Transcriptome (ESTs) of Avocado "Native" Mexicano Early Seed Development Shows Abundance of Regulatory, Antioxidant and Defense Genes

Luis María Suárez-Rodríguez,

Fernando Sánchez-Albarrán,

Huber León-Corona and Rodolfo López-Gómez

Additional information is available at the end of the chapter

Abstract

Avocado native "Mexicano" (*Persea americana* var drymifolia) has been a really important species in agricultural and indigenous medicine. In the agricultural world, it has been the germplasm source for the generation of economically important cultivars like Hass and it is the main source of rootstocks for the world production of Hass avocado fruit. In spite of its importance, little is known about the molecular network of seed-fruit development. The aim of this work was to know the expressed genes (ESTs) during the early avocado native "Mexicano" seed development. Using total RNA we constructed cDNA libraries of fourth months seed development, sequencing, assembling and bioinformatic analysis was made. For validation, a semi-quantitative PCR experiments with the most abundant genes were made. About 5005 ESTs from the 5' representing 1653 possible unigenes were isolated. After assembling process, we have 171 genes that are closely related to *Nelumbo nucifera* sequences. The transcriptome is dominating by one bHLH transcription factor, three metallothioneins and snakin, suggesting its main role in seed development. Until now, there are no molecular studies in avocado seed development.

Keywords: avocado fruit, transcriptome, transcription factor, stress, antimicrobials, cotyledon, alignment, domain, phylogenetic, metal homeostasis, cysteine, seed development, gene expression

1. Introduction

Avocado (*Persea americana* Mill.) is an oleaginous fruit produced by a tree belonging to the magnoliid clade, a basal linage of flowering plants. It belongs to the large plant family of Lauraceae, with approximately 2500–3000 species [1, 2]. Avocado has been rapidly incorporated as a component of human diet in many countries [3]. Due to low cost, vigor of seedling growth and easy propagation, most of the countries are still using seeds to produce rootstocks for grafted avocado trees despite their genetic variability [4]. Several Mexican varieties are derived from seeds that are resistant to attack by *Phytophthora cinnamomi* [5, 6] and are adapted to the soil and environmental conditions of the region. The Mexican state of Michoacán is the primary avocado-producing region in the world, and all the rootstocks used for the commercial production of cultivar Hass are obtained from *P. americana* var. drymifolia ("nativo mexicano") [7]. The principal consume form is as fresh fruit, but is really important in cosmetic industry [8]. Avocado plant has medicinal properties, including cancer prevention [9–11]. There is ethnopharmacological information on the use of avocado seeds for the treatment of health-related conditions, especially in America. Recent research has shown that the avocado seeds are rich in phenolic compounds and these maybe play a role in putative health effects [12]. The avocado fruit is a berry of one carpel containing a single seed. This large and very conspicuous seed is made up of two fleshy cotyledons and a central attached plumule, hypocotyl and radicle, the whole surrounded by two papery seed coats closely adherent to each other. There is no endosperm left in the seed at maturity. The cotyledons are formed of indifferentiated parenchyma tissue interspersed with occasional idioblasts. Starch is the main storage material of the cotyledons and is present in great abundance [13]. Despite its importance, avocado seed development remains uncharacterized. To date, little information is available regarding the molecular biology of the seed. Analysis of expressed sequence tags (ESTs) is a rapid and effective method to identify novel genes or to investigate gene expression in different tissues, organs and plants [14, 15]. Furthermore, EST libraries and databases could provide valuable resources for functional genomic studies [16].

In principle, the frequency with which the sequence of a given gene is read in ESTs sequencing projects should reflect the relative abundance of the corresponding mRNA. This approach uses EST counts to infer the relative level of expression of a gene [17–19].

In this work, we report the analysis of an ESTs collection from immature avocado nativo mexicano seeds and the analysis of expression of bHLH transcription factor, metallothioneins (MTs) and snakin like more abundantly expre3ssed.

2. Materials and methods

2.1. Biological material

Seeds from avocado nativo mexicano fruits of three stages of development (1, 4 and 8 months) were excised from fruits and frozen immediately in liquid nitrogen until use. The materials

were collected in the avocado Germplasm Bank of the Instituto Nacional de Investigaciones Forestales y Agropecuarias (INIFAP; Uruapan, Michoacán México).

2.2. cDNA library construction and sequencing by Sanger method

Total RNA from frozen seed tissue of 4 months of development was extracted using López-Gómez et al.'s protocol with some modifications [20]. The cDNA complementary library was built from 1 μg of total RNA using SMART™ cDNA Library Construction Kit (Clontech). The obtained cDNA sequences were cloned into pTripIEx2 vector. Cleavage experiments were made using *E. coli* BM25.8 cells to obtain the plasmid pTriplEx2. Sequencing reactions were performed using ABI PRISM BigDye Terminators v3.0 kit (Applied Biosystems), by 5′ end of plasmids extracted from random clones. The sequences obtained were filtered by quality using PHRED [21]; vector masked and trimming of poly A/T were performed using LUCY2 software [22] resulting in 5002 high-quality reads.

2.3. Assembly and identification of full-length cDNA

Sanger sequences were assembled using default parameters of the 454 Newbler-Assembler v1.1.03.24 (454 Life Science, Branford, CT) using 16,526 generated by the University of Florida and the Washington University Genome Sequencing Center.

Unigene set was generated by combining all assembled contigs and non-assembled reads (singlets). The consensus sequences of the Unigenes were analyzed with EuGeneHom [23] to identify the Unigenes that contained the components of a complete cDNA (5′ UTR, ORF and 3′ UTR): http://genoweb.toulouse.inra.fr/eugene/EuGeneHom/cgi-bin/EuGeneHom.pl.

2.4. Functional annotation

Stand-alone BLAST software was obtained from the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov). The unigenes were compared by BLAST against nucleotides and proteins plant databases. The BLAST results from different databases were used for gene ontology (GO) mapping and annotation. Blast2Go software was used to perform GO functional classification.

2.5. Transcript characterization and homologous search

The transcripts were analyzed by UGENE V1.26.1 software for the identification of ORF, CDS and hypothetical protein sequence and physicochemical parameters, using the standard genetic code.

2.6. Alignment and phylogeny

Mega7.0.14 software was used for alignment (Clustal W algorithm with Blossom 62 matrix) and reconstruction of the phylogenetic pattern (Neighbor Joining model with JTT matrix-based model for distance computing and 1000 replicates as bootstrap test).

GenBank Accession	Gene name	Primer sequence (5′–3′)	Tm (°C)
MF353071	PabHLHS1	**F:** TAGCGACAGGATTTGGCAGTTT	65°C
		R: TGCCCCGTCGTCTCTTTCTTCTACC	
MF353072	PaMT2a	**F:** TGAGAGGGAGATTGGAGGAG	54°C
		R: TCCCAACCACAGCATAGTACC	
MF353073	PaMT2b	**F:** CATGCACCTGCAAATGAGAG	60°C
		R: CACCCAGATACAGCAGGAGAA	
MF353074	PaMT3	**F:** TTGGTGTTGGTGTCCTATGC	57°C
		R: CATGACTCAACCACACACACC	
AGC92009	PaSn	**F:** TCCTTGCTTTCCTTCTCTTCAC	64°C
		R: AGATGTCCCGCAATCTATTTGT	
	SUMO	**F:** GATAAGAAGCCCACGGATCA	55°C
		R: GACGGCCATCGAATAAGAAC	

Table 1. Primers used in semi-quantitative PCR.

2.7. Semi-quantitative RT-PCR

The RNA from different seed stages (1, 4 and 8 months of development) digested with DNAse I amplification grade (Invitrogen) was used as a template in sqPCR reactions, and the synthesis of cDNAs was carried out with the First Strand cDNA Synthesis Kit (Thermo Scientific). In Veriti 96-well Thermal cycler (Applied Biosystems), in reactions of 10 µl (100 ng/µl cDNA,10X Reaction Buffer, 2 mM dNTPs, 50 mM $MgCl_2$, 1 U Taq DNA polymerase), sqPCR primers (**Table 1**) were designed by Primer3 webtool (http://bioinfo. ut.ee/primer3-0.4.0/); the primers for bHLH transcription factor, three metallothioneins, snakin and ubiquitin as reference gene were selected. The amplification procedures were 95°C - 10min., 30 cycles (95°C - 45 s, annealing 30 s, amplification 72°C - 45 s), 72°C - 7 min. Gene expression ratio between selected gene and endogenous control was calculated using band intensity measured with GelAnalyzer 2010. The semi-quantitative PCR was performed with three repeats.

3. Results

3.1. Sequencing and assembly of *Persea americana* var. drymifolia seed transcriptome

A total of 5005 sequences was assembled and obtained 3328 sequences (171 contigs of 3222 sequences + 106 singletons); these sequences were imported into BLAST analysis. In contig

assembly for the avocado seed ESTs, a total number of contigs was 171, derived from 3222 sequences and unique transcripts represented was 106 sequences. Contig length ranged from 84 to 1149 bp (**Figure 1**), the peak EST length was 84–195 bp with 478 sequences in range. The shortest sequence examined was 84 bp to known genes like GTP-binding nuclear protein.

3.2. Functional annotation

Functional interpretation is an important step in the analysis of transcriptomics which cannot be done without the availability of functional annotation. The most widespread and probably most extensive functional annotation schema for gene and protein sequences is the Gene Ontology (GO) [24] as standard in all public databases. Automatic functional annotation methods basically rely on sequence, structure, phylogenetic or co-expression relationships between known and novel sequences [25]. A total of 277 uniESTs sequences were manually annotated for a closer understanding of gene expression in avocado seed. The annotation proceeds through three basic steps: homologs search, GO term mapping and actual annotation. At the first step, NCBI-BLASTX and BLASTN are typically used, and for this work, the e-value $1xE^{-5}$, cut-off:33 and the number of 20 retrieved BLAST hits are used. These uniESTs were classified into five functional categories, including antioxidative protection (677, 20.34%), transcription regulatory (1013, 30.44%), defense (507, 15.23%), cellular structure and organization (287, 8.62%) and unknown (844, 25.36%).
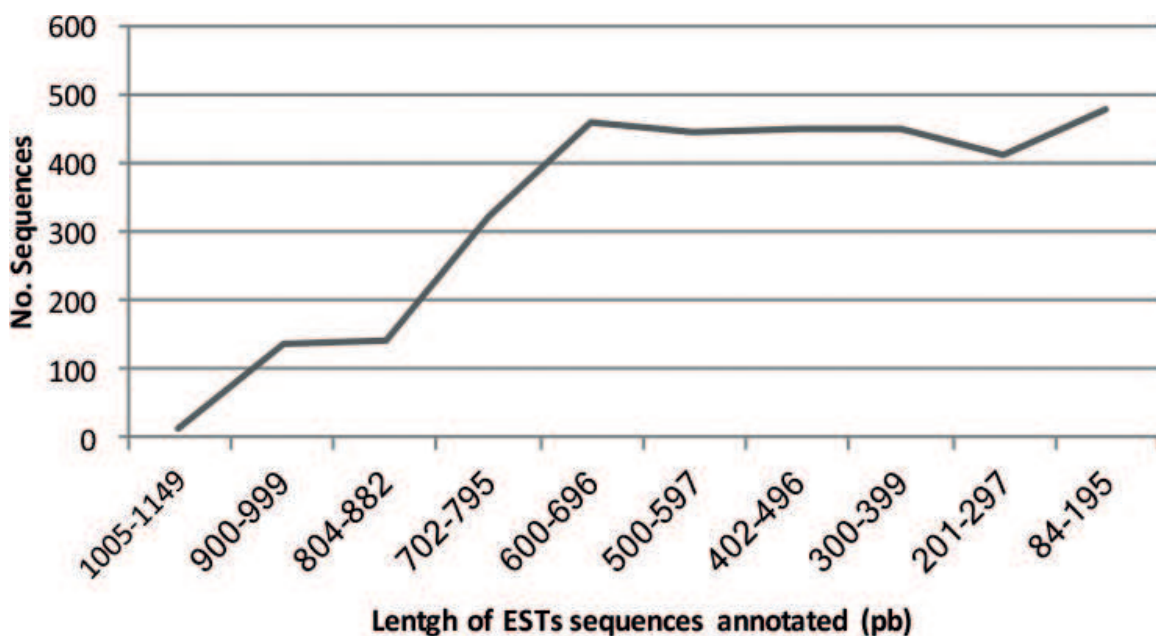

Figure 1. Sequence length distribution of ESTs from avocado native mexicano seed transcriptome. Total number of sequences analyzed were 3328. Data include both contigs and singletons.

Gene ontology annotations and functional analyses of avocado seed transcriptome were carried out with automated software Blast2GO. These were assigned into three standard classifications: biological processes, molecular functions and cellular components, and summarized according to GO criteria. The majority GO annotation was for biological process (65.34%), cellular component (19.34%) and molecular function (15.33%). In addition, the organisms closely related to the genetic load on avocado seed were reviewed within the databases, and the majority of analyzed sequences (23.47%) (**Figure 2**) were closely related to *Nelumbo nucifera* sequences (**Figure 3**). The 20 most abundant uniESTs and their annotations are shown in **Table 2**.

3.3. Avocado seed abundant genes and validation

The most abundant sequences match with metallothionein genes. This result suggests that metallothionein genes dominate the avocado seed transcriptome like avocado fruit [26]. Metallothioneins (MTs) were discovered by Margoshes and Vallee as cadmium-bound proteins isolated from the cortex of the equine kidney. These proteins were named for the high sulfur content and metals they are able to bind; depending on the metal species, these may possess more than 20% of its nature of metal ions [27, 28]. Mammalian metallothioneins are 60 amino acid peptides with 20 Cys residues and a molecular mass of about 6–7 kDa. Mammalian MTs are capable of binding up to 7 divalent metal ions via mercaptide bonds (sulfur-metal) with the Cys residues. By convention, any peptide or protein that resembles several characteristics of mammalian metallothioneins can be classified as metallothionein [28]. Plant metallothioneins have two (highly conserved) sequence similarity regions corresponding to the two Cys-rich terminal domains joined by a less conserved "spacer" (about 40 aa without Cys residues). In plants, the most distinctive feature is to have a large spacer, which differs from the MT of the
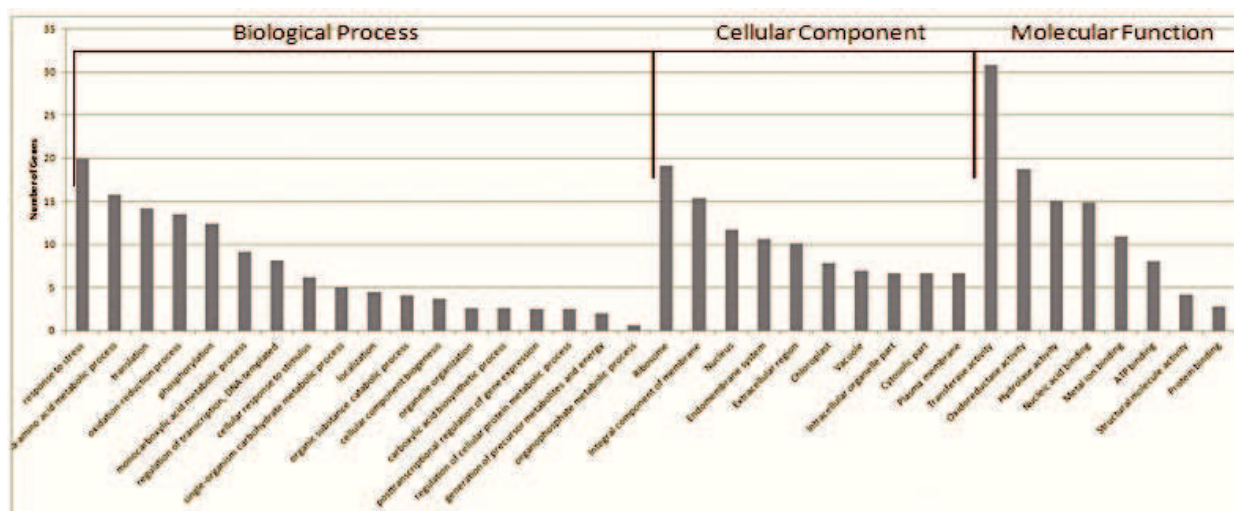

Figure 2. Gene ontology classification of *P. americana* var. drymifolia seed transcriptome. Unigenes with BLASTX matches were classified into three main GO categories: biological process, cellular components and molecular functions.
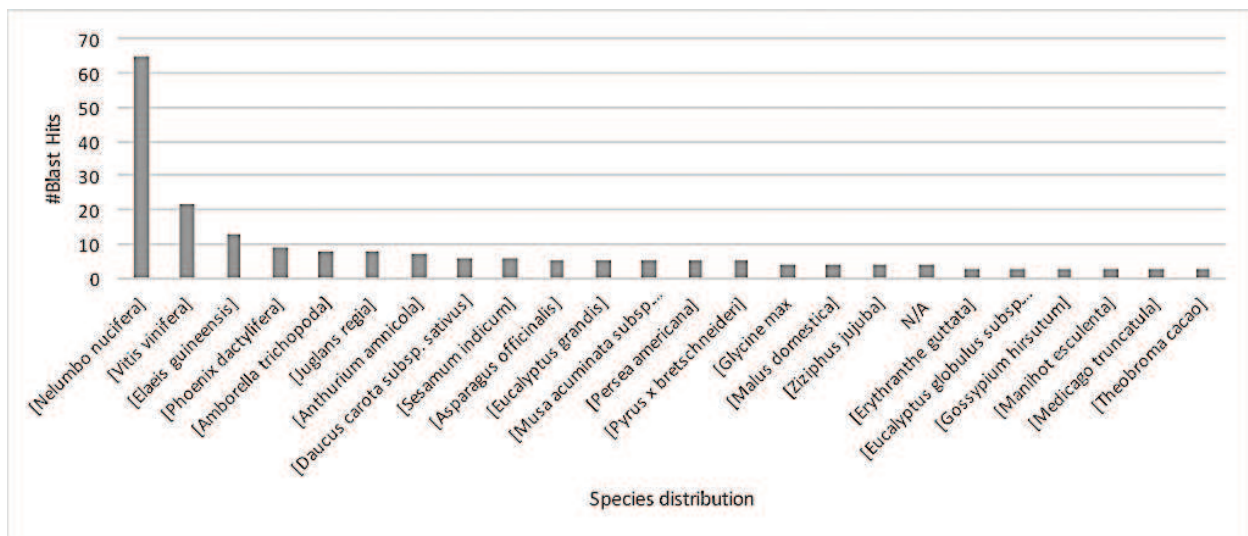
Figure 3. Main organisms with which the avocado seed transcriptome is closely related.

animals in which the Cys-rich domains are separated by a short spacer of less than 10 amino acids which do not include aromatic residues. The distribution of Cys residues as well as the length of the spacer region served to classify more MT of plants into four types, namely group 1, 2, 3 and 4 [29].

From the analysis of these abundant transcripts, we founded the existence of three MT's genes on avocado seed: PaMT2a, PaMT2b and PaMT3, which were registered in GenBank database with an accession code shown in **Table 3**. Characteristics predicted in silico for avocado metallothioneins, we have two sequences belonging to Methallotionein-2 super-family due to the two highly conserved Cys-rich motifs and the long spacer in the middle of them. PaMT3 keeps the spacer but Cys-rich motifs are not so conserved grouping this as part of the third family of plant metallothioneins (MT3). Alignment shown that the most conserved amino acids are around the Cys residues for both MT2 and MT3 groups, which are associated with the "metal binding clusters" (**Figures 4** and **5**). The alignments performed identify the amino and terminal carboxyl regions as having the most conserved Cysteine sequences, which correspond to the metal binding clusters. The intermediate spacer of about 40 amino acids is much more variable but has no cysteines. In mammalian metallothioneins, this spacer is very small (8 amino acids) and has no aromatic amino acids. However, in the family 2 to which the plants belong, we can find conserved tyrosine residues in the spacer, as well as several less conserved phenylalanines present in the metallothionein sequences of types 2 and 3 (**Figures 4** and **5**). Seed metallothioneins correspond with the three reported: NnMT2a, NnMT2b and NnMT3 [30]; the avocado seed metallothioneins are closely related with *Nelumbo nucifera* metallothioneins like 79% for NnMT2a-PaMT2a; 66% NnMT2b-PaMT2b and 71% for NnMT3-PaMT3 **Figure 6**. NnMT2a and NnMT3 were associated with processes of seed germination, tolerance to accelerated aging and salt on Arabidopsis [30]. However, avocado seeds have a low tolerance

ESTs	Description	Length	e-value	GO IDs	Abundance
arlgES1	Transcription factor bHLH96	942	1.35E-103	GO:0046983	123
arlgES153	Metallothionein type 3	195	4.52E-28	GO:0043167	123
arlgES154	Stress response nst1	696	1.68E-34	GO:0009507; GO:0010207	113
arlgES2	40S ribosomal S29	171	1.77E-32	GO:0003735; GO:0043167; GO:0005829; GO:0005840; GO:0006412	100
arlgES3	PREDICTED: uncharacterized protein LOC104611921	282	4.53E-37	N/A	96
arlgES4	Vesicle-associated membrane 726	579	5.71E-130	GO:0016192; GO:0005575	90
arlgES6	ras-related RIC2	525	2.61E-118	GO:0007165; GO:0043167; GO:0005622	78
arlgES7	Metallothionein 2	240	3.91E-30	GO:0043167	77
arlgES8	Isocitrate dehydrogenase [NADP]	588	1.02E-132	GO:0044281; GO:0006091; GO:0016491; GO:0043167	74
arlgES10	kDa proline-rich	483	4.35E-25	N/A	69
arlgES155	Early nodulin-93-like	318	8.41E-43	GO:0005575	69
arlgES156	PREDICTED: uncharacterized protein LOC103701850 isoform X3	552	2.83E-19	GO:0005739	66
arlgES157	Potassium transporter 12 isoform X1	366	2.97E-55	GO:0022857; GO:0009536	64
arlgES12	60S ribosomal L24	492	5.32E-63	GO:0003735; GO:0022618; GO:0005829; GO:0042254; GO:0005840; GO:0006412	58
arlgES13	60S ribosomal L7–2-like	738	1.65E-136	GO:0003735; GO:0005829; GO:0042254; GO:0005840; GO:0006412	57
arlgES15	Ethanolamine utilization eutQ	291	3.63E-61	N/A	54
arlgES158	Programmed cell death 4	636	1.00E-80	N/A	52
arlgES16	Type 2 metallothionein	243	5.56E-24	GO:0043167	51
arlgES159	Mitochondrial	178	4.72E-35	GO:0005739	50
arlgES161	Snakin	318	4.00E-57	GO:0009740	48

Table 2. Top 20 most abundant annotated ESTs of avocado nativo mexicano seed transcriptome.

to aging. Basic function of MTs is metal homeostasis and has been reported during biotic and abiotic stress conditions too [31]. Recently, it has been reported that one metallothionein interacts with a cytoskeleton protein in the nucleus of rice cells in response to salt stress [32].

Name	GenBank	Length (aa)	Weight (kDa)	Isoelectric Point	Cys Residues	% Cys
PaMT2a	MF353072	80	7.9	5.11	14	17.5
PaMT2b	MF353073	80	7.9	4.54	14	17.5
PaMT3	MF353074	64	67	4.5	10	15.6

Table 3. Characterization of Persea americana var. drymifolia metallothioneins.
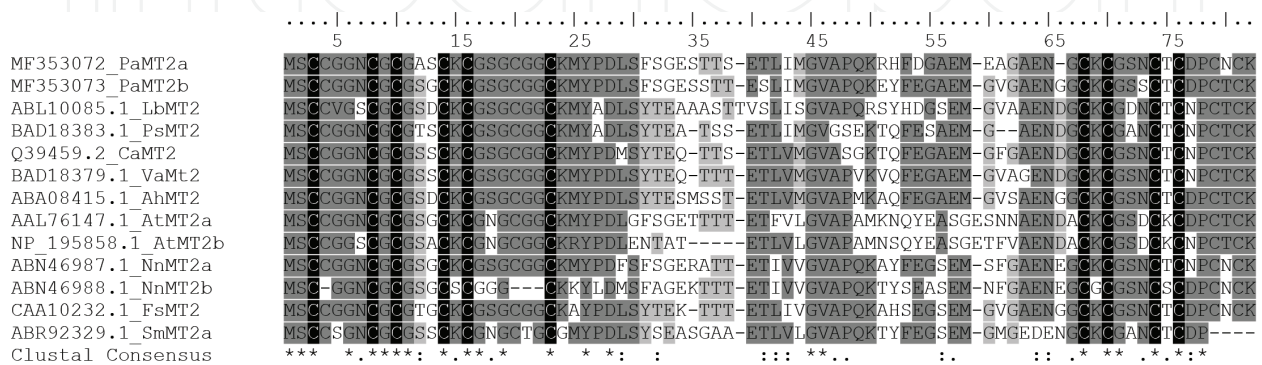

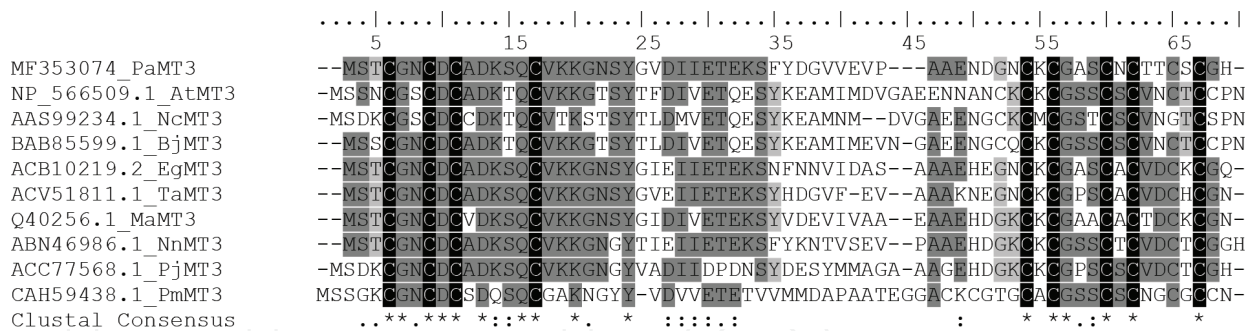**Figure 4.** Alignment of plant metallothioneins of subfamily 2.

Figure 5. Alignment of plant metallothioneins of subfamily 3.

Sequences of related proteins were obtained from Zhou et al. [30] and downloaded from NCBI-GenBank: BAD18376.1 GmMT1 (*Glycine max*); BAD18378.1 VaMT1 (*Vigna angularis*); AAF04584.1 MsMT1 (*Medicago sativa*); BAD18382.1 PsMT1 (*Pisum sativum*); BAD18380.1 VfMT1 (*Vicia faba*); ABL10085.1 LbMT2 (*Limonium bicolor*); BAD18383.1 PsMT2 (*Pisum sativum*); Q39459.2 CaMT2 (*Cicer arietinum*); BAD18379.1 VaMT2 (*Vigna angularis*); ABA08415.1 AhMT2 (*Arachis hypogaea*); AAL76147.1 AtMT2a (*Arabidopsis thaliana*); NP_195858.1 AtMT2B (*Arabidopsis thaliana*); ABN46987.1 NnMT2a (*Nelumbo nucifera*); ABN46988.1 NnMT2b (*Nelumbo nucifera*); CAA10232.1 FsMT2 (*Fagus sylvatica*); ABR92329.1 SmMT2a (*Salvia miltiorrhiza*); NP_566509.1 AtMT3 (*Arabidopsis thaliana*); AAS99234.1 NcMT3 (*Noccaea caerulescens*); BAB85599.1 BjMT3 (*Brassica juncea*); ACB10219.2 EgMT3 (*Elaeis guineensis*); ACV51811.1 TaMT3 (*Typha angustifolia*); Q40256.1 MaMT3 (*Musa acuminata*); ABN46986.1 NnMT3 (*Nelumbo*
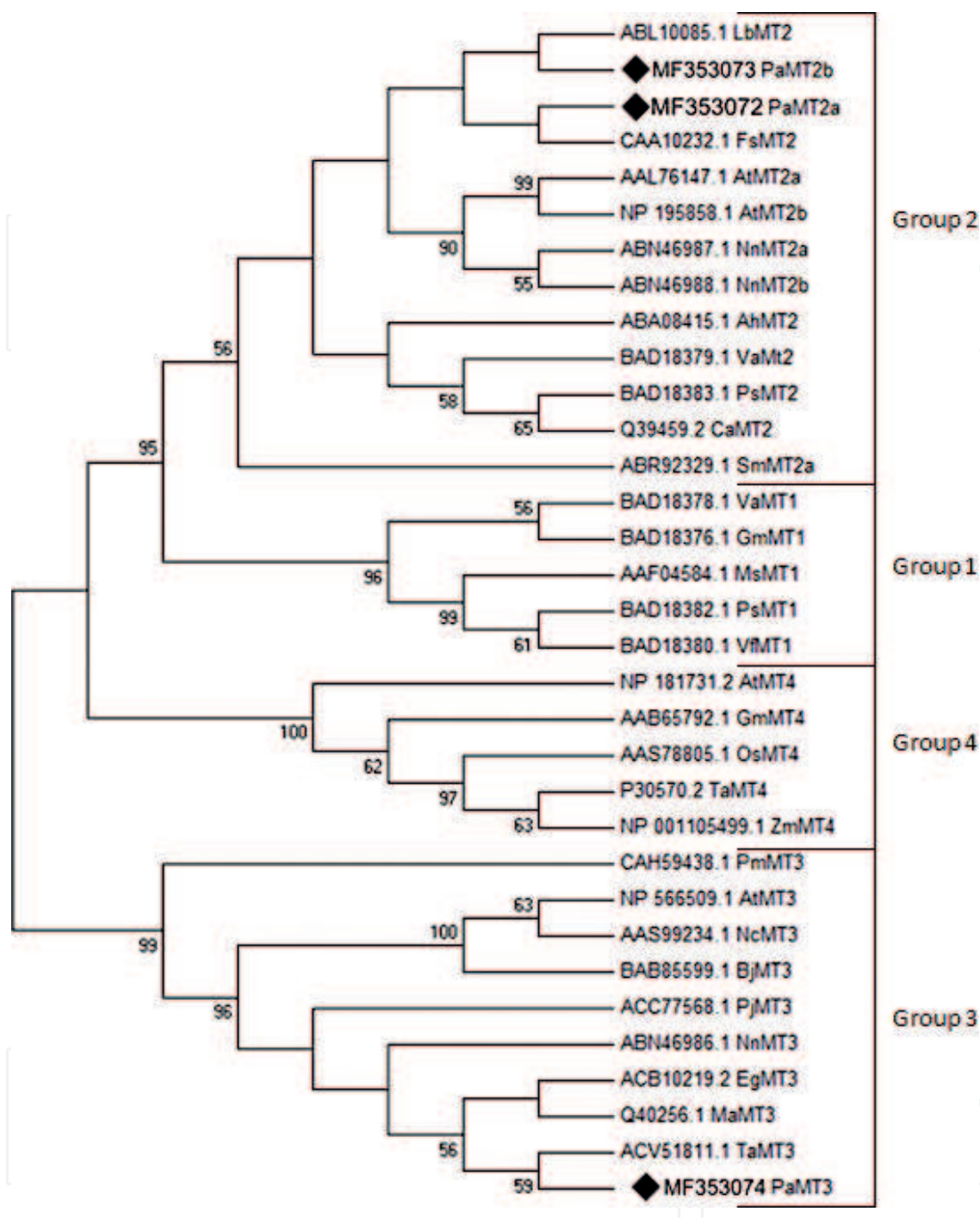
Figure 6. Phylogenetic analysis of avocado and plant metallothioneins. The evolutionary history was inferred using the neighbor-joining method. The bootstrap consensus tree inferred from 1000 replicates. The evolutionary distances were computed using the JTT matrix-based method. The analysis involved 33 amino acid sequences. All positions with less than 85% site coverage were eliminated.

nucifera); ACC77568.1 PjMT3 (*Prosopis juliflora*); CAH59438.1 PmMT3 (*Plantago major*); NP_181731.2 AtMT4 (*Arabidopsis thaliana*); AAB65792.1 GmMT4 (*Glycine max*); P30570.2 TaMT4 (*Triticum aestivum*); AAS78805.1 OsMT4 (*Oryza sativa* Japonica Group) and NP_001105499.1 ZmMT4 (*Zea mays*), [30].

Another abundant transcript is a messenger codified for a transcription factor type bHLH. The basic domain (bHLH) is a highly conserved amino acid motif that defines a group of transcription factors, which was initially described in animals and was soon discovered in all major eukaryotic lineages [33]. Proteins containing a bHLH domain (referred to as bHLH proteins) are involved in a variety of regulatory processes; their functions include the regulation of neurogenesis, myogenesis and the development of the heart in animals [34, 35], control of phosphate uptake and glycolysis in yeast [36] or modulation of secondary metabolism pathways, epidermal differentiation and environmental responses in plants [37]. The bHLH domain consists of two distinct segments composed by 50–60 amino acids, 10–15 mostly basic amino acids form the stretch (basic region) and approximately 40 amino acids form the two amphipathic helices separated by a loop (helix-loop-helix region). The analysis of the structure of bHLH proteins (yeast and mammalian) showed the basic region made in the DNA contact, while the two helices promote the formation of heterodimers between bHLH proteins [35]. These bHLH transcription factors are generally classified into six major groups (FAs) based on their ability to bind to DNA [35, 38, 39]. Most bHLH proteins are classified into group A or B; in group A, it is expected to bind to E-box consensus sequences (CACCTG or CAGCTG), in group B, it is specifically bind to G-Box consensus sequences (CACGTG or CATGTTG) and in group C, bHLH proteins share a PAS domain and bind to the recognized sequences without a need a E-box (ACGTG or GCGTG) sequences. The E group includes bHLH proteins containing a conserved Pro or Gly residue at a key position within the basic region, preferably bind to sequences referred to as N-boxes (CACGCG or CACGAG), and further share an additional WRPW motif. Groups D and F represent particularly proteins which were separated in the basic region. Some group D proteins have been described as being unable to bind to DNA and could form heterodimers that function as negative regulators of bHLH binding to DNA [40]. Group F includes so-called COE proteins. A phylogenetic study indicated that group A contained mammalian bHLH proteins and lacked bHLH plant proteins. The other groups had a mixture of different species and most of the bHLH proteins of plants belonged to group B [41, 42], It has been shown that the bHLH family of proteins in plants is monophyletic and subjected to significant radiation before the evolution of mosses; bHLH groups established in terrestrial plants during the first 400 million years were conserved during the later evolution of plants, although there were many duplications of genes. The transcription factors are very varied since it does not have many amino acids conserved throughout its sequence; nevertheless in the sites of union to the DNA like the case of the bHLH, the great majority of its amino acids is conserved within its main motive. Due to their propensity to form homodimers or heterodimers, bHLH proteins can participate in an extensive set of a combinatorial interactions leading to the regulation of multiple transcriptional programs. The development of fleshy fruits involves complex physiological and biochemical changes. Recent studies have described the involvement of bHLH proteins in the determination of plant organ size. The SPATULA protein was shown to control cotyledon, leaf and petal expansion by affecting cell proliferation in *Arabidopsis thaliana* [43]. Nicolas et al. [44] described a bHLH transcription factor preferentially expressed in grape berry fruit, but is weakly detected in seeds. This gene is involved in cell size determination. Three basic helix-loop-helix transcription factors (bHLH) were also found to be involved in *Arabidopsis* fruit dehiscence process: ALCATRAZ (*ALC*), SPATULA (*SPT*) and INDEHISCENT (IND); they form a regulatory network that orchestrates

the differentiation of the valve margin, allowing seed dispersal [45]. A protein Blast was performed and the sequences that were selected for a multiple alignment were the ones that presented greater coverage and identity with the transcription factor bHLH of Mexican native avocado seed. High sequence variability was found for the nearest bHLH motif, which presenting a large number of conserved amino acids **(Figure 7)**.

However, since they do not have any additional information regarding the function and/or tissue in which the function is performed, we analyzed our sequence with bHLH sequences of which their function or organospecificity is known; two bHLH factors were chosen with these characteristics of *Arabidopsis thaliana*, bHLH Zoupi (GenBank Accession: OAP16519) involved in seed development and SPATULA (GenBank Accession: AT4G36930) involved in
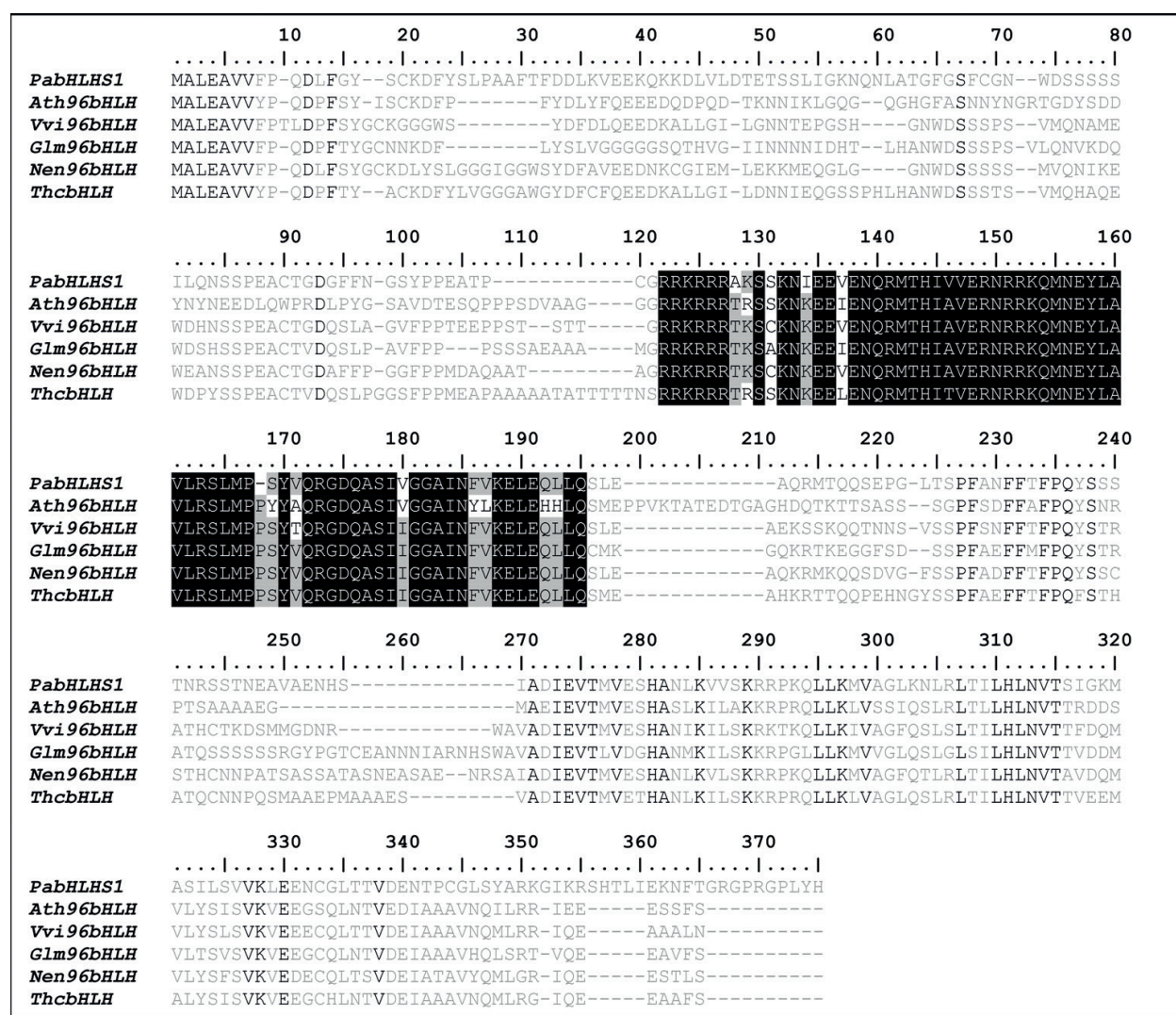
Figure 7. bHLH (PabHLHS1) avocado seed transcription factor alignment with bHLH transcription factors of some plant species. *Arabidopsis thaliana* (Ath96bHLH), *Vitis vinifera* (Vvi96bHLH), *Glicine max* (Glm96Bhlh), *Nelumbo nucifera* (96bHLH), *Theobroma cacao* (ThcbHLH), in shadow black are the bHLH motif represented in black letters the amino acids that are conserved in these proteins and gray letters the amino acids differs between them.

cell development; however, no conserved amino acids were found with the avocado bHLH (data not shown). To try to infer the function of this transcription factor highly expressed in avocado seed, it will be necessary to carry out research studies of the recognition boxes inside in DNA to infer the possible association to the group to which it belongs. It is probable that the bHLH transcription factor of the avocado seed has a principal paper in seed differentiation and development.

Multicellular organisms produce small cysteinerich antimicrobial peptides (AMPs) as an innate defense against pathogens. Native Mexican avocado seed abundantly express Snakin (PaSn) gene. These kind of AMPs were initially isolated from potato but were later found to be ubiquitous. Novel plant APs isolated include in *Arabidopsis* (family of 12-cysteine peptides).

We identified a single cDNA sequence for snakin/GASA (gibberellic acid-stimulated), which contains a coding sequence of 318 bp and encodes a predicted 106 amino acid peptide. This molecule comprises a 26 amino acid signal peptide (residues 1–26), identified by SignalP (http://www.cbs.dtu.dk/services/SignalP/) and a 79 amino acid mature peptide (**Figure 8**). An amino acid alignment of avocado snakin with other similar APs (**Figure 9**) showed that PaSn has the longest sequence compared with the previously reported StSN1 and StSN2 genes from potato. In addition, PaSn has the 12 characteristic Cys residues of this type of AP. In addition to the highly conserved 12 Cys residues, the other motifs in the PaSn protein consist of residues that are mostly polar, non-polar and basic (**Figure 9**). From these analyses, we hypothesized that the Mexican avocado snakin gene could be involved in plant defense in a similar way to that of the StSN1 and StSN2 genes in potato [7]. Until now, this is the first Snakin gene isolated from a seed.

```
  1 5'CGTGCCGGGGAAGCTAATATACGCTCCAATTGTCTATCTCTACAATACCCTTTCCTCATT
 21                                      M  A  L  S  K  T  F  L  A  F  L
 61 TCCTCCATCTGTTGAGTTCTCCTTCTTGAATGGCTCTGTCAAAGACTTTCCTTGCTTTCC
 41    L  F  T  I  L  A  L  N  L  M  D  D  A  L  G  F  Q  S  H  S
121 TTCTCTTCACTATCCTTGCTCTCAATCTCATGGATGATGCTCTTGGATTTCAGTCTCATT
 61       L  P  T  P  A  D  E  R  N  L  L  Q  Q  I  D  C  G  T  S  C
181 CGCTGCCTACACCTGCGGATGAGAGAAATCTCCTTCAACAAATAGATTGCGGGACATCTT
 81       S  A  R  C  R  L  S  S  R  P  R  L  C  K  R  A  C  G  T  C
241 GTTCTGCGAGGTGTCGATTGTCATCAAGGCCTAGACTCTGCAAGAGGGCTTGTGGGACAT
101       C  A  R  C  N  C  V  P  S  G  T  A  G  N  L  D  E  C  P  C
301 GTTGTGCGAGGTGCAACTGCGTACCATCAGGAACTGCAGGCAACTTAGACGAGTGCCCTT
121       Y  A  N  M  T  T  H  G  N  K  R  K  C  P  *
361 GCTACGCCAACATGACTACTCATGGGAACAAACGCAAATGTCCTTAAAAGACCTTATCTC
421 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTATGCTTGCTCTGTCCTTCTTTAGTA
481 TTCTGTGAGATATCTGTAAGAGAAATAAGGATGATTTGCTCTTGAGTGTGTGAATGATGA
541 AAACTGTGAGATGTGTGTAAACTTCTCTATCTGAGTAAGAGAAGTTATTCGAAAAAAAAA
601 AAAAAAAAAAAAAAAAAAAA3'
```

Figure 8. PaSn nucleotide sequence with amino acid frame translation (GenBank: KC012806). ORF in gray shaded; signal peptide (bold-underlined) is followed by the amino acid sequence of the mature peptide; the signal peptide prediction was performed using SignalP program (http://www.Cbs.Dtu.Dk/services/SignalP/).
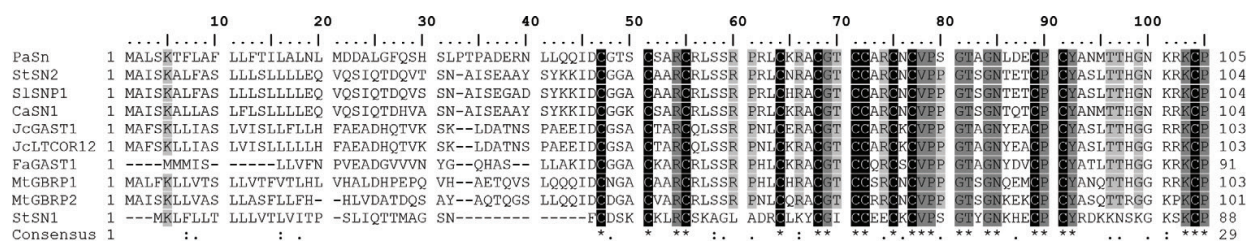
```
           10        20        30        40        50        60        70        80        90        100
        ...|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|
PaSn       1 MALSKTFLAF LLFTILALNL MDDALGFQSH SLPTPADERN LLQQIDGGTS CSARCRLSSR PRLCKRACGT CCARCNGVPS GTAGNLDECV CIANMTTHGN KRKCP 105
StSN2      1 MAISKALFAS LLLSLLLLEQ VQSIQTDQVT SN-AISEAAY SYKKIDGGGA CAARCRLSSR PRLCNRACGT CCARCNGVPP GTSGNTETCP CIASLTTHGN KRKCP 104
SlSNP1     1 MAISKALFAS LLLSLLLLEQ VQSIQTDQVS SN-AISEGAA SYKKIDGGGA CAARCRLSSR PRLCHRACGT CCARCNGVPP GTSGNTETCP CIASLTTHGN KRKCP 104
CaSN1      1 MAISKALLAS LFLSLLLLEQ VQSIQTDHVA SN-AISEAAY SYKKIDGGGK CSARCRLSSR PRLCKRACGT CCARCNGVPP GTSGNTQTCP CIANMTTHGN RRKCP 104
JcGAST1    1 MAFSKLLIAS LVISLLFLLH FAEADHQTVK SK--LDATNS PAEEIDGGSA CTARCQLSSR PNLCERACGT CCARCKGVPP GTAGNYEACP CIASLTTHGG RRKCP 103
JcLTCOR12  1 MAFSKLLIAS LVISLLFLLH FAEADHQTVK SK--LDATNS PAEEIDGGSA CTARCQLSSR PNLCKRACGT CCARCNGVPP GTAGNYEACP CIASLTTHGG RRKCP 103
FaGAST1    1 ----MMMIS- -----LLVFN PVEADGVVVN YG--QHAS-- LLAKIDGGGA CKAFCRLSSR PHLCKRACGT CCQRCSGVPP GTAGNYDVCP CVATLTTHGG KRKCP 91
MtGBRP1    1 MALFKLLVTS LLVTFVTLHL VHALDHPEPQ VH--AETQVS LQQQIDGNGA CAARCRLSSR PHLCHRACGT CCGSRCNGVPP GTSGNQEMCP CIANQTTHGG RRKCP 103
MtGBRP2    1 MAISKLLVAS LLASFLLFH- -HLVDATDQS AY--AQTQGS LLQQIDGDGA CVAFCRLSSR PNLCQRACGT CCGRRCNGVPP GTSGNKEKCP CIASQTTRGG KPKCP 101
StSN1      1 ---MKLFLLT LLLVTLVITP -SLIQTTMAG SN-------- -----FCDSK CKLFCSKAGL ADRCLKYCGI CCEECKGVIS GTYGNKHECP CIRDKKNSKG KSKCP 88
Consensus  1             :.          : .                       *.  *  **  :.   * : ** ** .*.***. ** ** . ** **   ..  : *** 29
```

Figure 9. Plant snakin amino acid sequences alignment. Amino acids conserved are indicated by an asterisk in the consensus sequence. The black-shaded cysteine residues are present in all sequences. The sequences included are: Avocado PaSn (KC012806); potato: StSN1 (Q948Z4), StSN2 (Q93X17); *Medicago truncatula:* MtGBRP1 (XP_003603759), MtGBRP2 (XP_003589486); *Capsicum annum:* CaSN1 (ACC91329); *Jatropha curcas:* JcGAST1 (ACV70139), JcLTRCOR12 (ACU30848); *Solanum lycopersicum:* SlSNP1 (ADR32106) and *Fragaria x ananassa:* FaGAST1 (AAB97006).

3.4. Expression patterns of selected genes measured by sqPCR

Expression patterns of five genes from the seed library were studied by semi-quantitative PCR (**Figure 10**); bHLH transcription factor, three metallothioneins, antimicrobial peptide snakin and SUMO like reference gene during avocado seed development. These genes could be divided into three stages according to the time of growth of the seed in avocado fruit. The bHLH gene has an expression pattern comparable to the endogenous gene SUMO (Ubiquitin), suggesting a role throughout the formation and development of the avocado fruit seed possibly modulating the biogenesis of the seed or embryo; since from the first month of formation (E1) to ripening (E8), similar expression levels were present. For the Metallothionein gene group, PaMT3 presented a pattern of constant expression in the three stages of seed
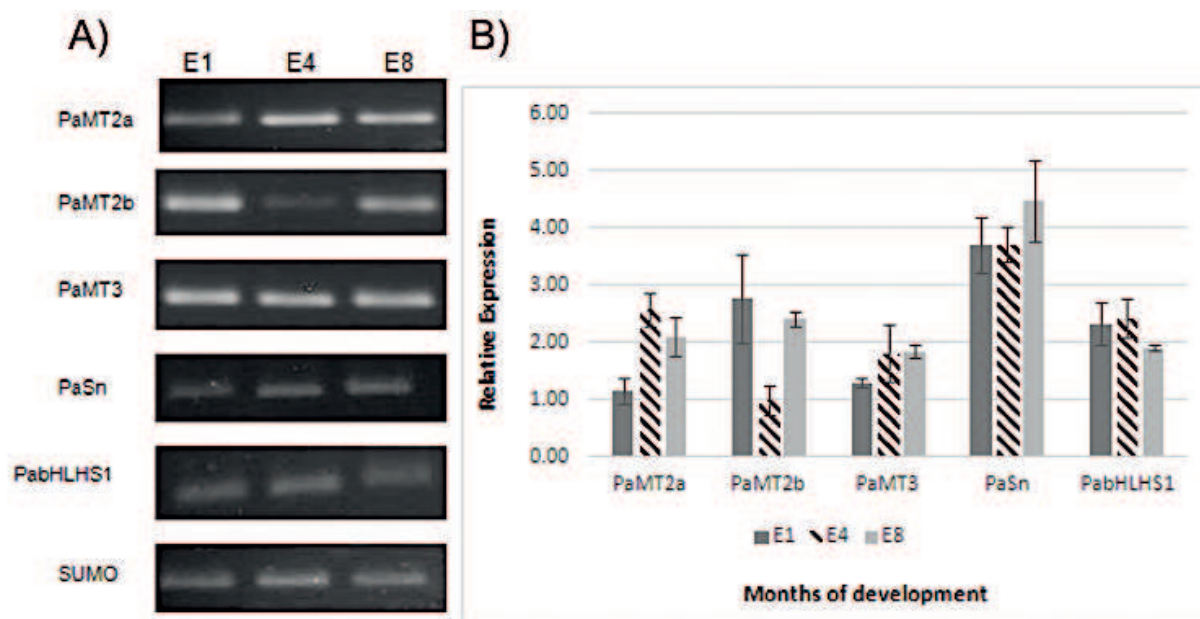

Figure 10. Transcript profiling of PaMT2a, PaMT2b, PaMT3, PaSn and PabHLHS1 during seed development. (A) Total RNA was extracted from avocado seeds of 1, 4 and 8 months of development. cDNAs were synthetized and used for sqPCR , the normalizing gene was avocado ubiquitin SUMO. (B) The amount of mRNA of different genes is expressed as the ratio of the densitometric measurement of the sample RT-PCR product to the SUMO corresponding product.

development used slightly above at the peak of expression compared to the endogenous gene but not for PaMT2a which has an initial level of expression low in first month of development, having its maximum expression peak in the stage of 4 months; PaMT2b has its maximum expression peak at the beginning of the fruit formation in the first month, decaying this by month 4 and recovering expression levels for ripening; it should be noted that metallothioneins have been directly involved within various roles within the functions. Some of them as carriers or facilitators of metal ions for processes of defense, synthesis or hydrolysis of reserve components to make them more bioavailable [46]; however, the authors do not reach an agreement to say that the different types or families of metallothioneins play a specific role. The snakin gene has a similar behavior throughout the development of the fruit emphasizing its role within the defense against pathogens or as the first barrier of protection or signaling of attack, making it therefore important to maintain the levels of expression throughout development and possibly after this for fruit protection (**Figure 10**). The expression patterns of the selected genes identified by sqPCR and the different expression patterns of avocado seed transcriptome suggested various roles of these genes in response to seed development and protection in avocado fruit.

4. Conclusions

In this work, we identified and characterized three novel metallothioneins and one transcription factor gene from avocado nativo mexicano seeds, which are expressed abundantly during seed development. This suggests that they can have a protagonic paper during seed development and probably form a network to protect the embryo for drought stress. More studies are necessary to elucidate the paper of these genes during avocado seed-fruit development.

Author details

Luis María Suárez-Rodríguez, Fernando Sánchez-Albarrán, Huber León-Corona and Rodolfo López-Gómez*

*Address all correspondence to: lgomez@umich.mx

Institute of Chemical-Biological Studies, Michoacana University of Saint Nicolás of Hidalgo, Morelia Michoacán, México

References

[1] Chanderbali AS, Albert VA, Ashworth VE, Clegg MT, Litz RE, Soltis DE, Soltis PS. *Persea americana* (avocado): Bringing ancient flowers to fruit in the genomics era. BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology. 2008;**30**:386-396. DOI: 10.1002/bies.20721

[2] Rohwer JG. Lauraceae. In: Kubitzki K, Rohwer JG, Bittrich V, editors. The Families and Genera of Flowering Plants. Springer-Verlag; 1993. p. 366-391. DOI: 10.1007/978-3-662-02899-5

[3] Ibarra-Laclette E, Méndez-Bravo A, Pérez-Torres CA, Albert VA, Mockaitis K, Kilaru A, López-Gómez R, Cervantes-Luevano J, Herrera-Estrella L. Deep sequencing of the Mexican avocado transcriptome, an ancient angiosperm with a high content of fatty acids. BMC Genomics. 2015;**16**:599. DOI: 10.1186/s12864-015-1775-y

[4] Ben-Ya'acov A, Michelson E. Avocado rootstocks. Horticultural Review. 1995;**17**:381-429

[5] Barrientos PAF. El Aguacate. Biodiversitas. 2010;**88**:1-7

[6] Andrade HP, De León C, Espíndola BMC, Alvarado RD, López JA, García ER. Selección de portainjertos de aguacate para tolerancia-resistencia a *Phytoptora cinnamomi* Rands. usando temperaturas controladas. Spanish Journal of Rural Development. 2012;**4**:1-8 ISSN: 2171-1216

[7] Guzmán-Rodríguez JJ, Ibarra-Laclette E, Herrera-Estrella L, Ochoa-Zarzosa A, Suárez-Rodríguez LM, Rodríguez-Zapata LC, Salgado-Garciglia R, Jimenez-Moraila B, López-Meza JE, López-Gómez R. Analysis of expressed sequence tags (ESTs) from avocado seed (*Persea americana* var. drymifolia) reveals abundant expression of the gene encoding the antimicrobial peptide snakin. Plant Physiology and Biochemistry. 2013;**70**:318-324. DOI: 10.1016/j.plaphy.2013.05.045

[8] Rivas-Gil A, Lares VC, Amaro-Luis JM, Bahsas BA. 2-Alquil-4-Hidroxi-Tetrahidrofuranos de la pulpa del aguacate. *Avances en Química*. 2006;**1**(3):35-37

[9] Birkbeck J. Health benefits of avocado oil. Food New Zealand. 2002;40-42

[10] QY L, Arteaga JR, Qifeng Z, Huerta S, Go VLW, Heber D. Inhibition of prostate cancer cell growth by an avocado extract: Role of lipid-soluble bioactive substances. The Journal of Nutritional Biochemistry. 2005;**16**:23-30. DOI: 10.1016/j.jnutbio.2004.08.003

[11] Ding H, Chin YW, Kinghorn AD, D'Ambrosio SM. Chemopreventive characteristics of avocado fruit. Seminars in Cancer Biology. 2007;**17**:386-394. DOI: 10.1016/j.semcancer.2007.04.003

[12] Dabas D, Shegog RM, Ziegler GR, Lambert JD. Avocado (*Persea americana*) seeds as source of bioactive phytochemicals. Current Pharmaceutical Design. 2013;**19**:6133-6140. DOI: 10.2174/1381612811319340007

[13] Cummings K, Schroeder CA. Anatomy of avocado fruit. California Avocado Society Yearbook. 1942;**27**:56-64 Record Number: 19430300995

[14] Peng FY, Reid KE, Liao N, Schlooser J, Lijavetzky D, Holt R, Martínez Zapater JM, Jones S, Marra M, Bohlmann J. Generation of ESTs in *Vitis vinifera* wine grape (cabernet sauvignon) and table grape (Muscat Hamburg) and discovery of new candidate genes with potential roles in berry development. Gene. 2007;**402**:40-50. DOI: 10.1016/j.gene.2007.07.016

[15] Teh SL, Chan WS, Abdullah JO. Development of expressed sequence tag resources for Vanda Mimi palmer and data mining for EST-SSR. Molecular Biology Reports. 2011;**38**:3903-3909. DOI: 10.1007/s11033-010-0506-3

[16] Low ET, Alias H, Boon SH, Shariff E, Tan CY, Ooi L, Cheah SC, Raha AR, Wan KL, Singh R. Oil palm (*Elaeis guineensis* Jacq.) tissue culture ESTs: Identifying genes associated with callogenesis and embryogenesis. BMC Plant Biology. 2008;**8**:62-81. DOI: 10.1186/1471-2229-8-62

[17] Okubo K, Hori N, Matoba R, Niiyama T, Fukushima A, Kojima Y, Matsubara K. Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. Nature Genetics. 1992;**2**:173-179. DOI: 10.1038/ng1192-173

[18] Franco GR, Adams MD, Soares MB, Simpson AJ, Venter JC, Pena SD. Identification of new *Schistosoma mansoni* genes by the EST strategy using a directional cDNA library. Gene. 1995;**15**:141-147. DOI: 10.1016/0378-1119(94)00747-G

[19] Lee H, Lee HK, An G, Lee YK. Analysis of expressed sequence tags from the red alga *Griffithsia okiensis*. Journal of Microbiology. 2007;**45**(6):541-546

[20] López-Gómez R, Gómez-Lim MA. A method for extracting intact RNA from fruits rich in polysaccharides using ripe mango mesocarp. HortScience. 1992;**27**(5):440-442

[21] Ewing B, Green P. Base-calling of automated sequencer traces using *Phred.* II. Error probabilities. Genome Research. 1998;**8**:186-194. DOI: 10.1101/gr.8.3.186

[22] Li S, Chou HH. Lucy2: An interactive DNA sequence quality trimming and vector removal tool. Bioinformatics. 2004;**20**(16):2865-2866. DOI: 10.1093/bioinformatics/bth302

[23] Foissac S, Bardou P, Moisan A, Cros MJ, Schiex T. EuGene'Hom: A generic similarity-based gene finder using multiple homologous sequences. Nucleic Acids Research. 2003;**31**:3742-3745. DOI: 10.1093/nar/gkg586#

[24] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium Nat. Genet. 2000;**25**:25-29. DOI: 10.1038/75556

[25] Frishman D. Protein annotation at genomic scale: The current status. Chemical Reviews. 2007;**107**:3448-3466. DOI: 10.1021/cr068303k

[26] López-Gómez R, Suáres-Rodríguez LM, Ibarra-Laclette E, Guzmán-Rodríguez JJ, López-Meza JE, Ochoa-Zarzosa A, Salgado-Garciglia R, Rodríguez-Zapata LC, Jimenéz-Moraila B, Herrera-Estrella L. Transcriptome (ESTs) of native Mexican avocado fruit is dominated by stress and innate immunity genes. Acta Horticulturae. 2016;**1144**:43-48. DOI: 10.17660/ActaHortic.2016.1144.6

[27] Margoshes M, Vallee BL. A Cadmium protein from equine kidney cortex. Journal of the American Chemical Society. 1957;**79**(17):4813-4814. DOI: 10.1021/ja01574a064

[28] Kagi JHR, Schaffer A. Biochemistry of metallothionein. Biochemistry. 1988;**27**(23):8509-8515. DOI: 10.1021/bi00423a001

[29] Cobbett C, Goldsbrough P. Phytochelatins and metallothioneins: Roles in heavy metal detoxification and homeostasis. Annual Review of Plant Biology. 2002;**53**(1):159-182. DOI: 10.1146/annurev.arplant.53.100301.135154

[30] Zhou Y, Chu P, Chen H, Li Y, Liu J, Ding Y, Tsang EWT, Jiang L, Wu K, Huang S. Overexpression of *Nelumbo nucifera* metallothioneins 2a and 3 enhances seed germination vigor in Arabidopsis (electronic material). Planta. 2012;**235**(3):523-537. DOI: 10.1007/s00425-011-1527-4

[31] Mir G, Doménech J, Huguet G, Guo WJ, Goldsbrough P, Atrian S, Molinas M. A plant type 2 metallothionein (MT) from cork tissue responds to oxidative stress. Journal of Experimental Botany. 2004;**55**:2483-2493. DOI: 10.1093/jxb/erh254

[32] Soda N, Sharan A, Gupta BK, Singla-Pareek S, Pareek A. Evidence for nuclear interaction of a cytoskeleton protein (OsIFL) with metallothionein and its role in salinity stress tolerance. Scientific Reports. 2016;**6**:34762. DOI: 10.1038/srep34762

[33] Murre C, McCaw P, Vaessin H, Caudy M, Jan L, Jan Y, Cabrera C, Buskin J, Hauschka S, Lassar A, Weintraub H, Baltimore D. Interactions between heterologous helix-loop-helix proteins generates complexes that bind specifically to a common DNA sequence. Cell. 1989;**58**:537-544. DOI: 10.1016/0092-8674(89)90434-0

[34] Massari ME, Murre C. Helix-Loop-Helix proteins: Regulators of transcription in eukaryotic organisms. Molecular and Cellular Biology. 2000;**20**(2):429-440

[35] Jones S. An overview of the basic helix-loop-helix proteins. Genome Biology. 2004;**5**:226

[36] Robinson KA, Lopes JM. *Saccharomyces cerevisiae* basic helix-loop-helix proteins regulate diverse biological processes. Nucleic Acids Research. 2000;**28**(7):1499-1505

[37] Ramsay NA, Glover BJ. MYB–bHLH–WD40 protein complex and the evolution of cellular diversity. Trends in Plant Science. 2005;**10**:63-70

[38] Atchley WR, Fitch WM. A natural classification of the basic helix-loop-helix class of transcription factors. Proceedings of the National Academy of Sciences. 1997;**94**:5172-5176

[39] Ledent V, Vervoort M. The basic helix–loop–hélix protein family: Comparative genomics and phylogenetic analysis. Genome Research. 2001;**11**:754-770

[40] Fairman R, Chao HG, Mueller L, Lavoie TB, Shen L, Novotny J, Matsueda GR. Characterization of a new four-chain coiled-coil: Influence of chain length on stability. Protein Science. 1995;**4**:1457-1469

[41] Stevens JD, Roalson EH, Skinner MK. Phylogenetic and expression analysis of the basic helix-loop-helix transcription factor gene family: Genomic approach to cellular differentiation. Differentiation. 2008;**76**(9):1006-1022

[42] Skinner MK, Manikkam M, Guerrero-Bosagna C. Epigenetic transgenerational actions of environmental factors in disease etiology. Trends in Endocrinology and Metabolism. 2010;**21**:214-222

[43] Ichihashi Y, Horiguchi G, Gleissberg S, Tsukaya H. The bHLH transcription factor SPATULA controls final leaf size in *Arabidopsis thaliana*. Plant and Cell Physiology. 2010;**51**:252-261

[44] Nicolas P, Lecourieux D, Gomès E, Delrot S, Lecourieux F. The grape berry-specific basic helix–loop–helix transcription factor VvCEB1 affects cell size. Journal of Experimental Botany. 2013;**64**(4):991-1003. DOI: 10.1093/jxb/ers374

[45] Tani E, Tsaballa T, Stedel C, Kalloniati C, Papaefthimiou D, Polidoros A, Darzentas N, Ganopoullos I, Flemetakis E, Katinakis P, Tsaftaris A. The study of a *SPATULA*-like bHLH transcription factor expressed during peach (*Prunus persica*) fruit development. Plant Physiology and Biochemistry. 2011;**49**:654-663

[46] Hassinen VH, Tervahauta AI, Schat H, Kärenlampi SO. Plant metallothioneins – Metal chelators with ROS scavenging activity? Plant Biology. 2011;**13**:225-232. DOI: 10.1111/j.1438-8677.2010.00398.x