Proceedings VI World Avocado Congress (Actas VI Congreso Mundial del Aguacate) 2007.
Viña Del Mar, Chile. 12 – 16 Nov. 2007. ISBN No 978-956-17-0413-8.

GENOMICS OF AVOCADO CRIOLLO FRUIT (*Persea americana* Mill. VAR. DRYMIFOLIA)

R. López-Gómez[1], Y. Torres-Cárdenas[1], M. Chávez-Moctezuma[1], R. Salgado-Garciglia[1], B. Jiménez-Moraila[2], G. Corona-Armenta[2] and L. Herrera-Estrella[2]

[1] Instituto de Investigaciones Químico-Biológicas, Universidad Michoacana de San Nicolás de Hidalgo. Francisco J. Mujica s/n. Edificio B-1. Morelia, Mic. México CP58060. Correo electrónico: lgomez@zeus.umich.mx
[2] Unidad de Servicios Genómicos. Laboratorio Nacional de Genómica y Biodiversidad (LANGEBIO). Km 9.5 Libramiento Norte carretera León-Irapuato. Irapuato Gto. México

México is the main consumer and producer of avocado in the world with an approximate annual production of 1.127.574.3 tons. Michoacán is the most important producer with an annual production of 1.012.667.6 tons. Today, México is the main exporter of this fruit in the world. Despite of the economic importance of avocado, little information is available on its genetics. It is significant that most of the important problems of production have a genetic base. The basic knowledge of how an organism works provides invaluable information for the biotechnological development. As complement to the knowledge of plant genomes, ESTs (Expressed Sequence Tags) projects have been generated, which basically consist of sequencing a great number of obtained cDNAs obtained from cDNA libraries, generated from different structures and stages of plant development. Our group has generated cDNA libraries of fruit and seed, and also genomic libraries of creole avocado (*Persea americana* var. Drymifolia). The cDNA libraries are currently being sequenced and to date our preliminary results show that 42% of the sequenced genes are related to metabolism, 20% are related to unknown function, 14% are related to fruit ripening, 8% are related to lipid synthesis, 6% to pathogens response, interestingly, 6% of the genes showed no similarity to any sequence reported in the databases. Finally 4% of the genes are involved in senescence process.

Key words.- Avocado criollo, fruit, pulp, cDNA library, genes, expression.

GENOMICA DEL FRUTO DE AGUACATE CRIOLLO (*Persea americana* Mill. VAR. DRYMIFOLIA)

R. López-Gómez[1], Y. Torres-Cárdenas[1], M. Chávez-Moctezuma[1], R. Salgado-Garciglia[1], B. Jiménez-Moraila[2], G. Corona-Armenta[2] y L. Herrera-Estrella[2]

[1] Instituto de Investigaciones Químico-Biológicas, Universidad Michoacana de San Nicolás de Hidalgo. Francisco J. Mujica s/n. Edificio B-1. Morelia, Mic. México CP58060. Correo electrónico: lgomez@zeus.umich.mx
[2] Unidad de Servicios Genómicos. Laboratorio Nacional de Genómica y Biodiversidad (LANGEBIO). Km 9.5 Libramiento Norte carretera León-Irapuato. Irapuato Gto. México

México, es el principal productor y consumidor de aguacate en todo el mundo con una producción anual aproximada de 1.127.574.3 toneladas anuales. El estado de Michoacán es el productor más importante con una producción anual de 1.012.667.6 toneladas. Actualmente México es el principal exportador de aguacate a nivel mundial. A pesar de su importancia poco se sabe de su genética y es significativo que la mayoría de los problemas de producción más importantes tienen una base genética. El conocimiento básico de cómo funciona un organismo provee una información invaluable para el desarrollo biotecnológico. Como complemento al conocimiento de los genomas de plantas han surgido los proyectos de ESTs (Expressed Sequence Tags) los cuales consisten básicamente en secuenciar un gran número de cDNAs obtenidos de librerías de cDNAs generadas de diferentes estructuras y estadios de desarrollo. Nuestro grupo ha generado librerías de cDNA de fruto y semilla y también librerías genómicas de aguacate criollo (*Persea americana* var. drymifolia). Las librerías de cDNA se están secuenciando actualmente y a la fecha de un análisis preliminar tenemos que el 42% de los genes secuenciados están relacionados con metabolismo, 20% son de función desconocida, 14% genes de maduración de fruto, 8% síntesis de Ac. Grasos, 6% de respuesta a patógenos, interesantemente un 6% son genes no reportados y 4% genes involucrados en senescencia

Palabras Clave.- Aguacte criollo, fruto, pulpa, biblioteca cDNA, genes, expresión.

1. Introduction

Avocado (Persea americana Mill.) belongs to the Lauraceae family, one of the oldest in our planet. It includes a little more than 50 genera and about 2, 200 species. Avocado is clasified in the genus *Persea* and subgenus *Persea*. There are three species known in the subgenus *Persea*: *P. americana* Mill., *P. schiedeana* Nees and *P. parvifolia* Williams. The majority of the known members of the subgenus Persea are mainly located in a same area that begins in the center of Mexico to Panama in Centralamerica. The findings of some primitives avocados from the Sierra Madre Oriental in Nuevo Leon, Mexico, to Costa Rica in Central america, support the hypothesis that this is an avocado origin center and maybe from all the subgenus *Persea* (Sánchez – Pérez, 2007).

Mexico is the main producer and consumer of avocado around the world, with an estimated production of 1, 127, 574.3 tons per year. The state of Michoacan is the most important producer in the country with a yearly production of 1, 012, 667.6 tons and a cultivated area of 78, 530 has. Nowadays Mexico is the principal avocado exporter worldwide (SAGARPA 2007). These days, the avocado is one of the most important crops in Mexico, not only for the amount of tons produced, which make it the most important producer in all the world, but because its cultivation produces thousands of direct and indirect jobs, allowing a high entrance of money because the exportation of the fruit.

The avocado has valuable nutritious properties thanks to its oil (from 12 to 30%) and protein (from 3 to 4%) contents; besides the different amounts of

carbohidrates, vitamins and minerals. These properties give it great possibilities of increasing its uptake in the human diet. In the last years its industrialization in food production , oil extraction and pharmaceutical products has been developed (Rodríguez, 1992; Ortiz et al., 2004; Kritchevsky et al., 2003). As any other fruit tree, it is exposed to different biotic and abiotic factors (grasses, plags, illnesses and post harvest damage) since its cultivation, harvest and storage; that can affect the production and adding up all the damages that these agents produce per year, the losses are huge, 14% in production and 10% in fruit´s quality (SAGAR-INIFAP, 1996). Despite its importance, a little is known about its genetics and it is significant that the most of the production problems have a genetic base. This lack of knowledge has contributed to endless difficulties in the improving, production and storage of the fruit.

Although is a big tree, avocado genome is small, 907 Mbp aproximately, just six times the one of Arabidopsis thaliana and 2.5 times the one of papaya (Arumuganathan and Earle, 1991). The wrong fruit post harvest management originates great economic losses. This is created for the lack of knowledge in the fruit physiology. Nevertheless there are technologies that reduce the losses, the succsess has been varable. Other methodologies application, like genetic manipulation, has had satisfactory results to extend post harvest life and increases the quality of crops such as tomato, melon and papaya. This has been possible thanks to the availability of identify genes and the in vitro culture system to all these crops. The development and ripenind of the fruit are unique processes in plants and represent an important component in the human and animal diet (Giovanoni 2004). In high plants, biological processes such as fruit ripening and senescence are regulated by a complex differential expression of genes. To understand these processes is indispensable identify, clone and characterize the genes involved.

The basic knowledge of an organism operating gives a valuableless information to the biotechnological development. As a complement to the plants genomes understanding the ESTs (Expressed Sequence Tags) projects have arised, which basically consist in sequencing a great number of cDNAs produce by different structures in different stages of development. This technique exploits the most recent advences in the automatic sequencing and handling of DNA. ESTs have proved to be essential in the human genome sequencing and they give important information about the genetic levels of expression. The comparative analisys of the ESTs data bases make easier the detection of conserved sequences among different organisms which lets us know essential sequences to the living beens. In 2005 there were 22 ESTs projects reported in the bibliography from human feeding important plants, more than 100 000 ESTs are reported for oat, soybean and corn, and there are 400 000 for wheat (Olmedo et al., 2005). Unlike tomato fruit, that has become a model in fruit's development at molecular level and for which there is a wide range of information (Yamamoto et al., 2005), there are a few studies at molecular level in avocado. In this work, it is presented a preliminar analisys of a cDNA library (ESTs) from the mesocarp of avocado criollo fruit at 8 months of development.

2. Materials and Methods

2.1.- Eight months avocado criollo fruits (*Persea americana* Mill. *var drymifolia*) were collected in the experimental camp of INIFAP Uruapan, which is located in the National Park "Barranca del Río Cupatitzio" in Uruapan, Michoacan, Mexico. The tree from were the fruits were collected is part of the germoplasma of avocado criollo (code 020 – 03).

2.2.- Total RNA extraction.- Total RNA was extracted from the mesocarp of avocado fruit (1 g) with some modifications on the López – Gómez et al., 1992 method.

2.3.- Construction of the cDNA library.- The SMART cDNA library construction kit (Clontech) was used following the given protocols and using the PCR option.

2.4.- λTriplEx2 phage conversion to the pTriplEx2 plasmid.- The conversion of a phage clone to a plasmid involves the excision an circularization in vivo of a recombinant phage to a complet plasmid. The plasmid is realised bas a result of the recombination mediated by the Cre recombinase specific site in the loxP places that flank the plasmid (Fig.1). The plasmid deliverance occurs automatically when the recombinat phage is translated in a bacterial host where Cre recombinase is being expressed. In this system the *E. coli* BM25.8 strain (that grows at 31 ºC) gives the Cre recombinase activity that is needed.

2.5.- DNA plasmid extraction.- The DNA plasmid extraction was done using the Miniprep technique (Sambrook and Russell, 2001).

2.6.- PCR.- To amplification and size valuation of the clone inserts, the λTriplEx5' LD – insert and λTriplEx3' LD – insert primers were used.

2.7.- Sequencing.- Clone inserts were sequenced with the method described by Sanger et al. (1997), using 5'λTriplEx2 and 3'λTriplEx2 sequencing primers in the istallatios of LANGEBIO, Cinvestav, IPN. Sequences analysis was made using the software MAZORKA.

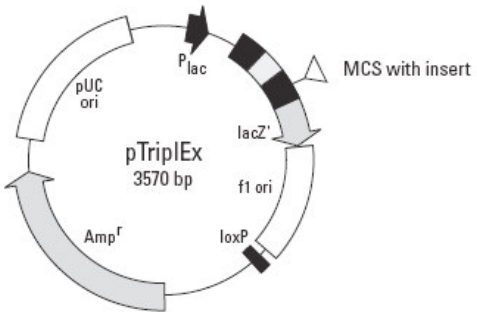


Figure 1.- pTriplEx2 plasmid map. Plasmid is obtain by recombination of λTriplEx2 recombinant phage to the corresponding plasmid using *E. coli* BM25.8 strain (Clontech). λTriplEx2 Multiple cloning site (MCS) is located inside an inner plasmid , which is flanked by LoxP sites in the joint with λ. pTriplEx2 carries the bla gene (ampicilin resistant) and the pUC origin to its autonomous replication in *E. coli*.

3. Results and Discussion

3.1. cDNA library construction.- Avocado mesocarp toatl RNA extrated was from an exellent quality. Following the protocols of SMART cDNA library construction kit, the cDNA library was manufactured. Library title was of 1 X 10^6 UFP and a 98% of the clones were recombinant, both of them, really good values for cDNA libraries. The library was amplified getting a 2.7 X 10^{10} UFP title.

3.2. PCR library analysis.- Excision random colonies were grown to extract plasmid DNA and analize the presence and size of inserts by PCR. Figure 2 shows the agarose gel electrophoresis from 26 PCR productsof the same colonies number. It is important to notice that the most of the colonies have inserts and their sizes go from 500 to 1000 bp and bigger., which suggests the presence of complete mRNAs and an heterogeneous population.

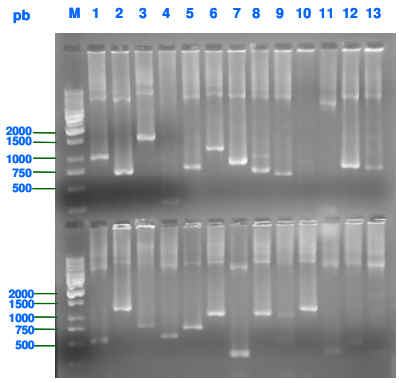


Figure 2.- Recombinant colonies DNA PCR from the avocado mesocarp cDNA library. Lane M molecular size marquer lambda – HindIII, lanes 1 – 13 recombinat colonies. Amplified DNA fragments sizes oscillate from 500 t0 1000 bp.

3.3. Library restriction analysis.- Size and presence was also verified by plasmid DNA restriction anlysis using ZIF enzyme, figure 3 shows diggestion results corroborating the presence of inserts from different sizes in the recombinant plasmids.

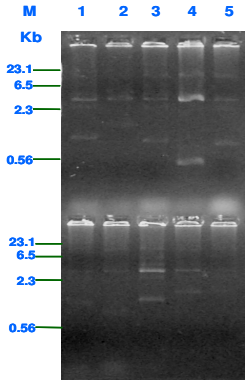


Figure 3.- Restriction analysis of DNA excised plasmids. Lane M reference of molecular size marquer lambda HindIII, lanes 1 – 5 restriction products. Inserts sizes go fron 500 to 1000 bp.

3.4. Sequences analysis.- Until now, 4612 sequences have been generated, with an average length of 721.63 bp and media quality average of 39.96. Figure 4 presents percentage realtions among the different genes clustered by function on the BLAST data base.

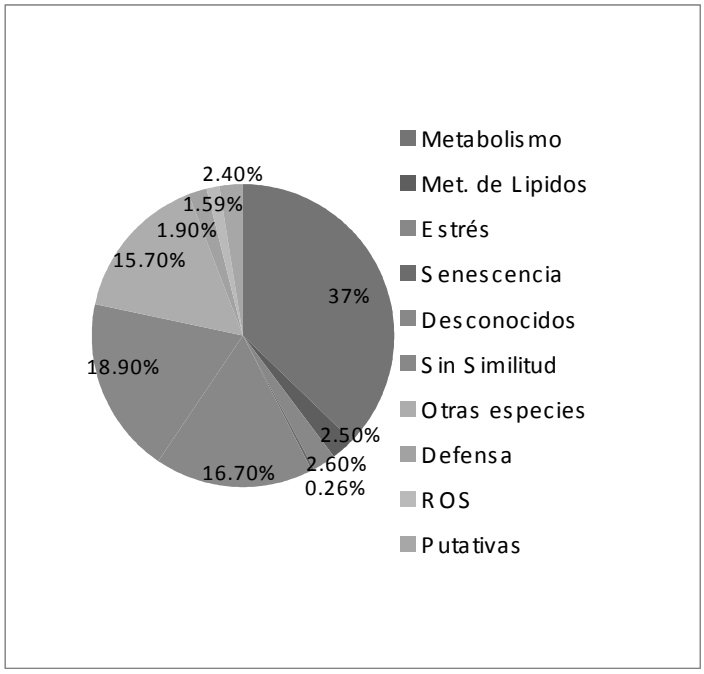
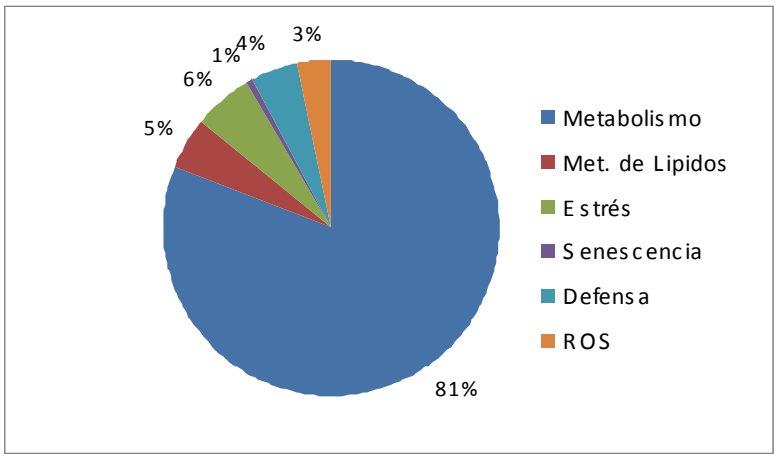


Figure 4.- Percentage relation of all the sequences clustered by function against sequences that did not represent a hit in the BLAST sequences libraries.

From this relation can be conclude that most of the sequences, represented by the unknown sequences (17%), sequences without similarity (19%) and similarity with other species (16%), represent 51% fron the total, suggesting the existance of a great number of new sequences and the lack of knowledge in a huge number avocado fruit's genes.

In figure 5 it is again presented in percentage, the genes codifying to different functions reported in the BLAST sequences libraries (NCBI)



Figue 5.- Percentage realtion of the sequenced genes clustered by function.

As it can be seen the majority of the genes are in the metabolism group (81%), being well represented the stress genes (6%) and lipids metabolism (5%), following with 4% defense involved genes, reactive oxigen species metabolism (ROS 3%) and senescence (1%).

Finally, a comparation of our sequences with data base libraries from other species was made to estimate the number of possible novel sequences. Figure 6 represents this analysis result, showing the high percentage of possible new sequences. It can be concluded that the possibility of having new sequences represent in this avocado criollo fruit library is really high.

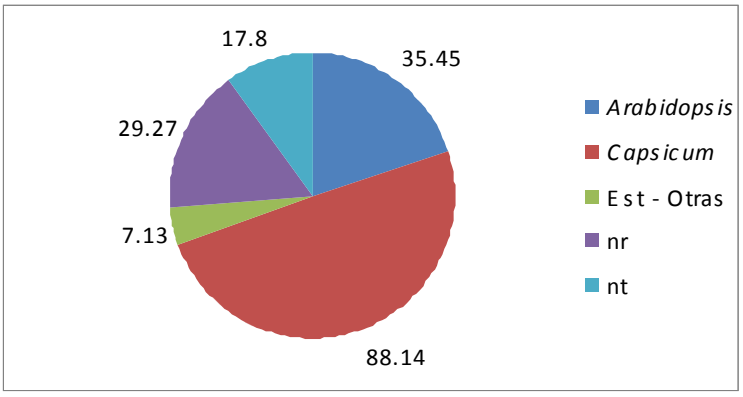


Figure 6.- Percentage representation of novel sequences found in the avocado criollo fruit library against other data base libraries.

4. Conclusion

The understanding of what is going on the ripening process of avocado has great importance to different aspects as grown, post harvest management and genetic improving programs. Generetion of trustable cDNA libraries will let us get the fruit ESTs which represents the base of these techniques improvement. From data get until now it can be conclude that there is a deep ignorance on the genes involved in avocado criollo (*Persea americana var drymifolia*) fruit development and perhaps lots of the sequences generated are novel genes. An interesting observation is there are not ripening genes present even when the age taken was 8 months, in which the fruit's physiological development has been completed. We are still working in the generation of other development and ripening stages libraries with the objective of understanding the processes happening in avocado criollo fruit.

5. References

Arumuganathan K., and Earle A.D. 1991. Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. 9: 208-219.

Giovanoni J. J. 2004. Genetic regulation of fruti development and ripening. The Plant Cell. Vol. 16 S10-S180.

Kritchevsky D., Tepper S.A., Wright S., Czarnecki S.K., Wilson T.A., Nicolosi R.J. 2003. Cholesterol Vehicle in experimetal atherosclerosis 24: avocado oil. J. Amer. Collage of Nutrition. 22 (1): 52-55.

López-Gómez R. and Gómez-Lim M.A. 1992. A method for extracted intact RNA from fruits rich in polysacharides using ripe mesocarp. Hort. Sci. 5:440-442.

Mazorka. Octavio Martínez de la Vega. 2000. Sistema automático de captura y análisis de Etiquetas de Secuencias Expresadas (ESTs). http://mazorka.ira.cinvestav.mx:8080/

Olmedo G., Parra S., Guzmán P. 2005. Genomic Basics for Food Improvement In: Food Biotechnology 2nd Edition. Kalidas Shetty, Anthony Pometto, Gopi Paliyathand Levin RE Eds. CRC Press, Boca Ratón FL.,USA, pp 627-647.

Ortiz M., Dorantes A.L., Galldnez M.J., Cárdenas E. 2004. Effect of a novel oil extraction method on avocado (*Persea americana* Mill) pulp microstructure. Plant Foods for Human Nutrition 59: 11-14.

Rodríguez-Suppo F. 1992. El aguacate. A.G.T. Editor S.A.México D.F. 167pp.
SAGAR-INIFAP (1996). Programa Nacional de Investigación de Aguacate. Grupo Interdisciplinario de Aguacate (GIA). En el Campo Experimental de Uruapan. Documento interno. 127pp.

SAGARPA 2007 http://www.siap.sagarpa.gob.mx/ar_comdeagr.html

Sambrook J. and Russell D.W. 2001. Molecular Cloning. A Laboratory Manual. Third Edition. Cold Spring Harbor Laboratory Press. 1-32, 1-34.

Sánchez-Pérez, J de la L. 2007. Identificación de marcadores asociados a la resistencia de aguacate raza mexicana (*Persea americana* Mill. var *drymifolia*) al oomiceto *Phytophtota cinnamomi* Rands. Tesis de Doctorado. UMSNH. 106 pp.
Sanger F., Miklen S., Coulson A.R. 1977. DNA sequencing with chain treminating ihibitors. Proc Natl Acad Sci USA 74: 5463-5467.

Torres-Cárdenas Y. 2006. Construcción de bibliotecas de ADN complementario del fruto de aguacate criollo Persea americana Mill. Var. drymifolia). Tesis de Maestría. UMSNH. 65 pp

Yamamoto N., Tsugane T., Watanabe M,, Yano K., Maeda F., Kuwata C., Torki M., Ban Y., Nishimura S., Shibata D. 2005. Expressed sequence tags from the laboratory-grown miniature tomato (*Lycopersicon esculentum*) cultivar Micro-Tom and mining for single nucleotide polymorphisms and insertions/deletions in tomato cultivars Gene 15 (356):127-34.