Genome
An International Journal
for Genetics and Genomics

OPEN ACCESS | Research Article

The genome sequence of the açaí berry (*Euterpe oleracea* Mart.) and RNA-Seq analysis of the fruit ripening

Maria Silvanira Ribeiro Barbosa[a,b], Sávio de Souza Costa[a], Davi Josué Marcon[a], Adan Rodrigues de Oliveira[a], Lucas da Silva e Silva[b], Maria Paula Cruz Schneider[a,b], Juarez Antônio Simões Quaresma[c], Diego Assis das Graças[a,b], Adonney Allan de Oliveira Veras[a], Maria do Socorro Padilha de Oliveira[d], Simone de Miranda Rodrigues[d], Elisa Ferreira Moura[d], Artur Silva[a,b], and Rafael Azevedo Baraúna[a,b]

[a]Laboratório de Engenharia Biológica, Espaço Inovação, Parque de Ciência e Tecnologia Guamá, 66075-750, Belém, PA, Brazil; [b]Centro de Genômica e Biologia de Sistemas, Instituto de Ciências Biológicas, Universidade Federal do Pará, 66075-110, Belém, PA, Brazil; [c]Núcleo de Medicina Tropical, Instituto de Ciências da Saúde, Universidade Federal do Pará, 66055-240, Belém, PA, Brazil; [d]Embrapa Amazônia Oriental, Tv Doutor Enéas Pinheiro, 66095-903, Belém, PA, Brazil

Corresponding author: **Rafael Azevedo Baraúna** (email: rabarauna@ufpa.br)

Abstract

Euterpe oleracea Mart. is a perennial wetland palm native to the eastern Amazon that produces a fruit called açaí, which is used to prepare a beverage of great social and economic importance to the region. We have sequenced the açaí genome using Oxford Nanopore long-read platform, and we have evaluated the gene expression of the fruit during the ripening process in both the white and purple varieties via RNA-Seq analysis. The assembled genome had a size of 3 066 969 163 bp, with the longest contig displaying 430 834 bp, an N50 of 54 646, and a GC content of 49.97%. The transcriptome analysis identified crucial genes that regulate the production of anthocyanins in purple açaí. For example, the upregulation of the flavonoid 3′5′-hydroxylase enzyme plays a crucial role in the synthesis of purple anthocyanins, such as delphinidin 3-glucoside, between 70 and 130 days after flowering. In contrast, a general downregulation of the enzymes responsible for the initial stages of the anthocyanin synthesis was observed in the white variety. Our findings provide a valuable contribution to the understanding of the molecular mechanisms that regulate the ripening process of açaí fruits and present the first genome of *E. oleracea*.

Key words: açaí, euterpe, transcriptomic, genomic, amazon

Introduction

Euterpe oleracea Mart. is a perennial wetland palm of the family Arecaceae native to the eastern Amazon region and the main bioeconomy product from the Brazilian amazon (Yamaguchi et al. 2015). The tree grows in multiple stems and produces a berry-like purple fruit called açaí, which is one of the most commercialized and appreciated products of regional agriculture. Some trees are capable of producing fruits with a low anthocyanin content, called white açaí (Oliveira et al. 2019; Laurindo et al. 2023). The molecular processes that control the production of this type of fruit are unknown, although a previous study suggests some important metabolic pathways (Darnet et al. 2023). The açaí beverage is produced by mechanical extraction of the fruit's mesocarp with water. The Brazilian state of Pará, located in the western Amazon, accounts for 96% of the açaí cultivation area in the country (about 188 000 ha) (Oliveira et al. 2019; Bezerra 2025).

Two other palm trees of the same genera have been documented in the Americas. *Euterpe precatoria* Mart. is a single-stemmed palm that naturally occurs in western Amazonia and Central America, while *Euterpe edulis* Mart. is found in the Atlantic Forest and evolved from a common ancestor with the Amazonian species (Pichardo-Marcano et al. 2019). Vicariant speciation due to the formation of the dry forest belt (emergence of Cerrado and Caatinga) is the main explanation for the current dispersal of these species. Thus, the Amazon is the geographical origin of these palm trees (Antonelli et al. 2009). Both *E. edulis* and *E. precatoria* have lower organoleptic quality and fruit pulp yield compared to *E. oleracea* and therefore are used to blend with *E. oleracea* for açaí beverages. Molecular studies using RAPD and SSR markers have shown that *E. oleracea* has high genetic variability (Oliveira et al. 2007). Using flow cytometry, the genomes of *E. edulis*, *E. oleracea*, and *E. precatoria* were predicted to be 4, 4.13, and 4.61 Gpb in size, respectively (Oliveira et al. 2016). All three species have a $2n = 36$ karyotype, with significant differences in chromosome morphology (Oliveira et al. 2016). This is the largest karyotype in the subfamily Arecoideae, a characteristic observed only in New World species (Oliveira et al. 2016). The plastomes of *E. edulis* and *E. oleracea* have already been sequenced (Lopes et al. 2021). Lopes et al. (2021) have shown that more than half of the protein-coding genes in the plastomes have undergone positive selection.

Transcriptomic studies have been widely applied across plant species of economic and environmental relevance. In *Vaccinium uliginosum* (Ericaceae), a high-value wild fruit, transcriptome analysis enabled the annotation of 42 837 genes, revealing that the absence of anthocyanins in white-fruited varieties is linked to the downregulation of specific genes (Yang et al. 2018). Similarly, in tomato fruits, transcriptional profiling identified 31 905 transcripts, with enrichment of carbon metabolism and amino acid biosynthesis pathways during ripening (Zhao et al. 2024). This developmental pattern is also observed in strawberry (*Fragaria* × *ananassa*), where the transition between green, white, and red stages involves the differential expression of 6608 genes (Galli et al. 2019). In contrast to these well-established species, genomic studies on açaí palm (*E. oleracea*) remain scarce. Recently, using the Ion Proton sequencing platform, Darnet et al. (2023) generated 255 million reads, identifying 22 486 transcripts.

From a genomic perspective, recent studies have expanded knowledge of the genus *Euterpe*. Francisconi et al. (2022) characterized the complete plastid genomes of *E. edulis*, *E. oleracea*, and *E. precatoria*, revealing the typical quadripartite structure of angiosperm chloroplasts, with two inverted repeat regions (IRA and IRB) flanking the large and small single-copy regions (LSC and SSC). Phylogenetic analyses indicated a closer relationship between *E. oleracea* and *E. precatoria*, while nuclear and mitochondrial genome analyses revealed 1077 SNPs and high interspecific divergence, particularly between *E. edulis* and *E. precatoria*, suggesting that, despite the few structural differences among plastomes, point mutations are sufficient to distinguish *E. edulis* from the other species in the genus. Complementarily, Francisconi et al. (2025), analyzing 160 individuals of *E. oleracea* from the eastern and western Amazon, identified signals of selection in the eastern region, characterized by a higher number of outlier SNPs and distinct gene profiles, possibly associated with more intense extraction pressure. Taken together, these studies reinforce the importance of integrated genomic approaches for understanding the diversity, population structure, and molecular basis of the açaí palm.

Due to the global socio-economic importance of açaí, especially the species *E. oleracea*, several research projects have been carried out to develop new technologies to support the fruit production chain. For example, Embrapa (the Brazilian Agricultural Research Corporation) launched a dryland cultivar with high productivity and an off-season fruit-producing cultivar in 2004 and 2019, respectively. The açaí fruit is rich in α-tocopherol, fiber, lipids, polyphenols (anthocyanins), and minerals (Rufino et al. 2010). The fruit extract showed anti-inflammatory, antinociceptive, and antioxidant activity, mainly due to polyphenols (especially anthocyanins and flavonoids) (Mertens-Talcott et al. 2008; Kang et al. 2012). These properties make açaí not only a globally appreciated beverage, but also one of the most important bioeconomy assets of the Amazonian biodiversity for the formulation of cosmetic and pharmaceutical products.

In this study, we present the first version of the *E. oleracea* genome, using the long-read DNA sequencing platform PromethION 2 Solo (Oxford Nanopore Technologies). We sequenced a representative strain of the eastern Amazon maintained in a plant nursery at Embrapa Amazônia Oriental. In addition, to improve gene annotation and shed light on the molecular process of fruit ripening in two açaí varieties (white and purple fruits), a comparative RNA-Seq analysis was performed at the three main stages of fruit ripening.

Materials and methods

Sampling

The strains of *E. oleracea* used in this study were maintained by the Empresa Brasileira de Pesquisa Agropecuária (Embrapa)—Amazônia Oriental. Access to Brazilian genetic heritage was registered in the SisGen web platform under No. A740F98. For the RNA-Seq analysis, the fruits of the two varieties were collected from one plant of each variety—one white açaí plant (accession 294-7) and one purple açaí plant (accession 33.1)—both from the açaí germplasm bank maintained at the experimental area of Embrapa Amazônia Oriental. A single plant per variety was intentionally selected to minimize excess genetic variation between samples. The following stages of ripeness were analyzed: unripe (70 days after flowering—70 DAF), intermediate (130 DAF), and ripe (160 DAF). In each group, biological triplicates were analyzed, totalizing 18 samples. The fruits were collected aseptically and immediately frozen in liquid nitrogen until RNA extraction. For the genome sequence, the DNA was extracted from a fragment of the root of a strain maintained at a plant nursery of the Embrapa Amazônia Oriental.

DNA and RNA extraction

The açaí fruits were macerated in liquid nitrogen, and total RNA was extracted using the RNeasy PowerPlant kit (Qiagen) according to the manufacturer's instructions. The extracted RNA was quantified using a Qubit 2.0 fluorometer (Invitrogen). Before preparing the libraries, the mRNAs were captured using the Dynabeads™ mRNA Purification kit (Thermo Fisher Scientific) according to the manufacturer's instructions. For genomic DNA extraction, 5 g of root was macerated in liquid nitrogen together with polyvinylpyrrolidone (PVP) and 2-mercaptoethanol. Total DNA was obtained using an extraction solution and chloroform-isoamyl alcohol method. Quantification was performed on a Qubit 2.0 fluorometer (Invitrogen), and DNA integrity was observed on a 1% agarose gel. NanoDrop was used to assess the purity of the extract.

Oxford nanopore long-read sequencing

Sequencing was performed on the PromethION 2 Solo platform (Oxford Nanopore Technologies). The genomic library was prepared with 1 ug of DNA according to the Ligation sequencing DNA v14 protocol (SQK-LSK114). DNA strands were repaired using Ultra End_prep Enzyme Mix, followed by purification and ligation of adapters to the fragments. The library concentration was adjusted to 20 fmol, and sequencing was performed using the R10 flow cell, according to the manufacturer's protocol.

Genome assembly and annotation

Trimmomatic v.0.39 was used to remove low-quality bases in the raw reads. The main parameters used in this step were SE, -phred33, LEADING:20, and -threads (Bolger et al. 2014). The BWA (Burrows-Wheeler Aligner) software v0.7.17 (Li and Durbin 2009) was used to map the trimmed reads against the chloroplast reference sequence of *E. oleracea* (accession No. NC_057603.1). Mapped reads were removed from the analysis.

The Shasta software v.0.11.1 (Shafin et al. 2020) was used to perform the genome assembly. The parameters used were –config, –assemblyDirectory, –threads, -Assembly.mode, –ReadGraph.strandSeparationMethod, –memoryBacking, and finally, –memoryMode. This process was executed on the Apollo 2000 computational cluster, located at the High-Performance Computing Center of the Federal University of Pará, using 200 processing cores and 480 TB of memory. Quast software v.5.2.0 (Gurevich et al. 2013) was used to evaluate the assembly results. The main metrics observed were N50, largest contig, total number of contigs, and total number of bases. Finally, gene completeness was assessed using BUSCO software, using the parameters "genome" and "-auto-lineage-euk" (Simão et al. 2015).

For genome annotation, the machine learning-based software Helixer v.0.3.4 with the land plant option was used (Stiehler et al. 2021). This method facilitated the prediction of genic regions, CDS, and UTR regions. Functional enrichment and COG prediction were conducted using emapper v.2.1.12 (Cantalapiedra et al. 2021) with the eggNOG database v.5.0.2, employing an ortholog genes seed e-value of 1e-2 and utilizing the genome method. All KEGG enrichments were retrieved for pathways through the web tool KEGG Mapper Reconstruct (https://www.genome.jp/kegg/mapper/reconstruct.html). The comparative genomic analysis was performed using the following species: *Phoenix dactylifera* (GCF_000413155.1), *Cocos nucifera* (GCA_008124465.1), *Elaeis guineensis* (GCA_000442705.2), and *Elaeis oleifera* (GCA_000441515.2). The figures were plotted using R and RStudio.

RNA-Seq analysis

The cDNA library was constructed using the Ion Total RNA Seq v2 kit (Thermo Fisher Scientific) according to the manufacturer's instructions. To this end, mRNA was fragmented using RNase III for 3 min, thereby generating fragments with a size of 200 bp. Subsequently, the reactions were purified using the Magnetic Bead Cleanup module. Thereafter, the mRNAs were ligated to adapters, followed by reverse transcription and purification. The library was then quantified using the QuantStudio™ 12K Flex system (Thermo Fisher Scientific), and concentrations were adjusted to 50 pmol/L. The chip 550 was automatically prepared on the Ion Chef™ Instrument (Thermo Fisher Scientific). Sequencing was carried out on the Ion GeneStudio™ S5 Plus platform (Thermo Fisher Scientific).

De novo transcriptome assembly and annotation

Unknown nucleotides (Ns), low-quality nucleotides, and hexamers were removed from the analysis. The treatment was performed using Trimmomatic v.0.39 (Bolger et al. 2014) removing reads with an average quality lower than 25, and final length below 50 bp. Quality control was conducted using FastQC.

De novo assembly was performed using Trinity v.2.15 (Grabherr et al. 2011). The assembly was conducted in three modes: (i) assembly of all samples into a single file; (ii) assembly of the white and purple açaí samples separately; (iii) assembly every two samples, according to pairwise comparison for differential expression analysis: 70 DAF versus 130 DAF and 130 DAF versus 160 DAF. To corroborate the significance of the gene that was found to be differentially expressed during the ripening process, a gene expression analysis between varieties at the same stage was also performed. The Trinity Statistics tool was utilized to assess assembly statistics.

The TransDecoder v.5.4 (available at https://github.com/TransDecoder/TransDecoder) was used for peptide prediction according to the following parameter: training with the top longest Open Reading Frames (ORFs) for each assembler used in the study. Gene prediction was performed with EggNOG-mapper database v.2.1 (Huerta-Cepas et al. 2019). Initial steps involved the acquisition and preparation of target genomic sequences for subsequent analysis. These sequences were then subjected to the EggNOG database to identify orthologs and paralogs, facilitating a thorough examination of gene groups and their evolutionary relationships. Subsequently, functional annotation of the predicted genes was conducted using the information provided by EggNOG. GhostKOALA web server v.3.1 was used for KEGG Ontology (KO) assignment and KEGG pathway mapping (Kanehisa et al. 2016).

Differential expression analysis

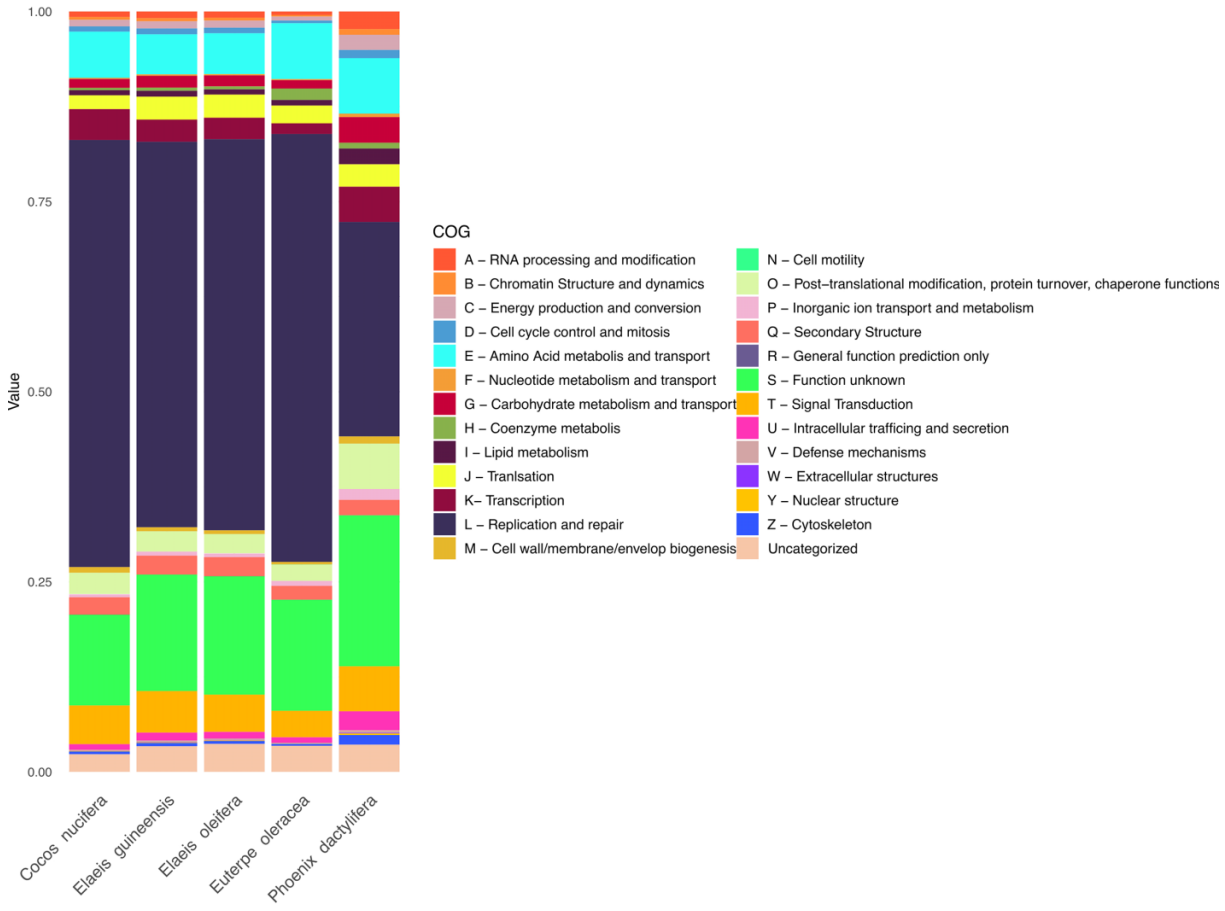
The Trinity pipeline was used to analyze differential expression. The Align Reads and Estimate Abundance tools were employed to perform abundance estimation using the Salmon method. Subsequently, isoform files were obtained. Following this step, mapping tables were generated through the Build Expression Matrix tool, employing the Salmon abundance estimation method. Differential Expression (DE) analysis was executed by estimating RNA-Seq fragment isoform counts using DESeq2 v.2.11.40 (Love et al. 2014). The DESeq2 package was used to extract and cluster differentially expressed transcripts with a p-value cutoff of approximately 0.001. Differentially expressed genes (DEGs) were filtered based on Log2FoldChange values, with a ≥ 2 cutoff to identify upregulated genes and ≤ 2 to identify downregulated genes. Graphical representations were obtained using the RStudio v.2024.12, with the DESeq2 (Love et al. 2014), edgeR (Robinson et al. 2010), limma (Ritchie et al. 2015), and ggplot2 (Wickham 2016) libraries.

Results

Whole genome sequencing

After long-read sequencing and raw data processing, a total of 12 202 098 reads were obtained, constituting 99.99% of all sequenced reads. The BWA (Burrows-Wheeler Aligner) software was used to identify and remove reads that exhibited a match with the plastome sequence of *E. oleracea*

Fig. 1. Distribution of genes in COG functional categories for each palm tree genome. The COG categorization took into account 26 different functional groups. The columns represent relative values, and the uncategorized genes are included as the last category.



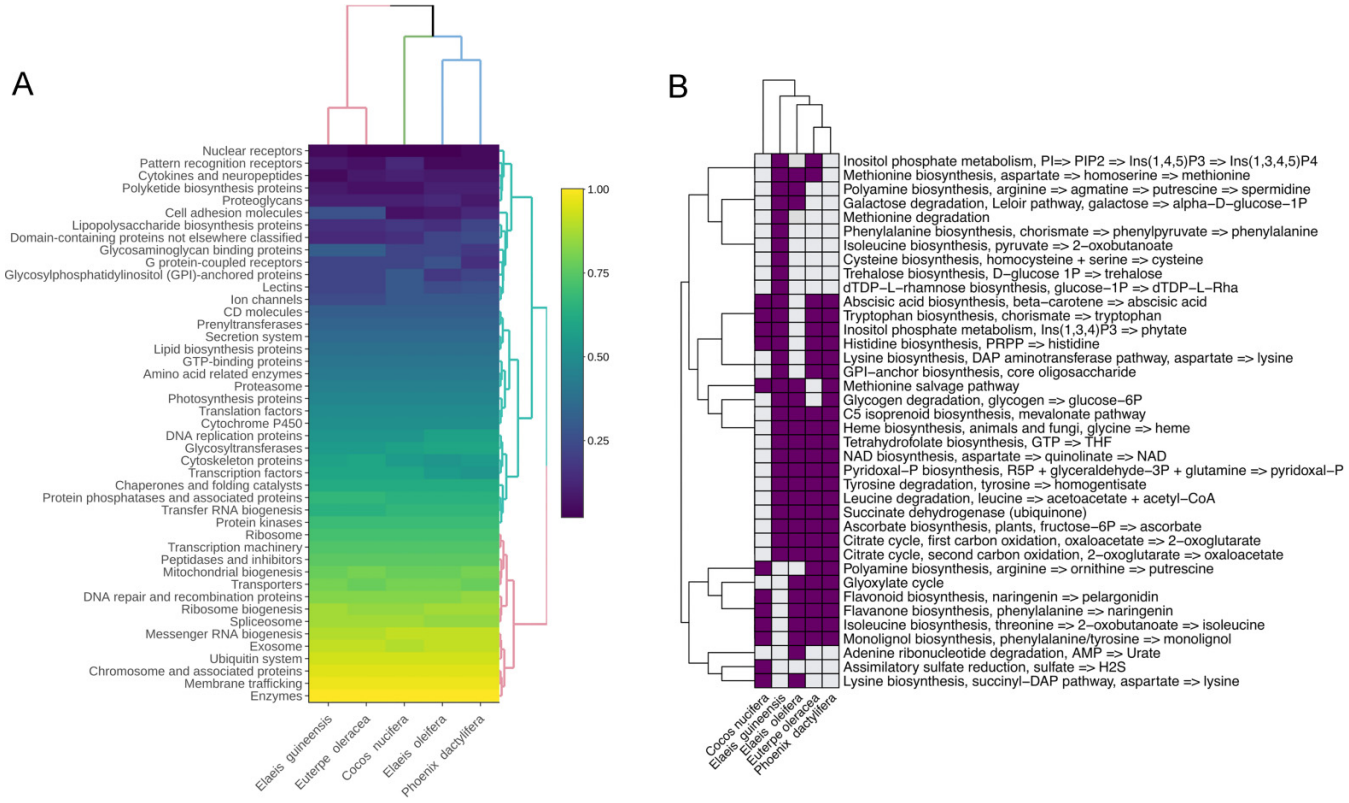
(NC_057603.1). A total of 64 802 reads aligned with the plastome reference and were consequently removed from the analysis, resulting in a dataset of 12 137 296 reads. Sequences smaller than 10 000 bp were discarded prior to assembly. Finally, the average read length was 17 931, and the N50 was 18 505. The assembled genome was assessed using QUAST software, revealing a size of 3 066 969 163 bp, with the longest contig displaying 430 834 bp, an N50 of 54 646, and a GC content of 49.97%. This size corresponds to approximately 75% of the expected açaí genome size, as determined by flow cytometry studies (Oliveira et al. 2016).

The açaí genome annotation through Helixer revealed 41.703 genes across the 3.1 Gbp genome. To understand the overall genome function of palm trees, we performed a COG-based functional annotation and identified a significant number of genes related to replication and repair in all five palm tree species. Notably, *P. dactylifera*, whose genome was sequenced and reported by Al-Mssallem et al. (2013), exhibited a reduced repertoire of genes in this functional category compared to the other species analyzed, while the other four species showed more than 50% of their genome related to the aforementioned COG (Fig. 1). A deeper analysis of the raw data revealed a high number of transposable elements in palm tree genomes, categorized as COG "L—Replication and repair", which biased the COG annotation. The second most abundant COG was "S—Function unknown", indicating a substantial genomic content without functional characterization in all palm trees analyzed. Interestingly, *P. dactylifera*, which has the smallest genome (573 Mbp) among the analyzed samples, displayed a significantly divergent profile, showing larger content related to intracellular trafficking, RNA processing, and carbohydrate and lipid metabolism (Fig. 1).

Aiming to classify açaí genes into functional hierarchies, we used KEGG gene enrichment to compare the five palm tree genomes by gene category enrichment (Fig. 2A). Our results showed a set of core cellular functions strongly enriched in all species, including mechanisms such as membrane trafficking, the ubiquitin system, and exosomes. Interestingly, the hierarchical categorization of genes grouped *E. oleracea* with *Elaeis guineensis*, the African oil palm, while the American oil palm, *Elaeis oleifera*, was grouped with *P. dactylifera*, one of the oldest and main staple crops in Southwest Asia and North Africa.

A detailed analysis of metabolic pathways (Fig. S1) revealed that the three palm trees with the largest genomes, *E. oleracea*, *Elaeis guineensis*, and *Elaeis oleifera* (Low et al. 2024), possess a high number of genes for terpenoid backbone

Fig. 2. Functional analysis of palm tree genomes through the categorization of genes and identification of unique complete pathways. (A) Heatmap illustrating the relative values of overall enriched metabolic categories, with warmer colors indicating high gene abundance and colder colors representing low gene abundance; the data has been scaled by column using empirical percentile transformation. (B) Presence or absence heatmap depicting the distribution of complete pathways, where rows represent specific complete pathways, columns correspond to the palm tree genomes, and the presence of color indicates a complete pathway within the genome. The top and side dendrograms represent hierarchical clustering of genomes and metabolic pathways, respectively.



biosynthesis, which are absent in the other two palms. They also possess genes related to the sulfur cycle, which are possibly important for sulphured amino acids such as methionine. Methionine was found to be produced solely from aspartate in the three aforementioned palm trees (Fig. 2B). Therefore, we identified a significant lack of complete pathways across the *C. nucifera* genome, which, despite having a low number of genes for sulfur metabolism, showed a particular assimilatory sulfate reduction pathway (Fig. 2B). In contrast, *Elaeis guineensis* exhibited a lack of some widely shared pathways and the exclusive presence of others, such as the alternative isoleucine biosynthesis pathway through pyruvate instead of threonine. *Elaeis oleifera*, on the other hand, was found to have a lysine biosynthesis pathway through the succinyl–DAP pathway rather than the DAP aminotransferase pathway.

Differential expression during fruit ripening

The correlation matrix and dendrogram in Fig. 3 show the degree of similarity among the samples in terms of the transcriptional profile of white and purple açaí at different stages of ripeness. In all cases, the replicates from each phase demonstrated increased similarity to each other and were grouped into clusters, with the exception of the 70 DAF and 130 DAF samples of white açaí, suggesting a greater relationship between the transcriptional profiles of these two phases in the white variant. The same result can be observed when the data are analyzed using Principal Component Analysis (PCA) (Fig. S2).

In terms of differential expression, the white açaí exhibited 76 541 transcript isoforms between the 70 DAF and 130 DAF, with 3972 upregulated genes and 8938 downregulated genes. Between the 130 DAF and 160 DAF, a total of 54 198 transcripts were detected, with 3188 upregulated and 4766 downregulated. For purple açaí between 70 DAF and 130 DAF, 62 577 transcripts were detected, with 4330 upregulated and 4812 downregulated. Finally, between 130 DAF and 160 DAF, 68 926 transcripts were described, with only 1208 genes upregulated and 4457 downregulated (Table 1).

The number of DEGs in each pairwise analysis is represented in the volcano plots of Figs. 4 and 5. The *x*-axis represents statistical significance (−log10 FDR), while the *y*-axis represents the magnitude of the differential expression (Log2FoldChange, FC). Upregulated genes are highlighted in red, while downregulated genes are in blue. It is important to note that the process of fruit ripening in both varieties is

Fig. 3. The analysis was conducted to ascertain the correlation of samples based on gene expression using log-transformed CPM values. The resulting correlation matrix between replicates for (A) white açaí 70 DAF vs. 130 DAF and (B) 130 DAF vs. 160 DAF; (C) purple açaí 70 DAF vs. 130 DAF and (D) 130 DAF vs. 160 DAF are illustrated in the form of heat maps. The similarity values are represented in a gradient from purple (low correlation) to yellow (high correlation). The phylogenetic trees reflect the hierarchical grouping of the samples based on the transcriptional profile.

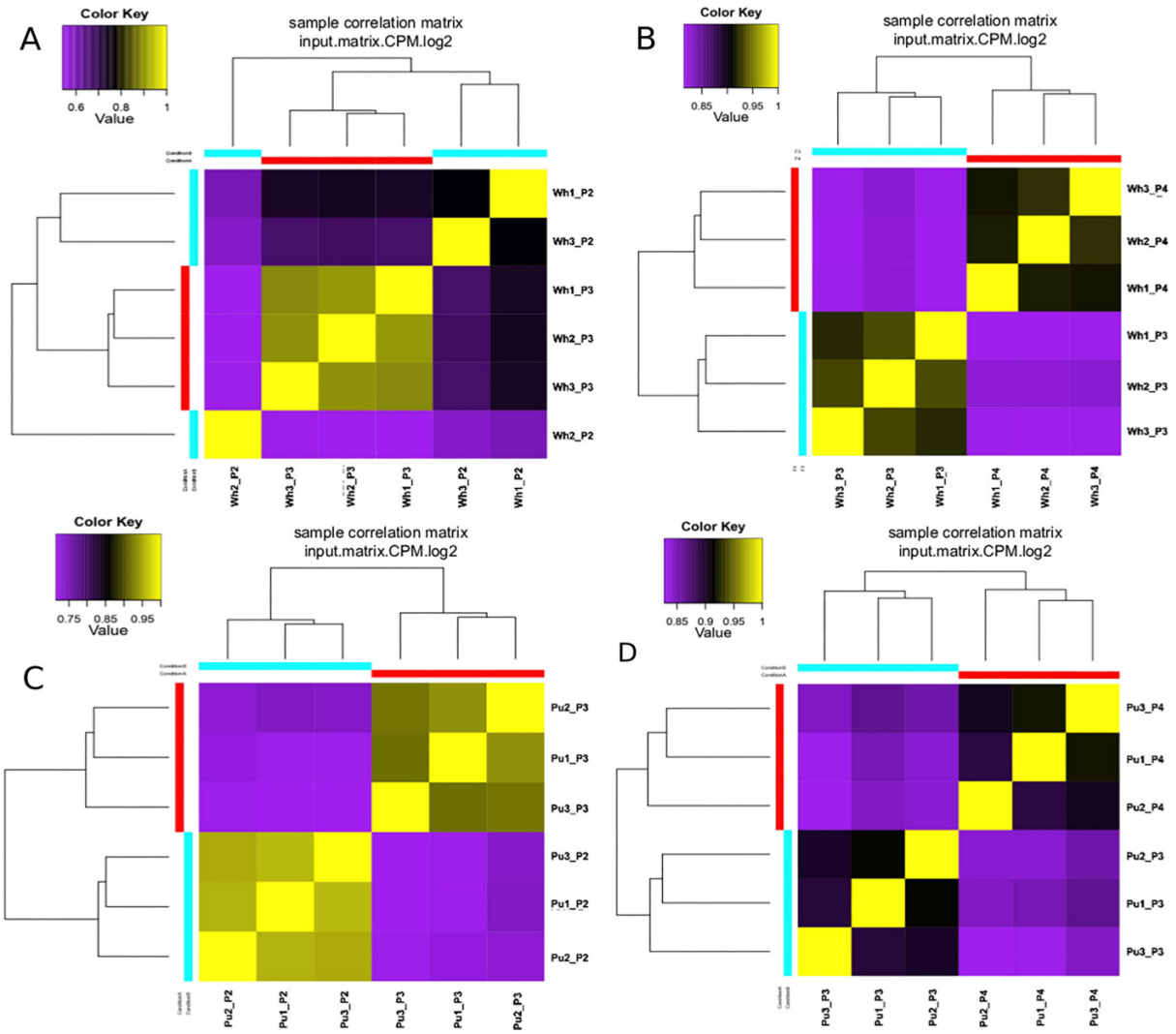


Table 1. Summary of the analyses of differential gene expression throughout the ripening phases in white and purple açaí.

Açaí variety	Condition	No. of transcripts	Upregulated genes	Downregulated genes
White	70 DAF vs. 130 DAF	76 541	3972	8938
White	130 DAF vs. 160 DAF	54 198	3188	4766
Purple	70 DAF vs. 130 DAF	62 577	4330	4812
Purple	130 DAF vs. 160 DAF	68 926	1208	4457

Note: The total number of transcripts identified, as well as the number of up and downregulated genes between the consecutive phases, are highlighted. Differentially expressed genes (DEGs) were determined based on log2foldchange values, with a ≥2 cutoff to identify upregulated genes and ≤2 to identify downregulated genes.

characterized by the regulation of numerous metabolic processes.

Among the most significant DEGs upregulated in white açaí between 70 DAF and 130 DAF, we can highlight phosphoenolpyruvate carboxykinase (PEPCK) (COG1866, Log2FC = 6.37), purple acid phosphatase (PAP) (COG1409, Log2FC = 9.95), and Tify domain protein (COG5641, Log2FC = 8.27) (Fig. 4A). On the other hand, the most significant downregulated genes encompass the BEL1-like homeodomain protein (KOG0773, Log2FC = −2.41), serine/threonine phosphatase (COG0631, Log2FC = −6.50), and saposin-like type B protein (KOG1339, Log2FC = −3.14) (Fig. 4A). The most significant DEGs that were found to be upregulated in purple açaí between 130 DAF and 160 DAF were as follows: Chaperonin (*HSP60*) (COG0459, Log2FC = 2.01), RCD1-SRO-TAF4 (RST) domain protein (28SQZ, Log2FC = 4.18),

Fig. 4. Volcano plots showing the differential expression of genes between different stages of white açaí fruit ripeness. (A) Comparison between stages 70 DAF and 130 DAF. (B) Comparison between stages 130 DAF and 160 DAF. The red dots represent upregulated genes, while the blue dots indicate downregulated genes. The grey dots represent genes with no significant change in expression. Genes with the highest or lowest Log2FC values were annotated and identified on the plots.

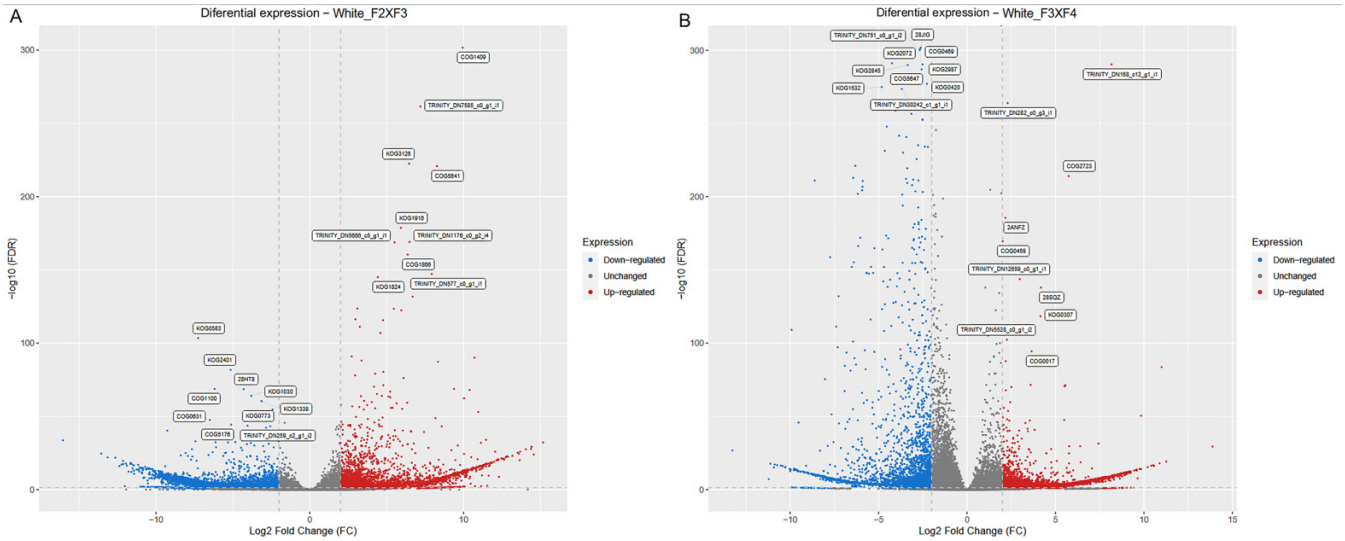
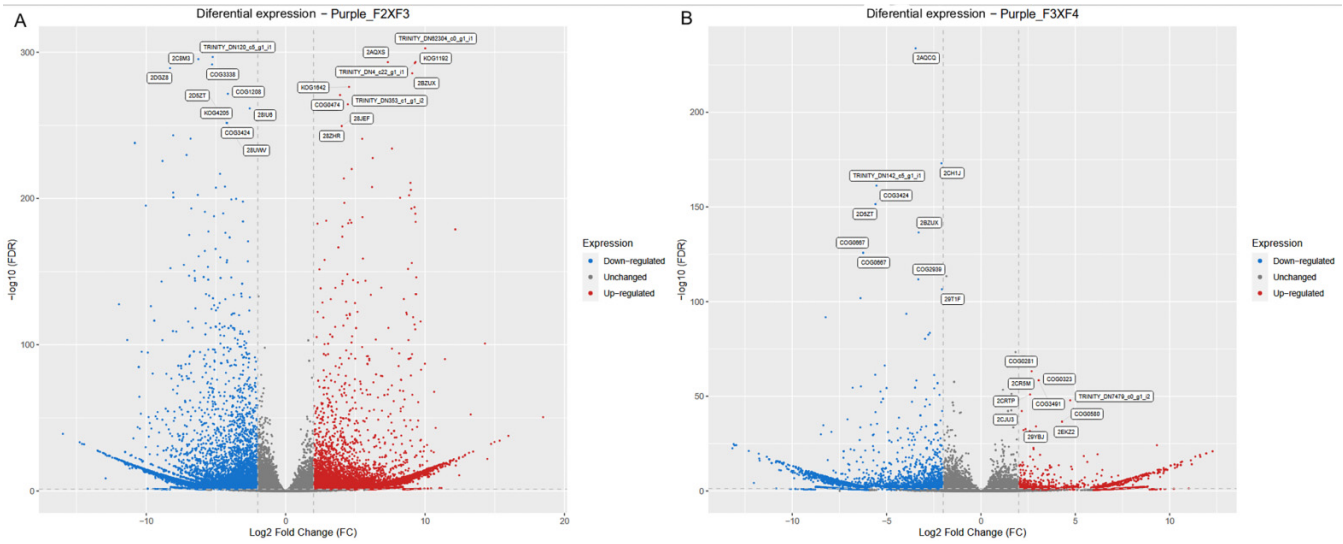


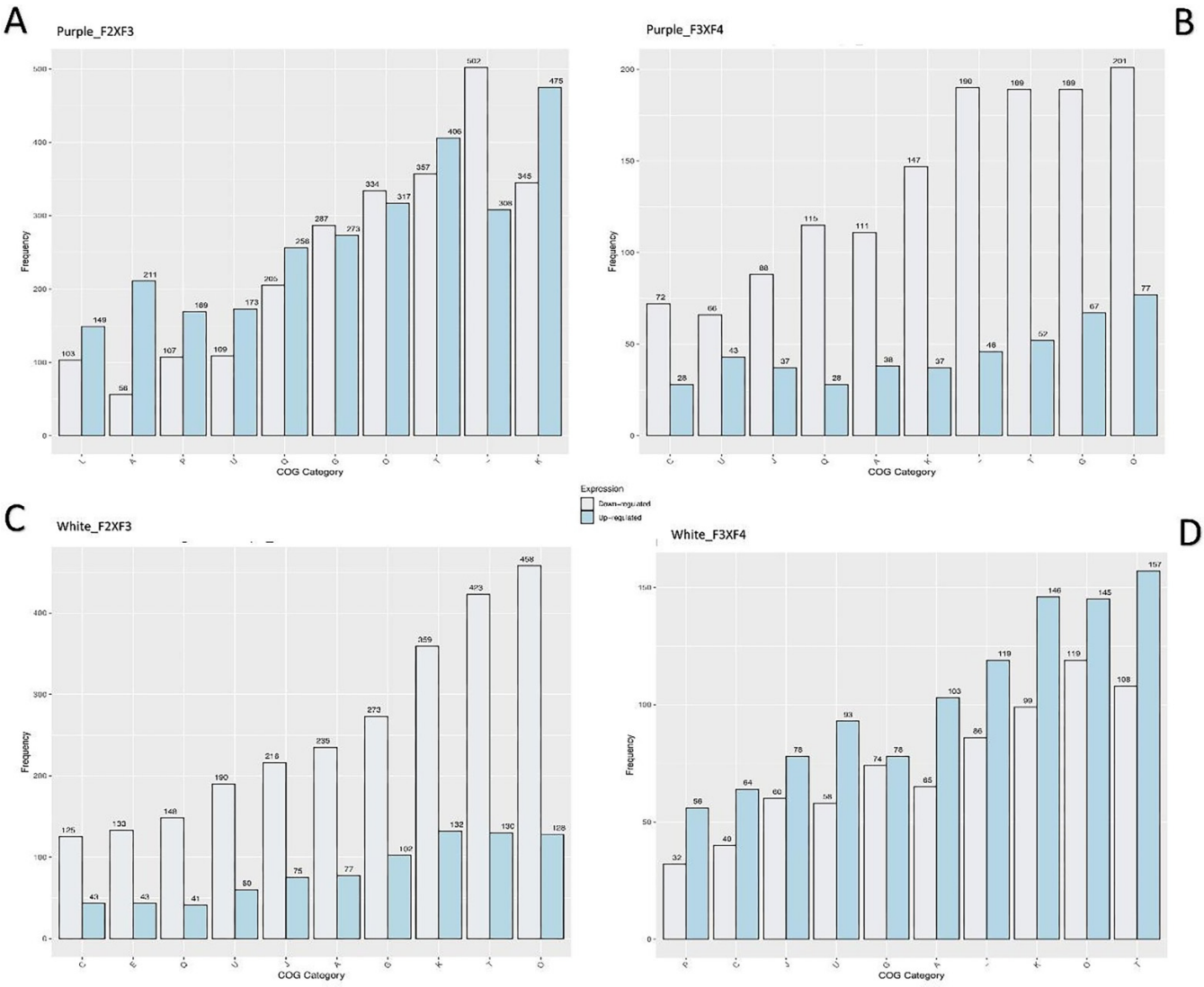
Fig. 5. Volcano plots showing the differential expression of genes between different stages of purple açaí fruit ripeness. (A) Comparison between stages 70 DAF and 130 DAF. (B) Comparison between stages 130 DAF and 160 DAF. The red dots represent upregulated genes, while the blue dots indicate downregulated genes. The grey dots represent genes with no significant change in expression. Genes with the highest or lowest Log2FC values were annotated and identified on the plots.



glycosyl hydrolase 1 family (COG2723, Log2FC = 5.74), and tRNA synthetases class II (COG0017, Log2FC = 3.65) (Fig. 4B). In contrast, the most significant downregulated DEGs were the WRKY transcription factor (28JIG, Log2FC = −2.65), pyruvate kinase (COG0469, Log2FC = −2.61), and RNA-binding component of eIF-3 complex (COG2072, Log2FC = −4.24) (Fig. 4B).

In the purple açaí, the most significant DEGs that were found to be upregulated between 70 DAF and 130 DAF were 2BZUX (log2FC = 9.08), invertase/pectin methylesterase (log2FC = 7.33), and a UDP-glycosyltransferase family protein (log2FC = 9.31) (Fig. 5A). The latter is potentially associated with the biosynthesis and modification of secondary metabolites, which play important roles in the flavor, color, and texture of the fruit. In contrast, the most significant downregulated genes were nectarine-3 (log2FC = −5.28) and 2C8M3 (log2FC = −6.26) (Fig. 5A). The former may be indicative of a decline in antioxidant defenses as the fruit ripens, while the latter has been linked to the protease inhibitor family and seed storage. Finally, the most significant DEGs of purple açaí that were found to be upregulated between 130 DAF and 160 DAF were the Iron ascorbate-dependent

Fig. 6. The bar charts illustrate the most significant categories of Clusters of Orthologous Groups (COGs) identified in the Differentially Expressed Genes (DEGs). The bar graphs show the most relevant categories of clustered orthologues (COGs) identified in the differentially expressed genes (DEGs). The most abundant COGs in purple açaí between 70 and 130 DAF and 130 and 160 DAF are represented in A and B, respectively. The most abundant COGs in white açaí between 70 and 130 DAF and 130 and 160 DAF are represented in C and D, respectively. The blue bars represent statistically significant DEGs, while the grey bars show the total frequency of genes per category.

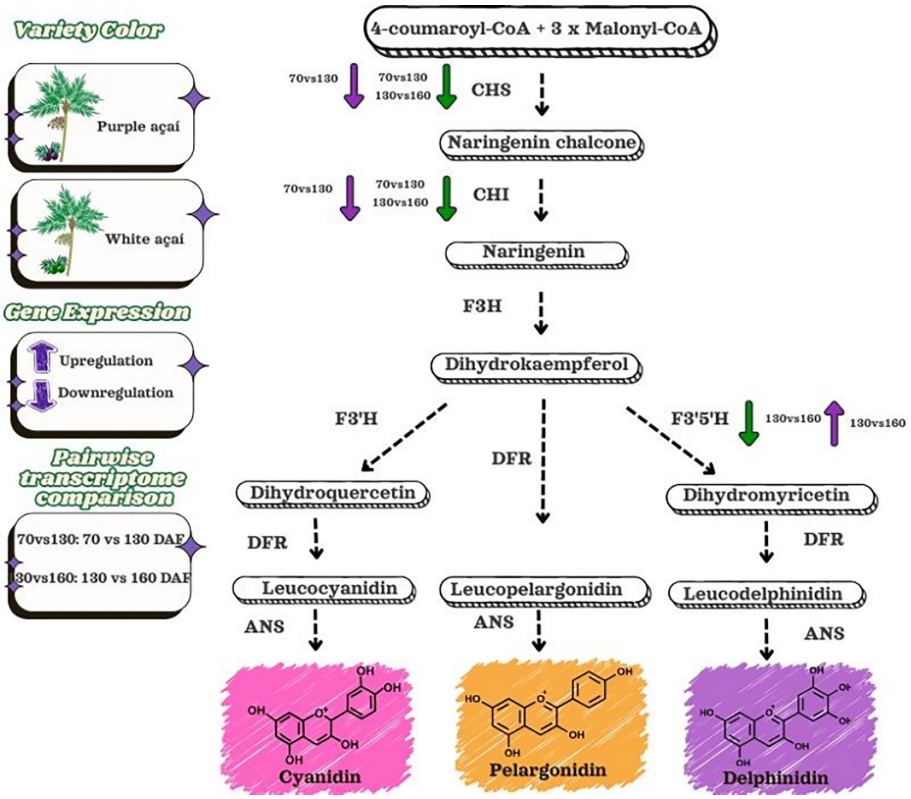


oxidoreductase family (COG3491, log2FC = 3.10) and the MIP aquaporin family (COG0580, log2FC = 4.19), while the COG0667, belonging to the aldose/ketose reductase family (log2FC = −6.24) and the chalcone stilbene synthase family gene (COG3424, log2FC = −5.59), which is involved in changes in flavonoid levels, were the most downregulated genes (Fig. 5B).

The DEGs were selected and classified into COGs. Figure 6 shows the most relevant COGs identified during the fruit ripening. In the purple variety, between 70 DAF and 130 DAF, the main functional categories identified were O (DNA replication and repair), K (protein translation, modification, and turnover), and S (unknown general functions). Between 130 and 160 DAF, significant changes were observed in the expression of genes classified into categories G (carbohydrate metabolism and transport) and Q (biosynthesis of secondary compounds). In the white variety, between 70 and 130 DAF, the most represented categories were O (DNA replication and repair) and E (Amino acid metabolism). Between 130 and 170 DAF, the following categories were most represented: X (Cell wall metabolism) and F (Nucleotide metabolism). In comparison with the purple variety, the white variety exhibited a reduced number of metabolic changes associated with pigmentation, indicative of a ripening pattern that is more focused on structural characteristics.

A significant disparity in the chemical composition of white and purple fruit is the concentration of anthocyanins. Consequently, a comprehensive analysis of the genes responsible for synthesizing these metabolites was conducted. No

Fig. 7. Schematic representation of the anthocyanin biosynthesis pathway in white and purple açaí. The substrates and products of the pathway are connected by arrows. The acronyms of the enzymes that catalyze the biochemical reactions are shown next to the arrows. The anthocyanin synthesis pathway begins with the reaction between 4-coumaroyl-CoA and three molecules of malonyl-CoA to form the chalcone naringenin, catalyzed by the enzyme chalcone synthase (CHS). The first part of the pathway ends with the production of dihydrokaempferol, which can serve as a substrate for several enzymes that initiate the synthesis of anthocyanins of different colors. In purple açaí, the enzyme flavonoid 3′5′-hydroxylase (F3′5′H) is upregulated, which may explain the development of the fruit's characteristic color. CHI: chalcone isomerase; F3H: flavanone 3-hydroxylase; F3′H: flavonoid 3′-hydroxylase; DFR: dihydroflavonol 4-reductase; ANS: anthocyanidin synthase.



gene involved in the anthocyanin biosynthesis pathway was upregulated in the white açaí. The majority of these genes, including the enzymes CHS (Chalcone Synthase), CHI (Chalcone Isomerase), F3H (Flavanone 3-Hydroxylase), DFR (Dihydroflavonol 4-Reductase), and ANS (Anthocyanidin Synthase), exhibited reduced levels of expression, indicating that the synthesis of anthocyanins is repressed in white açaí, thereby conferring its definitive characteristics (Fig. 7). Conversely, the expression of the enzyme UFGT (UDP-Glucose: Flavonoid 3-O-Glucosyltransferase) remained stable, indicating that its activity is predominantly influenced by the substrate availability in the preceding stages of the pathway. In purple açaí, the gene encoding for the enzyme F3′5′H (Flavonoid 3′,5′-Hydroxylase) was found to be upregulated, which may result in a substantial increase in the synthesis of dihydromyricetin and, consequently, a concomitant increase in the production of the anthocyanin delphinidin 3-glucosidase, whose derivatives are directly associated with the fruit's purple pigmentation (Fig. 7). The up regulation of the F3′5′H enzyme is primarily observed between the 70 DAF and 130 DAF, suggesting that these stages are pivotal in determining the appearance of the purple variety.

Discussion

The first genome sequence of the species *Euterpe oleracea*

This work presents the first version of the genome of *E. oleracea*. Our comparative results with other palm tree genomes show that the açaí genome is the largest among them. Despite the size difference, the overall genomic function of palm trees appears to be highly similar, with COG functions widely shared and only *P. dactylifera* exhibiting notable differences. Interestingly, all palm tree genomes contain a high number of transposable elements (TE), as clearly evidenced by COG analysis. This result is expected given that the oil palm has been reported to harbor a high number of transposable elements (TEs) (Kubis et al. 2003; Beulé et al. 2015). Other tree species, such as the Madagascar palm from the Apocynaceae family, have also been shown to possess TE-rich genomes (Cuello et al. 2024). Although some genomes of the genus *Elaeis* have already been published (Low et al. 2024), our results indicate that a significant proportion of the genes have unknown functions. This highlights the need for more

extensive efforts to characterize the function of the genes in these plants.

Despite phylogenetic evidence indicating a closer evolutionary relationship between the genera *Elaeis* and *Cocos*, both of which are distant from *E. oleracea* (Darnet et al. 2023), a more precise phylogenetic analysis has shown the strongest relationship between the genera *Elaeis* and *P. dactylifera*, along with highly similar genomic content between *Elaeis guineensis* and *Elaeis oleifera* (Singh et al. 2013; Low et al. 2024). This supports our findings from the genome enrichment analysis. A study that used transcripts to gather genomic information from *E. oleracea* detected structural changes between this species and both *P. dactylifera* and *Elaeis guineensis* (Darnet et al. 2023). However, this transcript-based approach must be validated through structural genomic analysis in future studies.

Searching for complete pathways in palm tree genomes revealed that *Elaeis oleifera*, *Elaeis guineensis*, *P. dactylifera*, and *E. oleracea* have unique, complete pathway profiles that are quite similar, with many shared pathways. By contrast, *C. nucifera* is positioned as an outgroup and lacks many important pathways, such as those involved in ascorbate and heme biosynthesis. It may therefore possess uncharacterized mechanisms to perform these functions due to its genomic adaptations to environmental and salt-related stress (Yang et al. 2021). However, while four palm trees possess the ascorbate biosynthesis pathway, only three of them possess the glyoxylate cycle pathway. Both pathways are important resources for reactive oxygen species (ROS) detox in plants (Zhang et al. 2020), and this result may suggest an improved detox activity in *Elaeis oleifera*, *E. oleracea*, and *P. dactylifera*. Other specific pathways were found, such as a pathway for isoleucine synthesis from pyruvate instead of threonine for *Elaeis guineensis* (Risso et al. 2008) or the lysine biosynthesis through succinyl-DAP instead of DAP aminotransferase, event known for bacteria (Liu et al. 2021) and found in *C. nucifera* and *Elaeis oleifera*.

General aspects of the RNA-Seq analysis

Intense pigmentation has been demonstrated to be associated with a few ecological advantages, including the attraction of seed dispersers and protection against environmental stresses, such as UV radiation and herbivory. During the process of açaí ripening, significant transcriptional changes were observed. The total number of bases that were sequenced amounted to 265 billion, which resulted in the assembly of 69 405 transcript isoforms. This number of isoforms reflects the high complexity of the açaí transcriptome, which showed an average GC content of 50% and a Phred score of 26. These data are consistent with those from other plants, such as *Spartina alterniflora* (Ye et al. 2023).

A comparison of the transcriptome profiles of white and purple açaí revealed that the white variety exhibited a greater number of DEGs between 70 and 130 days DAF. These results suggest that the white variety requires greater gene regulation during the early stages of ripening to produce the white fruit. On the other hand, the changes observed in the ripening process of purple fruit have yielded results that are like those in tomatoes (*Solanum lycopersicum*), grapes (*Vitis vinifera*), and apples (*Malus domestica*) (Klee and Giovannoni 2011). For example, studies on tomatoes have demonstrated that the initial stages of ripening are characterized by elevated expression levels of genes associated with primary metabolism, while the subsequent stages of ripening are associated with the activation of secondary metabolic pathways, including the biosynthesis of carotenoids and anthocyanins (Klee and Giovannoni 2011).

The differences in the main COG categories between the ripening stages of white and purple açaí reflect the marked distinctions in the gene regulation and metabolism of these two varieties (Fig. 6). The functional categories observed in purple açaí suggest a marked transition to the production of secondary compounds, especially anthocyanins. In contrast, the predominant COG categories in white açaí are associated with structural processes, including amino acid metabolism (E) and genomic maintenance (O). The absence of categories related to the biosynthesis of secondary compounds is indicative of the lack of significant pigmentation in white açaí. This phenomenon is consistent with the observations made in studies conducted on fruits such as white tomatoes, where ripening is more closely associated with changes in fruit texture than with changes in color (Klee and Giovannoni 2011).

Regarding the gene expression at different ripening stages (70 vs. 130 DAF and 130 vs. 160 DAF), the greatest similarity in the early stages observed for both varieties suggests the expression of common biological processes, such as cell expansion and carbohydrate synthesis. These processes have also been documented in other species, such as apple (*Malus domestica*) and strawberry (*Fragaria ananassa*) (Chen et al. 2022; Li et al. 2023). However, the expression pattern of enzymes involved in the anthocyanin biosynthetic pathway represents a notable exception to this trend. The most prominent DEGs associated with the production of purple pigments were detected in the comparison between 70 and 130 DAF. In the case of purple açaí, the marked down regulation of genes in the anthocyanin pathway likely reflects a developmental shift in the metabolic requirements as the fruit matures (Fig. 7). In particular, one of the most strongly downregulated genes was a member of the chalcone–stilbene synthase family (COG3424, $\log_2$FC = −5.59), which is a key enzyme in the early stages of flavonoid biosynthesis. Also, between 70 and 130 DAF, the gene 2BZUX ($\log_2$FC = 9.08), annotated as *invertase/pectin methylesterase*, was found to be highly upregulated. This gene is associated with the accumulation of sugars in ripe fruit and plays a role in fruit softening (Coculo and Lionetti 2022).

The marked difference in gene expression profiles at advanced stages (130 vs. 160 DAF) reflects, in the purple variety, the accumulation of anthocyanins and changes in the cell structure metabolism. Studies on grapes have shown that fruit ripening is associated with the activation of genes related to the synthesis of phenolic compounds and cell wall polysaccharides (Zhang et al. 2021). Corresponding patterns have been observed in pigmented orange (*Citrus sinensis*) and red pepper (*Capsicum annuum*), where differential expression of anthocyanin pathway genes has been correlated with fruit pigmentation (Butelli et al. 2008). Furthermore, in the advanced stages of ripening, the expression of the *HSP60* and *RCD1-SRO-TAF4* genes highlights a robust response to oxidative stress, a common phenomenon in ripening fruit that face high rates of respiration and production of reactive

oxygen species. These processes are vital for maintaining cell integrity, as demonstrated in studies with tomatoes and bananas (Bhadouria and Giri 2022; Gambhir et al. 2024).

In white açaí, the most upregulated DEGs were found to be associated with adaptation and resistance. For example, the up regulation of the *PEPCK* gene between 70 and 130 DAF, is indicative of a metabolic adaptation mechanism prevalent in fruits such as tomatoes and grapes, in which gluconeogenesis plays a key role in regulating the redistribution of sugars and organic acids. Another notable gene that encodes the enzyme Purple Acid Phosphatase (PAP) exhibited high levels of expression, suggesting an adaptive mechanism linked to phosphorus metabolism and the detoxification of reactive oxygen species (Walker et al. 2015). It has previously been demonstrated that tropical fruits such as mangoes and bananas also use this enzyme to respond to environmental stress (Bhadouria and Giri 2022).

Among the DEGs identified in purple açaí, those related to the synthesis of phenolic compounds and anthocyanins are of particular interest. One such example is the gene containing the TIFY domain (COG5641), which exhibited elevated expression levels in the advanced stages of ripening. This conserved domain characterizes a plant-specific family of transcription factor (TF) genes involved in florescence and plant development (Nishii et al. 2000). Also, the up regulation of this gene has been linked to the accumulation of anthocyanins (Qing et al. 2025). The anthocyanin concentration in açaí fruit increases progressively during ripening, reaching its peak in fully mature fruits (Gordon et al. 2012). A similar pattern has been reported in other fruits, such as jabuticaba (*Myrciaria cauliflora*), where anthocyanins have been shown to play key roles in pigmentation and protection against abiotic stresses (Lima et al. 2011). Furthermore, during the later stages of ripening, the up-regulation of genes associated with aquaporins underscores their pivotal function in water transportation and the preservation of osmotic cell pressure (Maurel et al. 2015).

Differential expression of enzymes from the anthocyanin synthesis pathway

The present study provides evidence to suggest that the low production of purple pigments in white açaí is directly associated with the generalized downregulation of genes encoding key enzymes of the anthocyanin biosynthesis pathway. The analysis revealed the downregulation of chalcone synthase (CHS) and chalcone isomerase (CHI) at all stages of white açaí development (Fig. 7). This downregulation has the potential to compromise the synthesis of vital intermediates, thereby impeding the metabolic flux towards the final products, i.e., anthocyanins (Petroni and Tonelli 2011; Xu et al. 2015). Further genetic studies are required to ascertain the extent to which epigenetic factors, other regulatory mechanisms, or structural mutations in specific promoter elements may influence the low synthesis of anthocyanins in the white variety (He et al. 2010). In addition, it is important to emphasize that environmental factors and local adaptations also play important roles in the observed differential expression (Jaakola 2013).

The absence of differential expression for the enzyme UDP-glucosyltransferase (UFGT) indicates that, although the enzymes downstream of CHS, CHI and DFR are not repressed, the lack of substrates from the initial steps limits its activity (Fig. 7). This finding underscores the necessity for further investigation into the interplay between transcriptional regulation and the availability of metabolites, as reported in other species exhibiting analogous regulatory patterns in metabolic pathways (Wang et al. 2014).

In contrast, the analysis of purple açaí revealed a gene expression profile that favors the synthesis of pigments responsible for the fruit's distinctive purple coloration. The up regulation of the enzyme flavonoid 3′5′-hydroxylase (F3′5′H) indicates a preferential synthesis of delphinidin, thereby substantiating the pivotal function of this enzyme in the diversification of anthocyanins across diverse species (Jaakola 2013) (Fig. 7). Although a reduction in the expression of initial enzymes in the pathway, such as CHS and CHI, was also observed, this effect may be compensated by post-transcriptional regulation mechanisms or by greater catalytic efficiency (Petroni and Tonelli 2011). Intense pigmentation has been demonstrated to be associated with a few ecological advantages, including the attraction of seed dispersers and protection against environmental stresses, such as UV radiation and herbivory.

Conclusion

The first version of the açaí genome has a size of 3.1 Gbp, which is about 75% of the expected size. With this vast amount of genetic information, it was possible to identify the main COG categories associated with the 41 703 annotated genes. The most abundant COGs were "L—Replication and repair", mostly represented by transposable genetic elements, and "S—Function unknown". Due to the lack of genomes from plants of the genus *Euterpe*, we compared the genome of *E. oleracea* with other palm trees of the family Aracaceae and found a high degree of functional similarity with the genus *Elaeis*. The difference between the number of annotated genes and the number of isoforms assembled from the transcriptome reveals a high degree of complexity in the post-transcriptional mechanisms. Thousands of DEGs were identified at different stages of ripening of white and purple fruits. The up regulation of the enzyme flavonoid 3′5′-hydroxylase seems to be crucial for the increase in anthocyanin synthesis in purple açaí. On the other hand, a general down regulation of the genes responsible for the initial stages of the anthocyanin synthesis pathway was observed in the white variety. Our findings provide a valuable contribution to the understanding of the molecular mechanisms that regulate the ripening process of açaí fruits and present the first genome of *E. oleracea*, which is one of the most significant symbols of the Amazonian biodiversity.

Acknowledgements

The authors would like to thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for the students' scholarships, and Centro de Computação de Alto

Desempenho e Inteligência Artificial (CCAD/UFPA) for providing access to the Apollo 2000 cluster for genome assembly.

Article information

History dates

Received: 26 October 2025
Accepted: 13 April 2026
Accepted manuscript online: 14 May 2026
Version of record online: 17 June 2026

Notes

This paper is part of a collection entitled, "Brazilian Excellence in Plant Genetics and Genomics." https://cdnsciencepub.com/topic/gen-brazil-plant-research.

Copyright

Data availability

The açaí genomes described in this study were submitted to GenBank under the accession number SAMN48671595 and transcriptome data were submitted to SRA under the accession numbers PRJNA1265271, PRJNA1265274, PRJNA1265420, and PRJNA1262982.

Author information

Author ORCIDs

Rafael Azevedo Baraúna https://orcid.org/0000-0001-6654-2118

Author contributions

Conceptualization: MPCS, JASQ, DAdG, SdMR, EFM, AS, RAB
Data curation: MSRB, DJM, DAdG, AAdOV, SdMR, EFM, AS, RAB
Formal analysis: MSRB, SdSC, DJM, ARdO, LdSeS, DAdG, AAdOV, EFM, RAB
Funding acquisition: MPCS, AS, RAB
Investigation: MSRB, SdSC, LdSeS, AAdOV, EFM, AS, RAB
Methodology: MSRB, SdSC, ARdO, DAdG, AAdOV, MdSPdO, SdMR, EFM, AS, RAB
Project administration: MPCS, JASQ, EFM, AS, RAB
Resources: RAB
Software: SdSC, DJM, ARdO, DAdG, AAdOV
Supervision: MPCS, JASQ, AAdOV, EFM, AS, RAB
Validation: MSRB, LdSeS, MdSPdO, SdMR, EFM, RAB
Visualization: MSRB, SdSC, JASQ, DAdG, AS, RAB
Writing – original draft: MSRB, SdSC, DJM, LdSeS, SdMR, RAB
Writing – review & editing: MPCS, JASQ, DAdG, AAdOV, MdSPdO, EFM, AS, RAB

Competing interests

The authors have no conflicts of interest to declare.

Funding information

Funder: Conselho Nacional de Desenvolvimento Científico e Tecnológico (grant number: 445350/2024-5). Funder: Fundação Amazônia de Amparo a Estudos e Pesquisas (grant numbers: 002/2023 and 005/2021).

Supplementary material

Supplementary data are available with the article at https://doi.org/10.1139/gen-2025-0105.

References

Al-Mssallem, I.S., Hu, S., Zhang, X., Lin, Q., Liu, W., Tan, J., et al. 2013. Genome sequence of the date palm *Phoenix dactylifera* L. Nat. Commun. **4**: 2274. doi:10.1038/ncomms3274. PMID: 23917264.

Antonelli, A., Nylander, J.A.A., Persson, C., and Sanmartín, I. 2009. Tracing the impact of the Andean uplift on Neotropical plant evolution. Proc. Natl. Acad. Sci. **106**(24): 9749–9754. doi:10.1073/pnas.0811421106. PMID: 19470489.

Beulé, T., Agbessi, M.D., Dussert, S., Jaligot, E., and Guyot, R. 2015. Genome-wide analysis of LTR-retrotransposons in oil palm. BMC Genom. [Electronic Resource], **16**: 1–14. doi:10.1186/s12864-015-2023-1.

Bezerra, F. 2025. A valorização do açaí no mercado global: impactos socioeconômicos e ambientais no estado do Pará na perspectiva do desenvolvimento sustentável. Rev. Econ. Reg. Urb. Trab. **14**(2). Available from https://periodicos.ufrn.br/rerut/article/view/41169.

Bhadouria, J., and Giri, J. 2022. Purple acid phosphatases: roles in phosphate utilization and new emerging functions. Plant Cell Rep. **41**: 33–51. doi:10.1007/s00299-021-02773-7. PMID: 34402946.

Bolger, A.M., Lohse, M., and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. **30**(15): 2114–2120. doi:10.1093/bioinformatics/btu170. PMID: 24695404.

Butelli, E., Titta, L., Giorgio, M., Mock, H.P., Matros, A., Peterek, S., et al. 2008. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nat. Biotechnol. **26**(11): 1301–1308. doi:10.1038/nbt.1506. PMID: 18953354.

Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol. Biol. Evol. **38**(12): 5825–5829. doi:10.1093/molbev/msab293. PMID: 34597405.

Chen, P., Li, H., Li, X., Zhou, X., Zhang, X., Zhang, A., et al. 2022. Transcriptomic analysis provides insight into defensive strategies in response to continuous cropping in strawberry (*Fragaria* × *ananassa* Duch.) plants. BMC Plant Biol. **22**: 476. doi:10.1186/s12870-022-03857-6. PMID: 36203126.

Coculo, D., and Lionetti, V. 2022. The plant invertase/pectin methylesterase inhibitor superfamily. Front. Plant Sci. **13**: 863892. doi:10.3389/fpls.2022.863892. PMID: 35401607.

Cuello, C., Jansen, H.J., Abdallah, C., Mbadinga, D.L.Z., Williams, C.B., Durand, M., et al. 2024. The Madagascar palm genome provides new insights on the evolution of Apocynaceae specialized metabolism. Heliyon, **10**(6): e28078. doi:10.1016/j.heliyon.2024.e28078. PMID: 38533072.

Darnet, E., Teixeira, B., Schaller, H., Rogez, H., and Darnet, S. 2023. Elucidating the mesocarp drupe transcriptome of açai (*Euterpe oleracea* Mart.): an Amazonian tree palm producer of bioactive compounds. Int. J. Mol. Sci. **24**(11): 9315. doi:10.3390/ijms24119315. PMID: 37298279.

Francisconi, A.F., Scaketti, M., Morales-Marroquín, J.A., de Carvalho, I.A.S., Fornazier, G.O., Malaquias Costa, F., et al. 2022. Complete chloroplast genomes and phylogeny in three *euterpe* palms (*E. edulis*, *E. oleracea* and *E. precatoria*) from different Brazilian biomes. PLoS One, **17**(7): e0266304. doi:10.1371/journal.pone.0266304. PMID: 35901127.

Francisconi, A.F., Scaketti, M., Morales-Marroquín, J.A., de Carvalho, I.A.S., Fornazier, G.O., Malaquias Costa, F., et al. 2025. Genomic consequences of açaí extraction in the Amazon: insights into selective

pressures, genomic diversity, and population structure. Front. Plant Sci. **16**: 1688760. doi:10.3389/fpls.2025.1688760. PMID: 41334190.

Galli, V., Messias, R.S., Guzman, F., Perin, E.C., Margis, R., and Rombaldi, C.V. 2019. Transcriptome analysis of strawberry (*Fragaria* × *ananassa*) fruits under osmotic stresses and identification of genes related to ascorbic acid pathway. Physiol. Plant. **166**: 979–995. doi:10.1111/ppl.12861. PMID: 30367706.

Gambhir, P., Raghuvanshi, U., Kumar, R., and Sharma, A.K. 2024. Transcriptional regulation of tomato fruit ripening. Physiol. Mol. Biol. Plants, **30**: 289–303. doi:10.1007/s12298-024-01424-x.

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. doi:10.1038/nbt.1883. PMID: 21572440.

Gordon, A., Cruz, A.P.G., Cabral, L.M.C., Freitas, S.C., Dib Taxi, C.M.A., Donangelo, C.M., et al. 2012. Chemical characterization and evaluation of antioxidant properties of açaí fruits (*Euterpe oleracea* Mart.) during ripening. Food Chem. **133**(1): 256–263. doi:10.1016/j.foodchem.2011.11.150. PMID: 25683393.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics, **29**(8): 1072–1075. doi:10.1093/bioinformatics/btt086. PMID: 23422339.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource. Nucleic Acids Res. **47**(D1): D309–D314. doi:10.1093/nar/gky1085. PMID: 30418610.

He, F., Mu, L., Yan, G.L., Liang, N.N., Pan, Q.H., Wang, J., et al. 2010. Biosynthesis of anthocyanins and their regulation in colored grapes. Molecules. **15**(12): 9057–9091. doi:10.3390/molecules15129057. PMID: 21150825.

Jaakola, L. 2013. New insights into the regulation of anthocyanin biosynthesis in fruits. Trends Plant Sci. **18**(9): 477–483. doi:10.1016/j.tplants.2013.06.003. PMID: 23870661.

Kanehisa, M., Sato, Y., and Morishima, K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J. Mol. Biol. **428**(4): 726–731. doi:10.1016/j.jmb.2015.11.006. PMID: 26585406.

Kang, J., Thakali, K.M., Xie, C., Kondo, M., Tong, Y., Ou, B., et al. 2012. Bioactivities of açaí (*Euterpe precatoria* Mart.) fruit pulp, superior antioxidant and anti-inflammatory properties to *Euterpe oleracea* Mart. in chemical-induced oxidative stress and in vivo inflammation models. Food Chem. **133**(3): 671–677. doi:10.1016/j.foodchem.2012.01.048.

Klee, H.J., and Giovannoni, J.J. 2011. Genetics and control of tomato fruit ripening and quality attributes. Annu. Rev. Genet. **45**: 41–59. doi:10.1146/annurev-genet-110410-132507. PMID: 22060040.

Kubis, S.E., Castilho, A.M., Vershinin, A.V., and Heslop-Harrison, J.S. 2003. Retroelements, transposons and methylation status in the genome of oil palm (*Elaeis guineensis*) and the relationship to somaclonal variation. Plant Mol. Biol. **52**: 69–79. doi:10.1023/A:1023942309092. PMID: 12825690.

Laurindo, L.F., Barbalho, S.M., Araújo, A.C., Guiguer, E.L., Mondal, A., Bachtel, G., et al. 2023. Açaí (*Euterpe oleracea* Mart.) in health and disease: a critical review. Nutrients. **15**: 989. doi:10.3390/nu15040989. PMID: 36839349.

Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics, **25**(14): 1754–1760. doi:10.1093/bioinformatics/btp324. PMID: 19451168.

Li, T., Zhang, X., Wei, Y., Xu, Y., Liu, W., Li, H., et al. 2023. Comparative transcriptome analysis of the climacteric of apple fruit uncovers the involvement of transcription factors affecting ethylene biosynthesis. Hortic. Plant J. **9 (** 4): 659–669. doi:10.1016/j.hpj.2022.12.002.

Lima, A.J.B., Corrêa, A.D., Saczk, A.A., Martins, M.P., and Castilho, R.O. 2011. Anthocyanins, pigment stability and antioxidant activity in jabuticaba [*Myrciaria cauliflora* (Mart.) O. Berg]. Rev. Bras. Frut. **33**(3): 877–887. doi:10.1590/S0100-29452011000300023.

Liu, N., Zhang, T.T., Rao, Z.M., Zhang, W.G., and Xu, J.Z. 2021. Reconstruction of the diaminopimelic acid pathway to promote L-lysine production in *Corynebacterium glutamicum*. Int. J. Mol. Sci. **22**(16): 9065. doi:10.3390/ijms22169065. PMID: 34445771.

Lopes, A.S., Pacheco, T.G., Silva, O.N., Vieira, L.N., Guerra, M.P., Mattar, E.P.L., et al. 2021. Plastid genome evolution in Amazonian açaí palm (*Euterpe oleracea* Mart.) and Atlantic Forest açaí palm (*Euterpe edulis* Mart.). Plant Mol. Biol. **105**(4–5): 559–574. doi:10.1007/s11103-020-01109-5.

Love, M.I., Huber, W., and Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. **15**: 550. doi:10.1186/s13059-014-0550-8. PMID: 25516281.

Low, E.T.L., Chan, K.L., Zaki, N.M., Taranenko, E., Ordway, J.M., Wischmeyer, C., et al. 2024. Chromosome-scale *elaeis guineensis* and *E. oleifera* assemblies: comparative genomics of oil palm and other arecaceae. G3: Genes, Genomes, Genetics, **14 (** 9): jkae135. doi:10.1093/g3journal/jkae135.

Maurel, C., Boursiac, Y., Luu, D.T., Santoni, V., Shahzad, Z., and Verdoucq, L. 2015. Aquaporins in plants. Physiol. Rev. **95**: 1321–1358. doi:10.1152/physrev.00008.2015. PMID: 26336033.

Mertens-Talcott, S.U., Rios, J., Jilma-Stohlawetz, P., Pacheco-Palencia, L.A., Meibohm, B., Talcott, S.T., and Derendorf, H. 2008. Pharmacokinetics of anthocyanins and antioxidant effects after the consumption of anthocyanin-rich açai juice in human subjects. J. Agric. Food Chem. **56**(17): 7796–7802. doi:10.1021/jf8007037. PMID: 18693743.

Nishii, A., Takemura, M., Fujita, H., Shikata, M., Yokota, A., and Kohchi, T. 2000. Characterization of a novel gene encoding a putative single zinc-finger protein, ZIM, expressed during the reproductive phase in *Arabidopsis thaliana*. Biosci. Biotechnol. Biochem. **64**(7): 1402–1409. doi:10.1271/bbb.64.1402. PMID: 10945256.

Oliveira, L.C., Oliveira, M.S.P.L., Davide, L.C., and Torre, G.A. 2016. Karyotype and genome size in Euterpe Mart. (Arecaceae) Species. Comp. Cytogenet. **10**(1): 17–25. doi:10.3897/CompCytogen.v10i1.5522. PMID: 27186334.

Oliveira, M.S.P.L., Amorim, E.P., Santos, J.B., and Ferreira, D.F. 2007. Genetic diversity among accessions of assai palm based on RAPD markers. Rev Bras Frutic. **31**(6): 1645–1653. doi:10.1590/S1413-70542007000600007.

Oliveira, A., Ribeiro, A., Oliveira, É., Garcia, M., Soares Júnior, M., and Caliari, M. 2019. Structural and physicochemical properties of freeze-dried açaí pulp (*Euterpe oleracea* Mart.). Food Sci. Technol. **40**: 282–289. doi:10.1590/fst.34818.

Petroni, K., and Tonelli, C. 2011. Recent advances on the regulation of anthocyanin synthesis in reproductive organs. Plant Sci. **181**(3): 219–229. doi:10.1016/j.plantsci.2011.05.009. PMID: 21763532.

Pichardo-Marcano, M., Nieto-Blázquez, M.E., MacDonald, A.N., Galeano, G., and Roncal, G. 2019. Phylogenetic, historical biogeography and diversification rates in an economically important group of neotropical palms: tribe Euterpe. Mol. Phylogenet. Evol. **133**: 67–81. doi:10.1016/j.ympev.2018.12.030.

Qing, H., Wu, Z., Mo, X., Wei, J., Shi, Y., Guo, H., et al. 2025. Genome-wide identification of the TIFY family in longan and their potential functional analysis in anthocyanin synthesis. Biology, **14 (** 4): 364. doi:10.3390/biology14040364. PMID: 40282229.

Risso, C., Van Dien, S.J., Orloff, A., Lovley, D.R., and Coppi, M.V. 2008. Elucidation of an alternate isoleucine biosynthesis pathway in *Geobacter sulfurreducens*. J. Bacteriol. **190**(7): 2266–2274. doi:10.1128/JB.01841-07. PMID: 18245290.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., et al. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. **43**(7): e47. doi:10.1093/nar/gkv007. PMID: 25605792.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, **26**(1): 139–140. doi:10.1093/bioinformatics/btp616. PMID: 19910308.

Rufino, M.S.M., Alves, R.E., Brito, E.S., Pérez-Jiménez, J., Saura-Calixto, F., and Mancini-Filho, J. 2010. Bioactive compounds and antioxidant capacities of 18 non-traditional tropical fruits from Brazil. Food Chem. **121**(4): 996–1002. doi:10.1016/j.foodchem.2010.01.037.

Shafin, K., Pesout, T., Lorig-Roach, R., Haukness, M., Olsen, H.E., Bosworth, C., et al. 2020. Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes. Nat. Biotechnol. **38**: 1044–1053. doi:10.1038/s41587-020-0503-6. PMID: 32686750.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, **31**(19): 3210–3212. doi:10.1093/bioinformatics/btv351. PMID: 26059717.

Singh, R., Ong-Abdullah, M., Low, E.T.L., Manaf, M.A.A., Rosli, R., Nookiah, R., et al. 2013. Oil palm genome sequence reveals divergence of interfertile species in old and new worlds. Nature, **500**: 335–339. doi:10.1038/nature12309. PMID: 23883927.

Stiehler, F., Steinborn, M., Scholz, S., Dey, D., Weber, A.P.M., and Denton, A.K. 2021. Helixer: cross-species gene annotation of large eukaryotic genomes using deep learning. Bioinformatics, **36**(22–23): 5291–5298. doi:10.1093/bioinformatics/btaa1044. PMID: 33325516.

Walker, R.P., Battistelli, A., Moscatello, S., Técsi, L., Leegood, R.C., and Famiani, F. 2015. Phosphoenolpyruvate carboxykinase and gluconeogenesis in grape pericarp. Plant Physiol. Biochem. **97**: 62–69. doi:10.1016/j.plaphy.2015.09.004. PMID: 26432988.

Wang, L., Albert, N.W., Zhang, H., Arathoon, S., Boase, M.R., Ngo, H., et al. 2014. Temporal and spatial regulation of anthocyanin biosynthesis provide diverse flower colour intensities and patterning in *Cymbidium* orchid. Planta, **240**(5): 983–1002. doi:10.1007/s00425-014-2152-9. PMID: 25183255.

Wickham, H. 2016. ggplot2: Elegant graphics for Data analysis. Available from https://link.springer.com/book/10.1007/978-0-387-98141-3.

Xu, W., Dubos, C., and Lepiniec, L. 2015. Transcriptional control of flavonoid biosynthesis by MYB–bHLH–WDR complexes. Trends Plant Sci. **20**(3): 176–185. doi:10.1016/j.tplants.2014.12.001. PMID: 25577424.

Yamaguchi, K.K., Pereira, L.F.R., Lamarão, C.V., Lima, E.S., and da Veiga-Junior, V.F. 2015. Amazon açaí: chemistry and biological activities: a review. Food Chem. **179**: 137–151. doi:10.1016/j.foodchem.2015.01.055. PMID: 25722148.

Yang, Y., Bocs, S., Fan, H., Armero, A., Baudouin, L., Xu, P., et al. 2021. Coconut genome assembly enables evolutionary analysis of palms and highlights signaling pathways involved in salt tolerance. Commun. Biol. **4**: 105. doi:10.1038/s42003-020-01593-x. PMID: 33483627.

Yang, Y., Cui, B., Tan, Z., Song, B., Cao, H., and Zong, C. 2018. RNA sequencing and anthocyanin synthesis-related genes expression analyses in white-fruited *Vaccinium uliginosum*. BMC Genom. [Electronic Resource], **19**: 930. doi:10.1186/s12864-018-5351-0.

Ye, W., Wang, T., Wei, W., Lou, S., Lan, F., and Zhu, S. 2023. The full-length transcriptome of *Spartina alterniflora* reveals the complexity of high salt tolerance in monocotyledonous halophyte. Plant Cell Physiol. **61**(5): 882–896. doi:10.1093/pcp/pcaa013.

Zhang, L., Li, X., Pang, Y., Cai, X., Lu, J., Ren, X., and Kong, Q. 2021. Phenolics composition and contents, as the key quality parameters of table grapes, may be influenced obviously and differently in response to short-term high temperature. LWT, **149**: 111791. doi:10.1016/j.lwt.2021.111791.

Zhang, Z., Liang, X., Lu, L., Xu, Z., Huang, J., He, H., and Peng, X. 2020. Two glyoxylate reductase isoforms are functionally redundant but required under high photorespiration conditions in rice. BMC Plant Biol. **20**: 1–12. doi:10.1186/s12870-020-02568-0.

Zhao, X., Li, F., Ding, J., Fu, X., Shang, J., Kong, X., et al. 2024. Comparative RNA-seq analysis of tomato (*Solanum lycopersicum* L.) provides insights into natural and postharvest ripening. Postharvest Biol. Technol. **216**: 113079. doi:10.1016/j.postharvbio.2024.113079.