

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

R. López-Gómez¹, Y. Torres-Cárdenas¹, M. Chávez-Moctezuma¹, R. Salgado-Garciglia¹, B. Jiménez-Moraila², G. Corona-Armenta² y L. Herrera-Estrella²

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.- Aguacate criollo, fruto, pulpa, biblioteca cDNA, genes, expresión.

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

R. López-Gómez¹, Y. Torres-Cárdenas¹, M. Chávez-Moctezuma¹, R. Salgado-Garciglia¹, B. Jiménez-Moraila², G. Corona-Armenta² and L. Herrera-Estrella²

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.

1. Introducción

El aguacate (*Persea americana* Mill.) pertenece a la familia Lauraceae, una de las más antiguas en nuestro planeta. Comprende poco más de 50 géneros y unas 2,200 especies. El aguacate se clasifica dentro del género *Persea* y subgénero *Persea*. En el subgénero *Persea* se reconocen tres especies: *P. americana* Mill., *P. schiedeana* Nees y *P. parvifolia* Williams. La mayoría de los miembros reconocidos del subgénero *Persea* se encuentran principalmente en una misma área que inicia del centro de México hasta Panamá en Centroamérica. Los hallazgos de aguacates primitivos desde la Sierra Madre Oriental en el Estado de Nuevo León, México, hasta Costa Rica en Centroamérica, apoyan la hipótesis de que se trata de un centro de origen del aguacate y probablemente de todo el subgénero *Persea* (Sánchez-Pérez, 2007). 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 mas importante con una producción anual de 1,012,667.6 toneladas y una superficie cultivada de 78,530 has. Actualmente México es el principal exportador de aguacate a nivel mundial (SAGARPA 2007).

El aguacate es en la actualidad uno de los cultivos más importantes en México, no solo por la cantidad de toneladas producidas que lo ubica como el productor más importante a nivel mundial, sino también porque es un cultivo que genera miles de empleos directos e indirectos, y permite una alta entrada de divisas por la exportación de su fruto.

El aguacate posee valiosas propiedades alimenticias por su alto contenido de aceite (de 12 a 30%) y proteína (de 3 a 4%), además de su contenido de hidratos de carbono, vitaminas y minerales. Estas propiedades le confieren

grandes posibilidades para el aumento en su consumo en la dieta humana. En los últimos años se ha desarrollado su industrialización en la producción de alimentos, extracción de aceites y productos farmacológicos (Rodríguez, 1992; Ortiz et al., 2004; Kritchevsky et al., 2003). Como cualquier frutal, esta expuesto desde su cultivo, cosecha y almacenamiento a diversos factores bióticos y abióticos (malezas, plagas, enfermedades y daños postcosecha) que pueden afectar la producción, y sumando los daños que causan todos estos agentes al año, las pérdidas son enormes, del 14% en la producción y un 10% en la calidad del fruto (SAGAR-INIFAP, 1996). A pesar de su importancia poco se sabe de su genética y es significativo que la mayoría de los problemas de producción mas importantes tienen una base genética. Esta carencia de conocimientos ha contribuido a dificultades perpetuas en el mejoramiento, la producción y almacenamiento del mismo.

A pesar de ser un árbol de gran talla el genoma de aguacate es relativamente pequeño aproximadamente 907 Mbp, sólo seis veces el tamaño de *Arabidopsis thaliana* y 2.5 veces el de papaya (Arumuganathan and Earle, 1991). El mal manejo poscosecha del fruto origina considerables perdidas económicas. Esto se origina por un escaso conocimiento de la fisiología del fruto. Aunque existen tecnologías que reducen las pérdidas el éxito ha sido variable. La aplicación de otras metodologías como la manipulación genética ha tenido resultados muy satisfactorios para extender la vida poscosecha y aumentar la calidad en cultivos como el tomate, el melón y la papaya. Lo que ha posibilitado estas aplicaciones es tanto la disponibilidad de genes identificados como el sistema de cultivo in vitro para todos estos cultivos. El desarrollo y la maduración del fruto son procesos únicos en las plantas y representan un componente importante en la dieta humana y animal (Giovanoni 2004). En las plantas superiores los procesos biológicos tales como la maduración del fruto y la senescencia están regulados por una compleja expresión diferencial de genes. Para entender estos procesos es indispensable identificar clonar y caracterizar los genes involucrados.

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. Esta técnica explota los avances recientes en la tecnología de secuenciación automatizada y de manejo del DNA. Los ESTs han probado ser de una importancia esencial en la secuenciación del genoma humano y proveen una información importantísima de los niveles de expresión genética. Los análisis comparativos de las bases de datos de los ESTs facilitan también la detección de secuencias conservadas entre diferentes organismos lo que permite conocer secuencias esenciales para los seres vivos. En el año 2005 existían reportados 22 proyectos de ESTs en la bibliografía relacionados con plantas de importancia para la alimentación humana, más de 100 000 ESTs estan reportados para avena, soya y maíz y para el caso de trigo existen 400 000 (Olmedo et al., 2005). A diferencia del fruto de tomate el cual ha surgido como un modelo para el estudio del desarrollo del fruto a nivel molecular y para el cual ya

existe amplia información en este tema (Yamamoto et al., 2005), para el caso de aguacate los estudios a nivel molecular son escasos. En este trabajo presentamos un análisis preliminar de un banco de cDNA (ESTs) del mesocarpo del fruto de aguacate criollo de 8 meses de desarrollo.

2. Material y metodos

2.1. Los frutos de aguacate criollo (*Persea americana* Mill. var *drymifolia*) de 8 meses de edad fueron colectados en el campo experimental del INIFAP unidad Uruapan el cual se encuentra dentro del Parque Nacional "Barranca del Río Cupatitzio" en la ciudad de Uruapan Michoacán, México. El árbol utilizado forma parte del banco de germoplasma de aguacate criollo (clave 020-03).

Extracción de RNA total. Se extrajo RNA total del mesocarpo del fruto de aguacate (1gr) con algunas modificaciones al método de López-Gómez et al., 1992.

2.2. Construcción de la biblioteca de cDNA.- Se utilizó el kit SMART cDNA Library Construction Kit (Clontech) siguiendo los protocolos proporcionados por el fabricante y utilizando la opción por PCR.

2.3. Conversión del fago λ TriplEx2 al plásmido pTriplEx2.- La conversión de una clona del fago a un plásmido involucra la escisión y circularización in-vivo de un fago recombinante a un plásmido completo. El plásmido es liberado como resultado de la recombinación mediada por el sitio específico de la recombinasa Cre en los sitios loxP que flanquean al plásmido (Fig.1). La liberación del plásmido ocurre automáticamente cuando el fago recombinante se traduce en un hospedero bacteriano en el cual la recombinasa Cre se esta expresando. En este sistema la cepa E. coli BM25.8 (que crece a 31°C) proporciona la actividad recombinasa Cre necesaria.

2.4. Extracción de DNA Plasmídico.- El DNA de los plasmidos de las células recombinantes se realizó siguiendo la técnica de Miniprep (Sambrook and Russell 2001)

2.5. PCR.- Se utilizaron los oligonucleótidos λ TriplEx5' LD-Insert y λ TriplEx3' LD-Insert Sequencing primer para la amplificación y estimación del tamaño de los insertos clonados.

2.6. Secuenciación.- Los insertos clonados se secuenciaron siguiendo el método descrito por Sanger et al. (1977), utilizando los primers 5' λ TriplEx2 Sequencing primer y 3' λ TriplEx2 en las instalaciones del LANGEBIO, Cinvestav, IPN. El análisis de las secuencias se realizó utilizando el programa informático MAZORKA.

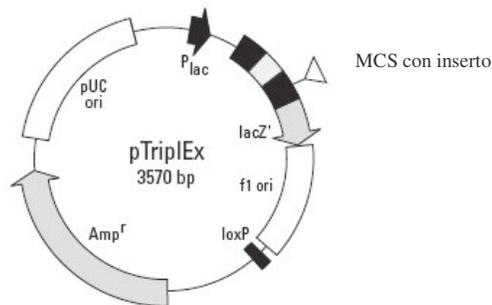


Figura 1.- Mapa del plásmido pTriplEx2. El plásmido se obtiene por la recombinación del fago λ TriplEx2 recombinante al plásmido correspondiente usando la cepa *E. coli* BM25.8 (Clontech). El sitio de clonación múltiple (MCS) de λ TriplEx2 se localiza dentro de un plásmido embebido, el cual es flanqueado por sitios loxP en la unión con λ . pTriplEx2 lleva el gen bla para resistencia a ampicilina y el origen de pUC para replicación autónoma en *E. coli*.

3. Resultados y discusión

3.1. Construcción de la biblioteca de cDNA.- El RNA total extraído del mesocarpo de aguacate fue de muy buena calidad. Siguiendo los protocolos del kit mencionado, se construyó la biblioteca de cDNA. La librería fue titulada obteniendo un valor de 1×10^6 UFP y un porcentaje de 98% de colonias recombinantes, los cuales son dos valores muy buenos para bibliotecas de cDNA. La biblioteca fue amplificada obteniéndose un valor de 2.7×10^{10} .

3.2. Análisis de la biblioteca mediante PCR.- Colonias al azar producto de la escisión fueron cultivadas con el fin de extraer DNA plasmídico y analizar la presencia y el tamaño de los insertos por medio de la reacción de PCR. La figura 2 muestra la electroforesis en gel de agarosa de 26 productos de PCR de un mismo número de colonias recombinantes. Cabe resaltar que la mayoría de las colonias presenta inserto y que el tamaño de este oscila entre 500 a 1000 pb y algunos de mayor tamaño, lo cual sugiere la existencia de mensajeros completos y de una población heterogénea de ellos.

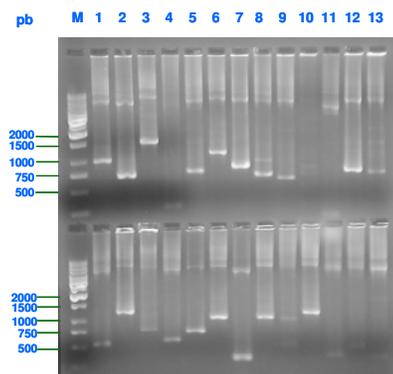


Figura 2.- PCR de ADN de colonias recombinantes de la biblioteca de cDNA de mesocarpo de aguacate. Carril M marcador de tamaño molecular lambda-HindIII, carriles 1-13 colonias recombinantes, cada carril contiene el producto de PCR de dos colonias. El tamaño de los fragmentos de ADN amplificados de la biblioteca de mesocarpo de aguacate oscilan entre 500 y 1000 pb.

3.3. Análisis de la biblioteca por restricción.-El tamaño y la presencia de insertos también fue verificado por restricción del DNA plasmídico utilizando la enzima ZIF, la figura 3 muestra el resultado de la digestión corroborándose nuevamente la presencia de insertos de diferentes tamaños en los plásmidos recombinantes.

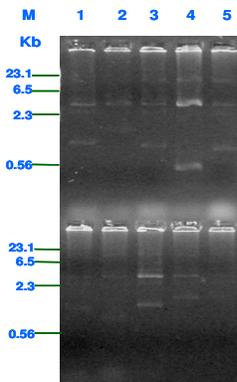


Figura 3.- Análisis de restricción del ADN de plásmidos obtenidos por escisión. M referencia del tamaño molecular lambda-HindIII, carriles 1-5 productos de la restricción. Los tamaños de los insertos oscilan entre 500 y 1000 pb.

3.4. Análisis de Secuencias.- Hasta el momento se han generado 4612 secuencias, con una longitud promedio de 721.63 pb con una calidad media promedio de 39.96. La figura 4 esquematiza las relaciones porcentuales entre los diferentes genes agrupados por función en una búsqueda en las bases de datos del BLAST.

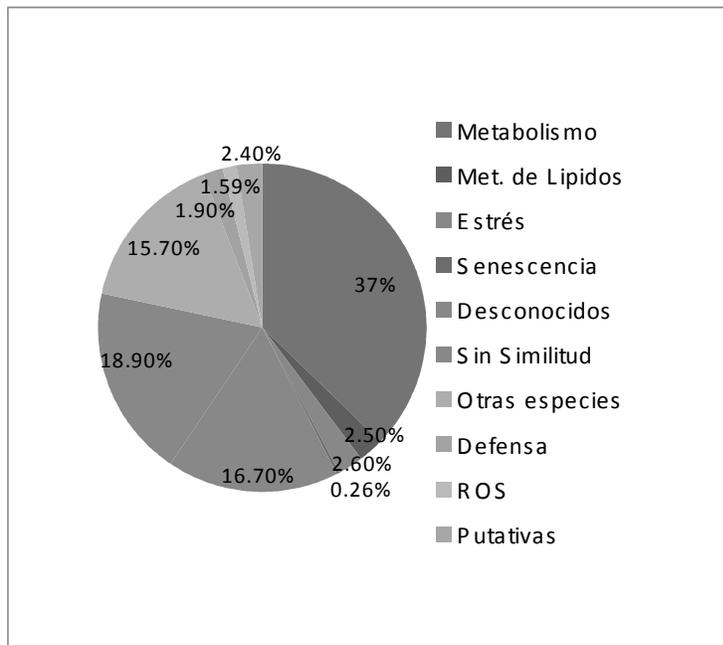


Figura 4.- Relación en por ciento del numero total de secuencias agrupadas por función con relación al por ciento de secuencias que no presentaron ningún hit en los bancos de secuencias de BLAST.

De esta relación se puede concluir que la mayoría de las secuencias representadas por el grupo de secuencias desconocidas (17%), secuencias sin similitud (19%) y con alguna similitud con otras especies (16%) representan el 51% del total, sugiriendo la existencia de un gran numero de secuencias novedosas y la ausencia del conocimiento de un gran numero de genes involucrados en el fruto de aguacate.

En la figura 5 representamos nuevamente en porcentaje los genes que codifican para diferentes funciones reportadas en los bancos de secuencias del sitio BLAST (NCBI).

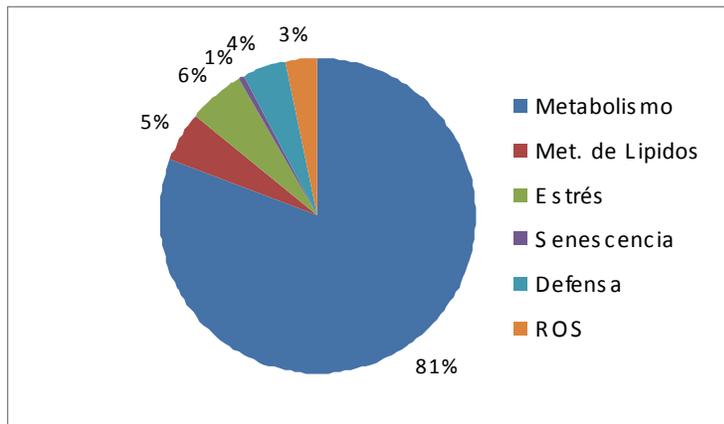


Figura 5.- Relacion porcentual de los genes secuenciados agrupados por función.

Como puede verse claramente la mayoría de los genes están contenidos en el grupo de metabolismo (81%), estando también bien representados los genes que participan en estrés (6%) y metabolismo de lípidos (5%), le siguen con el 4% genes involucrados en defensa (4%), metabolismo de especies reactivas de oxígeno (ROS 3%) y senescencia (1%).

Finalmente realizamos una comparación de nuestras secuencias con bancos de datos de otras especies para estimar el posible número de secuencias novedosas. La figura 6 respresenta el resultado de este analisis mostrando los posibles porcentajes de secuencias novedosas, como puede concluirse de este análisis la posibilidad de tener secuencias novedosas representadas en esta biblioteca de fruto de aguacate criollo es muy alta.

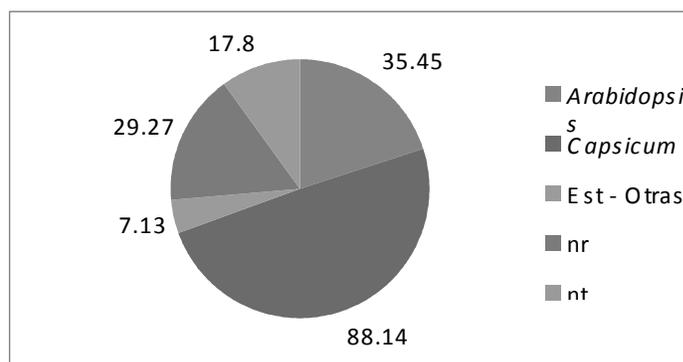


Figura 6.- Representación porcentual de posibles secuencias novedosas encontradas en la biblioteca de aguacate criollo comparado con otros bancos de datos.

4. Conclusiones

El entendimiento de lo que sucede a nivel molecular en el proceso de maduración de aguacate tiene una gran importancia para diferentes aspectos del cultivo, principalmente para el manejo postcosecha y para programas de mejoramiento genético. El generar librerías confiables de cDNA permitirá obtener los ESTs del fruto lo cual representa la base para el desarrollo de las investigaciones antes mencionadas. De los datos obtenidos hasta el momento podemos concluir que hay un gran desconocimiento de los genes involucrados en el proceso de desarrollo del fruto de aguacate criollo (*Persea americana* var *drymifolia*) y que posiblemente muchas de las secuencias generadas sean genes novedosos. Una observación interesante es que la presencia de genes involucrados en el proceso de maduración (reportados como tal) aun no está presente, a pesar de que el estadio elegido es de 8 meses, en el cual el desarrollo fisiológico del fruto es completo. Nuestro grupo sigue trabajando en generar bibliotecas de otros estadios de desarrollo y maduración del fruto con el objetivo de entender los procesos llevados a cabo en el fruto de aguacate criollo.

5. Literatura Citada

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 fruit 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 experimental 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 polysaccharides 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., Galdnez 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 *Phytophthora cinnamomi* Rands. Tesis de Doctorado. UMSNH. 106 pp.

Sanger F., Miklen S., Coulson A.R. 1977. DNA sequencing with chain terminating inhibitors. 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.