



# OPEN Conserved stress-responsive genes involved in the early development of *Euterpe Edulis*

Layra de Medeiros Cardozo<sup>1,2</sup>, Francine Alves Nogueira de Almeida<sup>1,2</sup>, Vinicius Sartori Fioresi<sup>1</sup>, Guilherme Bravim Canal<sup>1</sup>, Liana Hilda Golin Mengarda<sup>1</sup>, Adésio Ferreira<sup>1</sup> & Marcia Flores da Silva Ferreira<sup>1✉</sup>

The palm tree *Euterpe edulis* (juçara) plays a crucial ecological and socioeconomic role in the Atlantic Forest. Its wide distribution suggests the presence of adaptive mechanisms for tolerance to abiotic factors, particularly during early developmental stages, such as water availability and shading. This study aimed to identify conserved and differentially expressed genes (DEGs) involved in the early development of *E. edulis*, analyzing different tissues and two divergent plant matrices. Using eight RNA-Seq libraries, the strategy was to obtain the first transcriptome for the species through reference mapping against the *Elaeis guineensis* genome, with focus on conserved genes, followed by differential expression and functional annotation analyses. Among the 32,000 conserved genes identified, 1,133 were differentially expressed, with 11 showing differential expression in both tissues, 678 exclusively in leaves, and 444 in roots. Genes responsive to critical stress factors during early development were identified, revealing matrix-specific environmental adaptations, as well as genes associated with metabolism, light stimuli, and structural development. Leaf tissues exhibited the highest number of exclusive DEGs, indicating greater gene modulation in this tissue. The identification of conserved and highly expressed genes constitutes the first genomic insights for *E. edulis*, providing a foundation for studies aimed at its management, conservation, and genetic improvement for this non-model and endangered organism.

**Keywords** Arecaceae, Differential expression, Juçara, RNA-seq, Transcriptome

The juçara palm (*Euterpe edulis* Mart.) is endemic to the Atlantic Forest, with a natural range that spans the entire biome and extends into some regions of the Cerrado<sup>1,2</sup>. This species plays a vital role in supporting these ecosystems by providing fruits and seeds as a food source for various animals, particularly during periods of scarcity<sup>3,4</sup>. Along with its ability to adapt to diverse environments, *E. edulis* also displays significant phenotypic and genetic variation<sup>4–6</sup>. In recent years, juçara cultivation for fruit production has gained economic significance, driven by the growing demand for açaí (*Euterpe oleracea*), a processed pulp made from the fruits of a co-generic species<sup>7</sup>. However, effective strategies for establishing and managing this species in the field still need to be developed, due to the fragility of the species in its juvenile phase.

The early development of *E. edulis* poses one of the main challenges for cultivating and managing juçara in the field, due to the high mortality rate of seedlings, which require shading and have low tolerance to water deficits<sup>8,9</sup>. Additionally, the species has a long developmental period, taking approximately six years from planting to flowering and fruiting<sup>10</sup>. Naturally, *E. edulis* thrives in the shaded understory of Atlantic Forests, where water availability is typically high<sup>11</sup>. Although it is shade-tolerant, studies have shown that a moderate increase in light availability—such as in areas with less dense canopy—can promote better seedling growth<sup>12</sup>. Another critical factor is the species' recalcitrant seeds, which limit storage and hinder seedling production<sup>13,14</sup>. These factors underscore the urgency of developing genomic resources to support more efficient strategies for seedling production and field cultivation. Therefore, our study contributes can aid in candidate genes that may be used to develop more effective strategies for establishing and managing the species in the field<sup>15,16</sup>. To achieve this, generating genomic knowledge for the species is essential.

*Euterpe edulis* is a non-model organism without a reference genome, making transcriptome analysis an essential strategy for annotating genes and identifying gene expression patterns under different conditions<sup>17–19</sup>.

<sup>1</sup>Centro de Ciências Agrárias e Engenharias, Universidade Federal do Espírito Santo, Alto Universitário, S/N, Guararema, Alegre 29550-000, ES, Brasil. <sup>2</sup>Layra de Medeiros Cardozo and Francine Alves Nogueira de Almeida contributed equally. ✉email: marcia.ferreira@ufes.br

In transcriptome analysis, leveraging genomic data from available *Arecaceae* species is particularly valuable for identifying conserved genes across species<sup>20–24</sup>. In this context, the genome of *Elaeis guineensis* Jacq. (oil palm) stands out as a high-quality reference, as it is chromosome-based and offers high coverage<sup>21</sup>, representing a promising alternative for reference mapping in the transcriptome analysis of *E. edulis*. Moreover, the phylogenetic proximity between these species<sup>25,26</sup> provides a robust framework for identifying conserved genes. Using a well-established reference genome in transcriptome analyses also enhances the precision and reliability of gene annotation<sup>27,28</sup>.

Orthologous genes, which are evolutionarily conserved across different species and maintain their biological function<sup>26,29</sup>, play fundamental roles and serve as indicators of essential biological processes, particularly when highly expressed. Identifying these genes is crucial, as they reveal conserved mechanisms and transcripts key to critical biological processes specific to a botanical family<sup>26,30,31</sup>.

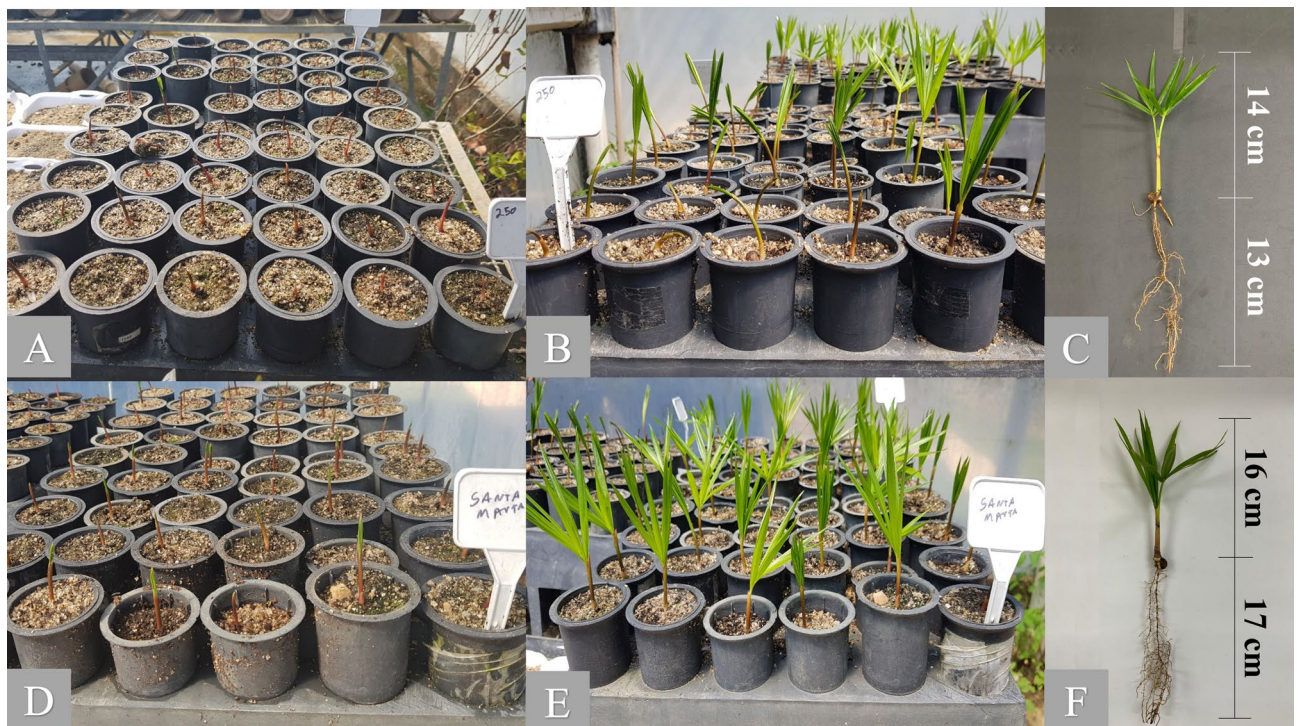
In various studies with different species, conserved genes involved in plant development and stress response have been identified through reference mapping, along with differentially expressed genes (DEGs) indicating specific adaptations to environmental conditions<sup>32–37</sup>.

Given the significance and limited availability of genomic data for *juçara*, particularly regarding the molecular mechanisms involved in its early development, the present study aims to investigate whether phenotypically two divergent *E. edulis* matrices—UFES\_250 and Santa Marta (SM)—exhibit distinct gene expression profiles that may reflect phenotypes of interest for the species production. Additionally, the study seeks to provide conserved genes related to its early development. The research explores the relationships between divergent phenotypic traits and the differential expression of leaf and root transcriptomes to identify conserved genes associated with development and environmental adaptation. By maximizing the detection of genes and transcripts, we aim to contribute to the knowledge of conservation strategies, management, and genetic improvement, ensuring a sustainable future for *E. edulis* and related species.

## Materials and methods

### Plant material and morphological characterization

Seeds from two morphologically divergent matrices, UFES\_250 (Fig. 1A, B, C) and SM (Fig. 1D, E, F), were collected. Plants of these matrices are located in a private planting area for *juçara* fruit production in the municipality of Rio Novo do Sul, in the state of Espírito Santo (Brazil) (latitude  $-20.807.598$ , longitude:  $-40.934.519$ ). The region has a tropical climate with a dry season, an average annual temperature of around 22 °C, and an altitude of 470 m.



**Fig. 1.** *Euterpe edulis* matrices UFES\_250 (Figures A, B, and C) and Santa Marta (SM) (Figures D, E, and F). Roots and leaves from both matrices were collected after germination for RNA extraction. **(A)** *Euterpe edulis* UFES\_250 during germination (three months); **(B)** *Euterpe edulis* UFES\_250 after six months of germination; **(C)** Root (mean 13 cm) and aerial part (mean 14 cm) of *E. edulis* UFES\_250. **(D)** *Euterpe edulis* SM during germination (three months), **(E)** *Euterpe edulis* SM after six months of germination, **(F)** root (mean 17 cm) and aerial part (mean 16 cm) of *E. edulis* SM.

The UFES\_250 matrix was selected for its high pulp yield (31.08%)<sup>38</sup> and for displaying typical species characteristics, including a greenish-yellow leaf sheath, cream-colored inflorescence, green immature fruits, and short bracts<sup>39</sup>. The SM matrix exhibits distinct morphological traits compared to the typical *E. edulis* plants found in Espírito Santo, thriving better in cooler environments<sup>40</sup>. According to local growers, SM plants are distinguished by their thicker stems, larger fruits (though with lower pulp yield), bigger fruit clusters, wider spacing between leaf scars, and longer heart of palm. In the field, the UFES\_250 genotype exhibited notable differences in fruit and seed characteristics compared to the SM genotype. UFES\_250 produced an average of 4 fruit bunches, while SM had a slightly higher average of 4.20 bunches. However, UFES\_250 had a significantly higher pulp yield (31.08%) than SM (16.42%). The fruits of UFES\_250 were also larger, with an equatorial diameter of 14.50 mm and a longitudinal diameter of 15.21 mm, compared to 13.17 mm and 13.01 mm, respectively, in SM. Similarly, seed dimensions were greater in UFES\_250, with an equatorial diameter of 13.02 mm and a longitudinal diameter of 13.60 mm, whereas SM seeds measured 12.59 mm and 11.94 mm, respectively.

The fruits were collected and de-pulped, and 120 seeds from each matrix were obtained in May 2021. The seeds were immediately immersed in warm water (32 °C) for 40 min before being placed in tubes with substrate and maintained in a greenhouse for germination and early development, without temperature control (Fig. 1). After six months, samples of leaves and roots were collected from 50 seedlings of each matrix for phenotypic measurements and transcriptome analysis.

The lengths of the leaflets, stem, aerial part, and root were recorded in centimeters to be used as phenotypic measurements. Fresh masses of the aerial part and root were measured in grams using a precision balance. To obtain dry masses, samples were dried in a forced-air circulation oven at 65 °C until reaching a constant weight, after which they were weighed using a precision balance. Descriptive statistical analyses of these traits were performed using RStudio v.4.2.1 with the “ggplot2” package<sup>41</sup>. Additionally, an analysis of covariance (ANCOVA) was conducted (Supplementary File 1, Table 3) using the “stats” package<sup>42</sup>.

## RNA extraction and sequencing

Leaf and root samples were collected from the 50 seedlings grown in the greenhouse for six months for RNA extraction (Supplementary File 1, Fig. 1). Five tissue pools were created for each sample, with each pool consisting of tissue from 10 different plants, collected and mixed in equal proportions. These pooled samples, totaling 100 mg of fresh tissue, were placed in microtubes and immediately frozen in liquid nitrogen. The RNA was then extracted from the pooled samples.

The frozen plant material was ground in a mortar using liquid nitrogen. RNA extraction was performed using the cetyltrimethylammonium bromide (CTAB) method, adapted<sup>43</sup> with modifications<sup>44</sup>. Initially, 900 µL of pre-warmed (65 °C) extraction buffer was added to the tissue powder, and the mixture was agitated until homogeneous. The mixture was then transferred to a 2 mL microtube and incubated at 65 °C for 10 min. Next, an equal volume of chloroform/isoamyl alcohol (24:1, v/v) was added, and the tube was vigorously shaken. The microtube was centrifuged at  $7,000 \times g$  for 20 min at 4 °C. The supernatant was collected, transferred to a new 1.5 mL microtube, and re-extracted with an equal volume of chloroform/isoamyl alcohol (~ 650 µL).

Next, 0.5 volume of 5 M lithium chloride (LiCl) was added to the supernatant, followed by incubation at –20 °C for four hours. The RNA was selectively pelleted by centrifugation at  $16,000 \times g$  for 30 min at 4 °C. The pellet was washed with 75% (v/v) ethanol and air-dried. The RNA was then solubilized in 30 µL of RNase-free ultrapure water and stored in an ultra-low temperature freezer at –80 °C for subsequent analyses. The extracted RNA was treated with DNase (DNase Treatment of RNA Samples Prior to RT-PCR – Promega) following the manufacturer’s protocol.

After RNA isolation, quantification was performed using a NanoDrop ND-1000 spectrophotometer, and the integrity of the samples was assessed by electrophoresis on a 1% agarose gel, using the GelRed intercalant (Biotium) for documentation. The quality and quantity of the total RNA were calculated using the TapeStation System (Agilent) and Qubit (Thermo Fisher Scientific), respectively.

The library was prepared using the Illumina TruSeq Stranded Total RNA Library Prep Plant Kit, following the TruSeq Stranded Total RNA Reference Guide (1000000040499 v00). This kit utilizes RiboZero beads to deplete ribosomal RNA from cytoplasmic, mitochondrial, and chloroplast sources in plant samples and a PCR master mix to transcribe RNA into cDNA. A normalization step was performed after obtaining the cDNA following a published protocol<sup>45</sup> and using the Trimmer and Trimmer-Direct kit (Evrogen) to increase the discovery of more genes. The libraries were quantified using Qubit (Thermo Fisher Scientific), and fragment sizes were estimated using the TapeStation System (Agilent). Eight libraries (two for each tissue type and matrix) were subjected to total RNA-seq utilizing the Illumina NovaSeq 6000 platform in transcriptome Analysis mode, generating paired-end reads of 200–400 bp.

## Mapping of the reads to the reference genome

Regarding the raw Illumina data in the FASTQ file format, data visualization was performed using FastQC v. 0.11.844<sup>46</sup>, followed by the removal of adapters and low-quality sequences using Trimmomatic v. 0.38<sup>47</sup> with the following settings: Phred 30, Leading 3, Trailing 3, Slidingwindow 4:15, and Minlen 36. The cleaned sequence files were re-assessed using FastQC to ensure the effectiveness of the quality control process.

The filtered sequences were mapped to the reference genome of *Elaeis guineensis* (GCF\_000442705.1)<sup>48</sup>. The statistical analysis was performed using HISAT2 version 2.1.0<sup>49</sup>, based on the read mapping conducted with Bowtie2 version 2.3.4.1<sup>50</sup>. Contiguous transcript sequences were assembled using StringTie version 2.1.3b, with the reference genome as the basis<sup>51</sup>.



## Differential gene expression analysis

After assembly, gene/transcript abundance was calculated based on read counts, with values normalized using the counts-per-million (CPM) metric. Genes with very low counts across all libraries were filtered out to avoid negatively impacting statistical analyses, as such genes might be underrepresented in the samples. This filtering was performed using the `filterByExpr` command from the “`edgeR`”<sup>52</sup> package, with a minimum count (min. count) of 100. Samples were grouped in a 2D space using principal component analysis (PCA) plots generated with the `plotMDS` function from the “`limma-voom`”<sup>53</sup> package (50). These plots were instrumental in providing information about the variability of biological replicates for each evaluated tissue. The analysis was conducted to find low variability within samples from the same group compared to the variability observed between different groups.

An initial analysis was performed using all eight libraries, followed by separate analyses of leaf and root tissue libraries, to assess differential expression between matrices. Contrasts for the comparisons were created using the `makeContrasts` function from the “`limma-voom`” package. CPM values were employed to calculate fold changes as the ratio between treatment and control on a logarithmic scale, e.g.,  $\log_2(\text{CPM}_{\text{SM}}/\text{CPM}_{\text{UFES\_250}})$ . Fold changes greater than zero were considered up-regulated, while those less than zero were classified as downregulated relative to the control sample.

A statistical threshold of  $P \leq 0.05$  was used to identify significant results. The “`limma-voom` package” was also applied for differential gene expression analysis. P-values less than 0.05 were adjusted using the Benjamini-Hochberg method to control the false discovery rate (FDR)<sup>54</sup>.

DEG hierarchical clustering was conducted using the `coolmap` function from the “`limma-voom`” package, applying the average linkage clustering method. Heatmaps were used to display the genetic abundance levels. These heatmaps were constructed using log-transformed and normalized values of genes, based on uncentered Pearson distances and the unweighted pair group method with arithmetic mean (UPGMA). The color scheme represented the logarithmic intensity of gene expression linked to z-score measurements. A quantitative palette ranging from blue to red was applied, where relatively higher gene expression levels were shown in red, and lower expression levels in blue.

DEG volcano plots were generated using the `EnhancedVolcano` package<sup>55</sup>, enabling the analysis of tissue-specific genes.

Functional annotation of DEGs, as well as non-differentially expressed and highly expressed genes, was performed using the Gene Ontology (GO) database<sup>56</sup> through the online tools DAVID<sup>57</sup> and GOSlimViewer from the AgBase Database<sup>58</sup>. A Venn diagram of DEGs was created to display tissue-specific and tissue-independent genes across all three analyses.

The methodology, spanning from the quality control of sequencing reads to the functional annotation analysis conducted in this study, is summarized in the flowchart below (Fig. 2).

## Results

### Morphological characterization

The seedlings from the SM matrix emerged earlier (Fig. 1D) than those from the UFES\_250 matrix (Fig. 1A). Among the morphological characteristics evaluated, seedlings from the SM matrix showed greater shoot length (SM: 16.43 cm; UFES\_250: 14.33 cm; p-value = 0.00229\*\*), stem length (SM: 7.802 cm; UFES\_250: 6.824 cm; p-value = 0.0442\*), and root length (SM: 17.5; UFES\_250: 13.15; p-value = 7.72<sup>e-09</sup>\*\*\*) (Fig. 3; Supplementary File 5) after six months. Overall, the average values for all measured variables were higher in the SM seedlings compared to the UFES\_250 seedlings.

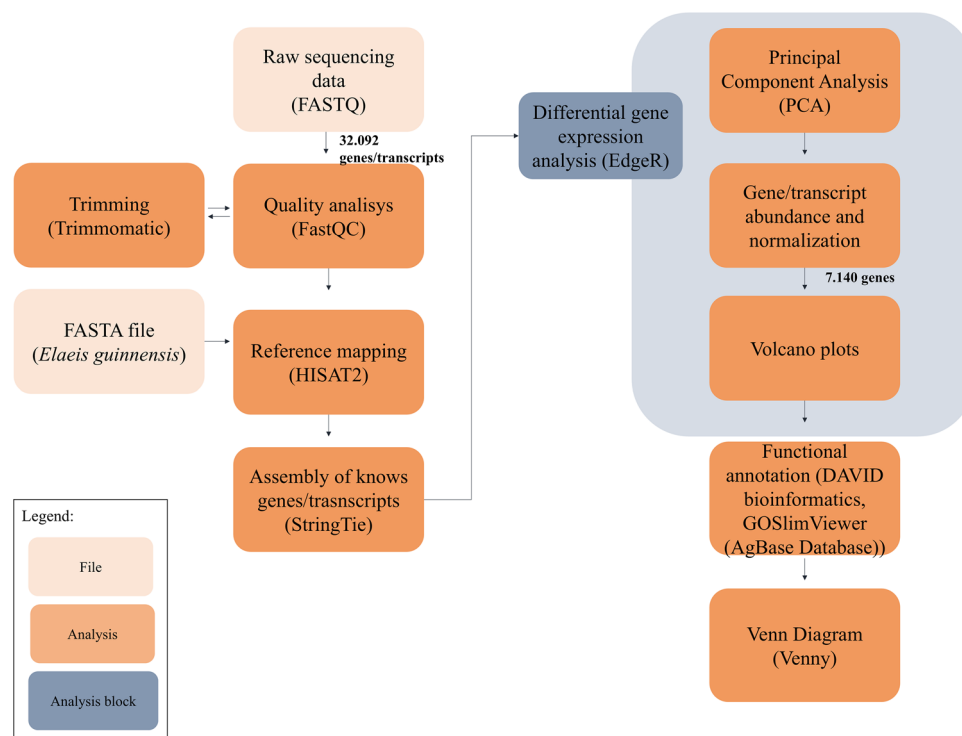
### Trimming of reads and sequencing data

The Illumina sequencing generated an average of 44,171,557 reads, totaling 6.3 Gbp across the eight libraries (Supplementary File 1, Table 1). After quality analysis, the libraries displayed an average GC content of 44.97% and a Q30 (percentage of bases with a Phred score above 30) of 94.92% (Supplementary File 1, Table 4). The average mapping rate of the reads to the *Elaeis guineensis* reference genome was 54% for leaf tissue and 12.45% for root tissue, with similar results observed across libraries from different matrices (Supplementary File 1, Table 5; Fig. 2). A total of 32,092 genes and transcripts were identified in the mapping process (Supplementary File 1, Table 2).

### Differences between libraries and normalization

Biological replicates clustered both by tissue type and matrix, demonstrating the experimental conditions' control and the distinct effects of matrix and tissue studied (Fig. 4 and Supplementary File 1, Fig. 3). Differential expression was observed between tissues, with greater variability in gene expression noted in leaf tissue compared to root tissue.

The low normalization of the eight libraries for leaf and root samples suggests the presence of a large number of highly up-regulated genes. A normalization factor below one indicates that a small subset of highly abundant genes dominates the sequencing output, leading to reduced counts for other genes compared to what would be expected given the library size. As a result, the effective library size for these samples was reduced. After normalization analysis across all eight libraries, 7,140 genes were identified (Supplementary File 2, Table 1). Of these, 670 were exclusively found in root samples (Supplementary File 2, Table 3), including 130 exclusive to the UFES\_250 matrix and 69 to the SM matrix (Supplementary File 2, Table 2). In leaf tissue, 92 exclusive genes were identified (Supplementary File 2, Table 3), with 44 exclusive to UFES\_250 and 15 to SM (Fig. 5A and Supplementary File 1, Table 7).



**Fig. 2.** Flowchart of the applied methodology. Reads obtained from the FASTQ file, resulting from Illumina sequencing, were analyzed for quality and mapped to reference, resulting in 32,092 genes. After filtering, 7,140 expressed genes were obtained, which were normalized. Differential gene expression analysis was then performed, highlighting 678 differentially expressed genes (DEGs) in leaves and 444 in roots. When both tissues were analyzed, 11 DEGs were identified. All expressed genes were functionally annotated.

In the evaluation of leaf libraries for both matrices ( $n = 4$ ), the process revealed a low normalization factor for the SM leaves. From this analysis, 4,239 genes were obtained (Supplementary File 3, Table 1), of which 368 were exclusive to UFES\_250 and 174 to SM (Fig. 5B and Supplementary File 3, Table 2). Regarding the root tissue libraries ( $n = 4$ ), the SM samples also exhibited a low normalization factor, revealing 6,052 genes (Supplementary File 4, Table 1), with 274 genes exclusive to UFES\_250 and 130 to SM (Fig. 5C and Supplementary File 4, Table 2).

### Gene differential expression analysis

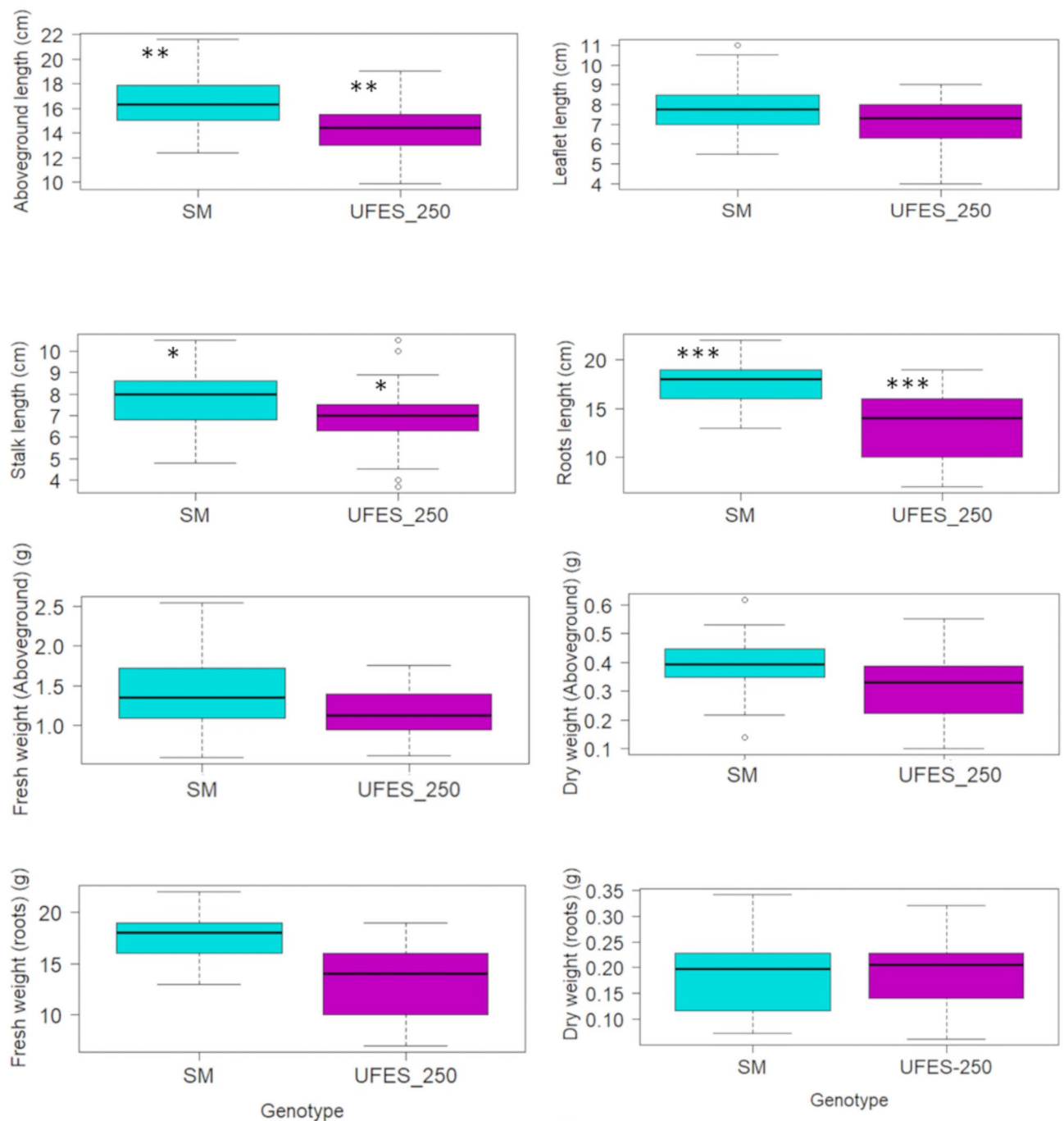
The heatmaps and volcano plots of the DEGs are shown in Fig. 6A (all tissues), Fig. 6B (leaf), and Fig. 6C (root). The red and blue colors represent DEGs that are up-regulated and down-regulated in UFES\_250 compared to SM, respectively. In the analysis between the matrices UFES\_250 and SM, considering all tissues together, 11 DEGs were identified (Supplementary File 2, Table 4), of which five were downregulated and six were up-regulated in UFES\_250 (Supplementary File 1, Table 6). In the analysis using only the leaf libraries, 678 DEGs were identified, with 285 up-regulated and 393 downregulated in UFES\_250 (Supplementary File 3, Table 3). For the roots, 444 DEGs were identified (Supplementary File 4, Table 3), with 119 up-regulated and 325 down-regulated in UFES\_250. The lists of all DEGs and the 50 most highly expressed genes for leaf and root tissues can be found in Table 5 of Supplementary File 3 and Table 5 of Supplementary File 4, respectively.

### Functional annotation of differentially expressed genes

All DEGs were functionally assigned to GO terms within the primary ontologies: 'biological process,' 'cellular component,' and 'molecular function.' Among the 11 DEGs identified in the analysis of all tissues (Fig. 7 and Supplementary File 2, Table 5), three GOs for 'biological process' functions and five for 'molecular function' stood out (Fig. 7A). Based on the GO results, the genes Hsp70 and LOC105056959 are related to responses to chemical stimuli and stress, regulated negatively and positively, respectively, in UFES\_250.

The functional annotation of the 678 DEGs in the leaf tissue revealed 25 GOs related to 'biological processes,' 12 to 'cellular components,' and 15 to 'molecular functions' (Supplementary File 3, Table 4). Genes involved in responses to stress, chemical stimuli, light stimuli, and anatomical structural development were annotated, with some positively regulated in UFES\_250 (e.g., LOC105042857, LOC105040294, LOC105049137, ClpB1, Hsp70, and aquaporin PIP2) and others negatively regulated (e.g., rpl2, rpl22, LOC105039602, LOC105039245, LOC105039517, and WRKY24) (Fig. 7B).

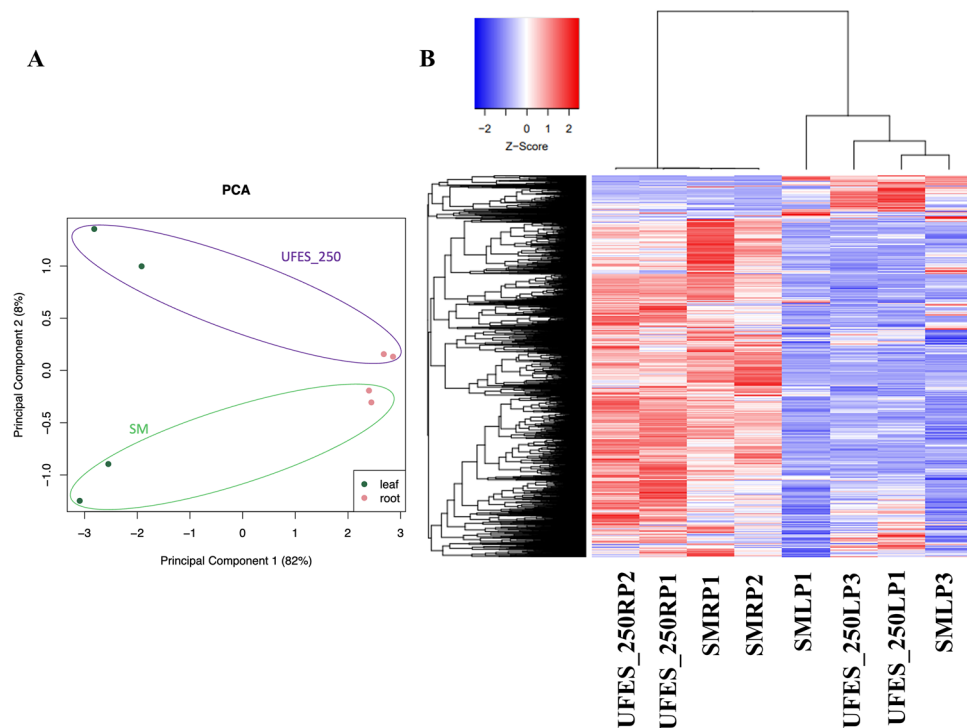
For the 444 DEGs found in the root tissue analysis, 16 GOs were related to 'biological processes,' seven to 'cellular components,' and 13 to 'molecular functions' (Supplementary File 4, Table 4). Genes involved in



**Fig. 3.** Comparative boxplot of early development data for traits such as lengths of aerial part, leaflet, stem, and root, as well as fresh and dry weight of aerial part and root from 50 seedlings of each UFES\_250 and SM matrix, after six months in a greenhouse. Significant differences are indicated by asterisks (\*\*\* 0.001 \*\* 0.01 \* 0.05).

responses to stress, chemical stimuli, and abiotic stimuli were identified, with some positively regulated in UFES\_250 (e.g., CLB3, TPS, aquaporin PIP2, LOC105042909, and LOC105041389) and others negatively regulated (e.g., LOC105039584 and WRKY24) (Fig. 7C).

Overall, the genes were similarly categorized by GO across both tissues. Of the 31 GOs shared between the two tissues (Fig. 7D), in addition to those involved in biological, metabolic, catabolic processes, and DNA binding activities commonly found in plants, GOs related to responses to stress, chemical stimuli, and abiotic stimuli were annotated for both tissues (Fig. 7A). Most of the 21 tissue-specific GOs for leaf are involved in 'biological processes,' such as reproduction, response to light stimulus, post-embryonic development, photosynthesis, and anatomical structure development. The five root-specific GOs are related to cellular protein modification processes, cell communication, vacuole, plasma membrane, and lipid binding.



**Fig. 4.** (A) Principal component analysis (PCA) showing the general differences between expression profiles of the libraries from the different matrices UFES\_250 (purple ellipse) and Santa Marta (SM) (green ellipse). (B) The relative abundance of the 7,000 normalized genes in leaves and roots of the UFES\_250 and SM matrices in individual and hierarchical clustering, describing the clustering between expression profiles in the analyzed tissues. Legend: RP1- Root pool 1; RP2- Root pool 2; LP1- Leaf pool 1; LP3- Leaf pool 3.

### Functional annotation of genes with no differential expression and highly expressed genes

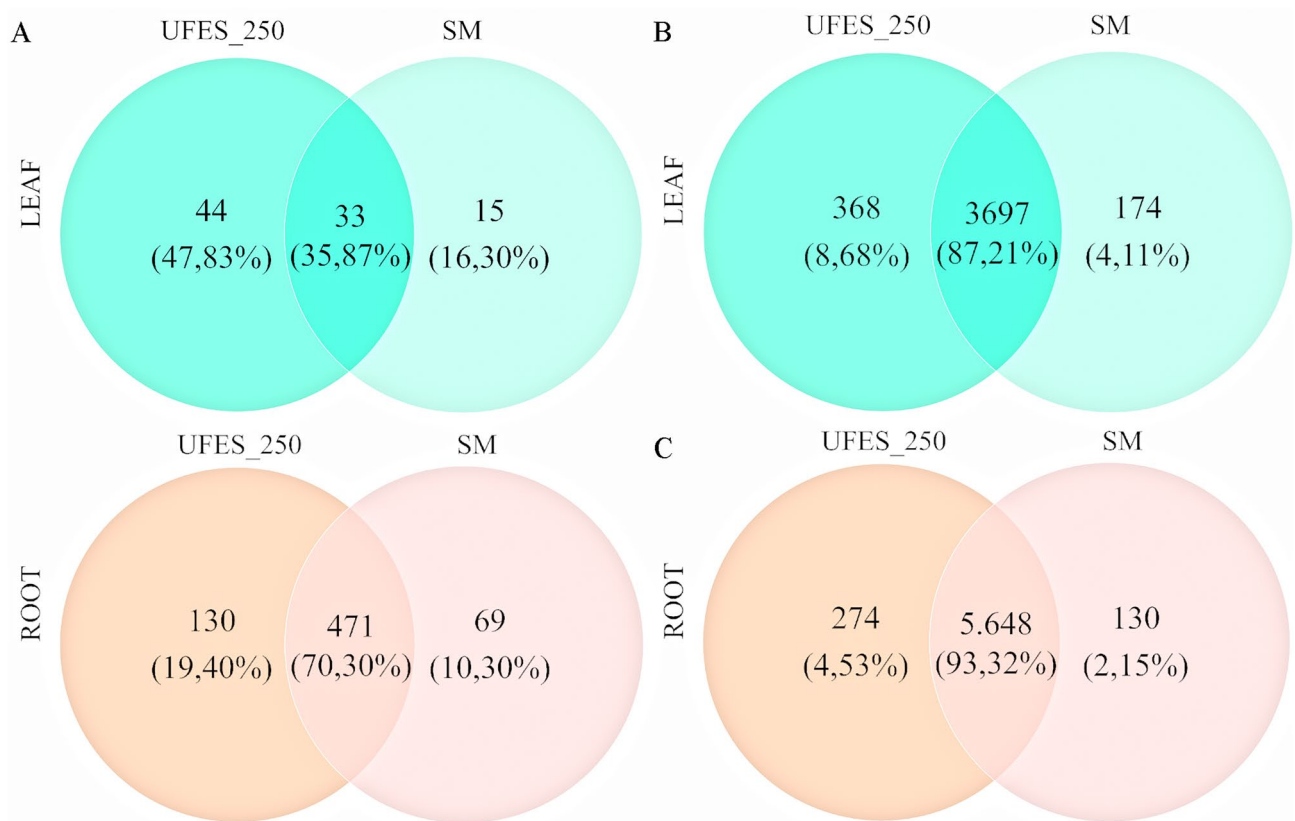
In the GO analysis of genes with no differential expression (Supplementary File 2, Table 6; Supplementary File 3, Table 6; Supplementary File 4, Table 6), the main 'biological processes' identified were cellular process, metabolic process, biosynthetic process, metabolic process of nitrogenous base-containing compounds, transport, cellular protein modification process, organization of cellular components, protein metabolic process, catabolic process, and stress response. The principal 'cellular components' included membrane, nucleus, cytoplasm, intracellular anatomical structure, cytosol, and chloroplast. For the 'molecular function' component, functions identified included binding, nucleotide binding, catalytic activity, protein binding, molecular function, hydrolase activity, and transferase activity.

Furthermore, genes with functions of interest to the present study were detected, including those related to abiotic stimulus-response, external stimulus-response, cellular homeostasis, photosynthesis, reproduction, light stimulus-response, biotic stimulus-response, cellular differentiation, post-embryonic development, growth, circadian rhythm, cell growth, gene expression regulation, epigenetics, cell-to-cell signaling, pollination, tropism, floral development, and embryonic development (Fig. 8).

When analyzing the highly expressed genes in the two tissues together (Supplementary File 2, Table 5), we found that many of the genes are related to metabolic processes and stress response mechanisms in the UFES\_250 matrix. For example, aquaporin PIP2-4 is involved in water transport, the large subunit of ribulose-1,5-bisphosphate carboxylase-oxygenase (ruBisCO) binding protein plays a role in photosynthesis, and heat shock proteins (HSPs) like Hsp 70 kDa protein 14 help protect against thermal stress. Genes related to the regulation of energy metabolism and antioxidant protection, such as glutathione peroxidase and 6-phosphofructokinase, are also highly expressed in this matrix.

In the SM matrix, the expressed genes show a greater predominance of transcription factors, like bHLH148 and MYB78, as well as genes related to hormonal signaling and defense processes, including cytochrome P450 and chitinase 10. The SM matrix also includes genes involved in carbohydrate degradation and synthesis, such as xyloglucan endotransglucosylase/hydrolase. Therefore, while UFES\_250 exhibits greater modulation of genes related to cellular metabolism and stress responses, SM stands out for gene regulation and defense mechanisms.

The highly expressed genes in the leaf tissue of UFES\_250 show a predominance of functions related to thermal stress, such as HSPs, which play crucial roles in the response to environmental stress and protein folding. Additionally, genes involved in flavonoid biosynthesis, like chalcone synthase and naringenin-dioxygenase, suggest a potential role in defense against pathogens and growth regulation. Genes related to energy metabolism and protein transport regulation, such as exosome complexes and ribosomes, are also present. In SM, the expressed genes display diverse functions, many related to cellular growth regulation and



**Fig. 5.** Venn diagram representing exclusive and shared genes by matrix and tissue. **(A)** Result of the 7,140 genes identified in the analysis of all eight libraries combined, showing the number of genes exclusive to leaf tissue (92) and root tissue (670) and by matrix (UFES\_250 and Santa Marta - SM). **(B)** Result of the 4,239 genes identified in the analysis of only the leaf libraries, showing the number of exclusive genes in the UFES\_250 (368) and SM (174) matrices. **(C)** Result of the 6,052 genes identified in the analysis of only the root libraries, showing the number of exclusive genes in the UFES\_250 (274) and SM (130) matrices.

response to environmental signals. Serine/threonine kinases and cell wall-related proteins, like xyloglucan endotransglucosylase/hydrolase, indicate an important role in cell wall remodeling and development.

Genes such as ABC transporters and oxidases indicate involvement in metabolite transport and oxidative stress response. Comparing the gene modulation between the matrices, UFES\_250 stands out for its genes related to thermal stress response and cellular protection, while SM is associated with signaling processes and structural development. Both matrices share genes that regulate growth and respond to environmental stresses.

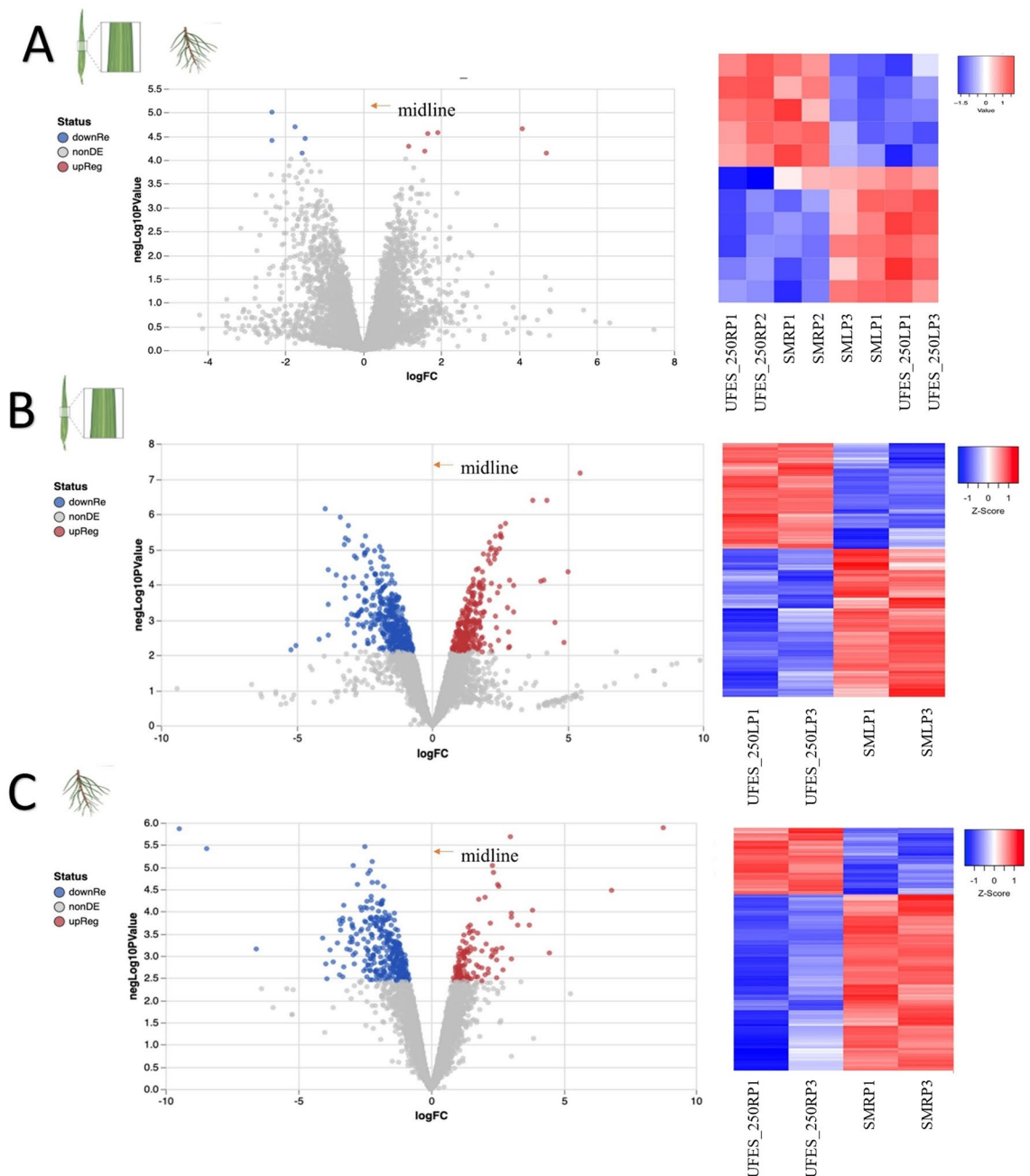
The highly expressed genes in the root tissues of the UFES\_250 matrix are associated with structural and cellular maintenance functions, such as tubulins and cytoskeleton-associated proteins. Genes involved in carbohydrate metabolism, including beta-glucosidase and phosphoglucosmutase, were also identified. The presence of genes related to autophagy and transport proteins indicates a role in cellular recycling processes and nutrient transport.

On the other hand, in the SM matrix, there is a prevalence of genes associated with hormonal and environmental responses, including bHLH transcription factors and proteins responsive to ethylene and auxin, suggesting greater involvement in growth regulation and stress response. Additionally, genes related to the synthesis of oxidative enzymes (e.g., aldo-keto reductase) indicate detoxification activities and regulation of secondary metabolism. Therefore, while the UFES\_250 root shows greater modulation for structural maintenance and transport genes, the SM root exhibits greater modulation for hormonal regulation and adaptive environmental responses.

## Discussion

This study presents the first genomic insights for *E. edulis*, with the annotation of 1,133 DEGs during the early development of seedlings from two divergent matrices (UFES\_250 and SM). The findings highlight: (1) differences in gene expression modulation between seedlings from different matrices in both leaf and root tissues; (2) greater conservation of genes in the root and differential modulation of leaf-specific genes; (3) DEGs associated with stress, early development, environmental stimuli, photosynthetic efficiency, and cellular integrity; (4) the predominance of biological and molecular processes specific to leaves and roots, with distinct GO terms for each tissue; (5) DEGs involved in flowering and responses to biotic and abiotic stresses with

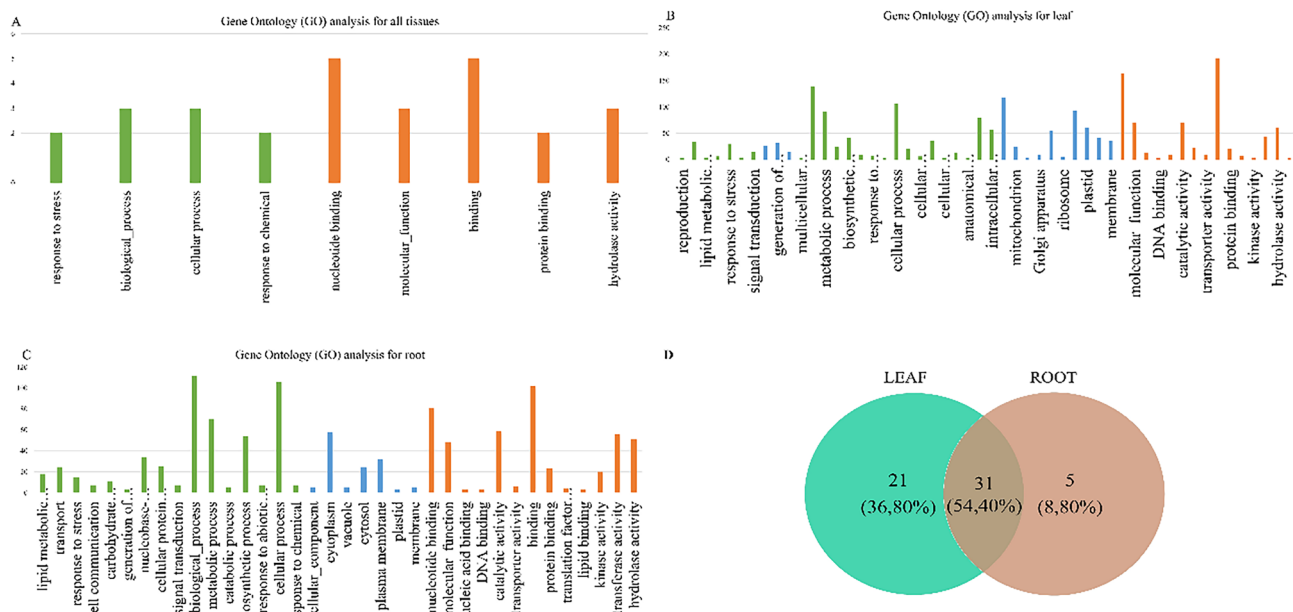




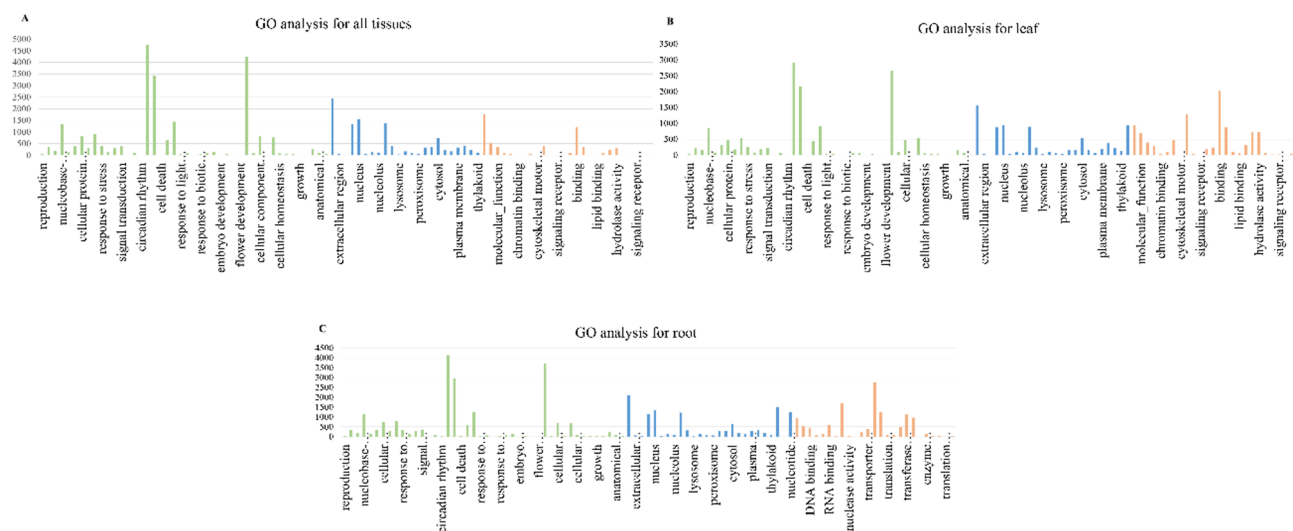
**Fig. 6.** Differentially expressed genes (DEGs) between UFES\_250 and Santa Marta (SM) matrices. On the left, volcano plots of the DEGs. The y-axis represents the significant gene expression level between samples, measured by the p-value, while the x-axis represents the DEGs' fold change (logFC). Blue and red points indicate down-regulated and up-regulated genes, respectively. Gray points represent genes with no significant differential expression. On the right, heatmaps of down-regulated and up-regulated genes with higher log<sub>2</sub> FC values. The color indicates the expression level of DEGs with log<sub>2</sub>. **(A)** Result of the 11 DEGs obtained from the analysis of all tissues together, six up-regulated and five down-regulated in UFES\_250. **(B)** Result of the 678 DEGs obtained from the analysis of only leaf tissue, 285 up-regulated and 393 down-regulated in UFES\_250. **(C)** Result of the 444 DEGs obtained from the analysis of only root tissue, 119 up-regulated and 325 down-regulated in UFES\_250. Legend: RP1- Root pool 1; RP2- Root pool 2; LP1- Leaf pool 1; LP3- Leaf pool 3.

differential expression between matrices; and (6) key DEGs such as TPS, Hsp70, aquaporin PIP2, ClpB, SERK, and WRKY, which are responsive to stress and regulate development.

The seedlings exhibited phenotypic differences at six months of age, with those from the SM matrix showing higher average values. These outcomes indicate that, in addition to the genetic differentiation of SM individuals<sup>40</sup>, there is also distinct modulation of gene expression during the early development of seedlings from different matrices.



**Fig. 7.** Analysis of the main Gene Ontology (GO) of differentially expressed genes (DEGs) from the UFES\_250 and Santa Marta (SM) matrices of *Euterpe edulis*, categorized into 'biological process' (green), 'cellular component' (blue), and 'molecular function' (orange). **A.** All tissues combined. **B.** Leaf tissue. **C.** Root tissue. **D.** Venn diagram of DEGs showing the distribution of tissue-specific and tissue-independent GOs in the two tissue sets (leaf and root).



**Fig. 8.** Gene Ontology (GO) analysis of genes that are not differentially expressed from the UFES\_250 and Santa Marta (SM) matrices of *Euterpe edulis*, categorized into 'biological process' (green), 'cellular component' (blue), and 'molecular function' (orange), considering **A.** both tissues; **B.** leaf tissue; **C.** root tissue.

Although more than 50% of the reads from the foliar transcriptome were mapped to the *Elaeis guineensis* reference genome, the highest number of genes was detected in roots (6,052) compared to those detected in leaves (4,239). The root genes also varied less in expression. However, the occurrence of only 12% mapping of the root reads indicates that a large part of the genes and the expression variation in this tissue are species-specific.

The identification of root-specific genes in the species and the modulation of their expression may help explain their occurrence in the Atlantic Forest, with significant distribution and variation in phytophysiology<sup>4</sup>. On the other hand, searching for conserved DEGs in plants through transcriptomes is a strategy for identifying candidate genes as promising molecular markers linked to agronomic traits of interest<sup>59</sup>. The present study corroborates and expands information from studies on Arecaceae species regarding the conservation of genes related to early development<sup>60–63</sup> and responses to abiotic stresses<sup>24,64–66</sup>.

The UFES\_250 matrix exhibited about two-thirds of the exclusive genes in both leaf and root tissues (368 in leaf and 274 in root) than the SM matrix, which had fewer exclusive genes (174 in leaf and 130 in root). However, the SM matrix showed more significant initial development.

Genes related to responses to different stresses, seedling viability, and somatic embryogenesis were detected in the leaf and root tissues of the seedlings, including trehalose-6-phosphate synthase (TPS), 70-kDa Hsp70, aquaporin PIP2, casein lytic proteinase B (CLpB), somatic embryogenesis receptor kinase (SERK), and WRKY transcription factor (WRKY).

Considering all libraries, we detected 11 DEGs in both matrices in the analyses of individual tissues. Among these genes, those down-regulated in UFES\_250 and up-regulated in SM were involved in transcription regulation (LOC105056959: probable mediator of RNA polymerase II transcription subunit 37c; LOC105044900: NAC domain-containing protein 68; and LOC105052255: WRKY transcription factor WRKY24-like)<sup>67–69</sup>; biosynthesis of sphingolipids, which aid in membrane formation and cell signaling (LOC105051063: serine C-palmitoyltransferase)<sup>70</sup>; negative regulation of gibberellin (GA) and abscisic acid (ABA) signaling in aleurone cells (LOC105052255: WRKY transcription factor WRKY24-like)<sup>69,71,72</sup>; and protection of plant cells against oxidative stress, maintenance of the redox balance in the mitochondrial electron transport chain to facilitate photosynthetic metabolism, and regulation of photorespiration (LOC105058492: mitochondrial uncoupling protein 5)<sup>73–75</sup>.

We also detected genes involved in chloroplast development and seedling viability, which were up-regulated in UFES\_250 and down-regulated in SM, including genes involved in mediating thylakoid membrane formation, chloroplast thermotolerance during heat stress (LOC105049549: chaperone protein ClpB3)<sup>76</sup>, catalyzing the interconversion of glyceraldehyde 3-phosphate and dihydroxyacetone phosphate in the glycolytic and gluconeogenic pathways (LOC105056934: triosephosphate isomerase)<sup>77</sup>, and protein degradation during the cell cycle (U-box domain-containing protein 52, transcript variant X4)<sup>78</sup>.

These differences in gene regulation between seedlings from different matrices suggest that UFES\_250 may have higher efficiency in photosynthesis, chloroplast development, and thermotolerance<sup>76</sup> processes. In contrast, SM may have an advantage in gene transcription<sup>67–69</sup>, membrane integrity, and hormonal signaling<sup>70,79,80</sup>. These variations indicate that UFES\_250 could be better adapted to environments with thermal stress and conditions that require high metabolic efficiency<sup>76</sup>, whereas SM may have an advantage in environments where transcription regulation and cellular integrity are more critical<sup>69</sup>. These differences in gene regulation can significantly influence the initial development of *E. edulis* plants, affecting aspects such as growth, environmental adaptation, and responses to abiotic stress. Notably, SM typically occurs naturally at higher altitudes.

Although the most genes were detected in roots compared to leaves, the greatest number of DEGs occurred in leaves (678) compared to roots (444), indicating a higher modulation of expression in the conserved leaf genes. However, studies with other species indicate that the number and type of DEGs can vary between these organs, particularly in response to different environmental stresses. For example, there was a predominance of differential gene expression in roots in *Brassica campestris* under cold stress<sup>81</sup>, *Medicago sativa* under saline stress<sup>82</sup>, *Baphicacanthus cusia* treated with methyl jasmonate<sup>83</sup>, *Lolium* and *Festuca* spp. under water stress<sup>84</sup>, and *Prunus persica* under different soil conditions<sup>85</sup>. In contrast, *Linum usitatissimum* under water stress showed a predominance of differential gene expression in the leaves, related to lignin and proline biosynthesis<sup>86</sup>, and there was also a predominance of differential gene expression in the leaves of *Phoenix dactylifera* (Arecaceae) under saline stress<sup>64</sup>.

GO characterization showed a predominance of 'biological processes,' followed by 'molecular functions' in the tissues. The Venn diagram analysis for the GOs found in the leaf and root tissues revealed both common and tissue-specific genes. Highly enriched GO terms across all tissues included cell, cell wall, membrane, membrane part, catalytic activity, binding, metabolic, and cellular processes.

Exclusive GO terms in leaves were related to functions such as reproduction, response to light stimuli, post-embryonic development, photosynthesis, and anatomical structure development. The genes LOC105040294, which expresses the GIGANTEA protein (up-regulated in UFES\_250), and the genes LOC105039602 and LOC105039245, which express zinc finger domain proteins (down-regulated in UFES\_250), are involved in the functions of reproduction, response to light stimuli, post-embryonic development, and anatomical structure. The GIGANTEA protein is important in regulating the flowering time of plants. Photoperiod-controlled flowering is a vital developmental process directly related to the plant's reproductive success<sup>87</sup>. Mutations in the GIGANTEA gene delay flowering on long days, but the effects are minimal on short days<sup>88</sup>. This suggests that GIGANTEA plays a crucial role in regulating the expression of flowering time genes, promoting photoperiod-induced flowering, and participating in the plant's circadian feedback cycle<sup>88,89</sup>.

Zinc finger domain proteins also play important roles in plant development and reproduction. These proteins respond to light stimuli and post-embryonic development, influencing the plant's anatomical structure. CONSTANS, a zinc finger protein, is a transcription factor that acts in the long-day flowering pathway and may mediate the interaction between the circadian clock and flowering control<sup>90</sup>. Reducing the flowering time is essential for the species studied, with GIGANTEA playing a critical role in this adjustment. Zinc finger proteins are involved in responding to light stimuli, and the plant requires shade for initial development, with an appropriate light response being crucial for healthy growth and development. Both protein types are involved in post-embryonic development and anatomical structure formation, essential for plant viability and environmental adaptation. Therefore, these genes are of great importance for study due to their importance in fundamental biological processes that support plant development and reproductive success.

Similar genes were also identified as critical for regulating flowering and anatomical development in the oil palm *Elaeis guineensis*<sup>91</sup>. Specifically, the positive regulation of the GIGANTEA gene in UFES\_250 may indicate a photoperiod-specific adaptive response, promoting flowering under certain light conditions<sup>92–95</sup>. Meanwhile, the negative regulation of zinc finger protein genes could be associated with mechanisms adjusting the plant's

circadian cycle<sup>96</sup>. Growth conditions specific to UFES\_250, such as light intensity and duration, as well as soil and climatic factors, may influence this differential regulation compared to SM plants<sup>91</sup>. Therefore, the regulation of these genes in UFES\_250 may represent an adaptive response that provides reproductive or developmental advantages suited to the local environment.

The exclusive GOs identified in roots are related to cellular protein modulation, cell communication, vacuole function, plasma membrane, and lipid binding processes. GOs involved in stress response, abiotic stimuli, and chemical stimuli were the most prominent, appearing frequently across the analyses and present in leaf and root tissues. The most significant DEGs involved in these processes included TPS, Hsp70, aquaporin PIP2, ClpB, SERK, and WRKY.

TPS (up-regulated in UFES\_250) is responsible for the biosynthesis of trehalose, a sugar critical in protecting cells against abiotic stresses such as dehydration and heat<sup>97</sup>. Hsp70 (up-regulated in UFES\_250) is an HSP expressed under stress conditions (strictly inducible), while some are present in cells under normal growth conditions and are not heat-inducible (constitutive or cognate)<sup>98</sup>. Aquaporin PIP2 (up-regulated in UFES\_250) facilitates water transport across the cell membrane, playing a vital role in the response to water stress<sup>99</sup>. ClpB (up-regulated in UFES\_250) is a member of the molecular chaperone family essential for chloroplast development and seedling viability, mediating the formation of internal thylakoid membranes and providing thermotolerance to chloroplasts during thermal stress<sup>76</sup>. SERK (up-regulated in UFES\_250) is involved in stress response signaling and somatic embryo development<sup>100</sup>. WRKY (down-regulated in UFES\_250) is a family of transcription factors regulating responses to biotic and abiotic stresses<sup>80</sup>. Studies with other palm species have shown similar genes playing critical roles in regulating stress responses and plant development<sup>101–103</sup>.

The increased expression of TPS, Hsp70, aquaporin PIP2, ClpB, and SERK in UFES\_250 may indicate an adaptation to more stressful environments, promoting effective survival under such conditions. Conversely, the decreased regulation of WRKY might suggest a specific adaptive response to different types of environmental stress. These traits could reflect unique adaptations in the SM or UFES varieties, leading to distinct gene expression patterns and responses to environmental stresses.

The gene expression profile of leaves and roots, as presented in this study, helps in understanding the interaction of *E. edulis* with its environment during the early developmental phase in the field. The findings here are novel and provide valuable information about conserved, differentially expressed, and highly expressed genes responsible for early development, often highlighting stress-response genes. Data normalization emphasized the high expression of essential genes during this critical developmental phase, offering a solid foundation for future management strategies of the species in various environmental conditions. The conserved genes identified in this study may also be applicable to other Arecaceae species. The reference analysis with the genome of *Elaeis guineensis* was significant, revealing variations among contrasting *E. edulis* matrices, even within conserved genes. These results support the hypothesis that conserved molecular mechanisms related to environmental responses are activated during the early developmental stage of *E. edulis* seedlings, demonstrating the species' adaptation to variable environmental conditions.

Future studies, such as *de novo* assembly, will enable the identification of species-specific genes, expanding our understanding of the *E. edulis* transcriptome and potentially uncovering new mechanisms related to development and adaptation. Ultimately, the data obtained corroborate existing literature and pave the way for practical applications in sustainable management and genetic improvement of the species.

## Conclusion

This study provided agronomically relevant genes, identified DEGs between morphologically divergent matrices, and proposed candidates for molecular markers in *E. edulis*. The transcriptome analysis of leaves and roots from genetically distinct backgrounds (UFES\_250 and SM) identified 32,000 genes, annotating 1,133 DEGs and other expressed genes, creating an important genomic database for the species. The results revealed gene expression profiles across different tissues and matrices, with greater expression divergence in the leaves. A higher number of exclusive genes was identified in the UFES\_250 genetic background, although overall expression was more pronounced in SM. DEGs such as GIGANTEA, zinc finger protein, CONSTANS, TPS, Hsp70, aquaporin PIP2, ClpB, SERK, and WRKY were identified, influencing matrices' development and morphological differences. The study also highlighted DEGs as potential molecular markers with applications in genetic improvement, while identifying genes with relevant molecular functions for further research. These findings contribute to understanding *E. edulis* adaptations and promote strategies for improving the species' productivity and resistance to adverse environmental conditions.

## Data availability

The RNAseq datasets generated during the current study are available in the Short Read Archive on the NCBI database. The accession number is PRJNA1200110. Data are also provided within the manuscript or supplementary information files.

Received: 19 December 2024; Accepted: 6 May 2025

Published online: 02 July 2025

## References

1. Henderson, A., Galeano, G. & Bernal, R. *Field Guide To the Palms of the Americas* (Princeton University Press, 2019).
2. Martinelli, G. & Moraes, M. A. *Livro Vermelho Da Flora Do Brasil* (Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, 2013).
3. Galetti, M. & Aleixo, A. Effects of palm heart harvesting on avian frugivores in the Atlantic rain forest of Brazil. *J. Appl. Ecol.* **35**, 286–293 (1998).



4. Pereira, A. G. et al. Patterns of genetic diversity and structure of a threatened palm species (*Euterpe Edulis* Arecaceae) from the Brazilian Atlantic forest. *Heredity (Edinb)*. **129**, 161–168 (2022).
5. Coelho, G. M. et al. Genetic structure among morphotypes of the endangered Brazilian palm *Euterpe Edulis* Mart (Arecaceae). *Ecol. Evol.* **10**, 6039–6048 (2020).
6. Carvalho, M. S. et al. Genetic diversity and population structure of euterpe *Edulis* by REML/BLUP analysis of fruit morphology and microsatellite markers. *Crop Breed. Appl. Biotechnol.* **20**, 1–9 (2020).
7. Vannuchi, N., Jamar, G., Pisani, L., Braga, A. R. C. & de Rosso, V. V. Chemical composition, bioactive compounds extraction, and observed biological activities from Jussara (*Euterpe edulis*): The exotic and endangered Brazilian superfruit. *Compr. Rev. Food Sci. Food Saf.* **20**, 3192–3224 (2021).
8. Klein, R. M. Ecology of the flora and vegetation of the Vale of Itajai (continuation). *Sellowia* **32**, 165–389 (1980).
9. Fisch, S. T. V., Nogueira Jr, L. R. & Mantovani, W. Fenologia reprodutiva de *Euterpe edulis* Mart. na Mata Atlântica (Reserva Ecológica do Trábiju, Pindamonhangaba-SP) 1. *Revista Biotécnicas*. **6**, n. 2 (2000).
10. Carvalho, L. M. J., Esmerino, A. A. & de Carvalho, J. L. V. Jussai (*Euterpe edulis*): a review. *Food Sci. Technol.* **42**, e08422, 1–11 (2022).
11. Campanili, M. & Prochnow, M. *Mata Atlântica: uma rede pela floresta*. (2006).
12. Nakazono, E., Costa, M. C. D., Futatsugi, K. & Paulilo, M. T. Crescimento inicial de *Euterpe Edulis* Mart. Em diferentes regimes de Luz. *Braz J. Bot.* **24**, 173–179 (2001).
13. de Mello, T. et al. Reserve mobilization and in vitro germination of *Euterpe Edulis* (Martius) seeds at different maturation stages. *Trees* **35**, 1–12 (2021).
14. Panza, V., Lainez, V. & Maldonado, S. Seed structure and histochemistry in the palm *Euterpe Edulis*. *Bot. J. Linn. Soc.* **145**, 445–453 (2004).
15. Zhang, Y. M. et al. Genome-Wide identification and evolutionary analysis of NBS-LRR genes from *Dioscorea rotundata*. *Front. Genet.* **11**, 484 (2020).
16. Kumar, S., Seem, K., Kumar, S. & Mohapatra, T. RNA-seq analysis reveals the genes/pathways responsible for genetic plasticity of rice to varying environmental conditions on direct-sowing and transplanting. *Sci. Rep.* **12**, 1–22 (2022).
17. Ekblom, R. & Galindo, J. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity (Edinb)*. **107**, 1–15 (2011).
18. Chalifa-Caspi, V. RNA-Seq in nonmodel organisms. *Methods Mol. Biol.* **2243**, 143–167 (2021).
19. Wang, Y. et al. Comprehensive RNA-Seq analysis pipeline for Non-Model organisms and its application in *Schmidtea mediterranea*. *Genes (Basel)* **14**, 989–1003 (2023).
20. Al-Mssallem, I. S. et al. Genome sequence of the date palm *Phoenix dactylifera* L. *Nat. Commun.* **4**, 1–7 (2013).
21. Singh, R. et al. Oil palm genome sequence reveals divergence of interfertile species in old and new worlds. *Nature* **500**, 335–339 (2013).
22. Zhao, H. et al. The chromosome-level genome assemblies of two rattans (*Calamus simplicifolius* and *Daemonorops jenkinsiana*). *Gigascience* **7**, 1–11 (2018).
23. Wang, S. et al. High-quality reference genome sequences of two coconut cultivars provide insights into evolution of monocot chromosomes and differentiation of fiber content and plant height. *Genome Biol.* **22**, 1–25 (2021).
24. Yang, Y. et al. Chromosome-scale genome assembly of areca palm (*Areca catechu*). *Mol. Ecol. Resour.* **21**, 2504–2519 (2021).
25. Francisconi, A. F. et al. Complete Chloroplast genomes and phylogeny in three *Euterpe* palms (*E. edulis*, *E. oleracea* and *E. precatoria*) from different Brazilian biomes. *PLoS One*. **17**, 1–26 (2022).
26. Darnet, E., Teixeira, B., Schaller, H., Rogez, H. & Darnet, S. Elucidating the mesocarp Drupe transcriptome of Açai (*Euterpe Oleracea* Mart.): an Amazonian tree palm producer of bioactive compounds. *Int. J. Mol. Sci.* **24**, 9315–9329 (2023).
27. Strickler, S. R., Bombarely, A. & Mueller, L. A. Designing a transcriptome next-generation sequencing project for a nonmodel plant species. *Am. J. Bot.* **99**, 257–266 (2012).
28. Zhao, Q. Y., Gratten, J., Restuadi, R. & Li, X. Mapping and differential expression analysis from short-read RNA-Seq data in model organisms. *Quant. Biol.* **4**, 22–35 (2016).
29. Fitch, W. M. Distinguishing homologous from analogous proteins. *Syst. Zool.* **19**, 99–113 (1970).
30. Rekik, I. et al. Transcriptome assembly and abiotic related gene expression analysis of date palm reveal candidate genes involved in response to cadmium stress. *Comp. Biochem. Physiol. Part. - C Toxicol. Pharmacol.* **225**, 108569–108580 (2019).
31. Xiao, Y. et al. Genome-wide identification of WRKY genes and their expression profiles under different abiotic stresses in *Elaeis guineensis*. *PLoS One*. **12**, 1–18 (2017).
32. Bhargava, A. et al. Identification of cytokinin-responsive genes using microarray meta-analysis and RNA-seq in *Arabidopsis*. *Plant. Physiol.* **162**, 272–294 (2013).
33. Hübner, S., Korol, A. B. & Schmid, K. J. RNA-Seq analysis identifies genes associated with differential reproductive success under drought-stress in accessions of wild barley *Hordeum spontaneum*. *BMC Plant. Biol.* **15**, 1–14 (2015).
34. Li, Y. et al. Comparative transcriptomic analysis reveals gene expression associated with cold adaptation in the tea plant *Camellia sinensis*. *BMC Genom.* **20**, 1–17 (2019).
35. Mota, A. P. Z. et al. Evolutionarily conserved plant genes responsive to root-knot nematodes identified by comparative genomics. *Mol. Genet. Genomics*. **295**, 1063–1078 (2020).
36. Reboledo, G., Agorio, A. D., Vignale, L. & Batista-García, R. A. Ponce de León, I. Transcriptional profiling reveals conserved and species-specific plant defense responses during the interaction of *Physcomitrium patens* with *Botrytis cinerea*. *Plant. Mol. Biol.* **107**, 365–385 (2021).
37. Hao, Z. et al. Conserved, divergent and heterochronic gene expression during *Brachypodium* and *Arabidopsis* embryo development. *Plant. Reprod.* **34**, 207–224 (2021).
38. Canal, G. B. et al. Genomic studies of the additive and dominant genetic control on production traits of *Euterpe Edulis* fruits. *Sci. Rep.* **13**, 1–12 (2023).
39. Wendt, T. et al. An evaluation of the species boundaries of two putative taxonomic entities of *Euterpe* (Arecaceae) based on reproductive and morphological features. *Flora Morphol. Distrib. Funct. Ecol. Plants*. **206**, 144–150 (2011).
40. Mengarda, L. H. G., Canal, G. B., Carvalho, M. S., Ferreira, A. & Ferreira, M. F. Genetic diversity of Juçara palm: an alternative for selection and conservation in cash crop for fruit production. *Front. Glob Chang.* **5**, 1–11 (2022).
41. Wickham, H. ggplot2. Wiley Interdiscip. Rev. Comput. Stat. **3**, 180–185 (2011).
42. Chambers, J. M. Errata and notes for software for data analysis: programming with R. *Springer*. **2**, 499–504 (2009).
43. Guzman, F., Kulcheski, F. R., Turchetto-Zolet, A. C. & Margis, R. De Novo assembly of *Eugenia uniflora* L. transcriptome and identification of genes from the terpenoid biosynthesis pathway. *Plant. Sci.* **229**, 238–246 (2014).
44. de Carpinetti, P. Efficient method for isolation of high-quality RNA from *Psidium guajava* L. Tissues. *PLoS One*. **16**, 1–19 (2021).
45. Bogdanova, E. A. et al. Normalization of Full-Length-Enriched cDNA. *Methods Mol. Biol.* **729**, 85–98 (2011).
46. Andrews, S. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics. (2010). Available at: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
47. Bolger, A. M., Lohse, M., Usadel, B. & Trimmomatic A flexible trimmer for illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
48. Benson, D. A. et al. *GenBank Nucleic Acids Res.* **46**, D41–D47 (2018).

49. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *12*, n. 4, pp. 357–360 *Nature methods* (2015).
50. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with bowtie 2. *Nat. Methods*. **9**, 357–359 (2012).
51. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, stringtie and ballgown. *Nat. Protoc.* **11**, 1650–1667 (2016).
52. Chen, Y., Lun, A. T. L. & Smyth, G. K. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research* **5**, 1438 (2016).
53. Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
54. Dunkler, D., Sánchez-Cabo, F. & Heinze, G. Statistical analysis principles for Omics data. *Bioinformatics for Omics Data: Methods and Protocols*, pp. 113–131 (2011).
55. Blighe, K., Rana, S., Lewis, M. & EnhancedVolcano Publication-ready volcano plots with enhanced colouring and labeling. at (2018). <https://github.com/kevinblighe/EnhancedVolcano>
56. Aleksander, S. A. et al. The gene ontology knowledgebase in 2023. *Genetics* **224**, 1–14 (2023).
57. Sherman, B. T. et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* **50**, W216–W221 (2022).
58. McCarthy, F. M. et al. AgBase: A functional genomics resource for agriculture. *BMC Genom.* **7**, 1–13 (2006).
59. Balti, I. et al. Identification of conserved genes linked to responses to abiotic stresses in leaves among different plant species. *Funct. Plant. Biol.* **48**, 54–71 (2020).
60. Lin, H. C. et al. Transcriptome analysis during somatic embryogenesis of the tropical monocot *Elaeis guineensis*: evidence for conserved gene functions in early development. *Plant. Mol. Biol.* **70**, 173–192 (2009).
61. Rajesh, M. K. et al. De Novo assembly and characterization of global transcriptome of coconut palm (*Cocos nucifera* L.) embryogenic calli using illumina paired-end sequencing. *Protoplasma* **253**, 913–928 (2016).
62. Khan, F. S. et al. Transcriptional Regulations and Hormonal Signaling during Somatic Embryogenesis in the Coconut Tree: An Insight. *Forests* **14**, (2023).
63. Meira, F. S. et al. Differential expression of genes potentially related to the callogenesis and in situ hybridization of SERK gene in Macaw palm (*Acrocomia aculeata* Jacq. *Lodd Ex. Mart Protoplasma*. **261**, 89–101 (2024).
64. Yaish, M. W. et al. Genome-wide expression profiling in leaves and roots of date palm (*Phoenix dactylifera* L.) exposed to salinity. *BMC Genom.* **18**, 1–17 (2017).
65. Lu, L. et al. Integrated transcriptomic and metabolomics analyses reveal molecular responses to cold stress in coconut (*Cocos nucifera* L.) seedlings. *Int. J. Mol. Sci.* **24**, 1–40 (2023).
66. Yusheng, Z., Jin, Z., Xingqi, Y., Jixin, Z. & Dongdong, L. Molecular mechanism of the effect of PEG and UV stress on the metabolism of Arecoline in *Areca catechu* callus. *Sci. Hortic. (Amsterdam)*. **321**, 112299 (2023).
67. Blazek, E., Mittler, G. & Meisterernst, M. The mediator of RNA polymerase II. *Chromosoma* **113**, 399–408 (2005).
68. Nakashima, K., Takasaki, H., Mizoi, J., Shinozaki, K. & Yamaguchi-Shinozaki, K. NAC transcription factors in plant abiotic stress responses. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1819**, 97–103 (2012).
69. Zhang, L. et al. Three WRKY transcription factors additively repress abscisic acid and Gibberellin signaling in aleurone cells. *Plant. Sci.* **236**, 214–222 (2015).
70. Hojjati, M. R., Li, Z. & Jiang, X. C. Serine palmitoyl-CoA transferase (SPT) deficiency and sphingolipid levels in mice. *Biochim. Biophys. Acta - Mol. Cell. Biol. Lipids*. **1737**, 44–51 (2005).
71. Xie, Z. et al. Annotations and functional analyses of the rice WRKY gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells. *Plant. Physiol.* **137**, 176–189 (2005).
72. Zhang, J., Peng, Y. & Guo, Z. Constitutive expression of pathogen-inducible OsWRKY31 enhances disease resistance and affects root growth and auxin response in Transgenic rice plants. *Cell. Res.* **18**, 508–521 (2008).
73. Brandalise, M., Maia, I. G., Borecký, J., Vercesi, A. E. & Arruda, P. Overexpression of plant uncoupling mitochondrial protein in Transgenic tobacco increases tolerance to oxidative stress. *J. Bioenerg Biomembr.* **35**, 203–209 (2003).
74. Sweetlove, L. J. et al. Mitochondrial uncoupling protein is required for efficient photosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 19587–19592 (2006).
75. Begcy, K. et al. An Arabidopsis mitochondrial uncoupling protein confers tolerance to drought and salt stress in Transgenic tobacco plants. *PLoS One* **6**, 23776–23790 (2011).
76. Lee, U. et al. The Arabidopsis ClpB/Hsp100 family of proteins: chaperones for stress and Chloroplast development. *Plant. J.* **49**, 115–127 (2007).
77. Wierenga, R. K., Kapetanious, E. G. & Venkatesan, R. Triosephosphate isomerase: A highly evolved biocatalyst. *Cell. Mol. Life Sci.* **67**, 3961–3982 (2010).
78. Rasool, K. G. et al. Differential proteomic analysis of date palm leaves infested with the red palm weevil (Coleoptera: Curculionidae). *Fla. Entomol.* **101**, 290–298 (2018).
79. Phukan, U. J., Jeena, G. S. & Shukla, R. K. WRKY transcription factors: molecular regulation and stress responses in plants. *Front. Plant. Sci.* **7**, 1–14 (2016).
80. Rushton, D. L. et al. WRKY transcription factors: key components in abscisic acid signalling. *Plant. Biotechnol. J.* **10**, 2–11 (2012).
81. Wang, C. et al. Transcriptome analysis and differential gene expression profiling of Wucai (*Brassica campestris* L.) in response to cold stress. *BMC Genom.* **23**, 1–16 (2022).
82. Bhattarai, S. et al. Transcriptomic analysis of differentially expressed genes in leaves and roots of two alfalfa (*Medicago sativa* L.) cultivars with different salt tolerance. *BMC Plant. Biol.* **21**, 1–16 (2021).
83. Lin, W. et al. Comparative transcriptome analyses revealed differential strategies of roots and leaves from Methyl jasmonate treatment *Baphicacanthus cusia* (Nees) Bremek and differentially expressed genes involved in Tryptophan biosynthesis. *PLoS One*. **14**, 1–17 (2019).
84. Fu, Y. et al. A comparison of shared patterns of differential gene expression and gene ontologies in response to water-stress in roots and leaves of four diverse genotypes of *Lolium* and *Festuca* spp. Temperate pasture grasses. *PLoS One*. **16**, 1–21 (2021).
85. Lawrence, B. T., Calle, A., Saski, C. A. & Melgar, J. C. Differential gene expression patterns in Peach roots under Non-Uniform soil conditions in response to organic matter. *Genes (Basel)* **15**, 70–84 (2024).
86. Wang, N. et al. Evaluation of differentially expressed genes in leaves vs. Roots subjected to drought stress in flax (*Linum usitatissimum* L.). *Int. J. Mol. Sci.* **24**, 12019–12033 (2023).
87. Hayama, R., Yokoi, S., Tamaki, S., Yano, M. & Shimamoto, K. Adaptation of photoperiodic control pathways produces short-day flowering in rice. *Nature* **422**, 719–722 (2003).
88. Fowler, S. et al. GIGANTEA: A circadian clock-controlled gene that regulates photoperiodic flowering in Arabidopsis and encodes a protein with several possible membrane-spanning domains. *EMBO J.* **18**, 4679–4688 (1999).
89. Park, D. H. et al. Control of circadian rhythms and photoperiodic flowering by the Arabidopsis GIGANTEA gene. *Science*. **285**, 1579–1582 (1999).
90. Suárez-López, P. et al. CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis. *Nature* **410**, 1116–1120 (2001).
91. Low, E. T. L. et al. Oil palm (*Elaeis guineensis* Jacq.) tissue culture ESTs: identifying genes associated with callogenesis and embryogenesis. *BMC Plant. Biol.* **8**, 1–19 (2008).

92. Mizoguchi, T. et al. Distinct roles of GIGANTEA in promoting flowering and regulating circadian rhythms in Arabidopsis. *Plant. Cell.* **17**, 2255–2270 (2005).
93. Sawa, M. & Kay, S. A. GIGANTEA directly activates Flowering Locus T in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 11698–11703 (2011).
94. Andrés, F. & Coupland, G. The genetic basis of flowering responses to seasonal cues. *Nat. Rev. Genet.* **13**, 627–639 (2012).
95. Almeida, F. A. N. *De. Population Genomics and Phylogeography of Euterpe Edulis* (Universidade Federal do Espírito Santo, 2023).
96. Valverde, F. CONSTANS and the evolutionary origin of photoperiodic timing of flowering. *J. Exp. Bot.* **62**, 2453–2463 (2011).
97. Elbein, A. D., Pan, Y. T., Pastuszak, I. & Carroll, D. New insights on Trehalose: A multifunctional molecule. *Glycobiology* **13**, 17–27 (2003).
98. Vierling, E. The roles of heat shock proteins in plants. *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* **42**, 579–620 (1991).
99. Maure, C., Verdoucq, L., Luu, D. T. & Santoni, V. Plant Aquaporins: membrane channels with multiple integrated functions. *Annu. Rev. Plant. Biol.* **59**, 595–624 (2008).
100. Hecht, V. et al. The arabidopsis SOMATIC EMBRYOGENESIS RECEPTOR KINASE 1 gene is expressed in developing ovules and embryos and enhances embryogenic competence in culture. *Plant. Physiol.* **127**, 803–816 (2001).
101. Lee, F. C., Yeap, W. C., Appleton, D. R., Ho, C. L. & Kulaveerasingam, H. Identification of drought responsive Elaeis guineensis WRKY transcription factors with sensitivity to other abiotic stresses and hormone treatments. *BMC Genom.* **23**, 1–17 (2022).
102. Osorio-Montalvo, P. et al. A peak in global DNA methylation is a key step to initiate the somatic embryogenesis of coconut palm (*Cocos nucifera* L.). *Plant. Cell. Rep.* **39**, 1345–1357 (2020).
103. Verde, G. L. et al. Members of the WRKY gene family are upregulated in Canary palms attacked by red palm weevil. *Arthropod Plant. Interact.* **13**, 109–116 (2019).

## Acknowledgements

Financial support for this study was provided by the Conselho Nacional de Pesquisa (CNPq, Brazil) - Finance Code 200329/2021-9, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil) - Finance Code 001, and the Fundação de Amparo à Pesquisa do Espírito Santo (FAPES, Vitória - ES, Brazil) - Finance Code 483/2021. Thanks to the companies Açai Juçara and Bonaloti for supporting research development and to Pedro and Vicente Bortoloti and their families, who are the owners of the managed area. We would like to thank all colleagues at the Biometrics and Genetics and Plant Improvement laboratories from the Universidade Federal do Espírito Santo for helping to process and organize the samples.

## Author contributions

L.M.C. analyzed the data, co-led the manuscript writing, and created the images. F.A.N.A. designed the research, analyzed the data, co-led the manuscript writing, and created the images. V.S.F. assisted with molecular biology work and created the images. G.B.C. was involved in prospecting, collecting, and selecting vegetation material, as well as in phenotypic analysis. L.H.G.M. was involved in prospecting and collecting vegetation material, as well as in phenotypic analysis. A.F. was responsible for the project conception and design and securing funding. M.F.S.F. contributed to the project conception and design, secured funding, and made critical contributions to the manuscript. All authors reviewed the manuscript.

## Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-01436-x>.

**Correspondence** and requests for materials should be addressed to M.F.S.F.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025