Genome-wide identification and characterization of the late embryogenesis abundant (LEA) protein-encoding gene family related to water deficit response in Theobroma cacao

Velasco-Anacona Ginna Patricia [1,2*], Giraldo-Murcia Julian Alejandro[1,3], Noriega-Ortega Jhon Eduar[1], Rojas-Cruz Alexis Felipe[4]

[1]Laboratorio de Ciencias Básicas, Servicio Nacional de Aprendizaje SENA, Centro de Formación Agroindustrial La Angostura, Campoalegre, Huila, Colombia
[2]Universidad Nacional de Colombia, Sede Bogotá, Facultad de Ciencias Agrarias, Bogotá, Colombia
[3]Escuela de Ciencias Agrícolas, Pecuarias y del Medio Ambiente, Universidad Nacional Abierta y a Distancia UNAD, Popayán, Cauca, Colombia
[4]Departamento de Nutrición y Bioquímica, Facultad de Ciencias, Pontificia Universidad Javeriana, Bogotá D.C, Colombia

*Corresponding author
E-mail: velascoginna@gmail.com
ORCID: https://orcid.org/0000-0002-8686-7136

Online Resource 3 Detailed protocol for RNA extraction and RT-qPCR assays

- **Total RNA extraction from cacao leaf tissue**

For RNA extraction from *Theobroma cacao* leaf tissue, a protocol was optimized based on the use of the GeneJET™ Plant RNA Purification Mini Kit (Thermo Fisher Scientific, USA) with the following steps: Young apical leaves from cacao seedlings were collected under different treatments, immediately submerged in RNAlater® Stabilization Solution (Thermo Fisher Scientific, United States), and stored at -80°C until use. The frozen plant material was ground to a fine powder in a mortar pre-treated with RNase AWAY® Surface Decontaminant (Thermo Fisher Scientific, United States) and liquid nitrogen.

Approximately 20 mg of ground tissue was transferred into sterile 2 ml microcentrifuge tubes, and 1 ml of extraction buffer (0.18 M Tris-HCl, 0.09 M LiCl, 4.5 mM EDTA, 1% SDS, pH 8.2) was added, followed by 300 µl of phenol equilibrated with TLE solution (0.2 M Tris-HCl, 0.1 M LiCl, 5 mM EDTA, pH 8.2). The mixture was homogenized by gentle inversion for 2 minutes. Subsequently, 300 µl of chloroform was added, mixed by inversion for 1 minute, and the samples were incubated at 50°C for 20 minutes. Phase separation was achieved by centrifugation at 13,000 rpm for 20 minutes at 4°C.

The aqueous phase was carefully transferred to a new sterile tube, avoiding disturbance of the interphase. Three sequential phenol extractions were performed, followed by a final extraction with chloroform. RNA was precipitated by adding one volume of isopropanol and incubating at -20°C for 12 hours. The RNA precipitate was collected by centrifugation at 13,000 rpm for 20 minutes at 4°C and washed three times with 500 µl of 70% ethanol prepared with DEPC-treated water.

The RNA pellet was solubilized in 100 µl of RNA Storage Solution, and RNA concentration was determined using fluorometry with a NanoPhotometer® NP80 (Implen GmbH. Munich, Germany). RNA quality was assessed through electrophoresis on 1% agarose gels stained with Safe DNA Dye, confirming integrity and suitability for downstream analyses.

- **cDNA Synthesis and Primer Specificity Testing**

Complementary DNA (cDNA) was synthesized from 100 ng of total RNA using M-MuLV reverse transcriptase (Thermo Fisher Scientific, USA) and oligo(dT)18 primers, following the manufacturer's recommendations. The resulting cDNA was preserved at −80 °C until quantitative PCR (qPCR) analysis. Primer specificity for each candidate gene was evaluated by endpoint PCR using Taq PCR Master Mix (Qiagen, Germany). Reactions were carried out in a final volume of 25 µL containing 10 ng of cDNA (corresponding to the initial RNA input) and 0.35 µM of each primer. Optimal annealing temperatures were established through gradient PCR assays, testing temperature intervals of 0.5 °C within a 5–7 °C range surrounding the predicted melting temperature (Tm) of each primer pair. Amplification products were resolved on 1.5% (w/v) agarose gels to verify the expected fragment size.

- **Quantitative Real-Time PCR (qPCR)**

Quantitative PCR assays were performed using the DyNAmo Flash SYBR Green qPCR Kit (Thermo Fisher Scientific, CA, USA) with 0.4 µM of each primer pair in a 20 µL reaction volume. Serial cDNA dilutions (15.00, 1.50, 0.15 and 0.015 ng) were used to generate standard curves in triplicate for each primer pair.

Amplifications were carried out on a QuantStudio 1 Real-Time PCR System (Applied Biosystems, Thermo Fisher Scientific, CA, USA). The slopes of the standard curves and corresponding linear regression coefficients (R^2) were calculated to determine the linear detection range and amplification efficiencies (E) (Online Resource 4). Amplification efficiency was calculated according to Pfaffl (2004). The ACPB gene was selected as the reference for normalization of differential expression analysis, based on its stable expression across sampling days during the water stress treatment (Osorio et al. 2021).