, proteasome complexes, and protein-containing complexes in general, located in different organelles prevalently at the membranes and in vesicles. Interestingly, GO term enrichment analyses did not detect terms directly associated to carotenoid biosynthesis, transportation or regulation (Figs. 5 and 6).

Finally, worthy of note among the other sub-clusters, overrepresented biological processes of sub-cluster 3.2 are related to ncRNA in line with the overexpression of ncRNA reported above (Fig. 4).

3.7. Metabolite analysis of storage roots

Storage roots of six of the genotypes analysed (A, C, G, O, I, L), spanning from white to deep yellow (Table 1), were subject to metabolite analysis by gas chromatography mass spectrometry (GC-MS) and liquid chromatography mass spectrometry (LC-MS). 149 metabolites were identified in total whereas 1378 features remained unmatched in the LC-MS data (Figs. S4 and S5). In both LC-MS and GC-MS analyses the white and pale-yellow landraces clustered together, albeit quite high variance between biological replicates was observed in yellow genotypes (Fig. 7). To assess the principal sources of variation in metabolites and gene expression data sets, we also performed a multi-omics factor analysis (MOFA) which, despite the predominance of the GC-MS data, revealed interesting relationships among data sets. In particular, we focused on factors 1 and 8 that showed the best association to yellow and white genotypes (Fig. S6A). While factor 1 is positively influenced by yellow genotypes (L, J, I, O) and negatively influenced by white and pale-yellow genotypes (A, P, R, B, C), factor 8 is positively influenced by mainly white genotypes (A, P, R) and negatively influenced by yellow and pale-yellow genotypes (L, J, G, D, B, C; Fig. S6B).

Amino acids asparagine, glutamic acid (Glu), isoleucine, lysine, methionine, serine and threonine were more abundant in white rooted genotypes (Figs. S4 and S5). This observation is in line with previous findings, comparing metabolites in cassava lines with varying β-carotene content, which showed higher abundance of several amino acids in the white-fleshed cassava compared to the yellow-fleshed (Drapal et al., 2019; Xiao et al., 2021).

Interestingly, in the MOFA Factor 1 active in carotenoid rich genotypes (Fig. S6), the reduced accumulation of Glu was connected with the increased expression of a gene encoding a Rubber elongation factor protein (REF, Manes.08G117800, Fig. S7) having amyloid properties and directly acting on the isoprenoid biosynthetic pathway (Berthelot et al., 2012). Factor 1, also tagged the protease inhibitor cystatin B gene (Manes.02G134700, Fig. 4 and Fig. S7) that together with the above-mentioned features seem to contribute to carotenoid accumulation in the genotypes analysed.

A positive correlation of arginine with the high carotene trait has

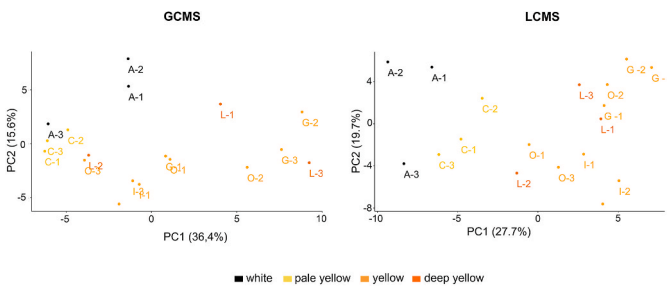

Fig. 7. Metabolites relationships among cassava genotypes. Principal component analysis (PCA) of metabolites analysed by gas chromatography mass spectrometry (GC-MS, left panel) and liquid chromatography mass spectrometry (LC-MS, right panel). The clustering of the replicates of the different genotypes and the separation between genotypes is visualized with PC1 vs PC2.

been reported (Drapal et al., 2019), however, in our study, arginine was found to be abundant in both white line A and deep yellow line L (Fig. S5). Indeed, a study of carotenoids and primary metabolites of targeted genotypes showed positive correlation between the levels of amino acids, including γ-aminobutyric acid (GABA) and glutamine (Gln), and carotenoids (Park et al., 2017).

In our study, higher accumulation of glutamine was observed in the yellow genotype G, and higher accumulation of proline and GABA amino acids, both related to stress responses (Signorelli et al., 2021), was observed both in the white line A and the yellow line G (and of GABA in O) compared to other genotypes (Fig. S4). In MOFA Factor 8, associated to the white and the I and O genotypes (Fig. S6), Gln accumulation is negatively correlated with the expression of a Thioredoxin-encoding gene (Manes.13G141400, Fig. S7). As suppression of thioredoxin expression induces an increase of carotenoid content in cyanobacteria (López-Maury et al., 2018), the role of this gene in carotenoid accumulation of genotypes I and O is worth further investigations.

Overall comparison of our and previous results could not identify a clear connection between specific amino acids and carotenoid root content. However, amino acids contribute to quality and the different tastes of fruits (Tieman et al., 2006)Shiraishi et al., 2009). Thus, depending on their relative abundance in white-rooted cassava cultivars, threonine, serine, and glutamic acid, known to contribute to sweetness, may improve the taste and flavour of cassava while isoleucine and arginine contribute to the bitterness. Ornithine, which is also known to contribute to bitterness in fruits, was found to be abundant in both white cultivar A and deep-yellow cultivar L, albeit in both cultivars it is less abundant compared of the other amino acids.

Contrary to previous findings (Wang et al., 2021; Xiao et al., 2021), we detected larger amount of organic acids such as fumaric acid and malic acid in the yellow lines compared to white lines, and the largest amount in genotype G compared to the other yellow lines (Figs. S4 and S5). Interestingly, malic acid and aspartic acid are influencing carotenoid-rich genotypes and are positively correlated to both cystatin B and REF gene expression (MOFA Factor 1, Fig. S7).

Pipecolic acid, a known signalling molecule critical for plant immunity (Návarová et al., 2013), was more abundant in yellow lines G and L compared to other cultivars (Figs. S4 and S5).

Salicylic acid (SA) and phenylalanine, a substrate of the phenylalanine ammonia lyase (PAL) pathway for SA biosynthesis (Chen et al., 2009) were both in high abundance in the white cultivar A compared to the yellow cultivars. Proline and betaine also accumulated more in white cultivar A and pale-yellow cultivar C and are negatively influencing MOFA Factor 1 (Fig. S7). This is in line with previous findings showing that SA positively influences the abundance of osmolytes proline and betaine (Arif et al., 2020). Interestingly, L-carnitine, another osmolyte active in the ABA signalling pathway (Jacques et al., 2018), was most abundant in cultivars A and C. Therefore, it can be hypothesised that together with the expression of *NCED*, *GEM* and *ABAH* genes (Fig. 3 and Table S4) it could influence β-carotene accumulation in cassava.

Consistently with previous results showing a positive relationship between sugars and carotenoids (Park et al., 2017), the GC-MS analysis revealed that sugars, such as sucrose, fructose and glucose were more abundant in yellow cultivars compared to white cultivars (Fig. S4). Our findings also suggest a positive relationship between abundance of sugars and organic acids in the yellow cultivars (Fig. S4, S5 and S7) as previously observed between accumulation of monosaccharides and the organic acids of the tricarboxylic acid (TCA) cycle among cassava varieties (Drapal et al., 2019). The abundance of reducing sugars in yellow cultivars may influence the taste and sweetness of cassava and cassava products since high culinary quality is thought to correlate with higher levels of monosaccharides in cassava cultivars (Bechoff et al., 2018; Drapal et al., 2019).

4. Conclusions

In this study, we expanded our knowledge on the molecular basis of carotenoid accumulation in cassava storage roots, providing original information about the accumulation of metabolites and the expression of genes involved in this process in field-grown landraces, which differ in root colour and β-carotene content. Our results confirm the existence of several putative regulators of carotenoid accumulation even in a single tissue of a single species like cassava roots. However, the significant number of landraces tested, and the comparative and integrated analysis of gene expression and metabolite accumulation provided new insights into molecular mechanisms specifically active in yellow rooted cassava genotypes. Our targeted analysis revealed that carotenoid-related gene expression alone cannot exhaustively and conclusively explain root colour and carotenoid composition of the analysed genotypes. Nevertheless, we found that expression of *NCED* (and other) candidate genes and reduced accumulation of L-carnitine were highly correlated with the root colour trait. Therefore, this gene is a strong candidate to be used as a molecular marker for root colour and β-carotene content and could be implemented in expression marker-assisted selection to increase the efficiency of cassava breeding programs by the early selection of new genotypes with high carotenoid root content. Furthermore, *NCED*, less expressed in all yellow rooted landraces, is also a promising target to be suppressed with biotechnological approaches (e.g. RNAi, CRISPR-Cas9) to stimulate provitamin A accumulation in suitable cassava genotypes.

While cassava genotypes with reasonably high content of starch and β-carotene are preferred candidates to engineer vitamin A fortified plants, the potential application of the identified candidate genes and metabolites to engineer cassava plants with high carotenoid and starch needs to be validated in the specific genotype to be improved.

Our findings suggest the involvement of several protein-related catabolic processes and organonitrogen catabolic processes enriched in yellow cultivars. Among them, increased ABA catabolism supported by the activation of related genes and metabolites indicates a possible negative correlation between ABA and carotenoids accumulation in cassava roots that deserves further study. Likewise, the underrepresentation of photosynthetic, light reactions and plastid membrane functions in yellow varieties compared to the other lines, suggests further investigations of plastid-to-nucleus retrograde signalling that may somehow regulate gene expression in the carotenoid biosynthesis pathway.

Author contribution

Conceptualization: PO, AG, EA, LS; Methodology: PO, AG, EA, LS; Validation: PO, LS; Formal analysis: PO, AG; Investigation: PO; Resources: EA, AG, LS; Writing-Original draft Preparation: PO, LS; Writing-Review and editing: all authors; Supervision: LS. AG, EA; Funding: LS. AG, EA.

Funding

This work was supported by the Vetenskapsrådet [grant number 2016-05703] and the Vetenskapsrådet: [grant number 2018-04660].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We thank Drs. Elizabeth Parkes and Ismail Y Rabbi for providing high carotenoid cassava genotypes; Peter Illuebey for providing cassava stem cuttings, field labor, and management throughout the experiment; Anetor Omonuwa and Ayo Akinwumi for providing field support for planting, sampling, and harvesting. We thank Linnea Stridh and Thuong Ha for assistance with samples for metabolite analysis. Swedish Metabolomics Centre Umeå, Sweden (www.swedishmetabolomicscentre.se) is acknowledged for metabolic profiling by GCMS and LCMS.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.plaphy.2023.107713.

References

Andrews, S., 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Anoman, A.D., Flores-Tornero, M., Rosa-Telléz, S., Muñoz-Bertomeu, J., Segura, J., Ros, R., 2016. The specific role of plastidial glycolysis in photosynthetic and heterotrophic cells under scrutiny through the study of glyceraldehyde-3-phosphate dehydrogenase. Plant Signal. Behav. 11, e1128614 https://doi.org/10.1080/15592324.2015.1128614.

Argelaguet, R., Velten, B., Arnol, D., Dietrich, S., Zenz, T., Marioni, J.C., et al., 2018. Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. Mol. Syst. Biol. 14, 1–13. https://doi.org/10.15252/msb.20178124.

Arif, Y., Sami, F., Siddiqui, H., Bajguz, A., Hayat, S., 2020. Salicylic acid in relation to other phytohormones in plant: a study towards physiology and signal transduction under challenging environment. Environ. Exp. Bot. 175 https://doi.org/10.1016/j.envexpbot.2020.104040.

Barrero, J.M., Rodríguez, P.L., Quesada, V., Alabadí, D., Blázquez, M.A., Boutin, J.P., et al., 2008. The ABA1 gene and carotenoid biosynthesis are required for late skotomorphogenic growth in Arabidopsis thaliana. Plant Cell Environ. 31, 227–234. https://doi.org/10.1111/j.1365-3040.2007.01759.x.

Bechoff, A., Tomlins, K., Fliedel, G., Becerra Lopez-lavalle, L.A., Westby, A., Hershey, C., et al., 2018. Cassava traits and end-user preference: relating traits to consumer liking, sensory perception, and genetics. Crit. Rev. Food Sci. Nutr. 58, 547–567. https://doi.org/10.1080/10408398.2016.1202888.

Berthelot, K., Lecomte, S., Estevez, Y., Coulary-Salin, B., Bentaleb, A., Cullin, C., et al., 2012. Rubber elongation factor (REF), a major allergen component in hevea brasiliensis latex has amyloid properties. PLoS One 7, 1–12. https://doi.org/10.1371/journal.pone.0048065.

Carluccio, A., David, L., Claußen, J., Sulle, M., Adeoti, S., T, A., et al., 2022. Set up from the beginning: the origin and early development of cassava storage roots. Plant Cell Environ. https://doi.org/10.1111/pce.14300.

Carvalho, L.J.C.B., Agustini, M.A.V., Anderson, J.V., Vieira, E.A., de Souza, C.R.B., Chen, S., et al., 2016. Natural variation in expression of genes associated with carotenoid biosynthesis and accumulation in cassava (Manihot esculenta Crantz) storage root. BMC Plant Biol. 16, 1–23. https://doi.org/10.1186/s12870-016-0826-0.

Ceballos, H., Davrieux, F., Talsma, E.F., Belalcazar, J., Chavarriaga, P., Andersson, M.S., 2017. Carotenoids in cassava roots. Carotenoids. https://doi.org/10.5772/intechopen.68279.

Ceballos, H., Morante, N., Sanchez, T., Ortiz, D., Aragon, I., Chavez, A.L., et al., 2013. Rapid cycling recurrent selection for increased carotenoids content in cassava roots. Crop Sci. 53, 2342–2351. https://doi.org/10.2135/cropsci2013.02.0123.

Chawade, A., Alexandersson, E., Levander, F., 2014. Normalyzer: a tool for rapid evaluation of normalization methods for omics data sets. J. Proteome Res. 13, 3114–3120. https://doi.org/10.1021/pr401264n.

Chayut, N., Yuan, H., Saar, Y., Zheng, Y., Sun, T., Zhou, X., et al., 2021. Comparative transcriptome analyses shed light on carotenoid production and plastid development in melon fruit. Hortic. Res. 8 https://doi.org/10.1038/s41438-021-00547-6.

Chen, Z., Zheng, Z., Huang, J., Lai, Z., Fan, B., 2009. Biosynthesis of salicylic acid in plants. Plant Signal. Behav. 4, 493–496. https://doi.org/10.4161/psb.4.6.8392.

Cheng, C., Wang, Z., Ren, Z., Zhi, L., Yao, B., Su, C., et al., 2017. SCFAtPP2-B11modulates ABA signaling by facilitating SnRK2.3 degradation in Arabidopsis thaliana. PLoS Genet. 13 https://doi.org/10.1371/journal.pgen.1006947.

Chiu, M., Dillon, A., Watson, S., 2016. Vitamin A deficiency and xerophthalmia in children of a developed country. J. Paediatr. Child Health 52, 699–703. https://doi.org/10.1111/jpc.13243.

D'Andrea, L., Rodriguez-Concepcion, M., 2019. Manipulation of plastidial protein quality control components as a new strategy to improve carotenoid contents in tomato fruit. Front. Plant Sci. 10, 1071. https://doi.org/10.3389/fpls.2019.01071.

Darnton-Hill, I., 2019. Public health aspects in the prevention and control of vitamin deficiencies. Curr. Dev. Nutr. 3, 1–14. https://doi.org/10.1093/cdn/nzz075.

Diamanti, K., Cavalli, M., Pan, G., Pereira, M.J., Kumar, C., Skrtic, S., et al., 2019. Intra- and inter-individual metabolic profiling highlights carnitine and lysophosphatidylcholine pathways as key molecular defects in type 2 diabetes. Sci. Rep. 9 https://doi.org/10.1038/s41598-019-45906-5.

Dinant, S., Clark, A.M., Zhu, Y., Vilaine, F., Palauqui, J.C., Kusiak, C., et al., 2003. Diversity of the superfamily of phloem lectins (Phloem protein 2) in angiosperms. Plant Physiol. 131, 114–128. https://doi.org/10.1104/pp.013086.

Drapal, M., Barros De Carvalho, E., Ovalle Rivera, T.M., Becerra Lopez-Lavalle, L.A., Fraser, P.D., 2019. Capturing biochemical diversity in cassava (Manihot esculenta Crantz) through the application of metabolite profiling. J. Agric. Food Chem. 67, 986–993. https://doi.org/10.1021/acs.jafc.8b04769.

Esuma, W., Rubaihayo, P., Pariyo, A., Kawuki, R., Wanjala, B., Nzuki, I., et al., 2012. Genetic diversity of provitamin A cassava in Uganda. J. Plant Stud. 1 https://doi.org/10.5539/jps.v1n1p60.

Fan, S., Kind, T., Cajka, T., Hazen, S.L., Tang, W.H.W., Kaddurah-Daouk, R., et al., 2019. Systematic error removal using random forest for normalizing large-scale untargeted lipidomics data. Anal. Chem. 91, 3590–3596. https://doi.org/10.1021/acs.analchem.8b05592.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., et al., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652. https://doi.org/10.1038/nbt.1883.

Guan, L., Ma, X., Zhou, X., Tan, B., Wang, Z.Y., 2019. An optimized method to obtain high-quality RNA from cassava storage root. 3 Biotech 9, 118. https://doi.org/10.1007/s13205-019-1608-9.

Guo, Y.-J., Luo, X.-L., Wei, M.-G., Liu, Z.-L., Fan, W.-J., Zhai, R.-N., et al., 2017. RNA sequence analysis of cassava varieties with high-starch content using de novo assembly. J. Plant Growth Regul. 37, 517–529. https://doi.org/10.1007/s00344-017-9749-8.

Hait, T.A., Maron-Katz, A., Sagir, D., Amar, D., Ulitsky, I., Linhart, C., et al., 2019. The EXPANDER integrated Platform for transcriptome analysis. J. Mol. Biol. 431, 2398–2406. https://doi.org/10.1016/j.jmb.2019.05.013.

Hajieghrari, B., Farrokhi, N., 2021. Plant RNA-mediated gene regulatory network. Genomics 114, 409–442. https://doi.org/10.1016/j.ygeno.2021.12.020.

Han, G., Lu, C., Guo, J., Qiao, Z., Sui, N., Qiu, N., et al., 2020. C2H2 zinc finger proteins: master regulators of abiotic stress responses in plants. Front. Plant Sci. 11, 1–13. https://doi.org/10.3389/fpls.2020.00115.

Hannoufa, A., Hossain, Z., 2012. Regulation of carotenoid accumulation in plants. Biocatal. Agric. Biotechnol. 1, 198–202. https://doi.org/10.1016/j.bcab.2012.03.004.

Imakumbili, M.L.E., Semu, E., Semoka, J.M.R., Abass, A., Mkamilo, G., 2021. Managing cassava growth on nutrient poor soils under different water stress conditions. Heliyon 7, e07331. https://doi.org/10.1016/j.heliyon.2021.e07331.

Jacques, F., Rippa, S., Perrin, Y., 2018. Physiology of L-carnitine in plants in light of the knowledge in animals and microorganisms. Plant Sci. 274, 432–440. https://doi.org/10.1016/j.plantsci.2018.06.020.

Jaramillo, A.M., Sierra, S., Chavarriaga-Aguirre, P., Castillo, D.K., Gkanogiannis, A., López-Lavalle, L.A.B., et al., 2022. Characterization of cassava ORANGE proteins and their capability to increase provitamin A carotenoids accumulation. PLoS One 17, e0262412. https://doi.org/10.1371/journal.pone.0262412.

Kim, H.S., Ji, C.Y., Lee, C.J., Kim, S.E., Park, S.C., Kwak, S.S., 2018. Orange: a target gene for regulating carotenoid homeostasis and increasing plant tolerance to environmental stress in marginal lands. J. Exp. Bot. 69, 3393–3400. https://doi.org/10.1093/jxb/ery023.

Kim, S.H., Ahn, Y.O., Ahn, M.J., Jeong, J.C., Lee, H.S., Kwak, S.S., 2013. Cloning and characterization of an Orange gene that increases carotenoid accumulation and salt stress tolerance in transgenic sweetpotato cultures. Plant Physiol. Biochem. 70, 445–454. https://doi.org/10.1016/j.plaphy.2013.06.011.

Liu, H., Able, A.J., Able, J.A., 2021. Priming crops for the future: rewiring stress memory. Trends Plant Sci. xx, 1–18. https://doi.org/10.1016/j.tplants.2021.11.015.

Llorente, B., Torres-Montilla, S., Morelli, L., Florez-Sarasa, I., Matus, J.T., Ezquerro, M., et al., 2020. Synthetic conversion of leaf chloroplasts into carotenoid-rich plastids reveals mechanistic basis of natural chromoplast development. Proc. Natl. Acad. Sci. U. S. A 117, 21796–21803. https://doi.org/10.1073/pnas.2004405117.

López-Maury, L., Heredia-Martínez, L.G., Florencio, F.J., 2018. Characterization of trxc, an atypical thioredoxin exclusively present in cyanobacteria. Antioxidants 7. https://doi.org/10.3390/antiox7110164.

Martinez, M., Cambra, I., Gonzalez-Melendi, P., Santamaria, M.E., Diaz, I., 2012. C1A cysteine-proteases and their inhibitors in plants. Physiol. Plantarum 145, 85–94. https://doi.org/10.1111/j.1399-3054.2012.01569.x.

Mauri, N., Fernández-Marcos, M., Costas, C., Desvoyes, B., Pichel, A., Caro, E., et al., 2016. GEM, a member of the GRAM domain family of proteins, is part of the ABA signaling pathway. Sci. Rep. 6, 1–11. https://doi.org/10.1038/srep22660.

Metsalu, T., Vilo, J., 2015. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. Nucleic Acids Res. 43, W566–W570. https://doi.org/10.1093/nar/gkv468.

Morris, W.L., Ducreux, L., Griffiths, D.W., Stewart, D., Davies, H.V., Taylor, M.A., 2004. Carotenogenesis during tuber development and storage in potato. J. Exp. Bot. 55, 975–982. https://doi.org/10.1093/jxb/erh121.

Návarová, H., Bernsdorff, F., Döring, A.C., Zeier, J., 2013. Pipecolic acid, an endogenous mediator of defense amplification and priming, is a critical regulator of inducible plant immunity. Plant Cell 24, 5123–5141. https://doi.org/10.1105/tpc.112.103564.

Nisar, N., Li, L., Lu, S., Khin, N.C., Pogson, B.J., 2015. Carotenoid metabolism in plants. Mol. Plant 8, 68–82. https://doi.org/10.1016/j.molp.2014.12.007.

Nott, A., Jung, H.S., Koussevitzky, S., Chory, J., 2006. Plastid-to-nucleus retrograde signaling. Annu. Rev. Plant Biol. 57, 739–759. https://doi.org/10.1146/annurev.arplant.57.032905.105310.

Olayide, P., Large, A., Stridh, L., Rabbi, I., Baldermann, S., Stavolone, L., et al., 2020. Gene expression and metabolite profiling of thirteen Nigerian cassava landraces to elucidate starch and carotenoid composition. Agronomy 10. https://doi.org/10.3390/agronomy10030424.

Orek, C., Gruissem, W., Ferguson, M., Vanderschuren, H., 2020. Morpho-physiological and molecular evaluation of drought tolerance in cassava (Manihot esculenta Crantz). Field Crop. Res. 255, 107861 https://doi.org/10.1016/j.fcr.2020.107861.

Osorio, C.E., 2019. The role of orange gene in carotenoid accumulation: manipulating chromoplasts toward a colored future. Front. Plant Sci. 10 https://doi.org/10.3389/fpls.2019.01235.

Paolillo Jr., D.J., Garvin, D.F., Parthasarathy, M.V., 2004. The chromoplasts of or mutants of cauliflower (Brassica oleracea L. var. botrytis). Protoplasma 224, 245–253. https://doi.org/10.1007/s00709-004-0059-1.

Park, Y.J., Park, S.Y., Valan Arasu, M., Al-Dhabi, N.A., Ahn, H.G., Kim, J.K., et al., 2017. Accumulation of carotenoids and metabolic profiling in different cultivars of Tagetes flowers. Molecules 22. https://doi.org/10.3390/molecules22020313.

Pfister, B., Zeeman, S.C., 2016. Formation of starch in plant cells. Cell. Mol. Life Sci. 73, 2781–2807. https://doi.org/10.1007/s00018-016-2250-x.

Rabbi, I.Y., Udoh, L.I., Wolfe, M., Parkes, E.Y., Gedil, M.A., Dixon, A., et al., 2017. Genome-wide association mapping of correlated traits in cassava: dry matter and total carotenoid content. Plant Genome 10, 1–14. https://doi.org/10.3835/plantgenome2016.09.0094.

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., et al., 2019. g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. 47, W191–W198. https://doi.org/10.1093/nar/gkz369.

Sanchez, T., Ceballos, H., Dufour, D., Ortiz, D., Morante, N., Calle, F., et al., 2014. Prediction of carotenoids, cyanide and dry matter contents in fresh cassava root using NIRS and Hunter color techniques. Food Chem. 151, 444–451. https://doi.org/10.1016/j.foodchem.2013.11.081.

Sandmann, G., 2021. Diversity and origin of carotenoid biosynthesis: its history of coevolution towards plant photosynthesis. New Phytol. 232, 479–493. https://doi.org/10.1111/nph.17655.

Sayre, R., Beeching, J.R., Cahoon, E.B., Egesi, C., Fauquet, C., Fellman, J., et al., 2011. The biocassava plus program: biofortification of cassava for sub-Saharan Africa. Annu. Rev. Plant Biol. 62, 251–272. https://doi.org/10.1146/annurev-arplant-042110-103751.

Shim, K.C., Kim, S.H., Jeon, Y.A., Lee, H.S., Adeva, C., Kang, J.W., et al., 2020. A ring-type e3 ubiquitin ligase, osgw2, controls chlorophyll content and dark-induced senescence in rice. Int. J. Mol. Sci. 21 https://doi.org/10.3390/ijms21051704.

Shiraishi, M., Fujishima, H., Chijiwa, H., 2009. Evaluation of table grape genetic resources for sugar, organic acid, and amino acid composition of berries. Euphytica 174, 1–13.

Signorelli, S., Tarkowski, L., O'leary, B., Tabares-Da Rosa, S., Borsani, O., Monza, J., 2021. In: DK, G., FJ, C. (Eds.), 'GABA and Proline Metabolism in Response to Stress,'" in *Hormones and Plant Response*. Springer International Publishing), Cham, pp. 291–314.

Stanley, L., Yuan, Y.W., 2019a. Transcriptional regulation of carotenoid biosynthesis in plants: so many regulators, so little consensus. Front. Plant Sci. 10. ARTN 101710.3389/fpls.2019.01017.

Stanley, L., Yuan, Y.W., 2019b. Transcriptional regulation of carotenoid biosynthesis in plants: so many regulators, so little consensus. Front. Plant Sci. 10, 1017. https://doi.org/10.3389/fpls.2019.01017.

Sun, T., Yuan, H., Chen, C., Kadirjan-Kalbach, D.K., Mazourek, M., Osteryoung, K.W., et al., 2020. ORHis, a natural variant of OR, specifically interacts with plastid division factor ARC3 to regulate chromoplast number and carotenoid accumulation. Mol. Plant 13, 864–878. https://doi.org/10.1016/j.molp.2020.03.007.

Tieman, D., Taylor, M., Schauer, N., Fernie, A.R., Hanson, A.D., Klee, H.J., 2006. Tomato aromatic amino acid decarboxylases participate in synthesis of the flavor volatiles 2-phenylethanol and 2-phenylacetaldehyde. Proc. Natl. Acad. Sci. U. S. A 103, 8287–8292. https://doi.org/10.1073/pnas.0602469103.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M.J., et al., 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28, 511–515. https://doi.org/10.1038/nbt.1621.

Trujillo, M., 2018. News from the PUB: plant U-box type E3 ubiquitin ligases. J. Exp. Bot. 69, 371–384. https://doi.org/10.1093/jxb/erx411.

Velasco-Arroyo, B., Diaz-Mendoza, M., Gomez-Sanchez, A., Moreno-Garcia, B., Santamaria, M.E., Torija-Bonilla, M., et al., 2018. Silencing barley cystatins HvCPI-2 and HvCPI-4 specifically modifies leaf responses to drought stress. Plant Cell Environ. 41, 1776–1790. https://doi.org/10.1111/pce.13178.

Wang, Z., Zhang, L., Dong, C., Guo, J., Jin, L., Wei, P., et al., 2021. Characterization and functional analysis of phytoene synthase gene family in tobacco. BMC Plant Biol. 21, 32. https://doi.org/10.1186/s12870-020-02816-3.

Welsch, R., Arango, J., Bar, C., Salazar, B., Al-Babili, S., Beltran, J., et al., 2010. Provitamin A accumulation in cassava (Manihot esculenta) roots driven by a single nucleotide polymorphism in a phytoene synthase gene. Plant Cell 22, 3348–3356. https://doi.org/10.1105/tpc.110.077560.

Xiao, L., Cao, S., Shang, X., Xie, X., Zeng, W., Lu, L., et al., 2021. Metabolomic and transcriptomic profiling reveals distinct nutritional properties of cassavas with different flesh colors. Food Chem. Mol. Sci. 2, 100016 https://doi.org/10.1016/j.fochms.2021.100016.

Yu, B., Lydiate, D.J., Young, L.W., Schafer, U.A., Hannoufa, A., 2008. Enhancing the carotenoid content of Brassica napus seeds by downregulating lycopene epsilon cyclase. Transgenic Res. 17, 573–585. https://doi.org/10.1007/s11248-007-9131-x.

Zhou, X., Welsch, R., Yang, Y., Álvarez, D., Riediger, M., Yuan, H., et al., 2015. Arabidopsis OR proteins are the major posttranscriptional regulators of phytoene

synthase in controlling carotenoid biosynthesis. Proc. Natl. Acad. Sci. U. S. A 112, 3558–3563. https://doi.org/10.1073/pnas.1420831112.