

T  
1171

 XOXIMILCO SERVICIOS DE INFORMACION  
ARCHIVO HISTORICO

124165

UNIVERSIDAD AUTONOMA METROPOLITANA



**Casa abierta al tiempo**

Clonación y caracterización de fragmentos RAPD y *loci* microsatélite, asociados a *Psidium guajava* L. cultivada en 4 estados de la República Mexicana.

TESIS

Que para obtener el grado de  
Doctor en Ciencias Biológicas

P R E S E N T A

M. en C. Iris Angélica Feria Romero

Comité tutorial:

Dr. Héctor Fernando Serrano

Dra. Marisol López López

Dra. Sandra Adela Orozco Suárez

Mayo/2008

El Doctorado en Ciencias Biológicas de la Universidad Autónoma Metropolitana pertenece al Padrón de Posgrados de Excelencia del CONACYT y además cuenta con apoyo del mismo Consejo, con el convenio PFP-20-93.

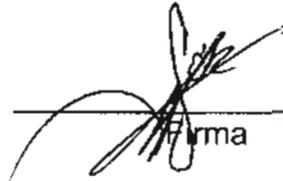
El jurado designado por las Divisiones de Ciencias Biológicas y de la Salud de las Unidades Iztapalapa y Xochimilco aprobó la tesis que presentó

M. en C. Iris Angélica Feria Romero

El día 12 de Mayo del año de 2008

Jurado:

Tutor: Héctor Fernando Serrano



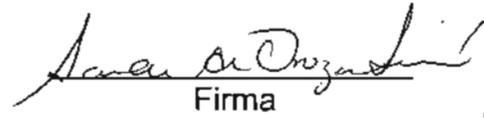
Firma

Asesor: Marisol López López



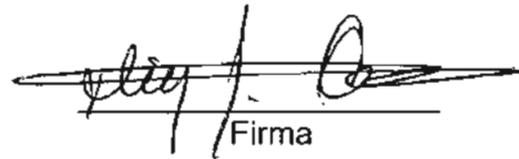
Firma

Asesor: Sandra Adela Orozco Suárez



Firma

Sinodal: Diego Julio Arenas Aranda



Firma

Sinodal: Rosenda Isabel Peñaloza Espinosa



Firma

## Dedicatorias y agradecimientos

Esta tesis doctoral fue realizada gracias al apoyo financiero CONACyT que fue otorgado mediante la beca con número de registro 95747 durante el periodo febrero/2004 a diciembre/2006, dentro de los programas de excelencia de dicho consejo; y por el proyecto IMSS-FOFOI FP-2003/039 otorgado por la Coordinación de Investigación del IMSS.

Agradezco a:

El Dr. Héctor Serrano por su apoyo, guía y paciencia.

Marisol, Sandy y Meli porque más que ser mis asesoras y grandes investigadoras, las considero unas buenas amigas que aprecio y respeto.

La Dra. Rosenda Peñaloza y al Dr. Diego Arenas por sus valiosos comentarios y observaciones que enriquecieron el trabajo.

Mis padres por estar siempre conmigo en cada uno de las etapas de mi vida, por su apoyo incondicional y gran amor.

Mi querida familia compuesta de seres grandes y únicos, de los cuales he aprendido cosas maravillosas.

A mis amigos cercanos y lejanos a los cuales llevo siempre en mi corazón y recuerdo con mucho cariño.

## RESUMEN.

Introducción. El gran número de habitantes en el mundo que utilizan a las plantas medicinales como recurso para solucionar sus problemas de salud, atrae la atención de instancias gubernamentales e internacionales que establecen métodos de control de calidad para garantizar su seguridad y eficacia. Los documentos oficiales reconocen como pruebas de autenticidad, a las de identificación taxonómica y de compuestos químicos que son característicos de la especie; sin embargo, en años recientes se están empleando marcadores de ADN para la identificación de la especie y discriminación de adulterantes. Entre los marcadores de ADN se encuentran los del tipo RAPD, RFLP, AFLP, secuencia específica y microsatelital. La hoja de guayaba (*Psidium guajava* L.) ampliamente cultivada en México, se utiliza popularmente en el tratamiento de trastornos estomacales y estudios clínicos recientes demuestran su eficacia en el tratamiento del síndrome de colon irritable.

Planteamiento del problema. Debido a la escasa información genética existente de *P. guajava*, el presente trabajo se enfocó en la obtención de marcadores genéticos que se relacionen con la acumulación de quercetina y permitan la autenticación de la especie, como una herramienta en el control de calidad en el desarrollo de fitofármacos.

Hipótesis. Algunas variaciones en regiones del ADN entre individuos de *P. guajava* cultivados en diferentes estados de la República Mexicana están asociadas a la acumulación de flavonoides mientras que las regiones conservadas lo están con la identificación de la especie.

Objetivos. a) Caracterizar las variaciones u homologías de secuencia en segmentos tipo RAPD provenientes del ADN genómico de *P. guajava* L.; b) Obtener la distribución de la frecuencia alélica de diferentes

muestras de *P. guajava* L. a partir de la construcción de una biblioteca genómica de *loci* microsatélite de la clase dinucleótida. Materiales y Métodos. Retoños limpios se colectaron en 40 árboles en cuatro estados de la República Mexicana. El material vegetal se procesó para obtener por un lado un extracto metanólico hidrolizado para la cuantificación de quercetina por HPLC usando columnas RP-18 y un detector PDA, y por otro lado se obtuvo ADN genómico. En éste se amplificaron regiones aleatorias tipo RAPD con seis oligonucleótidos de 10-mer (Amersham Bioscience, UK), y por otra parte se fraccionó con enzimas de restricción de acuerdo a lo informado previamente. Los amplificadores aleatorios RAPD se analizaron y correlacionaron con la acumulación de quercetina total, mientras que los fragmentos RFLP se hibridaron con sondas (GT)<sub>15</sub> biotiniladas, para capturar fragmentos con secuencias dinucleótidas repetidas. Las bibliotecas tipo RAPD y microsatelitales, se clonaron, secuenciaron, analizaron y de acuerdo con esto se diseñaron oligonucleótidos específicos. Los marcadores de ADN seleccionados se utilizaron en la autenticación de la especie y discriminación con otras plantas morfológicamente semejantes. Resultados. Los estados seleccionados producen el 84% de la guayaba que se cultiva en México. Las muestras de Aguascalientes mostraron la mayor acumulación de quercetina total en primavera, pero en el estado de México se encontró un árbol con la mayor acumulación de quercetina total. Las otras muestras de los estados de Querétaro y Michoacán tuvieron valores inferiores. En el amplificador aleatorio RAPD se usaron seis oligonucleótidos para generar 91 amplificadores con una homología entre 74 y 100% entre los individuos. El oligonucleótido 1 generó dos amplificadores polimórficos de 560 y 610pb diferenciando

los patrones de bandeo 1a, 1b y 1c; el **2** mostró un amplificado polimórfico de 370pb con lo que se diferenciaron los patrones 2a y 2b; el **3** dió un amplificado polimórfico de 690pb que distingue los patrones 3a y 3b; el **4** generó dos amplificados polimórficos de 460 y 480pb diferenciando los patrones de bandeo 4a, 4b y 4c, y los oligonucleótidos 5 y 6 mostraron un solo patrón de bandeo cada uno. El aumento en la acumulación de quercetina se correlaciona con la presencia de los patrones de bandeo 1a y 2a con 4c. De las bibliotecas se obtuvieron 36 secuencias; 5 de ellas con microsatélites dinucleótidos; una similar en un 91% a *Arabidopsis thaliana* y se logró diseñar oligonucleótidos específicos para 22 secuencias de *P. guajava*. Se establecieron las condiciones de amplificación para 11 secuencias usando 10 ng de ADN genómico y 1 mM MgCl<sub>2</sub> o para 12 secuencias usando 5 ng de ADN genómico y 1.5 mM MgCl<sub>2</sub>. Cinco de las secuencias seleccionadas fueron exclusivas para *P. guajava*. Conclusiones. El marcador químico de quercetina, permitió identificar a los individuos del estado de Aguascalientes en primavera como los de mayor acumulación de quercetina. Los marcadores genéticos tipo RAPD 1a y 2a con 4c se correlacionan con un aumento de quercetina con un valor LoD de 9.498 ( $p < 0.05$ ). En la autenticación de la especie los marcadores 1.2A, 4.1, 4.3, 5.8 y 5.5 fueron específicos para *P. guajava*, permitiendo discriminar la guayaba de otros individuos taxonómicamente similares de la familia *Mirtacea*.

## ABSTRACT.

Introduction. People around the world use medicinal plants to attend their health problems so government and international institutions are now taking attention in quality control methods to guarantee the security and efficiency of plant drug material, avoiding the incorporation of related species that lack the active compound or intrinsically have inhibitory compounds. Official guidelines recognize taxonomic identification and chemical fingerprinting with authenticity tests; however RAPD, RFLP, AFLP, specific-sequence and microsatellital DNA markers are used on species identification and adulterant discrimination in basic preparative research projects. Guava (*Psidium guajava* L.) leaf is an extensive crop in México and it is used in abdominal disorders, and recently in the treatment of irritable bowel syndrome.

Justification. Due to limited genetic information of *P. guajava*, the purpose of this work was to identify genetic markers related to quercetin accumulation and that allow the authentication of this species to be used in the development of phytodrugs, as a tool of quality control.

Hypothesis. Some genetic variations between individuals of *P. guajava* cultivated in different states from Mexico are associated with a variation in the accumulation of flavonoids while conserved regions are associated with the identification of the species.

Objectives. a) To characterize varieties or homologies in RAPD fragments sequences from *P. guajava* DNA; b) To obtain a dinucleotide microsatellite genomic library from *P. guajava* L.

Material and methods. Young clean leaves of 40 trees where collected from four Mexican states and processed for DNA genomic and quercetin extraction. The raw material was used to obtain a methanolic hydrolyzed extract for quercetin quantification by the HPLC technique using RP-18

columns and a PDA detector. On the other hand, it was used for DNA extraction, amplification of random DNA regions by the RAPD technique using 10-mer primers, and to obtain DNA fragments with restriction enzymes. RAPD analysis was correlated with high quercetin accumulation. RFLP fragments were hybridized with biotinylated tandem sequences. RAPD and microsatellites library were cloned, sequenced, analyzed and specific primers were designed. DNA markers were used in authentication and discrimination from other morphologically alike species. Results. Selected states account for 84% of the guava fruit production in Mexico. Spring samples collected from Aguascalientes showed a high quercetin accumulation, but one individual sample from the state of Mexico had the highest. Six decamer commercial RAPD primers generated a total of 91 fragments, sharing 74 – 100% of bands among collected trees. The oligonucleotide **1** amplified two polymorphic segments of 560 and 610 bp to recognize the 1a, 1b and 1c banding patterns, oligonucleotide **2** amplified one polymorphic segment of 370 bp to recognize the 2a, and 2b banding patterns, oligonucleotide **3** amplified one polymorphic segment of 690 bp to recognize the 3a and 3b banding patterns, oligonucleotide **4** amplified two polymorphic segments of 460 and 480 bp to recognize the 4a, 4b and 4c banding patterns, oligonucleotides **5** and **6** amplified a pattern band each. Quercetin accumulation increase was correlated between 1a and 2a with 4c pattern bands. Thirty-six sequences were obtained from genomic libraries, five of them with tandem regions, one with a similarity of 91% with *Arabidopsis thaliana*, and only 22 sequences were viable for specific primer design. Amplified conditions were developed for 11 sequences using 10 ng of genomic DNA and 1 mM MgCl<sub>2</sub> or 12

sequences using 5 ng of genomic DNA and 1.5 mM MgCl<sub>2</sub>. Five fragments were specific for *P. guajava*. Conclusions. The chemical marker quercetin allowed the identification of individuals from Aguascalientes state in spring with the highest accumulation of quercetin among selected regions. The banding patterns 1a and 2a with 4c are correlated with an increase of quercetin with a LoD value of 9.498 ( $p < 0.05$ ). In the authentication of species markers 1.A, 4.1, 4.3, 5.8 and 5.5 amplified for *P. guajava*, allowing to discriminate against guava from others individuals morphologically similar.

## ÍNDICE GENERAL

## I. INTRODUCCIÓN

I.1. Control de calidad de las plantas medicinales .....	17
I.1.1 Aspectos botánicos y morfológicos .....	18
I.1.2 Aspectos químicos .....	19
I.2. Aplicación de la biología molecular en la identificación del material vegetal .....	20
I.2.1 Extracción del material genético .....	21
I.2.2 Marcadores genéticos de tipo RFLP, RAPD y AFLP .....	22
I.2.3 Marcadores genéticos de secuencia específica .....	24
I.2.4 Marcadores genéticos microsatelitales .....	26
I.3. <i>Psidium guajava</i> L. ....	29
I.3.1 Descripción taxonómica .....	29
I.3.2 Estudios realizados en <i>P. guajava</i> .....	30
II. PLANTEAMIENTO DEL PROBLEMA .....	33
III. HIPÓTESIS .....	34
IV. OBJETIVOS .....	34
IV.1 Objetivos generales .....	34
IV.2 Objetivos particulares .....	34
V. DISEÑO EXPERIMENTAL .....	36
VI. MATERIALES Y METODOS .....	39
VI.1. Obtención de muestras de <i>P. guajava</i> .....	39
VI.2. Cuantificación de quercetina por cromatografía líquida de alta	

resolución (HPLC) .....	40
VI.2.1 Preparación de la muestra .....	40
VI.2.2 Cuantificación y análisis cromatográfico .....	40
VI.3. Extracción y purificación del ADN genómico .....	41
VI.4. Obtención de amplificadores aleatorios tipo RAPD .....	42
VI.5. Digestión del ADN genómico .....	42
VI.6. Hibridación con oligonucleótidos biotinilados y captura con estreptavidina .....	43
VI.7. Clonación y transformación de los amplificadores aleatorios tipo RAPD y fragmentos con regiones repetidas .....	43
VI.8. Escrutinio de las bibliotecas de amplificadores aleatorios RAPD y fragmentos con regiones repetidas .....	44
VI.9. Secuenciación de los amplificadores aleatorios RAPD y fragmentos con regiones repetidas .....	45
VI.10. Diseño de oligonucleótidos específicos .....	46
VI.11. Estandarización de la amplificación por PCR de secuencias de <i>P. guajava</i> empleando oligonucleótidos específicos .....	46
VI.12. Validación de las secuencias en muestras representativas y diferentes a <i>P. guajava</i> .....	47
VI.13. Análisis de los resultados .....	47
VII. RESULTADOS .....	49
VII.1. Selección de la materia prima .....	49

VII.2. Cuantificación del principio activo.....	49
VII.3. Selección de los individuos para el estudio genético en base a la acumulación de flavonoides .....	52
VII.4. Identificación de segmentos polimórficos en las muestras colectadas utilizando el fraccionamiento de amplificación aleatoria RAPD .....	53
VII.5. Identificación de secuencias similares y regiones repetidas .....	61
VII.6. Identificación de secuencias homólogas utilizando la base de datos Gene Bank de NCBI .....	65
VII.7. Diseño de oligonucleótidos específicos para la identificación de las clonas seleccionadas .....	65
VII.8. Estandarización en la amplificación de secuencias usando ADN genómico de <i>P. guajava</i> y oligonucleótidos específicos .....	66
VII.9. Validación de las secuencias en la identificación genotípica de <i>P. guajava</i> .....	66
VII.10. Identificación genotípica de <i>P. guajava</i> .....	69
VIII DISCUSIÓN .....	72
VIII.1. Selección de la materia prima.....	72
VIII.2. Cuantificación del principio activo .....	73
VIII.3. Selección de los individuos para el estudio genético en base a la acumulación de flavonoides .....	74
VIII.4. Identificación de segmentos polimórficos en las muestras colectadas usando amplificación aleatoria tipo RAPD .....	75

VIII.5. Identificación de secuencias similares y regiones repetidas .....	76
VIII.6. Identificación de secuencias homólogas utilizando la base de datos Gene Bank de NCBI .....	77
VIII.7. Diseño de oligonucleótidos específicos para la identificación de las clonas seleccionadas .....	78
VIII.8. Estandarización en la amplificación por PCR de secuencias de <i>P. guajava</i> empleando oligonucleótidos específicos.....	79
VIII.9. Validación de las secuencias en la identificación genotípica de <i>P. guajava</i> .....	80
VIII.10. Identificación genotípica de <i>P. guajava</i> .....	81
IX CONCLUSIONES .....	82
X BIBLIOGRAFÍA .....	84
ANEXO 1 .....	i
ANEXO 2 .....	iv
ANEXO 3 .....	vii
ANEXO 4 .....	xii
ANEXO 5 .....	xiii
ANEXO 6 .....	xv
ANEXO 7 .....	xxvi
ANEXO 8 .....	xxvii
ANEXO 9 .....	xxix
ANEXO 10 .....	xxx
Artículo de revisión:	

Aplicación de la biología molecular en el control de calidad de los  
fitomedicamentos

Artículo:

Molecular RAPD markers associated with quercetin accumulation in  
*Psidium guajava*

## **I. INTRODUCCIÓN.**

### **I.1. Control de calidad de las plantas medicinales.**

Actualmente las plantas medicinales tienen gran importancia por su aplicación en la medicina alternativa, captando la atención de instancias internacionales como la Organización Mundial de la Salud que señala que el 80% de la población de países en desarrollo resuelve sus problemas de salud por esta vía (Romero-Cerecero y cols., 2004); con ello, se ha incrementado de igual forma el número de remedios herbolarios y medicamentos elaborados a partir del material vegetal o sus extractos.

La demanda del material vegetal ha inducido el cultivo de especies con importancia terapéutica en diferentes zonas geográficas, propiciando una alteración en la acumulación de los principios activos relacionados con el entorno ambiental en el que crecen. En otros casos, la explotación comercial y necesidad económica ha favorecido la confusión, sustitución y adulteración de las plantas de interés (Barnes, 2002).

Debido a lo anterior, los países y sus estancias gubernamentales han elaborado documentos oficiales o farmacopeas herbolarias donde se recopila toda la información reglamentaria, los métodos de control de calidad y las monografías de la parte de la planta con uso medicinal (SSA, 2001). Particularmente, las monografías descritas por la farmacopea mexicana se enfocan en la descripción de la parte de la planta con interés medicinal, ensayos de identidad, análisis químicos y conservación; variando su contenido de acuerdo al conocimiento que se tenga de la misma (SSA, 2001).

Por otro lado, tanto la Organización Mundial de la Salud como diversas asociaciones científicas alrededor del mundo, relacionadas con la farmacognosia de las plantas medicinales y fitoterapia, han publicado monografías no oficiales con información técnica y científica.

Entre los aspectos sobresalientes en el control de calidad del material vegetal se encuentran los aspectos botánicos, morfológicos y químicos que en conjunto proporcionan los caracteres de autenticación de la materia prima (WHO, 1999; Vanaclocha y cols., 2003).

### **1.1.1 Aspectos botánicos y morfológicos.**

En una primera instancia, la información botánica y morfológica se obtiene de la identificación taxonómica proporcionada por el ejemplar de interés colectado en campo. Una vez corroborada la especie, se realizan una serie de descripciones macro y microscópicas de la parte de la planta con propiedades terapéuticas. El nivel de la descripción dependerá de su integridad como material vegetal y la similitud con sus adulterantes; entendiendo como adulterante al reemplazo o mezcla con otra especie que no posee propiedades terapéuticas, pero que es morfológicamente similar, generalmente tiene un costo menor y puede ser la misma parte de la planta que se está reemplazando (Harkey y cols., 2002). La información microscópica es más completa y detallada en comparación con la macroscópica, sobre todo en aquellas muestras incompletas. Los especialistas en esta técnica hacen uso de herramientas histológicas con la finalidad de describir las estructuras celulares propias de la especie (Ruzin, 1999). Esta información puede ser utilizada en un primer contacto con el material vegetal adquirido por terceros, que es lo común

cuando los laboratorios carecen de la opción de una sola agroindustria abastecedora. Sin embargo la autenticación de la especie presenta ciertas limitaciones cuando el material vegetal se encuentra adulterado con especies filogenéticamente muy cercanas.

### **1.1.2 Aspectos químicos.**

La identificación de adulterantes desde el punto de vista químico se realiza a través de técnicas de caracterización de los compuestos en el extracto de la planta o identificación por perfil cromatográfico. Las técnicas de análisis desarrolladas son la cromatografía en capa fina (TLC, *Thin layer chromatography*), cromatografía de líquidos de alta resolución (HPLC, *High Performance Liquid Chromatography*), cromatografía de gases y diversas técnicas espectroscópicas (Drasar y Moravcova, 2004). La información colectada se considera confiable debido a que los individuos de la misma especie y cultivados en la misma zona geográfica presentan un perfil cromatográfico muy similar, como consecuencia del medio ambiente que favorece la acumulación de principios activos. Las huellas químicas identificadas permiten determinar la autenticidad, calidad, seguridad y eficacia de la materia prima antes de convertirse en un medicamento de origen vegetal o fitomedicamento (Huang y Song, 2001). Por lo anterior, la detección de adulterantes se presenta cuando existe alguna alteración importante en las huellas químicas, siendo el aspecto más relevante la ausencia del principio activo. Estos compuestos también son utilizados en el control de calidad del extracto durante el desarrollo del fitomedicamento; cobrando relevancia al momento de cuantificar al principio activo o en su defecto, al compuesto

con mayor abundancia, que se relaciona indirectamente con éste (Schaneberg y cols., 2003).

A pesar de los avances tecnológicos en la identificación química del material vegetal, al igual que en la identificación macro y microscópica, esta técnica presenta limitaciones cuando el extracto a analizar se compone de individuos de diferentes especies con una estrecha cercanía filogenética. En estos casos, el extracto presenta perfiles cromatográficos muy similares, sobre todo si el adulterante se encuentra en cantidades inferiores.

## **1.2. Aplicación de la biología molecular en la identificación del material vegetal.**

La domesticación y explotación de plantas con interés económico impulsa a los agricultores a seleccionar y mejorar líneas o cultivos con características deseadas. Los individuos de un cultivo pueden presentar heterogeneidad en sus caracteres debido a que provienen de diferentes semillas (origen policlonal), por la acumulación de mutaciones somáticas presentes en la semilla (origen monoclonal) y/o por las alteraciones provocadas por una infección (origen patológico) (Pecetti y cols., 2004; Hoffmann-Sommergruber y cols., 1997).

El ADN se encuentra en el núcleo y en algunos organelos de las células (mitocondria y cloroplasto en los organismos vegetales), siendo similar en cualquier célula de la planta, lo que permite su identificación independientemente si se tiene el ejemplar de herbario completo o solo una parte (Singer y Berg, 1993). Esta molécula de doble cadena, súper enrollada y gran estabilidad contiene la información genética capaz de identificar la especie de un individuo. En ella se encuentra almacenada la información

necesaria para la expresión de las características físicas o morfológicas (fenotipo) que los botánicos utilizan para poder diferenciar entre los organismos, utilizando las claves de identificación taxonómica.

En la actualidad se ha recurrido a técnicas de biología molecular con el fin de obtener marcadores genéticos especie específicos, encontrar diferencias polimórficas y proporcionar la información necesaria para la identificación de variedades. Los marcadores moleculares permiten estimar la distancia genética, la identificación y discriminación de poblaciones, variedades, líneas puras e híbridos; además establecen relaciones de parentesco y localizan e identifican regiones del ADN cualitativos que afectan caracteres cuantitativos. El análisis de los marcadores genéticos es útil por el polimorfismo, herencia mendeliana sin epistaxis (o sea, sin interacción entre los genes), insensibilidad a los factores ambientales o al desarrollo de la planta y su fácil identificación y co-dominancia (Heckenberger y cols., 2005). Estos conocimientos son aplicados antes del desarrollo y elaboración de fitomedicamentos con la finalidad de autenticar la especie e identificar adulterantes. La corroboración de la especie se logra mediante la comparación de marcadores genéticos entre la especie identificada en campo y la muestra problema.

### **1.2.1 Extracción del material genético.**

El primer paso en la identificación de la especie es la extracción del material genético, el cual debe estar lo suficientemente puro para su manipulación, fraccionamiento, amplificación y secuenciación. Actualmente existen diferentes sistemas comerciales que permiten la extracción del ADN a partir del material vegetal, sin embargo no hay que descartar el uso de técnicas especiales que se han

desarrollado con el fin de disminuir la interferencia que pudiera presentarse por un alto contenido de polifenoles, polisacáridos o mucopolisacáridos en la muestra que disminuyen su pureza y rendimiento (Friar, 2005; Wulff y cols., 2002; Tel-Zur y cols., 1999). El método de purificación además, debe contemplar la integridad del material genético a fin de garantizar la reproducibilidad de los resultados, ya que un fraccionamiento parcial o total del ADN modifica la forma en que se distribuyen algunos marcadores genéticos al momento de someterlos a una separación electroforética, obteniéndose patrones diferentes a los que se presentan cuando las condiciones son óptimas.

### **1.2.2 Marcadores genéticos de tipo RFLP, RAPD y AFLP.**

En un primer acercamiento de la especie a nivel de material genético, varios grupos de investigación han optado por la segmentación controlada del ADN con la finalidad de obtener un patrón conservado a partir de dichos segmentos. De la misma manera, también se considera un grupo de marcadores genéticos, un gen, un sitio de restricción, una secuencia o una región del genoma que permita identificar su caracterización y/o ubicación en un cromosoma.

La partición del ADN se puede realizar mediante el fraccionamiento con enzimas de restricción (endonucleasas que cortan la doble cadena de ADN) y/o la reacción en cadena de la polimerasa (PCR, *polymerase chain reaction*) (Becerra y Paredes, 2000), y dependiendo de la técnica empleada se nombran marcadores tipo RFLP (*Restriction Fragment Length Polymorphisms* o Polimorfismo en los Fragmentos de Restricción), RAPD (*Random Amplified Polymorphic DNA* o Fragmentos polimórficos

de ADN Amplificados al Azar) o AFLP (*Amplified Fragment Length Polymorphisms* o Polimorfismo en los Fragmentos Amplificados de ADN).

La técnica RAPD ampliamente utilizada en plantas medicinales, fue propuesta por Williams y cols en 1990 como una alternativa a la técnica RFLP en la identificación de individuos de diferente especie y al igual que en la amplificación por PCR se compone de ciclos de desnaturalización del ADN, hibridación pero a temperaturas bajas (36°C) y amplificación usando oligonucleótidos cortos (10 nucleótidos), que se diseñaron inicialmente en ausencia de cualquier información de secuencia de las especies *Glycine max* variedad Bonus, *Glycine soja*, *Zea mays* líneas CM37 y T232, *Neurospora crassa*, *Homo sapiens* líneas Hu2 y Hu3, *Escherichia coli*, *Listeria monocytogenes*, *Staphylococcus aureus* y *Salmonella typhimurium*, una composición G+C entre un 50 – 80% y que no fueran palíndromes (Williams y cols., 1990). Al término de la reacción, el número de amplificados aleatorios separados electroforéticamente se ve modificado en función del tamaño y la secuencia del oligonucleótido utilizado (Abdel-Ghany y Zaki 2003), dando lugar a amplificados polimórficos que representan diferentes *loci* (varios *locus*) que en plantas indican su herencia dominante, su habilidad para detectar regiones de ADN altamente variables (5-10 *loci* por fragmento cebador), su potencialidad en la localización de genes, identificación de variedades, estudios de hibridación inter e intraespecífica y el estudio de la variación genética en poblaciones emparentadas (Gan y cols., 2003; Amarger y Mercier, 1995). Los marcadores tipo RAPD con una poca cantidad de ADN permiten obtener resultados con rapidez, a un bajo costo e inversión de equipos; sin embargo, presenta la desventaja en la inconsistencia de datos, ya que la

reproducibilidad se afecta con pequeñas variaciones en las condiciones de amplificación (Williams y cols., 1990). Esta desventaja se puede reducir con un alto grado de estandarización de las condiciones de reacción y de amplificación, usando un control interno para asegurar la reproducibilidad de los productos de la amplificación, y llevando un registro de las bandas nítidas y consistentes (Antoni, 1997)

La aplicación de esta técnica en la identificación de especies de plantas, con uso medicinal y sus adulterantes, ha sido informada por Cheng y cols. (1997) para la diferenciación de ocho rizomas del género *Coptis*, utilizadas en la medicina tradicional china para el tratamiento de la fiebre. Hosokawa y cols. (2000) lograron diferenciar entre tres especies del género *Scutellaria* muy similares morfológicamente, empleando partes aéreas secas de las plantas; que son utilizadas en Europa, Estados Unidos y Asia para combatir la fiebre causada por el paludismo. Cheng y cols. (2000) identificaron a la especie *Astragalus membranaceus* de su adulterante *Hedysarum polybotrys* en un grupo de 20 muestras con presentación comercial y adquiridas en diferentes establecimientos; ambas especies presentan el principio activo astragaloside IV con actividad diurética. Cheng y cols (2001) identificaron tres especies del género *Atractylodes* con una relación filogenética muy cercana a partir de muestras adquiridas en el mercado de plantas y que son utilizadas en el tratamiento de enfermedades gastrointestinales.

### **1.2.3 Marcadores genéticos de secuencia específica.**

El conocimiento cada vez más extenso sobre las proteínas y los genes involucrados en el funcionamiento de las plantas ha permitido identificar con mayor facilidad tanto

secuencias codificables (exones) como no codificables (intrones y regiones intergénicas) en un mayor número de especies (Lewin, 2000). Las regiones codificables son muy apreciadas en el campo de la biología, química, bioquímica y medicina ya que permiten dilucidar vías metabólicas y enfermedades relacionadas con la presencia o ausencia de una sustancia o proteína específica. En el caso de las regiones no codificables se desconocía o ignoraba su función, ahora se sabe que muchas de estas regiones permiten el ordenamiento de exones, como sucede con los intrones, y logran una distribución adecuada de los genes en el cromosoma, como es el caso de las regiones intergénicas. En la práctica, la amplificación de regiones no codificables ha permitido la identificación de especies cuando el ADN está fraccionado, como sucede en muestras secas de plantas. Esto se debe a tres razones principales: 1) el tamaño del producto amplificado no se ve afectado por la degradación parcial del ADN, 2) los oligonucleótidos flanqueantes son altamente conservados y 3) las regiones amplificadas presentan una variación franca entre secuencias de diferentes especies.

En el campo de las plantas medicinales la caracterización de la región intergénica ARNr-5S (ácido ribonucleico 5S) nuclear, la región ribosomal ITS1-5.8S-ITS2 que se encuentra en el plástido y la región *trnL/trnF* del cloroplasto se han empleado en la identificación de especies; sin embargo, la mayoría de los trabajos publicados estudian a la región intergénica ARNr-5S nuclear (Buchanan, 2000). En los eucariontes superiores, se compone de una región codificable de unidades repetidas (cistrón), de aproximadamente 120 pares de bases (pb), separadas por un espacio cercano a 300 pb. Este gen es considerado altamente conservado, pero el espacio

que separa las regiones es muy variable entre las diferentes especies; lo que permite utilizarlo como un método de identificación. La técnica consiste principalmente en la utilización de cebadores conservados que delimiten la región intergénica ARNr-5S de la planta de estudio, para llevar a cabo su amplificación y secuenciación. Posteriormente se pueden utilizar enzimas de restricción que relacionen al producto digerido con la identificación de la especie. En un estudio realizado a 25 individuos, en donde se incluían especies y variedades del género *Fritillaria*, la amplificación específica de la región intergénica ARNr-5S y su digestión con la enzima *EcoR* I, permitió identificar a *Fritillaria cirrosa* de *F. puqiensis*, *F. anhuiensis* y *F. thunbergii*, una vez que el fragmento de interés fue secuenciado en cada una de las especies. *F. cirrosa* es muy apreciada por sus propiedades antitusivas y expectorantes por lo que su bulbo llega a ser hasta 100 veces más caro que el de *F. thunbergii*, aunque la toxicidad de esta sea mucho mayor, de ahí la importancia en la identificación del material vegetal (Cai y cols., 1999).

#### **1.2.4 Marcadores genéticos microsatelitales.**

Los microsatélites son secuencias simples repetidas en tandem de ADN con unidades mono, di, tri, tetra, penta y hexanucleótidas, que se encuentran de forma abundante en genomas eucariontes, son altamente polimórficas debido al mecanismo mutacional conocido como deslizamiento de una hebra durante la replicación del ADN (*slippage* de la ADN polimerasa), y están presentes como regiones codificables y no codificables en todo el genoma (Toth y cols, 2000). Los *loci* microsatélites, desde el punto de vista de herencia Mendeliana, son codominantes y su identificación alélica es reproducible, lo que ha permitido

utilizarlos como marcadores genéticos altamente informativos en pruebas de paternidad, identificación de individuos, desarrollo de mapas genéticos, relación con enfermedades, estudios filogenéticos y en genética en poblaciones (Weissenbach y cols., 1992; Goldstein y Pollock, 1997; Hakki y cols., 2002). En plantas, la mayor frecuencia dinucleótida se presenta como (AT)<sub>n</sub> y (GT)<sub>n</sub> (Rakoczy-Trojanowska, 2004), la **Tabla 1** muestra las unidades mono y dinucleótidas en *Caenorhabditis elegans*, embriofitas, *Saccharomyces cerevisiae* y hongos de diferentes regiones del ADN. La aplicación de estas secuencias en especies tropicales y agronómicas se extiende a la identificación y caracterización de los genotipos y especies; conservación del germoplasma y generación de colecciones; análisis de la variación genética de la población; selección de variedades resistentes a enfermedades; identificación de especies de fácil o difícil propagación; construcción de árboles filogenéticos que permitan identificar relaciones de parentesco, cercanía evolutiva y formación de híbridos, así como la ventaja de identificar genes o regiones cromosómicas que están relacionadas con características poligénicas como la resistencia a enfermedades o el color de la fruta.

**Tabla 1** Distribución de la frecuencia de microsatélites de las clases mono, di y tetranucleótidos *C. elegans*, embriofitas, levadura y otros hongos (Tóth y cols, 2000)

Grupo taxonómico	Regiones del ADN	Unidad repetida					
		A	C	AC	AG	AT	CG
<i>Caenorhabditis elegans</i>	Intergénicas	252	321	249	353	219	1
	Intrones	335	177	151	173	224	1
	Exones	23	20	16	29	9	-
Embriofita	Intergénicas	1975	37	115	521	1077	2
	Intrones	1341	39	168	444	710	-
	Exones	17	1	4	39	7	-
<i>Saccharomyces cerevisiae</i>	Intergénicas	3121	19	266	57	1552	-
	Intrones	2994	18	26	26	1385	-
	Exones	36	-	5	-	14	-
Fungi	Intergénicas	2024	56	140	136	279	-
	Intrones	1868	207	511	140	362	-
	Exones	9	-	2	-	2	-

El éxito en el aislamiento y selección de microsatélites, obtenidos principalmente en especies de plantas de interés agronómico, se ha logrado gracias a la construcción y escrutinio de bibliotecas enriquecidas de secuencias simples repetidas. Este enriquecimiento se logra fraccionando el ADN y empleando sondas marcadas diseñadas de acuerdo a la región repetida tandem que se desea capturar. Una vez identificados, secuenciados y clonados, los marcadores microsatelitales pueden ser comparados intra e interespecíficamente, identificando la homología y variación en la secuencia dentro de las variedades o especies cercanas taxonómicamente (Rafalski y Tingey, 1993). En variedades de cebada, se ha descrito la frecuencia de microsatélites de la clase dinucleótida (AC)<sub>10</sub> y (AG)<sub>10</sub> cada 220 Kb (Lui y cols., 1996), similar a lo reportado en *Arabidopsis* (Bell y Ecker, 1994), arroz (Wu y Tanksley, 1993) y trigo (Roder y cols., 1995).

### **I.3. *Psidium guajava* Linn.**

#### **I.3.1 Descripción taxonómica.**

*Psidium guajava* (guayaba) pertenece al reino vegetal, subreino Fanerógamas, clase Angiospermas, subclase Dicotiledónea, subdivisión Lignosae, orden Myrtales, familia Myrtaceae, género *Psidium*, especie *guajava*. Dentro del género *Psidium* se encuentran aproximadamente 140 especies cuyos frutos son comestibles. *P. guajava* es un arbusto arborescente, su corteza es lisa, de color café rojizo oscuro, tersa escamosa. Las hojas son simples, opuestas, de forma ovaladas, tienen de 10 a 25 pares de nervaduras laterales. Las flores de color blanco se encuentran en grupos dos a tres. Los frutos son de color amarillo verdoso en su exterior o de color amarillo

claro en su plena madurez, rugosos o lisos, punteados densamente, brillantes y fragantes. La pulpa es jugosa, de color blanco amarillento, rosado o rojo subido, con sabor dulce notoriamente almizclado y aromático. Las semillas son numerosas, pequeñas, de color amarillo claro o café amarillento. Debido al alto contenido en vitamina C y azúcar, los frutos son consumidos en forma directa o utilizados en la fabricación de dulces y jugos (González y cols., 2002). El origen de esta planta es incierto pero se le ubica en Mesoamérica para ser propagada por los españoles y portugueses a todos los trópicos del mundo donde se ha naturalizado con la ayuda de los pájaros.

### **1.3.2 Estudios realizados en *P. guajava*.**

El guayabo es explotado comercialmente por sus frutos, también es utilizado en la medicina tradicional mexicana para el tratamiento de la diarrea y el malestar que ocasiona, empleando la infusión de sus hojas (Lozoya X, 1994). Debido a su interés comercial y aplicación terapéutica, grupos de investigación alrededor del mundo han estudiado su identidad anatómica y morfológica, composición química, efecto farmacéutico, estudios biotecnológicos y de biología molecular.

Los estudios realizados de identidad anatómica y morfológica, describen las características macro y microscópicas de la hoja de guayaba con la finalidad de obtener un parámetro de control de calidad que permita identificar a la droga vegetal utilizada con fines terapéuticos (Rivera-Arce y cols., 2003).

Con respecto a la composición química, Idstein y cols. (1985) identificaron en el fruto los aceites esenciales ácido cinamóico y ácido hexanóico; Lozoya y cols. (1994) identificaron en el extracto metanólico de la hoja de guayaba, un grupo de glicosidos

de quercetina con actividad espasmolítica; Mercadante y cols. (1999) identificaron 14 tipos de carotenoides de la pulpa rosada de una variedad Brasileña; Begum y cols. (2002) identificaron triterpenos característicos de la hoja. En relación al efecto farmacéutico, Cheng y cols. (1983) observan un efecto hipoglucemiante en ratones y humanos al consumir jugo de guayaba; Lozoya y cols. (1990) observaron el efecto antidiarreico del extracto metanólico de la hoja de guayaba, al emplear un modelo *in vitro* de ileon de cobayo estimulado eléctricamente; Jaiarj y cols (1999) observaron actividad anti tusígena con los extractos acuosos de la hoja de guayaba, al inducir la irritación con ácido capsaicinico en cobayos y ratas, paralelamente también observaron la inhibición del crecimiento de *Staphylococcus aureus* al emplear extractos acuosos, metabólicos y clorofórmicos de la hoja; Arima y cols (2002) aislaron flavonoides de la hoja de guayaba con actividad antimicrobiana a *Salmonella enteritidis* y *Bacillus cereus*; Conde-Garcia y cols (2003) detectaron un efecto inotrópico negativo (relajación del músculo cardiaco) producido por un extracto acuoso-etanólico de la hoja de guayaba en un modelo *in vitro* del atrio izquierdo de cobayo estimulado eléctricamente para su contracción. En referencia a los estudios biotecnológicos, en *P. guajava* se han empleado técnicas biotecnológicas para la micropropagación y cultivo de plántulas provenientes de callos generados a partir de tejido frutal (Madhavi y cols., 1992) y de anteras (Babbar y Gupta, 1986).

Con respecto a estudios de genética poblacional, Prakash y cols (2002) utilizaron 36 genotipos diferentes de *P. guajava* y 5 individuos del genero *Psidium*, cultivados en India, para determinar la distancia genética. El trabajo emplea la técnica de amplificación aleatoria con 8 oligonucleótidos diferentes. Los resultados, analizados

por el método de Distancia Cuadrada Euclidiana, presentaron un intervalo entre el 54% al 11%, donde la mayor diferencia se obtuvo en individuos de las especies *P. guajava* y *P. quadrangularis*, y su contraparte entre individuos de diferentes selecciones de guayaba.

En relación a la identificación de secuencias funcionales en *P. guajava* L. se identificó la secuencia de una enzima abundante en fruto, 13-hidroxiperoxidasa liasa, empleando oligonucleótidos cebadores homólogos en otras especies (Tijet y cols., 2000). También esta planta se ha utilizado como modelo para la identificación de secuencias funcionales en sistemas bioinformáticos QSAR (*Quantitative structure-activity relationship* o Relación cuantitativa estructura-actividad) para la identificación de miembros de la familia ACC oxidasa, estudiando estructuras secundarias de RNA (Gonzalez-Díaz y cols., 2005).

Al inicio del presente estudio, el Servicio de Información y Estadística Agroalimentaria y Pesquera (SIAP) de la SAGARPA, informó que los 3 estados de la República Mexicana con mayor producción de *P. guajava* en el 2002 fueron: Michoacán (106,854 Ton), Aguascalientes (46,168 Ton) y Zacatecas (31,620 Ton) (<http://www.siea.sagarpa.gob.mx>).

## II PLANTEAMIENTO DEL PROBLEMA.

Hasta el momento los estudios a nivel genético de *P. guajava* se limitan a la identificación de enzimas relacionadas al metabolismo de los azúcares y ácidos grasos, la obtención de marcadores tipo RAPD en variedades hindúes, así como la comparación morfológica del fruto con estos marcadores en árboles mexicanos. Sin embargo, esta información no se relaciona con el control de calidad o la acumulación de quercetina en la hoja del guayabo para la elaboración de fitomedicamentos. El presente trabajo propone utilizar las herramientas de la biología molecular para la obtención de marcadores genéticos polimórficos informativos, del tipo RAPD y *loci* microsatélite, que se relacionen con la acumulación de flavonoides, autentifiquen a la especie y/o variedades, y permitan la discriminación con otras especies utilizadas como posibles adulterantes.

### III HIPOTESIS

Algunas variaciones en regiones del ADN entre individuos de *P. guajava* cultivados en diferentes estados de la República Mexicana están asociadas con la acumulación de flavonoides mientras que las regiones conservadas con la identificación de la especie.

### IV OBJETIVOS

#### IV.1 Objetivos Generales.

IV.1.1 Caracterizar las variaciones u homologías de secuencia en los amplificadores aleatorios tipo RAPD provenientes del ADN genómico de *P. guajava* L.

IV.1.2 Construcción de una biblioteca genómica de *loci* microsatélite de la clase dinucleótida para *P. guajava* L.

#### IV.2. Objetivos Particulares.

IV.2.1 Obtención del ADN genómico de *P. guajava* para la amplificación aleatoria de segmentos y su secuenciación.

IV.2.2 Obtención de una biblioteca microsateletal rica en secuencias dinucleótidas repetidas.

IV.2.3 Identificar variaciones genéticas de las muestras colectadas de *P. guajava*.

IV.2.4 Identificar marcadores genéticos que correlacionen con la acumulación de quercetina.

IV.2.5 Emplear la base de datos NCBI para genomas de plantas en la identificación de secuencias homólogas.

**IV.2.6** Diseñar oligonucleótidos específicos para la identificación genética de *P. guajava*.

## V DISEÑO EXPERIMENTAL

La caracterización y análisis en las homologías y variaciones de las secuencias de *P. guajava* L. se realizó mediante la construcción de las bibliotecas genómicas tipo RAPD y *loci* microsatélite, empleando el ADN genómico de retoños provenientes de 40 árboles seleccionados al azar de los estados de Aguascalientes, estado de México, Querétaro y Michoacán.

Los amplificadores aleatorios tipo RAPD fueron obtenidos del ADN genómico con oligonucleótidos comerciales (Amersham Biosciences, England), separados por electroforesis en gel de agarosa al 1.8%, capturados y analizados por un documentador de geles. Aquellos individuos que mostraron diferencias en los perfiles de bandeo fueron analizados en la acumulación de quercetina, ya que este flavonoide es utilizado como huella química en la selección de hoja de guayaba para la elaboración de fitomedicamentos. Este compuesto se extrae de las hojas secas con metanol de forma exhaustiva, se hidrolizan sus derivados glucosilados y se cuantifica por cromatografía líquida de alta resolución (HPLC).

Los fragmentos tipo *loci* microsatélites se obtuvieron con la digestión del ADN genómico empleando las enzimas de restricción Rsa I, Alu I y Sau 3A y la hibridación con oligonucleótidos biotilizados que contienen secuencias repetidas deseadas.

Los segmentos de ADN con un tamaño entre 200 a 1200 pb fueron clonados en el vector *pCR-TOPO* y propagados en células *E. coli* (TOP10), químicamente competentes. Las colonias transformadas seleccionadas por el tamaño del inserto, se propagaron, los plásmidos se purificaron, se secuenciaron los insertos y las secuencias obtenidas se compararon con las descritas por el GenBank NCBI para

encontrar regiones homólogas. Las secuencias que no presentaron alguna homología se diseñaron sus cebadores de amplificación para comprobar su presencia en las muestras colectadas y especies similares a *P. guajava*, **Figura 1**.

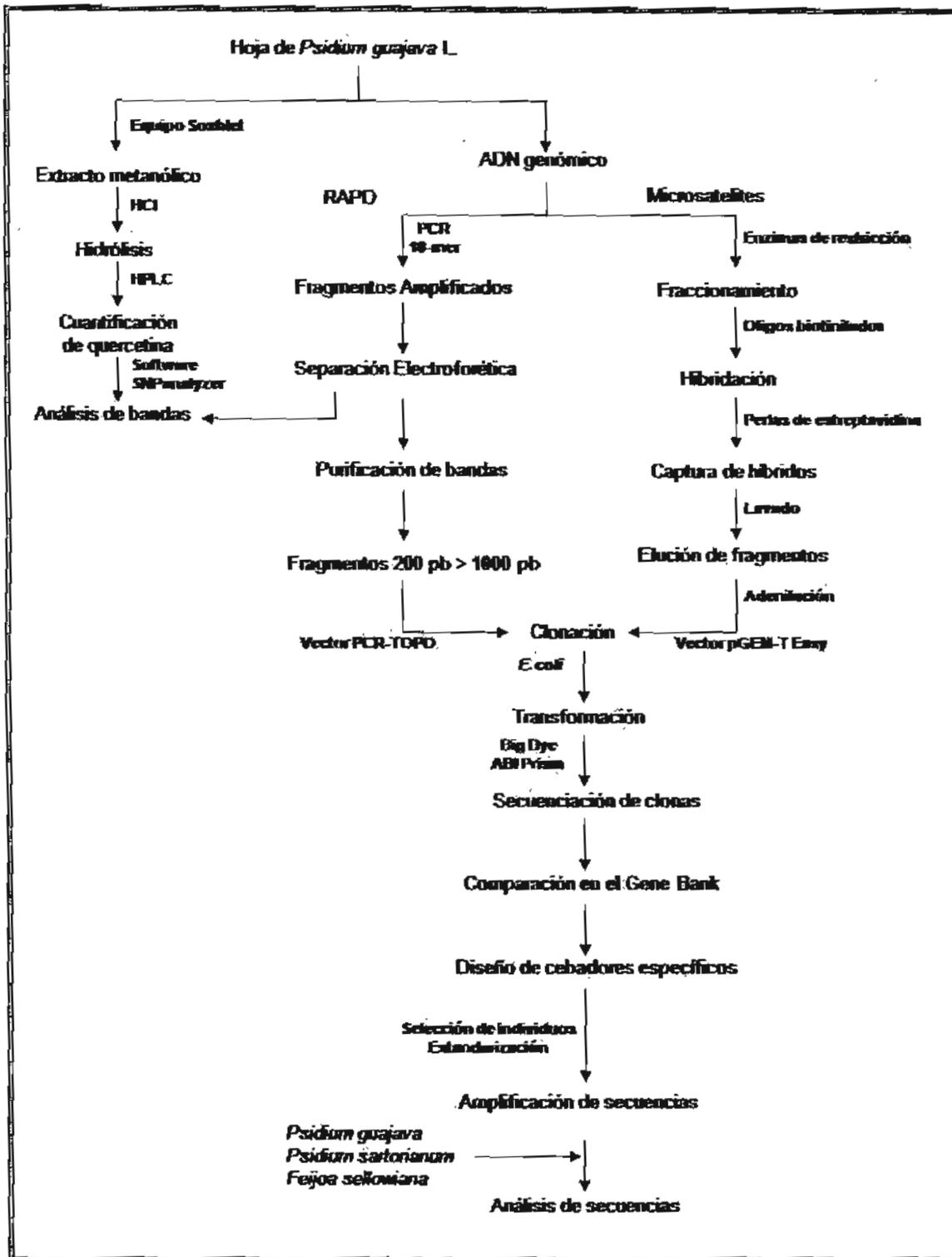


Figura 1. Obtención de marcadores asociados a la acumulación de quercetina y en la autenticación de *Psidium guajava* Linn.

## VI MATERIALES Y MÉTODOS

### VI.1. Obtención de muestras de *P. guajava*.

Para este estudio se seleccionaron 4 huertos de diferentes estados de la República Mexicana con diferente importancia en la producción de guayaba, y diferentes condiciones geográficas, climáticas y agronómicas: La Labor (Calvillo, Ags), Rancho Arizmendi (Cuatepec Harinas, Edo. Mexico), Los Pinzanes (Zitácuaro, Mich) y El Exilio (Jalpan, Qro). El criterio de inclusión más sobresaliente fue el agronómico, debido a que los 4 estados representaron el 80% de la producción nacional total entre 1998 – 2002 según reportes estadísticos de la secretaría de agricultura, ganadería, desarrollo rural, pesca y alimentación (SAGARPA).

El contacto con los agricultores se estableció a través de las delegaciones estatales de la SAGARPA, quienes a su vez refirieron a los principales productores de guayaba de cada estado. Cada productor fue contactado telefónicamente, se les explicó el proyecto y pidió una autorización para visitar sus cultivos. Los productores, una vez informados y con disponibilidad de su participación, se concertó una primera cita donde se realizó el muestreo aleatorio de 10 árboles. En esta visita, se seleccionaron cultivos al final de la floración inicio de fructificación, con el fin de homogenizar el estado fenológico de los árboles en la toma de muestra. Los árboles fueron localizados en un mapa del terreno y por estacas clavadas a un lado del árbol señalando su número correspondiente, esto permitió facilitar su localización en visitas posteriores (**ANEXO 1**).

En cada huerta se seleccionó de manera aleatoria 10 árboles de *P. guajava* L., obteniendo tanto ejemplares de herbario como retoños de hoja para su

procesamiento. Los ejemplares de herbario, fueron depositados en el Herbario de Plantas Medicinales del I.M.S.S. del Centro Médico Nacional Siglo XXI (IMSSM) para su determinación botánica.

## **VI.2. Cuantificación de quercetina por cromatografía de líquidos de alta resolución (HPLC).**

### **VI.2.1 Preparación de la muestra.**

Un gramo de hoja limpia y seca con aire caliente (40 °C), se pulverizó, empaquetó en cartucho y colocó en el aparato Soxhlet para su extracción exhaustiva con 25 ml de metanol por 3 hr. El extracto evaporado al vacío se resuspendió en 20 ml de solución metanólica HCl 0.5 N y se sometió a reflujo por 1 hr. La solución resultante se aforó con metanol en un matraz volumétrico de 25 ml. Posteriormente un mililitro de la solución se diluyó 10 veces con el mismo solvente y se filtró en una membrana de 0.45  $\mu\text{m}$  para su análisis cromatográfico.

### **VI.2.2 Cuantificación y análisis cromatográfico**

El equipo utilizado fue el sistema *Waters 2796* (Milford, USA), conformado de un inyector automático simple, una bomba binaria, un desgasificador continuo con vacío, una columna con sistema de calentamiento y un detector *PDA Waters 996*.

La separación cromatográfica se llevó a cabo en una columna de 4.6 mm X 150 mm, RP-18, 5  $\mu\text{m}$  (Metachem Technologies, USA) a 35 °C en una fase móvil compuesta de ácido fosfórico 0.01 M (A) y acetonitrilo (B). La velocidad del flujo fue constante, a 0.6 ml/min durante los 35 min que duró el corrimiento de la muestra y con la siguiente proporción en el gradiente: 0-2 min una relación 80%-20% B, 2-30 min una relación

40-60% B, 30-32 min una relación 40-60% B, 32-35 min una relación 80-20% B. La muestra se filtró con GHF Acrodisc (Pall Gelman, USA) de 0.45  $\mu\text{m}$  antes de su uso y el volumen de inyección fue de 10 $\mu\text{l}$ . La concentración de los cinco glicósidos de quercetina se determinó como la concentración final de su aglicona quercetina, interpolando del área del pico correspondiente en la curva de referencia. El pico de quercetina en los cromatogramas fue asignado en base a su tiempo de retención en el rango UV de 200 a 450 nm.

### **VI.3 Extracción y purificación del ADN genómico.**

Para la obtención del ADN, se emplearon 150 mg de hoja colectada en campo, limpia, congelada con nitrógeno líquido y pulverizada en un mortero estéril.

La hoja pulverizada se calentó a 60°C por 1 hr con 1 ml de amortiguador de extracción (100 mM Tris-HCl pH 8.0, 250 mM EDTA pH 8.0, 1.5 M NaCl, 0.2 %  $\beta$ -Mercaptoetanol, 0.1% PVP 40). Al término de la incubación, se agregó un volumen igual de cloroformo: alcohol isoamilico (24:1) y se extrajo la mezcla por 15 min. Pasado el tiempo de incubación, se centrifugó a 8,000 rpm por 10 min. La fase acuosa se transfirió a un tubo limpio, se agregó 0.5 vol de 5M NaCl, se mezcló, se agregaron 0.6 vol de isopropanol y se mezcló nuevamente. La solución se dejó toda la noche a 4°C. La purificación se complementó al utilizar el sistema *DNeasy Plant Mini Kit* (Qiagen, Alemania), según las recomendaciones del proveedor. La cuantificación del ADN se realizó mediante la lectura de la absorbancia espectrofotométrica a 260 nm y su integridad se verificó en un gel de agarosa al 0.8%. El material extraído se guardó a -70°C hasta su uso.

#### VI.4 Obtención de amplificadores aleatorios tipo RAPD.

El ADN individual de las 40 muestras colectadas en campo se amplificaron aleatoriamente usando la técnica RAPD. Para ello, se emplearon 15 ng de ADN genómico, una esfera del sistema *RAPD Analysis Bead* (Amersham Biosciences, England) y 25 pmoles de uno de los seis cebadores 10-mer, en un volumen total de 25  $\mu$ l. Las secuencias de los oligonucleótidos utilizados fueron: **1** 5'-d[GGTGCGGGAA]-3'; **2** 5'-d[GTTTCGCTCC]-3'; **3** 5'-d[GTAGACCCGT]-3'; **4** 5'-d[AAGAGCCCGT]-3'; **5** 5'-d[AACGCGCAAC]-3'; **6** 5'-d[CCCGTCAGCA]-3' (Amersham Biosciences, England) . El perfil de temperaturas que se programó en el termociclador *T-Gradient* (Biometra, Alemania) para la amplificación fue: 1 ciclo a 95°C por 5 min y 45 ciclos a 95°C por 1 min, 36°C por 1 min y 72°C por 2 min. Los amplificadores aleatorios se separaron electroforéticamente en un gel de agarosa al 1.8% junto con 1 $\mu$ g del marcador molecular de 100 bp (Invitrogen Life Technologies, USA). La imagen del gel se capturó con el sistema de documentación *Biolmaging* (UVP, UK) y las bandas se analizaron con el programa de cómputo *LabWorks v1.1.27* (UVP, UK).

#### VI.5 Digestión del ADN genómico.

La digestión del ADN genómico se realizó con las enzimas de restricción *Rsa I*, *Alu I* y *Sau 3A*, donde se obtuvieron extremos rasurados y pegajosos (New England Biolabs, USA). Las condiciones de digestión y el número de unidades de actividad empleadas fueron las recomendadas por el fabricante.

### **VI.6 Hibridación con oligonucleótidos biotinilados y captura con estreptavidina.**

La hibridación se realizó en una mezcla de reacción compuesta por ADN genómico desnaturalizado, un amortiguador ( $\text{NaPO}_4$  0.5M y SDS 0.5%) y el oligonucleotido marcado con biotina en el extremo 5'  $(\text{GT})_{15}$ , incubando toda la noche.

Después de la hibridación, la mezcla se incubó con esferas magnéticas ( $\text{Fe}^{++}$  y estreptavidina) y el amortiguador 1 (Tris 100 mM, pH 7.5 y NaCl 150 mM), por 45 minutos a temperatura ambiente. Los híbridos biotinilados se capturaron con un magneto y el sobrenadante se removió por centrifugación. El precipitado se lavó dos veces con el amortiguador 1 y una vez con TE (Tris-EDTA, pH 7.5).

### **VI.7 Clonación y transformación de los amplificadores aleatorios tipo RAPD y fragmentos con regiones repetidas.**

Se tomaron de 1 a 4  $\mu\text{l}$  de la reacción de amplificación aleatoria tipo RAPD, obtenida de la muestra del árbol 8 del estado de México (muestra con mayor acumulación de quercetina), se mezclaron con 1  $\mu\text{l}$  de solución salina y 1  $\mu\text{l}$  del vector *pCR-TOPO* (3.9 kb. Invitrogen Life Technologies, USA), obteniéndose un volumen final de 6  $\mu\text{l}$ . La reacción se incubó por 30 min a temperatura ambiente. Se tomaron 2  $\mu\text{l}$  de la mezcla de clonación, se colocaron en un tubo de células *E. coli* químicamente competentes cepa *TOP10* y se incubaron en hielo por 5 min. Posteriormente se dió un choque térmico de 45 seg a 42°C, se agregaron 250  $\mu\text{l}$  de medio líquido SOC y se incubó por 1 hr a 37°C sin agitación. Se tomaron 50, 100 o 150  $\mu\text{l}$  de la mezcla con las

células transformadas y se esparcieron en placas con medio agar Luria-Bertani (LB), 40  $\mu\text{l}$  de X-Gal (40 mg de X-Gal/ml en dimetilformamida) y kanamicina (50  $\mu\text{g}$  de kanamicina/ ml de medio).

Los híbridos biotinilados se adenilaron antes de su clonación, para ello, en 10  $\mu\text{l}$  de volumen de reacción compuesta por 6  $\mu\text{l}$  de la solución de elusión con los híbridos biotinilados, 1  $\mu\text{l}$  de solución amortiguadora PCR 10x, 0.2  $\mu\text{l}$  de  $\text{MgCl}_2$  50mM, 0.2  $\mu\text{l}$  de dATP 10 mM y 0.2  $\mu\text{l}$  de Taq ADN polimerasa 5 U/ $\mu\text{l}$  (Invitrogen Life Technologies, USA); se incubaron a 72°C por 30 min. Posteriormente se realizó el procedimiento de clonación y transformación descrito anteriormente.

#### **VI.8 Escrutinio de las bibliotecas de amplificadas aleatorias RAPD y fragmentos con regiones repetidas.**

La selección de las clonas transformadas se hizo en base al tamaño del inserto identificado por PCR (200 a 1200 pb). Las colonias blancas se picaron con una punta de 200  $\mu\text{l}$  y resuspendieron en 10  $\mu\text{l}$  de la mezcla de reacción compuesta por 1  $\mu\text{l}$  de solución amortiguadora PCR 10x, 0.2  $\mu\text{l}$  de  $\text{MgCl}_2$  50mM, 0.2  $\mu\text{l}$  de dNTPs 10 mM (Promega, USA), 0.2  $\mu\text{l}$  de cebadores M13 sentido y contrasentido 10 mM y 0.2  $\mu\text{l}$  de la enzima de amplificación Taq ADN polimerasa 5 U/ $\mu\text{l}$  (Invitrogen Life Technologies, USA). El perfil de temperaturas que se programó en el termociclador *T-Gradient* (Biometra, Alemania) para la amplificación fue: 1 ciclo a 95°C por 5 min y 30 ciclos a 95°C por 1 min, 53°C por 1 min y 72°C por 2 min. Posteriormente, se separaron por electroforesis en un gel de agarosa al 2% y se determinó su longitud cuando se comparó con el marcador de 100pb (Invitrogen Life Technologies, USA). Antes de

ser resuspendidas en la mezcla de reacción, se hizo un respaldo de cada colonia en una placa con agar LB tratada con kanamicina, cuadrículada y enumerada para su identificación. Las colonias seleccionadas se propagaron en 10 ml de medio líquido LB/Kanamicina a 37°C con agitación toda la noche. La purificación de los plásmidos se hizo con el sistema *QIAprep Spin Miniprep Kit* (Qiagen, Alemania) según las indicaciones del proveedor.

#### **VI.9 Secuenciación de los amplificadores aleatorios RAPD y fragmentos con regiones repetidas.**

Los plásmidos purificados con los insertos de tamaño requerido se secuenciaron en ambas direcciones con los oligonucleótidos flanqueantes. La solución de secuenciación se llevó a cabo con 500 ng del plásmido, 1 µl del amortiguador de secuenciación 5X, 1 µl del reactivo *BigDye V 3.0* y 1 µl de los cebadores flanqueantes 10 mM, en un volumen total de 5 µl. El perfil de temperaturas que se programó en el termociclador *T-Gradient* (Biometra, Alemania) para la amplificación fue: 25 ciclos de 96°C por 10 seg, 50°C por 5 seg y 60°C por 4 min. Después de la amplificación, las reacciones fueron purificadas con el sistema *DyeEx 2.0* (Qiagen, Alemania) según las indicaciones del proveedor y se analizaron con el secuenciador *ABI 3700* (ABI; USA). Los nucleótidos adenina, citosina, guanina y timina del sistema *BigDye*, están marcados con dR6G, dROX, dR110 y dTAMRA para su identificación. Las secuencias se compararon con los genomas de las plantas *Aegilops tauschii*, *Aegilops umbellulata*, *Allium cepa*, *Arabidopsis thaliana*, *Avena sativa*, *Beta vulgaris*, *Brassica juncea*, *Brassica napus*, *Brassica nigra*, *Brassica oleracea*, *Brassica rapa*,

*Capsicum annuum*, *Eragrostis tef*, *Glycine max*, *Hordeum vulgare*, *Lotus japonicus*, *Manihot esculenta*, *Medicago sativa*, *Oryza sativa*, *Phaseolus vulgaris*, *Populus trichocarpa*, *Prunus dulcis*, *Secale cereale*, *Setaria italica*, *Solanum lycopersicoides*, *Solanum lycopersicum*, *Solanum melongena*, *Solanum peruvianum*, *Sorghum bicolor*, *Theobroma cacao*, *Triticum aestivum*, *Triticum turgidum*, *Vigna radiata* y *Zea mays* subspecies *mays* empleando el programa de cómputo *BLAST Gene Bank* (Benson y cols., 2004) disponible en la dirección de Internet:

<http://www.ncbi.nlm.nih.gov/Genomes/index.html>.

#### **VI.10 Diseño de oligonucleótidos específicos.**

Las secuencias seleccionadas de las bibliotecas RAPD y *loci* microsatélite, se utilizaron en el diseño de cebadores específicos, utilizando el programa de cómputo *Web Primer DNA* disponible en la dirección de Internet: <http://seq.yeastgenome.org/cgi-bin/web-primer>. Los oligonucleótidos resultantes, fueron seleccionados por su temperatura de hibridación (55 °C), tamaño (18 – 22 pb) y secuencia única (**ANEXO 7**). Los cebadores fueron sintetizados por una casa comercial (Invitrogen Life Technologies, USA) a una concentración de 50nmoles, desalados y desprotegidos.

#### **VI.11 Estandarización de la amplificación por PCR de secuencias de *P. guajava* empleando oligonucleótidos específicos.**

Para establecer las condiciones de amplificación se utilizaron 5 - 20 ng de ADN plasmídico o 1 - 50 ng de ADN genómico proveniente de hoja fresca y seca de la muestra del árbol 8 del estado de México. La amplificación se realizó en 10 µl de la

mezcla de reacción compuesta por 1  $\mu$ l de solución amortiguadora PCR 10x, 0.2  $\mu$ l de  $MgCl_2$  50mM, 0.2  $\mu$ l de la mezcla de dNTPs (Promega, USA), 0.2  $\mu$ l de los oligonucleótidos específicos sentido y contrasentido 10 mM, 0.2  $\mu$ l de Taq ADN polimerasa (Invitrogen Life Technologies, USA) y el templado de ADN. El perfil de temperaturas que se programó en el termociclador *T-Gradient* (Biometra, Alemania) para la amplificación fue: 1 ciclo a 95°C por 5 min y 35 ciclos de 95°C por 1 min, 56°C por 1 min y 72°C por 2 min. Los amplificados se separaron electroforéticamente en un gel de agarosa al 2.0% y se determinó su longitud por comparación con el marcador de ADN, *100 bp ADN ladder* (Invitrogen Life Technologies, USA).

#### **VI.12 Validación de los fragmentos en muestras representativas y diferentes a *P. guajava*.**

La amplificación de las muestras de *P. guajava* y de las otras especies recolectadas se hizo empleando la concentración del ADN genómico,  $MgCl_2$ , mezcla de reacción y programa de amplificación, utilizadas durante la estandarización. Los fragmentos se separaron electroforéticamente en un gel de agarosa al 2.0% y se determinó su longitud por comparación con el marcador de 100 pb (Invitrogen Life Technologies, USA).

#### **VI.13 Análisis de los resultados.**

El análisis estadístico en la acumulación de quercetina total de las muestras colectadas como el de la amplificación de marcadores específicos para *P. guajava*, se hicieron por ANOVA de una vía empleando el programa de cómputo SPSS 10.0. Los parámetros de normalidad en la acumulación de quercetina fueron evaluados

estadísticamente con el programa de cómputo Sigma Stat y la prueba Kolmogorov-Smirnov. Posteriormente las muestras se agruparon considerando su distribución Gaussiana en los rangos establecidos, y se identificaron a las muestras que salieran de la normalidad, o sea, con mayor y menor acumulación de flavonoides.

La evaluación estadística de los amplificadores aleatorios RAPD y su correlación con la acumulación de quercetina se obtuvo con el programa de cómputo *SNP analyzer* (ISTECH, Corea) disponible en la dirección de Internet: [http://www.istech.info/istech/board/login\\_form.jsp](http://www.istech.info/istech/board/login_form.jsp). Para el análisis, los perfiles RAPD se clasificaron según las variaciones polimórficas que presentaron los segmentos en al menos 2 muestras utilizando el mismo oligonucleótido. Las variaciones se consideraron homocigotas, por la presencia o ausencia de una sola banda; o heterocigotas en varias bandas. La estimación de las frecuencias alélicas dominantes se obtuvo usando la prueba estadística Chi cuadrada y el análisis en el desequilibrio de unión de dichas frecuencias según lo descrito por Lewontin (1964, 1995). El análisis de asociación se realizó entre todos los perfiles RAPD obtenidos en cada muestra y su mayor acumulación de quercetina en primavera, para calcular el algoritmo de máxima expectación según lo descrito Excoffier y Slatkin (1995).

## VII RESULTADOS

### VII.1 Selección de la materia prima

En el presente estudio, se seleccionaron 4 estados de la República Mexicana por sus características agronómicas. Datos estadísticos recientes de la SAGARPA de 1980 a 2005 reportan que los estados de Aguascalientes y Michoacán son los principales productores de guayaba en el país, mientras que el estado de México y Querétaro son productores pequeños. Los 4 estados producen más del 70% de la fruta desde 1980, alcanzando un 86% de la producción total en septiembre del 2006. Al margen de estos datos, también se midió la altitud y condiciones climáticas (temperatura y precipitación pluvial) al momento de la recolecta. En cada estado se presentaron condiciones geográficas y climáticas diferentes, presentes en el **ANEXO 2**.

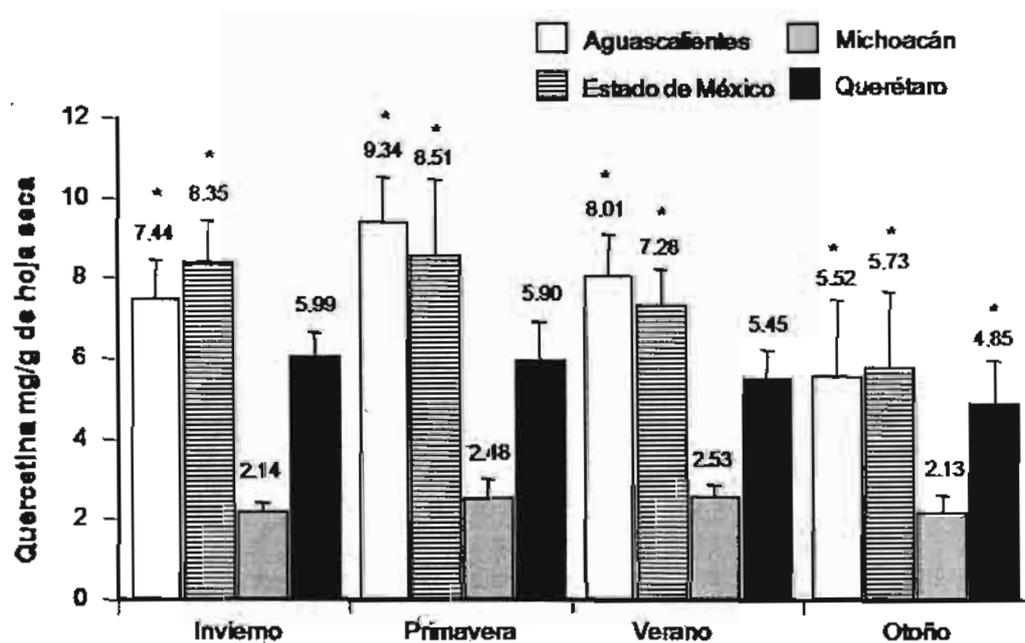
Los ejemplares representativos de cada región que fueron depositados en el Herbario de plantas medicinales del IMSS (IMSSM) se clasificaron como *P. guajava* Linn, *Myrtaceae* con los números de registro: 15,155; 15,158; 15,156 y 15,157 correspondientes a los estados de Aguascalientes, México, Querétaro y Michoacán, respectivamente.

### VII.2 Cuantificación del principio activo

A los extractos de etanol hidrolizados, obtenidos a partir de hojas limpias y secas se les determinó la concentración de quercetina (principio activo) por **cromatografía líquida de alta resolución**. Los datos obtenidos durante el año presentan diferencias significativas ( $p < 0.001$ ) en los promedios totales, entre los estados de Aguascalientes y México en comparación con Querétaro y Michoacán. La mayor concentración de quercetina se encontró en la estación de primavera correspondiente a los estados de

Aguascalientes y de México con valores de 9.34 y 8.51 mg/g de hoja seca, en comparación con Michoacán y Querétaro con valores de 2.48 y 5.90 respectivamente **(Figura 2)**.

## Concentración de Quercetina



\* $p < 0.01$

**Figura 2.** Acumulación de quercetina total en hojas secas de *P. guajava* L. de los estados de Aguascalientes, México, Michoacán y Querétaro, colectadas en Invierno (Enero y Febrero), Primavera (Abril y Mayo), Verano (Julio y Agosto) y Otoño (Octubre y Noviembre). El promedio + DS se obtuvo de los mismos 10 árboles aleatorios muestreados en cada región todo el año.

Al margen de las condiciones de cultivo y la etapa fisiológica de la planta, existen individuos que presentan una elevada y baja acumulación de quercetina. Específicamente, un árbol del estado de México presenta la mayor acumulación de quercetina de 12.62 mg/g de hoja seca en primavera, observándose una variación de hasta 2.12 veces con respecto a otros individuos de la misma zona.

### **VII.3 Selección de los individuos para el estudio genético en base a la acumulación de flavonoides.**

La característica fenotípica a considerar en este estudio fue la acumulación de quercetina que presentaron en el año de recolecta 2002.

Los árboles con mayor acumulación de quercetina en primavera fueron A2, A3 y A9 de Aguascalientes, EM5 del estado de México, M2 y M3 de Michoacán, y Q2 y Q6 de Querétaro, por otro lado, los árboles con menor acumulación de quercetina fueron A1, A7, EM1, EM7, M4 y Q9 de los respectivos estados (**ANEXO 3**). Por último, se seleccionaron muestras representativas con mayor y menor acumulación de quercetina en cada estado, adicionando a este grupo muestras con un polimorfismo característico por su amplificación aleatoria RAPD con los oligonucleótidos 1, 2 y 3. En el **ANEXO 4** se muestra la clave original de la muestra, el número reasignado y la mayor acumulación de quercetina obtenida en el año.

El árbol 5 del estado de México o muestra VIII, se toma como referencia genética para estudios posteriores por presentar la mayor acumulación del principio activo (12.62 mg/g hoja seca en primavera). Las muestras IV, V y XVI se integraron a este grupo, debido a que presentaron un polimorfismo evidente con los oligonucleótidos 1, 2 y 3 respectivamente.

Una vez seleccionadas las muestras se purificaron, cuantificaron y se les verificó la integridad del ADN total.

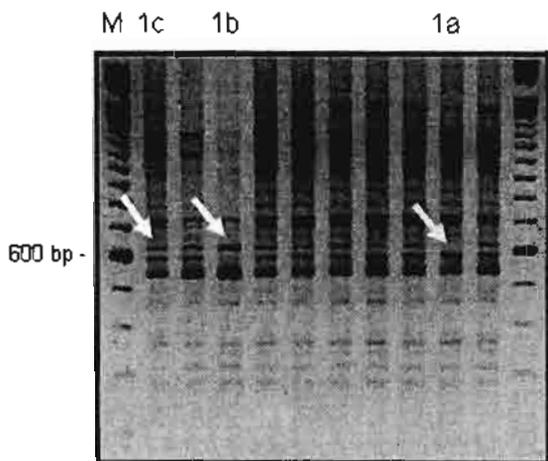
#### **VII.4 Identificación de fragmentos polimórficos en las muestras colectadas utilizando el fraccionamiento de amplificación aleatoria RAPD.**

La primera etapa del estudio genético de *P. guajava* se enfoca la amplificación aleatoria tipo RAPD utilizando 6 oligonucleótidos comerciales nombrados del 1 al 6. En el análisis de los segmentos se consideran a los que tienen un tamaño entre 200 a 1200 pares de bases (pb). Los seis cebadores generan un total de 91 segmentos, dando lugar a diferentes perfiles genéticos. Las diferencias se encuentran tanto al utilizar diferentes oligonucleótidos como al utilizar el mismo oligonucleótido en diferentes muestras.

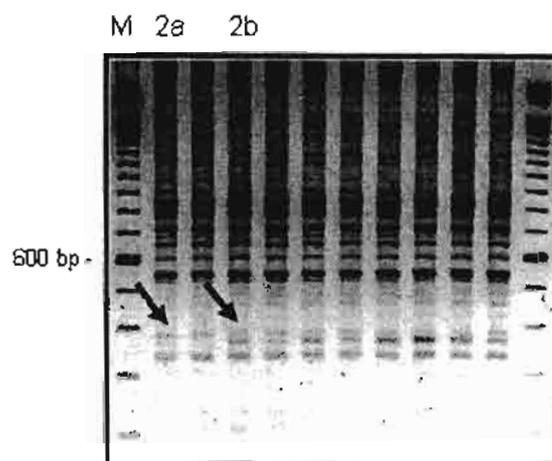
La **Figura 3** muestra la separación electroforética de los diferentes perfiles RAPD obtenidos en las muestras colectadas, donde se observan claras diferencias polimórficas (flechas) que se distinguen de las bandas conservadas o repetidas. De los 91 segmentos RAPD que se generan con los oligonucleótidos utilizados, el 77% se comparten con el oligonucleótido 1, 74% con el 2, 87% con el 3, 94% con el 4, con los oligos 5 y 6 se comparten casi en su totalidad las bandas. La amplificación aleatoria con el oligonucleótido 1 se representa en **A** con muestras del estado de Michoacán, donde se generan dos segmentos polimórficos y por consiguiente tres patrones de bandeo diferentes. La amplificación aleatoria con el oligonucleótido 2 se representa en **B** con muestras del estado de Aguascalientes, donde se genera un segmento polimórfico y dos patrones de bandeo diferentes. La amplificación aleatoria con el oligonucleótido 3 se representa en **C** con muestras del estado de Querétaro,

donde se genera un segmento polimórfico y dos patrones de bandas diferentes. La amplificación aleatoria con el oligonucleótido **4** se representa en **D** con muestras del estado de Aguascalientes, donde se generan dos segmentos polimórficos y tres patrones de bandeo diferentes. Las amplificaciones aleatorias con los oligonucleótidos **5** y **6** se representan en **E** y **F** con muestras del estado de México y Michoacán respectivamente, en ellos no se observan segmentos polimórficos fácilmente distinguibles.

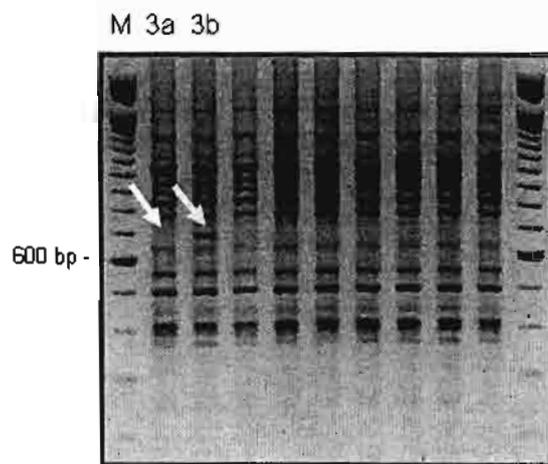
**A**



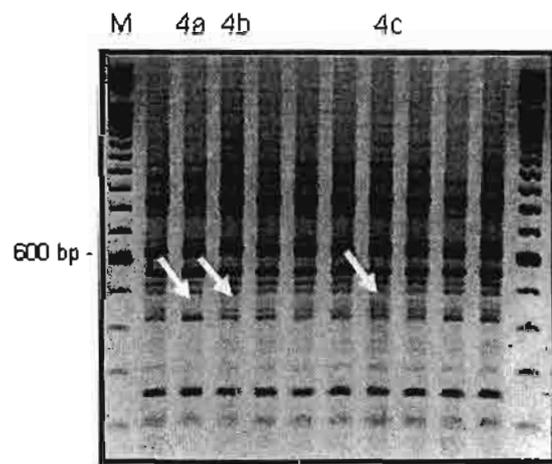
**B**



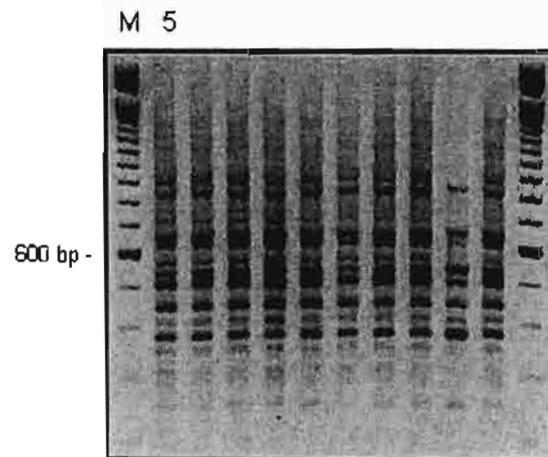
**C**



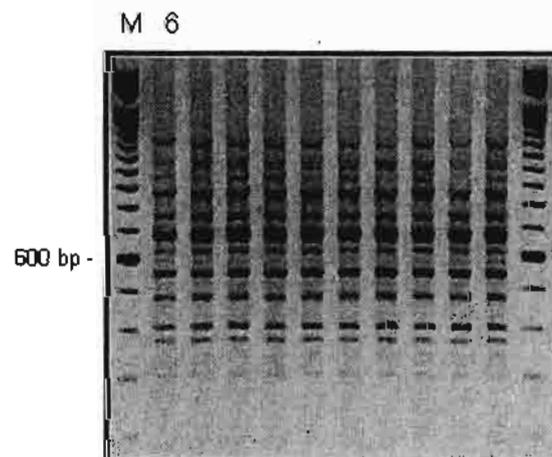
**D**



**E**



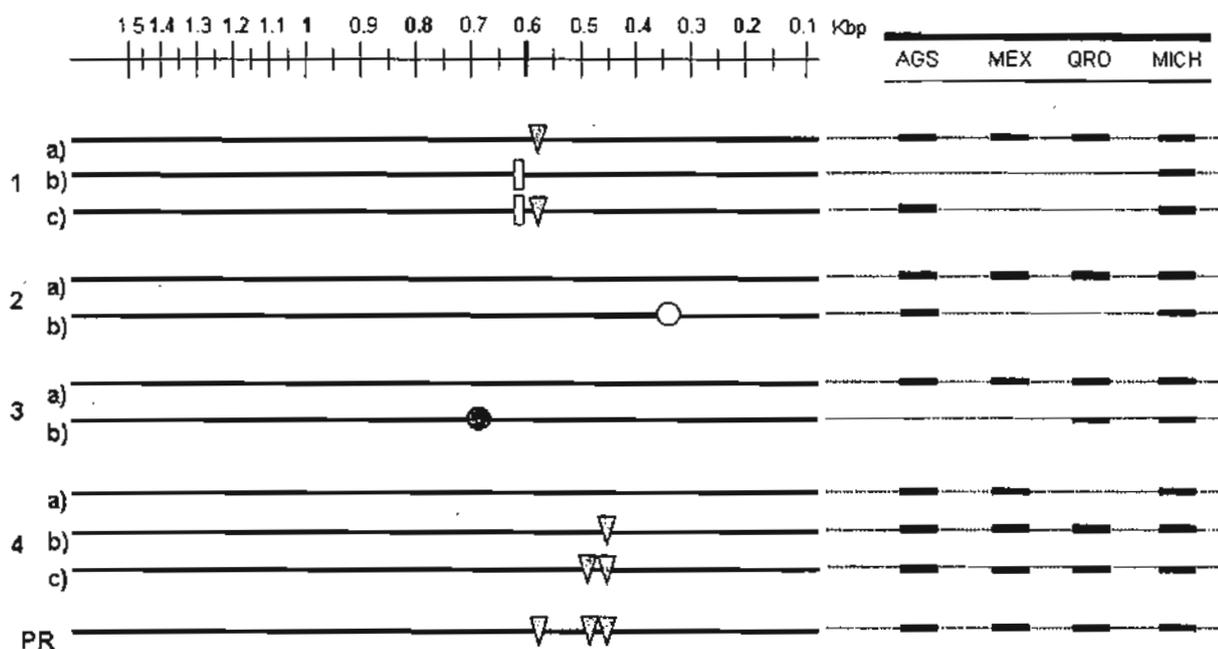
**F**



**Figura 3** Patrones de bandas tipo RAPD de *P. guajava* provenientes de diferentes estados.

Carril M: Marcador 100 bp. **A** Carriles 1a, 1b y 1c: representan perfiles polimórficos RAPD usando el cebador **1** (muestras del estado de Michoacán). **B** Carriles 2a y 2b: representan los perfiles polimórficos RAPD usando el cebador **2** (muestras del estado de Aguascalientes). **C** Carriles 3a y 3b: representan los perfiles polimórficos RAPD usando el cebador **3** (muestras del estado de Querétaro). **D** Carriles 4a, 4b y 4c: representan los perfiles polimórficos RAPD usando el cebador **4** (muestras del estado de Aguascalientes). **E**: Carril 5 representa el perfil RAPD usando el cebador **5** (muestras del estado de México). **F**: Carril 6 representa el perfil RAPD usando el cebador **6** (muestras del estado de Michoacán). Las flechas indican las variaciones polimórficas con comportamiento alélico.

La **Figura 4** es la representación esquemática de los segmentos polimórficos identificados en la **Figura 3**. En el esquema se muestran únicamente las bandas representativas obtenidas con los oligonucleótidos **1** al **4** que se localizan en los diferentes estados muestreados. El esquema nos indica que el patrón de bandeo **1a** se encuentra en árboles de todos los estados y se caracteriza por la presencia de una banda de 590 pb; el patrón **1b** se localiza en un árbol de Michoacán, presentando una banda de 610 pb y el patrón **1c** está presente en árboles de Aguascalientes y Michoacán, con las 2 bandas anteriores de 590 y 610 pb. Con la amplificación aleatoria del oligo **2**, el patrón de bandeo **2a** se encuentra en árboles de todos los estados y el patrón **2b** se localiza en Aguascalientes y Michoacán, este patrón se caracteriza por una banda de 370 pb. Con el tercer oligonucleótido, el patrón de bandeo **3a** se encuentra en árboles de todos los estados y el patrón **3b** únicamente se localiza en 2 árboles, uno en Querétaro y otro en Michoacán, este perfil se caracteriza por una banda de 690 pb. Por último, el cuarto oligonucleótido genera **4a** presente en árboles de los estados de Aguascalientes, México y Michoacán; el patrón **4b** que se caracteriza por una banda de 460 pb y el patrón **4c** por 2 bandas de 460 y 480 pb, estos dos últimos se encuentran en todos los estados. En cada amplificación aleatoria se determinaron los perfiles con mayor frecuencia, resultando en cada caso los patrones **1a**, **2a**, **3a** y **4b**. El patrón representativo (PR) o suma de los patrones con mayor frecuencia se identifica por la banda de 590 pb proveniente de **1a** y las 2 bandas de 460 y 480 pb provenientes de **4c**.



**Figura 4** Esquema de los fragmentos polimórficos RAPD usando oligonucleótidos 1 – 4.

La escala superior indica la longitud relativa de los segmentos, tomando como referencia un marcador de ADN (*100 bp ladder*). Cada símbolo representa los segmentos polimórficos con un comportamiento homocigoto o heterocigoto, presentes en los patrones de bandeo (combinación de los números y letras a la izquierda), y las barras ( — ) sobre las líneas continuas ( — ) a la derecha representan las regiones donde se encontraron. La línea inferior PR es la representación global de *P. guajava* resultado de los patrones de bandeo polimórficos con mayor frecuencia (1a, 2a, 3a y 4b). AGS: Aguascalientes, MEX: Estado de México, QRO: Querétaro, MICH: Michoacán.

En la **Tabla 2** se resume el oligonucleótido utilizado, los segmentos polimórficos y totales, los patrones con la banda o bandas características y el patrón con mayor frecuencia en cada fraccionamiento, para correlacionar por último con el desequilibrio genético entre los patrones con mayor frecuencia. Estos resultados se obtienen utilizando los datos del **ANEXO 5** en combinación con el programa de cómputo SNP analyzer.

**Tabla 2** Secuencias de los oligonucleótidos 10-mer usados para el análisis RAPD, número de segmentos polimórficos y totales (pb), su frecuencia, patrón de bandas dominante y su desequilibrio genético en individuos de *P. guajava* L.

Cebador	Secuencia (5' a 3')	Número de segmentos polimórficos (pb)	Número de segmentos totales	Patrones de bandeo generados de los segmentos polimórficos	Patrón de bandeo dominante (frecuencia)	Desequilibrio genético
1	GGTGCGGAA	2 (590, 610)	22	1a* (590) 1b (610) 1c (590/610)	0.82	4c <sup>A</sup>
2	GTTTCGCTCC	1 (370)	16	2a* 2b (370)	0.79	4c <sup>A</sup>
3	GTAGACCCGT	1 (690)	15	3a* 3b (690)	0.98	4b
4	AAGAGCCCGT	2 (460, 480)	14	4a 4b (460) 4c* (460/480)	0.69	
5	AACGCGCAAC	0	12	5		
6	CCCGTCAGCA	0	12	6		

\* $p < 0.05$

Los patrones de bandas con mayor frecuencia están marcados con asterisco (\*) y se señala su correspondiente desequilibrio genético.

Como se observa, los patrones de bandas 1a, 2a y 3a se encuentran en desequilibrio genético con 4c y 4b.

Posteriormente, se exploró la posibilidad de usar a los perfiles RAPD como indicadores de la acumulación de quercetina (característica fenotípica). El análisis de correlación entre las combinaciones posibles de dos perfiles RAPD con mayor frecuencia y una alta acumulación de quercetina, fue evaluado usando el programa *SNPanalyzer*. Los resultados indicaron una relación entre los perfiles 1a y 2a con 4c, con valor F de 5.184 y 5.652, respectivamente y a una diferencia significativa de  $p < 0.05$ , con el aumento en la acumulación del principio activo. Por último, el valor LoD que representa la característica cuantitativa (QTL) arrojó un valor de 9.498, afirmando como la presencia de 1a y 2a con 4c como marcadores genéticos, se correlacionan con un aumento en la acumulación de quercetina en los árboles muestreados.

#### **VII.5 Identificación de secuencias similares y regiones repetidas.**

En la construcción de las bibliotecas genómicas se utilizaron los cuatro oligonucleótidos que generan segmentos polimórficos, el oligonucleótido 5 que no genera segmentos polimórficos y los fragmentos obtenidos por captura de híbridos. En total se obtuvieron 350 colonias, de las cuales se seleccionaron 60 con un tamaño de inserto diferente y de estos solo se obtuvieron 36 secuencias.

En el **ANEXO 6** se presentan las secuencias con tamaños entre 200 a 1200 pb, obtenidas con la amplificación aleatoria RAPD; así como la digestión con enzimas de restricción e hibridación con las sondas  $(CA)_{15}$  y GATA del grupo de muestras seleccionadas del **ANEXO 3**. La nomenclatura de las secuencias en los segmentos

RAPD a emplear es: el número del oligonucleótido utilizado (a la izquierda del punto), el número asignado en la caja de respaldo y la letra A si proviene de la repetición en la construcción de la biblioteca. La nomenclatura de los fragmentos con repetición en tandem o microsatelitales a emplear es: las iniciales SK o CAG para indicar la primera o segunda biblioteca y el número que corresponde a su ubicación en la caja de respaldo.

En la selección de las clonas a partir de una misma biblioteca genómica se observa que varias de ellas presentan un peso molecular muy cercano y en ocasiones la misma secuencia con mínimas diferencias. Tal es el caso de las clonas 1.33 y 1.2A con un peso molecular de 315 y 316 pb respectivamente, que sólo se diferencian por una adenina en la posición 13 de la clona 1.2A y que no se presenta en 1.33. En el caso de las clonas CAG.37 y CAG.38, se observa una gran similitud entre sus secuencias excepto por el cambio de una adenina en la posición 60, una supresión en la posición 62, un cambio de una citosina en la posición 537 y el cambio de una guanina en la posición 550. La mayor diferencia entre clonas similares, se observa en la región repetida de las clonas CAG.1 y SK2 con la ausencia de 14 nucleótidos en las posiciones 186 a la 200 en SK2, y en las clonas 4.38 y 4.3 donde se observa una región homóloga de 166 pb al inicio y final de su secuencia respectivamente. Como caso particular, se observa una región de 24 pb (TCTAAGGCCTTGCTAGCAGAAGCA) que se encuentra repetida al inicio y dentro de las secuencias SK5, SK2, CAG.28, CAG.1, CAG.20, CAG.7, CAG.37 y CAG.38, pero especialmente en CAG.4 ya que se repite hasta 5 veces.

Entre las secuencias con una misma longitud pero diferente lectura, se encuentra la clona 1.1 con un tamaño de 316 pb similar a las clonas 1.33 y 1.2A, así como las clonas 2.2A y 2.1 con tamaños de 210 y 215 pb respectivamente. Las secuencias con regiones repetidas no son exclusivas de las bibliotecas generadas por captura de híbridos, ya que también se presenta la región  $(CT)_{20}CC(CT)_4$  en la secuencia 1.5A. Por otro lado, las secuencias capturadas con la sonda biotinilada  $(GT)_{15}$  que presentan un microsatélite claramente identificable son SK2, CAG1, SK3 y SK4, corresponden a un 33% de las clonas secuenciadas (**Tabla 3**).

**Tabla 3** Secuencias obtenidas de las bibliotecas genómicas tipo RAPD y microsatelitales, número de secuencias similares e identificación de regiones tipo tandem.

Oligonucleótido	Biblioteca genómica	# Secuencias	Secuencias similares	Clonas con regiones tipo tandem
1	RAPD	8	1	1.5A, (CT) <sub>20</sub> CC(CT) <sub>4</sub>
2	RAPD	7	0	
3	RAPD	2	0	
4	RAPD	4	1	
5	RAPD	3	0	
(GT) <sub>n</sub>	Microsatelite	12	2	SK2, (CA) <sub>4</sub> AA(CA) <sub>21</sub> CAG.1, (CA) <sub>4</sub> AA(CA) <sub>26</sub> SK3, (CA) <sub>26</sub> SK4, (CA) <sub>14</sub>

## VII.6 Identificación de secuencias homólogas utilizando la base de datos

### Gene Bank de NCBI.

Las secuencias del **ANEXO 6** analizadas con la base de datos Gene Bank de NCBI para plantas, no mostraron una homología significativa a excepción del fragmento de la clona 1.1A con 549 pb. Este inserto muestra una homología del 91% en un fragmento de 229 pb del cromosoma 2 de *Arabidopsis thaliana*.

### VII.7 Diseño de oligonucleótidos específicos para la identificación de las clonas seleccionadas.

De las 36 secuencias presentes en el **ANEXO 6**, se descartó el inserto 1.1A por su homología con *A. thaliana*. De las 35 clonas restantes, solo es posible diseñar oligonucleótidos flanqueantes específicos (cebadores) para 24 secuencias.

Al comparar las 35 secuencias únicas y las regiones flanqueantes diseñadas para su amplificación, observamos que se obtuvieron 6 pares de cebadores de las 8 secuencias obtenidas con la biblioteca genómica RAPD 1, se obtuvo solo un par de cebadores de las 7 secuencias obtenidas con la biblioteca genómica RAPD 2, se obtuvieron 2 pares de cebadores de las 2 secuencias obtenidas con la biblioteca genómica RAPD 3, se obtuvieron 4 pares de cebadores de las 4 secuencias obtenidas con la biblioteca genómica RAPD 4, se obtuvieron 3 pares de cebadores de las 8 secuencias obtenidas con la biblioteca genómica RAPD 5 y se obtuvieron 6 pares de cebadores de las 12 secuencias obtenidas con la biblioteca genómica (GT)<sub>15</sub>. La lista de los 22 pares de cebadores seleccionados por su tamaño, temperatura de hibridación, contenido de GC, mínimo alineamiento entre cebadores y secuencia única tanto de la clona como del cebador, se presentan en el **ANEXO 7**.

### **VII.8 Estandarización en la amplificación de secuencias usando ADN genómico de *P. guajava* y oligonucleótidos específicos.**

La concentración de magnesio utilizada en futuros experimentos, se determinó mediante el diseño de una matriz donde se probaron concentraciones de 1.0, 1.5, 2.0 y 2.5 mM de MgCl<sub>2</sub>, 10 ng de ADN plasmídico proveniente de las diferentes clonas y los oligonucleótidos diseñados. En cada caso, se obtuvo el amplificado a la longitud esperada entre 1.0 a 2.0mM MgCl<sub>2</sub>.

La concentración mínima de ADN plasmídico y ADN genómico proveniente de material vegetal fresco y seco se determinó al emplear 1.0 mM MgCl<sub>2</sub>, con los oligonucleótidos diseñados previamente. La muestra E5 nuevamente es utilizada en esta etapa del trabajo.

Los resultados muestran que la amplificación de los plásmidos seleccionados se obtiene desde 10 pg y que la cantidad de ADN genómico necesaria en la amplificación de estos mismos segmentos puede variar de 1 – 20 ng. Finalmente observamos que son 13 las regiones de ADN genómico que se amplifican a partir de los 16 pares de cebadores obtenidos de las bibliotecas genómicas tipo RAPD (ANEXO 8).

### **VII.9 Validación de las secuencias en la identificación genotípica de *P. guajava*.**

Los criterios y condiciones de amplificación que se obtuvieron durante la estandarización se utilizaron para validar la amplificación de los segmentos seleccionados de las bibliotecas genómicas tipo RAPD. Los productos de amplificación se obtuvieron empleando 10 pg de ADN plasmídico, 10 ng de ADN

genómico con 1.0 mM MgCl<sub>2</sub>. El ADN utilizado corresponde de las muestras seleccionadas del **ANEXO 3**, también se incluyó el plásmido control y un árbol ancestral obtenido del primer huerto establecido en Calvillo, Aguascalientes (**ANEXO 9**).

Los productos de amplificación obtenidos no fueron homogéneos en ninguno de los casos, observándose ausencia o baja intensidad en las muestras. En el amplificado 1.2A fueron las muestras VII y XVI; en 1.12A las muestras III - VI, X, XV y XVI; en 1.3A las muestras VII y IX; en 3.15 las muestras V - VII, IX, X, XIII y XVI; en 4.1 las muestras III y IX; en 4.3 las muestras III, IV y VII; en 4.9 las muestras IV, VIII y XII; en 5.8 las muestras G1, I, III - IX, XI y XIV; y en 5.39 las muestras II - IX y XII.

Para el análisis estadístico, se empleó ANOVA de una vía para comparar los marcadores contra la región geográfica y la concentración de quercetina. El análisis se hizo cualitativamente (presencia o ausencia del amplificado) como cuantitativamente (densitometría de la banda). Los resultados indican que aparentemente el marcador 1.3A pudo ser empleado como un indicador diferencial entre las regiones con una  $p < 0.02$  en el nivel cuantitativo; el marcador 4.3 pudo identificar a individuos con una alta producción de quercetina puesto que a nivel cualitativo muestra una diferencia superior del 330 al 340%, con una  $p < 0.007$ ; el alelo 5.39 no tiene una diferencia cuantitativa con respecto al estado y la cantidad de quercetina cuantificable, sin embargo cualitativamente podría ser considerado como potencial marcador; el alelo 5.5 a nivel cuantitativo es el más representativo, ya que está relacionado tanto con la concentración de quercetina, como las condiciones ambientales ( $p < 0.05$ ); el marcador 3.15 que identifica 2 alelos en el genoma, a pesar

de tener una relación del 100% con respecto a la regionalización y el 280% en la concentración de quercetina, lo debemos considerar solamente en el aspecto cualitativo.

Posteriormente, a estas mismas muestras se modificaron las condiciones de amplificación a 5 ng de ADN genómico y 1.5 mM de  $MgCl_2$  (**ANEXO 10**).

Los resultados presentan una amplificación no homogénea entre las muestras aunque se presenten en un mayor número de ellas. Los amplificadores 1.12A, 5.8 y 5.39 se favorecen especialmente y en menor grado la región intergénica 5S rRNA y la región de ADN 1.11A. Las muestras que no amplificaron o que presentan una bajo producto en la región intergénica 5S rRNA fueron la G1, VI - IX, XV y XVI; en 1.12A la muestra VIII, XV y XVI; en 1.11A el ADN plasmídico III- IX, XII, XIV y XVI; en 3.15 las muestras G1, III, IV, VII, IX y XIV; y en 5.8 la muestra 5. La muestra XIV del segmento 4.9, la muestra XI de 5.5 y la muestra IX de 5.39 no se consideran porque la solución se evaporó durante la reacción.

El análisis estadístico ANOVA de una vía realizado bajo estas nuevas condiciones de amplificación, mostraron que los marcadores 1.11A, 3.15a y 3.15b pueden ser utilizados como indicadores diferenciales cualitativos y cuantitativos por la región y concentración de quercetina, sin embargo se ven afectados por la calidad del ADN empleado.

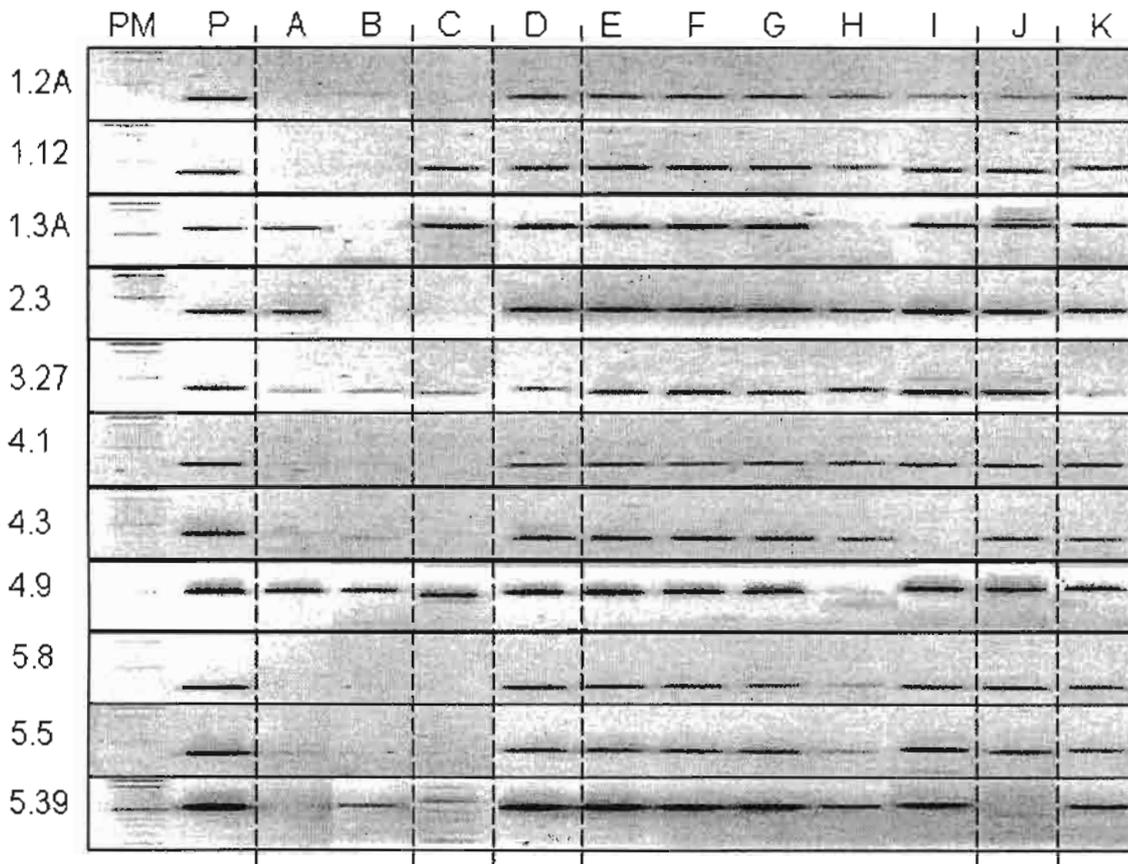
El dato más interesante obtenido en el análisis estadístico fue entre los valores densitométricos de los marcadores y su relación con la zona geográfica, donde los fragmentos 1.2A, 2.3, 4.3, 4.9, 5.5, 5.8 y 5.39 presentan una diferencia significativa sin que interfiera la calidad de material genético. Debido a lo anterior se decidió

hacer una tabla comparativa para conocer que marcadores eran representativos de la zona geográfica.

#### **VII.10 Identificación genotípica de *P. guajava*.**

Por último, se determinaron los segmentos específicos para autenticar con marcadores genéticos a la especie *P. guajava* L. y aquellos que puedan discriminar especies con características físicas semejantes o nombres populares similares.

La **Figura 5** muestra 11 de los 13 segmentos utilizados bajo diferentes concentraciones de ADN y  $MgCl_2$  durante la validación, excluyendo a 1.11A y 3.15 por su baja especificidad. Las muestras A y B se colectaron en Querétaro con el nombre popular de "guayabita" o "arrayán" e identificadas como *Psidium sartorium*, la muestra C fue obtenida de un árbol cultivado en la ciudad de México con el nombre popular de "guayaba brasileña" e identificada como *F. sellowiana*, la muestra D es una variedad michoacana de guayaba que se caracteriza por su forma tipo pomo, las muestras E - I fueron colectadas de diferentes árboles de guayabo silvestres del estado de Oaxaca; la muestra J fue obtenida de un árbol de guayabo cultivado de la ciudad de México y la muestra K se compró a la empresa (Hangzhou Singho Co.,Ltd.) de origen chino.



**Figura 5.** Amplificación de regiones específicas de ADN en *P. guajava* y su comparación con otras especies. PM: Marcador de 100pb (Invitrogen Life Technologies, USA), P: Plásmido, A y B: *P. sartorianum*, C: *F. sellowiana*, D: *P. guajava* (variedad), E - J: *Psidium* spp, K: *P. guajava* (origen Chino).

Los resultados muestran que los amplificadores 1.3A, 2.3, 3.27, 4.9 y 5.39 están presentes en *P. sartorianum* y *F. sellowiana* con bandas claras y definidas. De forma particular, *F. sellowiana* presenta un amplificado de menor longitud (500 pb) en 4.9 e inespecificidad en 5.39.

Los amplificadores 1.2A, 4.1, 4.3 y 5.5 están presentes en muestras de *P. sartorianum* como una banda muy tenue, en comparación a las muestras de *P. guajava*.

En las muestras D, E, F, G y K se amplificaron todos los segmentos de ADN; en la muestra H no se aprecia 1.3A y en la muestra I el amplificado 4.3.

De manera interesante, la muestra J proveniente de un árbol cultivado en la ciudad de México aparecen varias bandas en los amplificadores 1.3A y 5.39.

## VIII DISCUSIÓN.

### VIII.1 Selección de la materia prima.

En la obtención de la materia prima fue necesario seleccionar los árboles y lugares de recolecta para el desarrollo de este trabajo.

La decisión de muestrear árboles en cultivos comerciales de guayabo se tomó por dos factores principales. La primera fue la importancia agroindustrial que existe en México en la explotación del fruto de guayaba, lo que deja a la hoja como un subproducto de desecho potencialmente aprovechable. La segunda fue establecer las mejores condiciones de recolecta en cultivos controlados con el propósito de desarrollar un fitomedicamento a partir del extracto etanólico de hoja de guayaba, para el tratamiento del síndrome del colon irritable.

La selección de los estados representativos fue a través de la SAGARPA y su portal en Internet, ya que actualmente existe la ventaja de utilizar este medio como fuente de información. De esta forma se estableció que los árboles de los estados de Aguascalientes, Michoacán, México y Querétaro representaron en promedio el 79% de la producción nacional de guayaba, cinco años previos al muestreo. Cada estado presenta condiciones ambientales y un manejo agronómico diferente. Aguascalientes y Querétaro tienen un manejo orgánico, no usan fertilizantes e insecticidas químicos, y generan estrés hídrico a sus árboles; Querétaro en menor grado que Aguascalientes. Los estados de México y Michoacán, por su parte utilizan fertilizantes e insecticidas químicos en sus cultivos, y cuentan con sistema de riego.

### VIII.2 Cuantificación del principio activo

Los flavonoides además de su función en la planta (regulan el metabolismo de las vitaminas liposolubles y hormonas, favorecen la polinización, etc.), tienen propiedades antioxidantes y antiinflamatorias que se ven reflejadas en el tratamiento de diversos padecimientos (hipertensión, diabetes, problemas cardiovasculares, etc.) (Havsteen, 2002). Lo anterior justifica el interés de estos metabolitos secundarios en el campo de la farmacéutica y su aplicación en la medicina.

Tomando en cuenta las funciones fisiológicas de los flavonoides en la planta, se realizó el muestreo durante los meses con mayor incidencia de rayos solares, debido a su poder antioxidante y captura de radicales libres; durante el proceso de floración y fructificación, debido a su papel durante la polinización y desarrollo del fruto; y post cosecha, como un proceso inherente de acumulación. Al término del año, se observó que la mayor acumulación de quercetina se obtuvo en la etapa pos-cosecha y en los meses con una alta incidencia de rayos solares. En la comparación de resultados, obtuvimos el mismo comportamiento que Vargas-Alvarez y cols. (2006). Ellos reportan que la mayor acumulación de flavonoides (miricetina, quercetina, luteolina, kaempferol) en hojas maduras de guayaba y en árboles del estado de México, se presenta en Mayo y Julio del 2000 o para nosotros primavera y verano.

El siguiente dato interesante fue la selección de la zona geográfica. Los agricultores compran árboles propagados por estacas al municipio de Calvillo en Aguascalientes, por lo que se considera un origen común y relativamente homogéneo entre los cultivos.

Michoacán no presentó diferencias significativas a lo largo del año. Querétaro, aunque presentó el doble de acumulación de quercetina con respecto a Michoacán durante el año, la acumulación de quercetina se mostró lineal y poco atractiva. Aguascalientes y estado de México mostraron una cinética de acumulación esperada y altas concentraciones quercetina, Vargas-Alvarez y cols. (2006) reportan 2.9 mg/g de hoja seca como la mayor acumulación de quercetina total, mientras que nosotros obtuvimos valores 3 veces más de lo reportado, probablemente por el método de extracción exhaustiva y la hidrólisis total de flavonoides que se empleó en este trabajo para la cuantificación de quercetina. Sin embargo, para los intereses posteriores de este trabajo, entre los 2 estados se considera como mejor opción Aguascalientes por el manejo orgánico de sus cultivos, y para evitar agentes químicos tóxicos en el extracto y el fitomedicamento.

Por último, a pesar de la gran influencia que presentan condiciones ambientales en la acumulación de flavonoides; el análisis individual de las hojas dentro de los mismos cultivares, nos permitió indagar sobre la influencia de factores genéticos propiciados por los mismos agricultores.

### **VIII.3 Selección de los individuos para el estudio genético en base a la acumulación de flavonoides.**

Durante la primera visita, se observó que los agricultores de Aguascalientes trabajan continuamente en el mejoramiento de sus árboles, además de la explotación comercial de dos tipos de guayaba (media china y pulpa blanca) en la misma zona de cultivo. En referencia a la selección de cultivos por parte de los agricultores y las diferencias entre los frutos, Padilla-Ramírez y cols. (2002) hacen una comparación

genética de 12 árboles de guayaba. Estos ejemplares, fueron seleccionados por la calidad y rendimiento de sus frutos (característica fenotípica), para obtener un banco con 45 accesiones de la región Calvillo, Aguascalientes – Cañones, Zacatecas y realizar un agrupamiento por homología genética.

En el presente trabajo, la característica fenotípica fue la acumulación de quercetina presente en las hojas, seleccionandose 13 árboles de un grupo de 40. El análisis estadístico de normalidad, permitió identificar a los individuos con valores mínimos y máximos que posteriormente fueron correlacionadas con las diferencias genéticas.

#### **VIII.4 Identificación de segmentos polimórficos en las muestras colectadas usando amplificación aleatoria RAPD.**

La amplificación aleatoria y controlada por la técnica RAPD nos permite tener un primer acercamiento con las diferencias genotípicas presentes entre los individuos. En el análisis de todas las muestras con esta técnica y solo 6 oligonucleótidos decámeros, nos permitió identificar regiones polimórficas y monomórficas con un rango de homogeneidad entre el 74 al 94%, más amplio que el reportado por Padilla-Ramírez y cols (2002) de 88 al 96%. Los resultados muestran que el estado de México presenta la mayor homogeneidad genética entre sus individuos, a diferencia de los estados de Michoacán y Aguascalientes que presentan una mayor heterogeneidad por estar presentes todos o la mayoría de los patrones genéticos. Este fenómeno se puede deducir por el continuo mejoramiento de la especie, que los agricultores propician para lograr un fruto con características atractivas para el consumidor y al mismo tiempo obtener una mejor adaptación al medio ambiente que se ve reflejado en diferencias genotípicas.

En el análisis directo entre los individuos con mayor y menor acumulación de quercetina total, observamos que en Aguascalientes el árbol A1 con una baja concentración de quercetina, presentó un patrón genético diferente a A9; lo mismo se presentó en los demás estados. Esta observación preliminar, nos motivó a buscar alguna estrategia que permitiera correlacionar a los segmentos de ADN polimórficos con la propiedad fenotípica deseada. El análisis de resultados obtenido con el programa de cómputo *SNPanalyzer*, permitió obtener numéricamente la relación entre el patrón de bandeo o variación polimórfica y la acumulación de quercetina total.

#### **VIII.5 Identificación de secuencias similares y regiones repetidas.**

En el análisis individual de las secuencias se observan diferencias en la lectura que van desde el cambio de una base y corrimiento por eliminación o inserción de una base, principalmente. Sin embargo, no hay que descartar la posibilidad de errores de lectura durante el corrimiento de la muestra en el secuenciador. Es así como, la secuencia 1.5A que presenta la región  $(CT)_{20}CC(CT)_4$ , se puede pensar en un microsatélite  $(CT)_{25}$ . Sin embargo en el caso de las clonas SK2 y CAG.1, la región  $(CA)_n$  se ve interrumpida por una adenina en el mismo lugar, lo que sugiere fuertemente que la lectura de la secuencia es correcta. El diseño adecuado de esta región nos permitirá utilizarlo como un marcador que diferencie formas polimórficas y presencia o ausencia de heterocigotos, debido a que las regiones flanqueantes de los microsatélites son altamente conservadas y altamente polimórficos con repeticiones mayores de 12.

Como caso particular, se observó que una región de 24 pb (TCTAAGGCCTTGCT

AGCAGAAGCA) que se encuentra repetida al inicio y dentro de las secuencias SK5, SK2, CAG.28, CAG.1, CAG.20, CAG.7, CAG.37 y CAG.38 y especialmente en CAG.4 hasta 5 veces. Debido a la recurrencia de esta región, se buscó su homología en Gene Bank NCBI y se encontró una relación directa con secuencias microsatelitales. De manera particular esta región se encuentra flanqueando un microsatélite (GA)<sub>16</sub> presente en *Magnifera indica* (mango). Aunque esta región es importante para la identificación de microsatélites, no se puede utilizar como oligonucleótido específico en la identificación de la frecuencia alélica debido a su recurrencia en diferentes regiones del ADN genómico, sin embargo al realizar mapeo físico de comparación con mango, el diseño de primers permitirá si hay o no amplificación y por consiguiente determinar distribución en especies, géneros e incluso hasta familias en la escala taxonómica.

#### **VIII.6 Identificación de secuencias homólogas utilizando la base de datos Gene Bank de NCBI.**

La obtención de secuencias únicas de la especie, nos permite garantizar su autenticación y la discriminación con otras especies similares físicamente sin la intervención de un taxónomo o morfológico profesional. Esta forma de identificación genética también cobra importancia cuando el material vegetal no cuenta con los requisitos de un ejemplar de herbario, porque solo se presenta una parte de la planta o porque ha sufrido algún procesamiento post cosecha para su comercialización.

En un primer análisis, se empleó la base de datos nucleotídica de SWBIC y NCBI para el alineamiento de las secuencias, por el interés de identificar regiones codificantes en bibliotecas de ADNc. En todos los casos se encontró una homología

en los segmentos entre 10 a 50 pb; sin embargo la relación entre el segmento homólogo con el segmento obtenido no se considera significativamente similar, descartando la posibilidad de haber obtenido secuencias codificantes.

Por otro lado, el alineamiento de las secuencias contra las bibliotecas genómicas de 33 plantas diferentes, nos proporcionó una información más completa para determinar su autenticidad como secuencias únicas de *P. guajava*.

De esta forma se pudo excluir a la clona 1.1A, por su homología en 229 pb con el cromosoma 2 de *Arabidopsis thaliana* en un 91% de sus secuencia y por otro lado, incluir a las demás secuencias en el diseño de oligonucleótidos flanqueantes para el diseño de marcadores específicos.

#### **VIII.7 Diseño de oligonucleótidos específicos para la identificación de las clonas seleccionadas.**

El diseño de cebadores se considera el paso más importante en la obtención de marcadores, ya que de ellos depende la identificación y amplificación de las secuencias. Para su diseño, es importante tomar en cuenta un tamaño entre 18 a 22 pb para garantizar la especificidad en la amplificación; permitir una temperatura de hibridación menor a la de amplificación (72°C) o sea entre 50 a 55°C, para evitar una amplificación prematura; un contenido de GC de más del 50% para facilitar la hibridación; evitar un alineamiento entre ellos, para que no formen dímeros durante la amplificación y por supuesto que tengan secuencia única (Rychlik y Rhoads, 1989). El programa de cómputo *Web Primer DNA* se utilizó en su diseño, ya que permite tomar en cuenta todos estos aspectos.

En este paso también se hizo una selección de secuencias, ya que solo 24 de las 35 clonas restantes cumplieron satisfactoriamente con los parámetros descritos anteriormente para el diseño de oligonucleótidos flanqueantes. Si consideramos además, que dos de las secuencias están repetidas, solo quedan 22 pares de oligonucleótidos para establecer las condiciones de amplificación y utilizarlos en estudios posteriores.

En el caso particular de las secuencias 4.38 y 4.3, se presenta una región homóloga de 166 pb, es por ello que el diseño de uno de los oligonucleótidos es el mismo.

#### **VIII.8 Estandarización en la amplificación por PCR de *P. guajava* empleando oligonucleótidos específicos.**

Las secuencias que se utilizaron provienen de las bibliotecas tipo RAPD, con carácter conservado en la muestra de origen (E5). Es así como, la información que se generó nos permitió determinar las diferentes regiones conservadas a las obtenidas entre el árbol del estado de México y las demás muestras; lo que permite establecer marcadores únicos de especie; y discriminar posibles adulterantes.

En el establecimiento de las condiciones óptimas para la amplificación por PCR, es común probar diferentes concentraciones de cloruro de magnesio, debido a la propiedad del magnesio como coenzima. La abundancia o escasez de este compuesto, facilita o limita la actividad enzimática de la *Taq DNA polimerasa* (Rychlik, 1995). En un primer paso, se obtuvo el amplificado entre concentraciones de 1.0 a 2.0mM MgCl<sub>2</sub> (datos no mostrados), por lo que se seleccionó la menor concentración para permitir una amplificación más dirigida.

El siguiente aspecto que se considera en la PCR, es la temperatura de hibridación. Los oligonucleótidos diseñados a una temperatura de hibridación de 50°C, no presentaron problemas de múltiple bandeo o ausencia de amplificado a esta temperatura, por lo que se continuó con esta temperatura.

Por último fue importante establecer los rangos de concentración de ADN plasmídico y genómico para la amplificación, debido a que no se lleva a cabo la reacción con altas concentraciones de material genético. Por otro lado, también se considera el estado de la hoja (fresca o seca); debido a que en la práctica la adquisición de este tipo de material vegetal es seca y en trozos muy pequeños porque la concentración del principio activo no se modifica bajo estas condiciones. Aunque las concentraciones variaron dependiendo del fragmento, en general se logró una buena amplificación con 10 pg de ADN plasmídico y 10 ng de ADN genómico.

De forma complementaria, se diseñaron los cebadores de la región intergénica ARNr 5S, ya que ha sido utilizado en la identificación genotípica de diferentes especies (Cheng y cols 1997, 2000 y 2001), y que al mismo tiempo fuera un control interno en la identificación de las muestras de guayaba. Sin embargo el diseño de este fragmento mostró una inespecificidad no deseada para los propósitos del trabajo.

#### **VIII.9 Validación de las secuencias en la identificación genotípica de *P.***

##### ***guajava.***

Para la validación de un marcador cuya aplicación será la identificación de especie, se requiere su presencia en un número considerable de individuos. Los árboles que se emplearon en este experimento es una muestra representativa de diferentes regiones de la República Mexicana con variaciones considerables en su propiedad

fenotípica (**ANEXO 2**). En este trabajo no todas las secuencias con diseño de cebadores, amplificaron en todos los individuos, razón por la cual se excluyeron desde el inicio como regiones especie-específicas. Estas diferencias se pueden explicar desde un punto de vista polimórfico, ya que al no tener una referencia previa del fragmento seleccionado, esto puede repercutir en el diseño de cebadores; o también que las condiciones de amplificación no sean las más adecuadas.

#### **VIII.10 Identificación genotípica de *P. guajava*.**

La escasa información genética que existe sobre otras plantas, nos conduce a aplicar la amplificación de fragmentos en especies que pueden confundirse con guayaba.

En este experimento, se pudo corroborar la presencia de algunas de las regiones de ADN seleccionadas en otras dos especies diferentes a *P. guajava*, pero con características similares. Sin embargo, es clara la diferencia entre los patrones de bandeo obtenidos entre las 3 especies.

Adicionalmente, también observamos como estos amplificados no fueron capaces de diferenciar claramente entre variedades de guayaba, sin embargo podemos dudar de la especie de las muestras H, I y J por presentarse una alteración evidente en los amplificados.

Por último, estos resultados nos permitieron confirmar que la muestra de origen Chino si se comporta como *P. guajava* por lo que se decidió emplear la técnica de cuantificación de quercetina total para apoyar este resultado.

## IX CONCLUSIONES

Las hojas de *P. guajava* que fueron evaluadas como materia prima para la elaboración del fitomedicamento, se les determinó el contenido de quercetina total. De las cuatro regiones seleccionadas, se identificó a Calvillo Aguascalientes como una región geográfica, económicamente viable y con un manejo biológico, durante la primavera y una etapa pos cosecha con las condiciones ambientales adecuadas para la acumulación del principio activo, lo que la hace susceptible a ser seleccionada para la producción del fitomedicamento.

Como complemento a las técnicas de control de calidad, la amplificación aleatoria tipo RAPD nos permitió identificar patrones conservados de las regiones polimórficas y posteriormente correlacionarlo con la acumulación de quercetina. Por lo anterior, la presencia de los segmentos polimórficos 1a y 2a con 4c que se correlacionaron con un aumento del principio activo al obtener un valor LoD de 9.498 ( $p < 0.05$ ), permiten identificar genéticamente individuos con una materia prima de buena calidad.

En el análisis de las 36 secuencias obtenidas se encontraron dos repetidas, otra fue descartada por su similitud con *A. thaliana* y solo 16 fragmentos cumplieron las condiciones necesarias para el diseño de oligonucleótidos específicos. Una información útil para determinar cercanía filogenética usando regiones homólogas con otras especies o la autenticación de especie al amplificar secuencias diferentes. Se observó que una alta concentración de  $MgCl_2$  (1.5 mM) con una baja cantidad de ADN genómico (5 ng) favorece la amplificación a excepción de los fragmentos 1.11A y 3.15.

La biblioteca rica en regiones dinucleótidas tandem construida por el fraccionamiento e hibridación con sondas específicas, permitió la obtención de 12 secuencias de las cuales 2 fueron repetidas; 5 secuencias presentaron el microsatélite (CA)<sub>n</sub>; 9 secuencias presentaron una región de 24 pb (5'-TCTAAGGCCTTGCTAGCAGAAGCA-3') flanqueando o dentro de la secuencia que también es característica de microsatélites (GA)<sub>n</sub> de mango (*Mangifera indica*); se diseñaron marcadores específicos para 6 secuencias, sin lograr establecer las condiciones de amplificación de estas.

La caracterización, secuenciación y diseño de los 5 pares de oligonucleótidos específicos 1.2A, 4.1, 4.3, 5.8 y 5.5, permitieron discriminar a la especie *P. guajava* de las especies *P. sartorium* y *F. sellowiana*, morfológicamente similares a guayaba.

## BIBLIOGRAFÍA.

1. Abdel-Ghany AGA y Zaki A. 2003. DNA sequences of RAPD fragments in the Egyptian cotton *Gossypium barbadense*. *Afr J Biotechnol*, 2(5): 129-32.
2. Amarger V y Mercier L. 1995. Molecular analysis of RAPD DNA based markers: their potential use for the detection of genetic variability in jojoba (*Simmondsia chinensis* L Schneider). *Biochimie*, 77(12): 931-6.
3. Antoni J. 1997. Randomly amplified polymorphic DNA (RAPD) analysis. En: Caetano-Anollés G. y Gresshoff PM. (eds.), DNA Markers. Protocols, applications, and overviews. WILEY-LISS, New York, pp. 75-83.
4. Arima H, Danno G. 2002. Isolation of antimicrobial compounds from guava (*Psidium guajava* L.) and their structural elucidation. *Biosci Biotechnol Biochem*, 66(8): 1727-30.
5. Babbar SB y Gupta SC. 1986. Induction of androgenesis and callus formation *in vitro* cultured anthers of a myrtaceous fruit tree (*Psidium guajava* L.). *Bot Magazine*, 99: 75-85.
6. Barnes J. Herbal therapeutics. 2002. (1) An introduction to herbal medicinal products. *Pharmaceutical j*, 268: 804-6.
7. Becerra V y Paredes CM. 2000. Uso de marcadores bioquímicos y moleculares en estudios de diversidad genética. *Agricultura técnica*, 60(3): 270-81.
8. Begum S, Hassan SI, Siddiqui BS, Shaheen F, Nabeel Ghayur M, Gilani AH (2002) Triterpenoids from the leaves of *Psidium guajava*. *Phytochemistry*, 61(4): 399-403.

9. Begum S, Hassan SI, Siddiqui BS. 2002. Two new triterpenoids from the fresh leaves of *Psidium guajava*. *Planta Med*, 68(12): 1149-52.
10. Begum S, Siddiqui BS, Hassan SI. 2002. Triterpenoids from *Psidium guajava* leaves. *Nat Prod Lett*, 16(3):173-7.
11. Bell CJ, Ecker JR. 1994. Assignment of 30 microsatellite *loci* to the linkage map of *Arabidopsis*. *Genomics*, 19:137-44.
12. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GeneBank: update. *Nucleic Acids Res*. 2004; 32(Database issue):D23-6.
13. Buchanan BB, Gruissem W, Jones RL. 2000. Chapter 6 Nucleic Acids. En: *Biochemistry & Molecular Biology of Plants*. American Society of Plant Physiologists. Rockville, Maryland. pp. 260-310.
14. Cai ZH, Li P, Dong TTX, Tsim KWK. 1999. Molecular diversity of 5S-rRNA spacer domain in *Fritillaria* species revealed by PCR analysis. *Planta Med*, 65: 360-364.
15. Cheng JT y Yang RS. 1983. Hypoglycemic effect of guava juice in mice and human subjects. *Am J Chin Med*, 11(1-4): 74-6.
16. Cheng KT, Chang HC, Su CH, Hsu FL. 1997. Identification of dried rhizomes of *Coptis* species using random amplified polymorphic DNA. *Bot Bull Acad Sin*, 38: 241-4.
17. Cheng KT, Su B, Chen CT, Lin CC. 2000. RAPD Analysis of *Astragalus* medicines marketed in Taiwan. *Am J Chin Med*, 28(2): 273-8.

18. Cheng KT, Su YH, Lin JG, Hsin LH, Su YP, Su CH, Li SY, Cheng JH, Mao SJT. 2001. Identification of *Atractylodes* plants in Chinese herbs and formulations by random amplified polymorphic DNA. *Acta Pharmacol Sin*, 22(6): 493-7.
19. Conde Garcia EA, Nascimento VT, Santiago Santos AB. 2003. Inotropic effects of extracts of *Psidium guajava* L. (guava) leaves on the guinea pig atrium. *Braz J Med Biol Res*, 36(5): 661-8.
20. Drasar P y Moravcova J. 2004. Recent advances in analysis of Chinese medical plants and traditional medicines. *J Chromatogr B Analyt Technol Biomed Life Sci.*, 812: 3-21.
21. Excoffier, L. y M. Slatkin. 1995. Maximum likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.* 12: 921-7.
22. Friar EA. 2005. Isolation of DNA from plants with large amounts of secondary metabolites. *Methods Enzymol*, 395: 3-14.
23. Gan S., Shi J., Li M., Wu K., Wu J. y Bai J. 2003. Moderate-density molecular maps of *Eucalyptus urophylla* S. T. Blake and *E. tereticornis* Smith genomes based on RAPD markers. *Genetica*, 118(1): 59-67.
24. Goldstein DB y Pollock DD. 1997. Launching microsatellites: a review of process and methods of phylogenetic inference. *J Hered*, 88(5): 335-42.
25. Gonzalez-Diaz H, Aguero-Chapin G, Varona-Santos J, Molina R, de la Riva G, Uriarte E. 2005. 2D RNA-QSAR: assigning ACC oxidase family membership with stochastic molecular descriptors; isolation and prediction of a sequence from *Psidium guajava* L. *Bioorg Med Chem Lett*, 15(11): 2932-7.

26. González E, Padilla JS, Reyes L, Perales MA, Esquivel F. 2002. "Guayaba. Su Cultivo en México. Libro Técnico No. 1" Editorial CEPAB.
27. Hakki E, Peev C, Karagyozev L, Akkaya M. 2002. Isolation of wheat microsatellite DNA fragments by hybridization selection. *Bulg J Plant Physiol*, 28(1-2): 3-10.
28. Harkey MR, Henderson GL, Gershwin ME, Stern JS, Hackman RM. 2002. Variability in commercial ginseng products: an analysis of 25 preparations. *Am J Clin Nutr*. 75(3):600-1.
29. Havsteen BH. 2002. The biochemistry and medical significance of the flavonoids. *Pharmacol Ther*. 96(2-3): 67-202.
30. Heckenberger M, Bohn M, Frisch M, Maurer HP, Melchinger AE. 2005. Identification of essentially derived varieties with molecular markers: an approach based on statistical test theory and computer simulations. *Theor Appl Genet*, 111(3): 598-608.
31. Hoffmann-Sommergruber K, Vanek-Krebitz M, Radauer C, Wen J, Ferreira F, Scheiner O, Breiteneder H. 1997. Genomic characterization of members of the Bet v 1 family: genes coding for allergens and pathogenesis-related proteins share intron positions. *Gene*, 197 (1-2): 91-100.
32. Hosokawa K, Minami M, Kawahara K, Nakamura I, Shibata T. 2000. Discrimination among three species of medicinal *Scutellaria* plants using RAPD markers. *Planta Med*, 66: 270-2.
33. Huang WH y Song CQ 2001. A survey on chemistry and pharmacology of Chinese *Angelica* root. *China Journal Chinese Material Medicine*, 26: 147-51.

34. Idstein H, Bauer C, Schreier P. 1985. Volatile acids in tropical fruits: cherimoya (*Annona cherimolia*, Mill.), guava (*Psidium guajava*, L.), mango (*Mangifera indica*, L., var. Alphonso), papaya (*Carica papaya*, L.). *Z Lebensm Unters Forsch*, 180(5): 394-7.
35. Jaiarj P, Khoohaswan P, Wongkrajang Y, Peungvicha P, Suriyawong P, Saraya ML, Ruangsomboon O. 1999. Anticough and antimicrobial activities of *Psidium guajava* Linn. leaf extract. *J Ethnopharmacol*, 67(2): 203-12.
36. Lewin B. 2000. From genes to genomes. En: *Genes VII*. Oxford University Press, New York, pp. 37-79.
37. Lewonting, R. 1964. The interaction of selection and linkage II. Optimum models. *Genetics*. 30: 757-82.
38. Lewonting, R. 1995. The detection of linkage disequilibrium in molecular sequence data. *Genetics*. 140: 377-88.
39. Liu ZW, Biyasheu RM, Saghai Maroof MA. 1996. Development of simple sequence repeat markers and their integration into a barley linkage map. *Theor Appl Genet*, 93: 869-76.
40. Lozoya X, Becerril G, Martínez M. 1990. Intraluminal perfusion model of in vitro guinea pig ileum as a model of study of the antidiarrheic properties of la guava *Psidium guajava*. *Arch Med Res*, 21: 155-162.
41. Lozoya X, Meckes M, Abou-Zaid M, Tortoriello J, Nozzolillo C, Thor J. 1994. Quercetin glycosides in *Psidium guajava* L. leaves and determination of a Spsmolytic Principle. *Arch Med Res*, 25(1): 11-5.

42. Madhavi DL, Chand N, Rajalakshmi D, Patwardham MV. 1992. Effects of growth hormones and maturity of the fruits on the callus culture of guava (*Psidium guajava*) fruits using response surface methodology. *J Sci Food Agric*, 58(1): 29-34.
43. Mercadante AZ, Steck A, Pfander H. 1999. Carotenoids from guava (*Psidium guajava* L.): isolation and structure elucidation. *J Agric Food Chem*, 47(1): 145-51.
44. Padilla-Ramírez JS, González-Gaona E, Esquivel-Villagrana F, Mercado-Silva E, Hernández-Delgado S, Mayek-Pérez N. 2002. Caracterización de germoplasma sobresaliente de guayabo de la región Calvillo-Cañones, México. *Rev Fitotec Mex*, 25(4): 393-9.
45. Pecetti L, Piano E, Romani M. 2004. Effect of various plant factors on the expression of creeping-rootedness in lucerne (*Medicago sativa* L. complex). *Euphytica*, 135(1): 89-97.
46. Prakash PD, Narayanaswamy P, Sondar SN. 2002. Analysis of molecular diversity in guava using RAPD markers. *J Hort Sci Biotechnol*, 77(3): 287-93.
47. Rakoczy-Trojanowska M, Bolibok H. 2004. Characteristics and a comparison of three classes of microsatellite-based markers and their application in plants. *Cellular & molecular biology letters*, 9: 221-238.
48. Rafalski JA y Tingey SV. 1993. Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. *Trends Genet*, 9(8): 275-80.

49. Roder MS, Plaschke J, König SU, Börner A, Sorrells ME, Tanksley SD, Ganai MW. 1995. Abundance, variability and chromosomal location of microsatellites in wheat. *Mol Gen Genet*, 246: 327-33.
50. Romero-Cerecero O, Reyes-Morales H, Herrera-Arellano A, Lozoya-Legorreta X, Tortoriello J. 2004. Aceptación de los fitofármacos por médicos y pacientes en clínicas de atención primaria. *Revista Médica IMSS*, 42(2):125-130.
51. Rivera-Arce E, Gattuso M, Lozoya X. 2003. Anatomical identity parameters of the crude drug *Psidium guajavae folium*. *Pharmaceutical Biology*, 41(7): 516-21.
52. Ruzin SE. 1999. Infiltrating and Embedding Tissues. En: *Plant microtechnique and microscopy*, Oxford University Press, New York, pp. 61-85.
53. Rychlik W. 1995. Priming efficiency in PCR. *Biotechniques*, 18(1): 84-6, 88-90.
54. Rychlik W, Rhoads RE. 1989. A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. *Nucleic Acids Res*; 17(21): 8543-51.
55. Schaneberg BT, Crockett S, Bedir E, Khan IA. 2003. The role of chemical fingerprinting: aplicación to Ephedra. *Phytochemistry*, 62: 911-8.
56. Singer M y Berg P. 1993. Las bases moleculares de la herencia: una visión general. En: *Genes y genomas*. Ediciones Omega, S.A., Barcelona, pp. 2-72.
57. SSA. Extrafarmacopea Herbolaria. 2001 En: *Farmacopea herbolaria de los Estados Unidos Mexicanos*. Secretaría de Salud. FEUM, México, pp. 27-69.
58. SSA. Métodos generales de análisis de la Farmacopea Herbolaria. 2001 En: *Farmacopea herbolaria de los Estados Unidos Mexicanos*. Secretaría de Salud. FEUM, México, pp. 27-69.

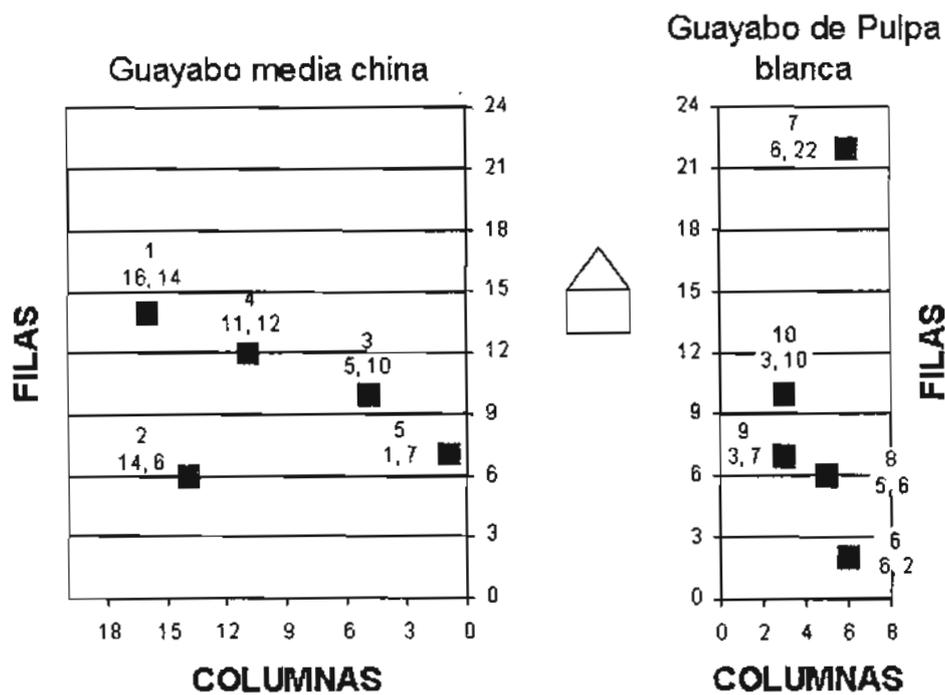
59. Tel-Zur N, Abbo S, Myslabodski D, Mizrahi Y. 1999. Modified CTAB procedure for DNA isolation from Epiphytic cacti of the genera *Hylocereus* and *Selenicereus* (Cactaceae). *Plant Mol Biol Rep*, 17: 249-54.
60. Tijet N, Waspi U, Gaskin DJ, Hunziker P, Muller BL, Vulfson EN, Slusarenko A, Brash AR, Whitehead IM. 2000. Purification, molecular cloning, and expression of the gene encoding fatty acid 13-hydroperoxide lyase from guava fruit (*Psidium guajava*). *Lipids*, 35(7): 709-20
61. Toth G, Gaspari Z, Jurka J. 2000. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res*, 10(7): 967-81.
62. Vanaclocha B, Cañigüeral S, Vila R, Risco E, Pérez F, Portillo A, Freixa B, Milo B, Alonso JR, Rios JL, Morales MA, Bachiller LI. 2003. Monografías de plantas medicinales. En: Vanaclocha B y Cañigüeral S. (eds.), *Fitoterapia. Vademécum de prescripción*. 4ª ed. MASSON, S.A., Barcelona, pp. 81-511.
63. Vargas-Alvarez D, Soto-Hernández M, González-Hernández VA, Engleman MA, Martínez-Garza Á. 2006. Cinética de acumulación y distribución de flavonoides en guayaba (*Psidium guajava* L.). *Agrociencia* 40: 109-115.
64. Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, Lathrop M. (1992). A second generation linkage map of human genome. *Nature*, 359: 794-801.
65. Welsh J., McClelland M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res*, 18(24): 7213-8.
66. WHO. WHO monographs on selected medicinal plants. 1999. En: World Health Organization, Geneva, vol 1 y 2, pp. 16-285 y 12-349.

67. Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res*, 18(22): 6531-5.
68. Wu KS y Tanksley SD. 1993. Abundance, polymorphisms and genetic mapping of microsatellites in rice. *Mol Gen Genet*, 241: 225-35.
69. Wulff EG, Torres S, Gonzáles VE. 2002. Protocol for DNA extraction from potato tubers. *Plant Mol Biol Rep*, 20: 187a-187e.

## ANEXO 1

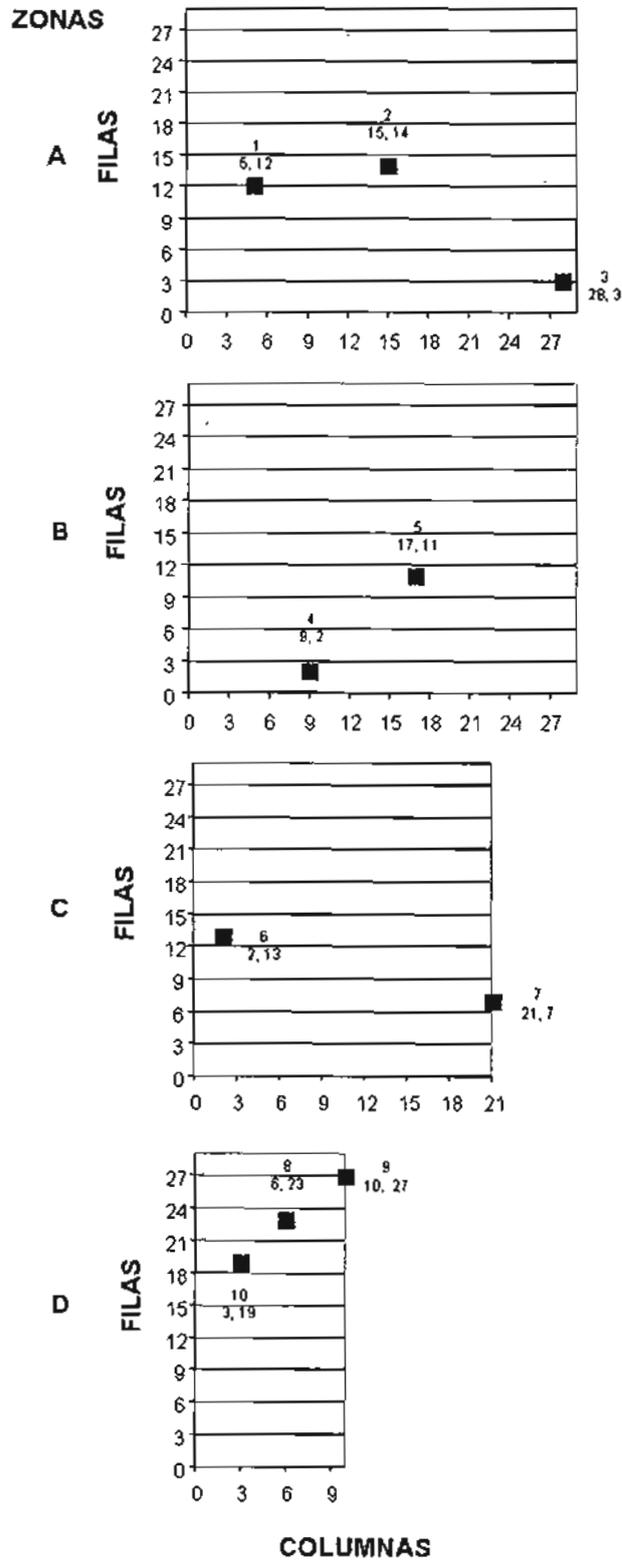
Mapas de los cultivos de guayaba con el número de árbol y coordenadas en los estados y municipios seleccionados.

Calvillo, Aguascalientes  
Huerta La Labor

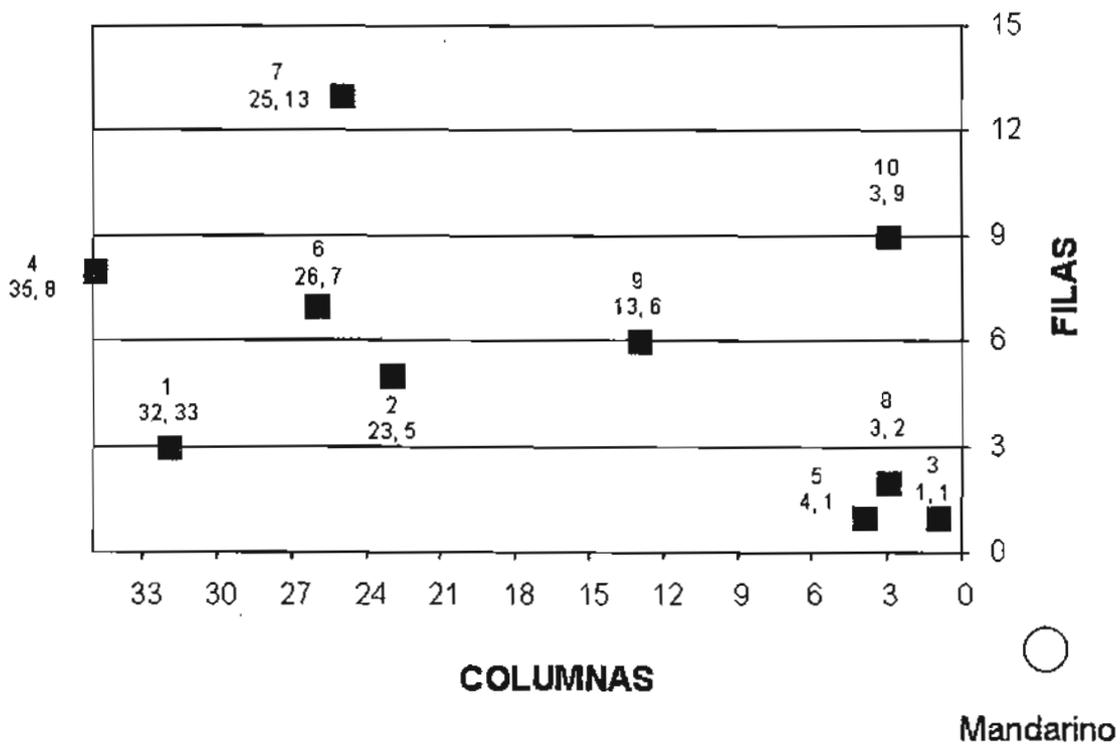


\* Nota: Las coordenadas se cuentan sin considerar árboles de otra especie.

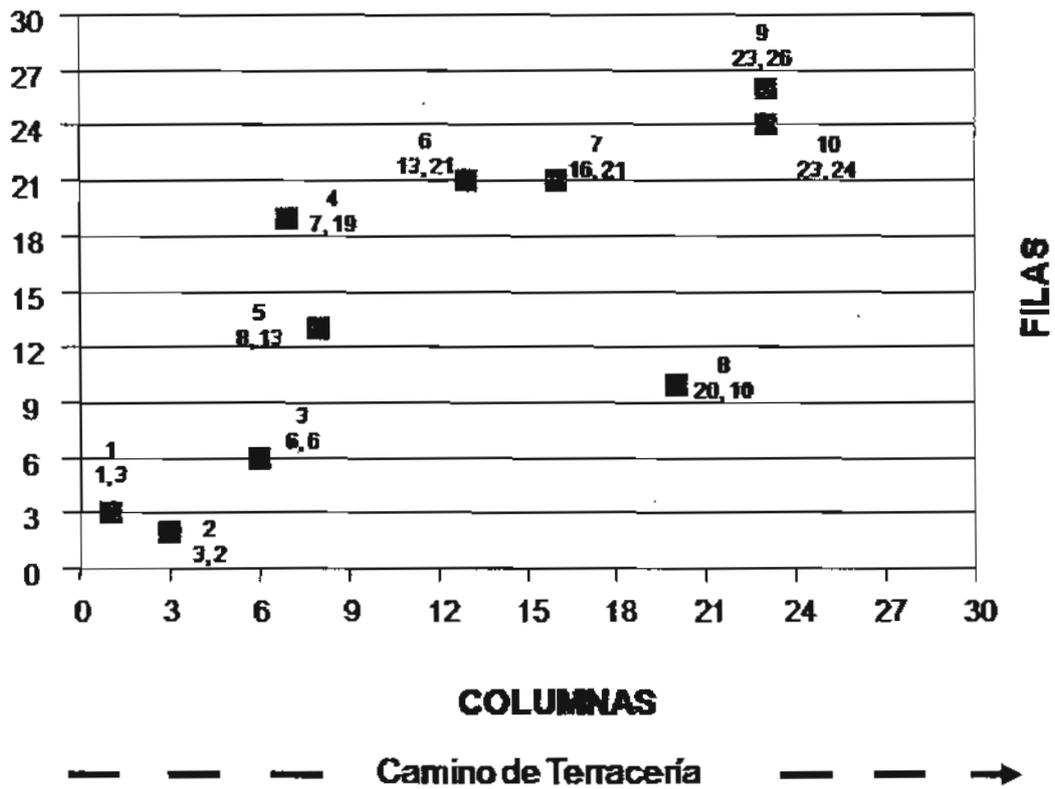
## Coatepec harinas, Estado de México Rancho Arizmendi



Zitácuaro, Michoacán  
Huerta Los Pinzanes



### Jalpan, Querétaro Huerta El Exilio



El anexo representa gráficamente los cuatro cultivares de guayaba que se seleccionaron para la recolectar hojas de guayaba durante de forma estacional durante el ciclo 2002. El huerto "La Labor" en Calvillo, Aguascalientes se compone de 2 zonas de cultivo separadas por terrasería y una pequeña bodega, con aproximadamente 600 árboles de selección media china (izquierda), pulpa blanca (derecha) y otras especies frutales. El cultivo es atendido por la familia y jornaleros, podan una vez al año los árboles y usan prácticas de cultivo orgánicas en la fertilización y control de infecciones. El rancho "Arizmendi" en Coatepec harinas, estado de México se compone de 4 zonas de cultivo (A, B, C y D) con aproximadamente 1,000 árboles guayaba sin especificación en la selección por parte del agricultor. El cultivo es atendido por jornaleros, podan los árboles una o dos veces al año y no usan prácticas de cultivo orgánicas. El huerto "Los Pinzanes" en Zitácuaro, Michoacán con aproximadamente 400 árboles guayaba media china, forma parte de un grupo de huertos dentro de un ejido que perteneces a un mismo dueño. El cultivo es atendido por jornaleros, podan los árboles varias veces al año y no usan prácticas de cultivo orgánicas. El huerto "El Exilio" en Jalpan, Querétaro tiene aproximadamente 400 árboles de selección media china (izquierda), pulpa blanca (derecha) y otras especies frutales. El cultivo es atendido por el dueño que poda una vez al año los árboles y usa prácticas de cultivo orgánicas.

La identificación de los árboles se hizo con estacas etiquetadas clavadas a un costado del árbol. La localización (cuadros grises) se hizo numerando las líneas de árboles (filas) y contando el número de árboles en cada línea (columnas) en cada cultivo.

## ANEXO 2

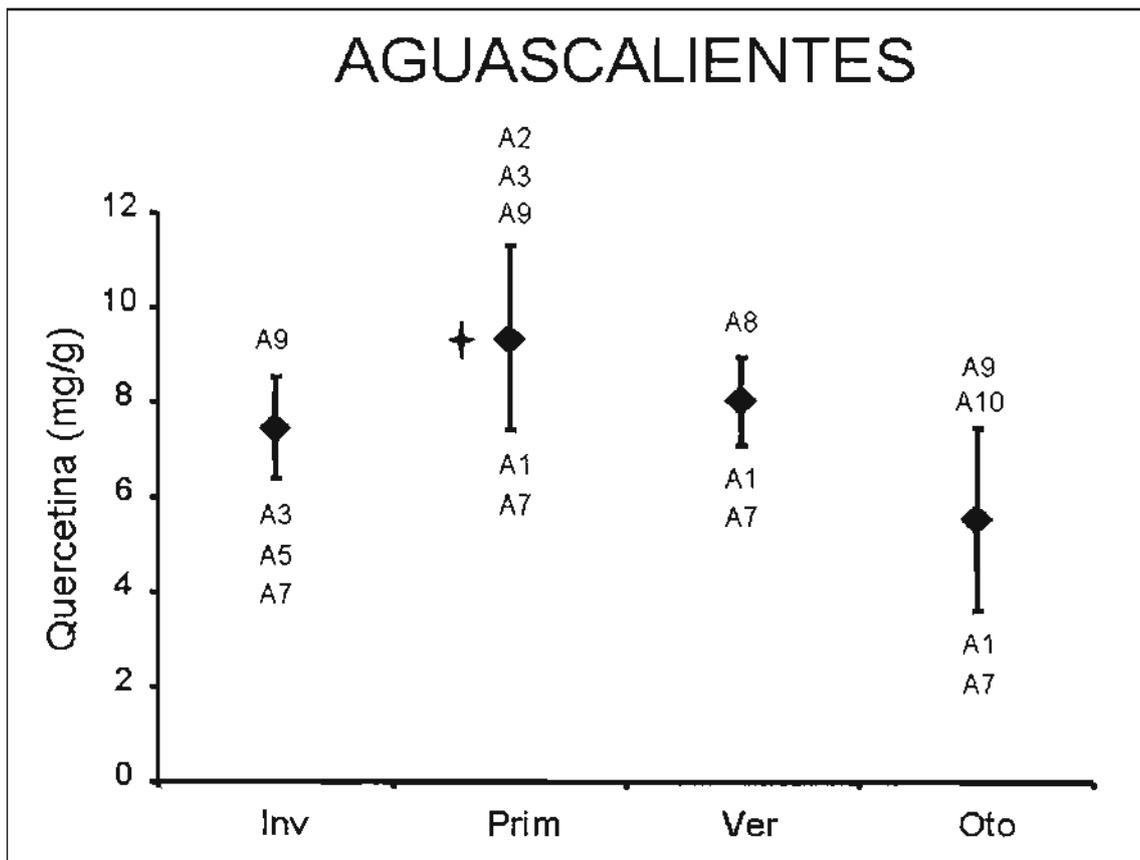
Parámetros climáticos en las regiones muestreadas durante el ciclo 2002.

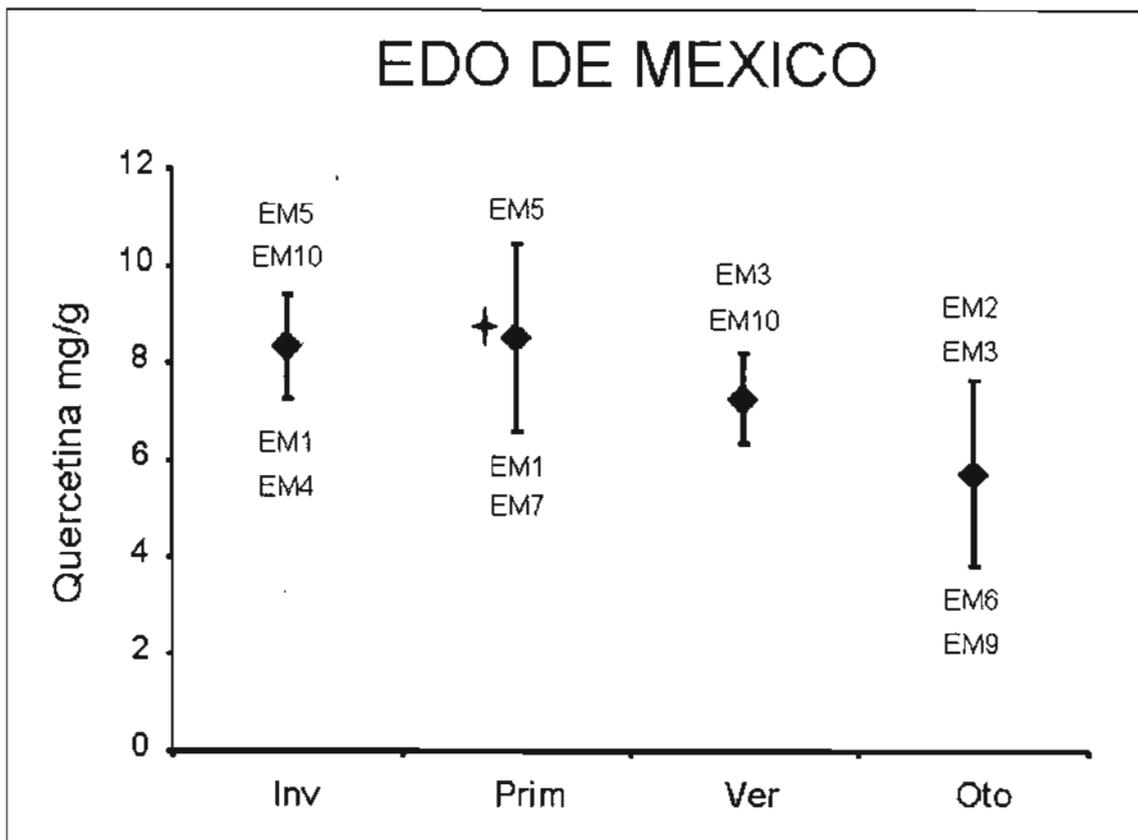
Estado de la República	Altitud (m.s.n.m*)	Temperatura promedio anual (°C)	Precipitación pluvial promedio anual (mm)
Aguascalientes	2,700	19.3	600
Estado de México	790	24	1,600
Michoacán	1,100	21	813
Querétaro	830	23.9	965

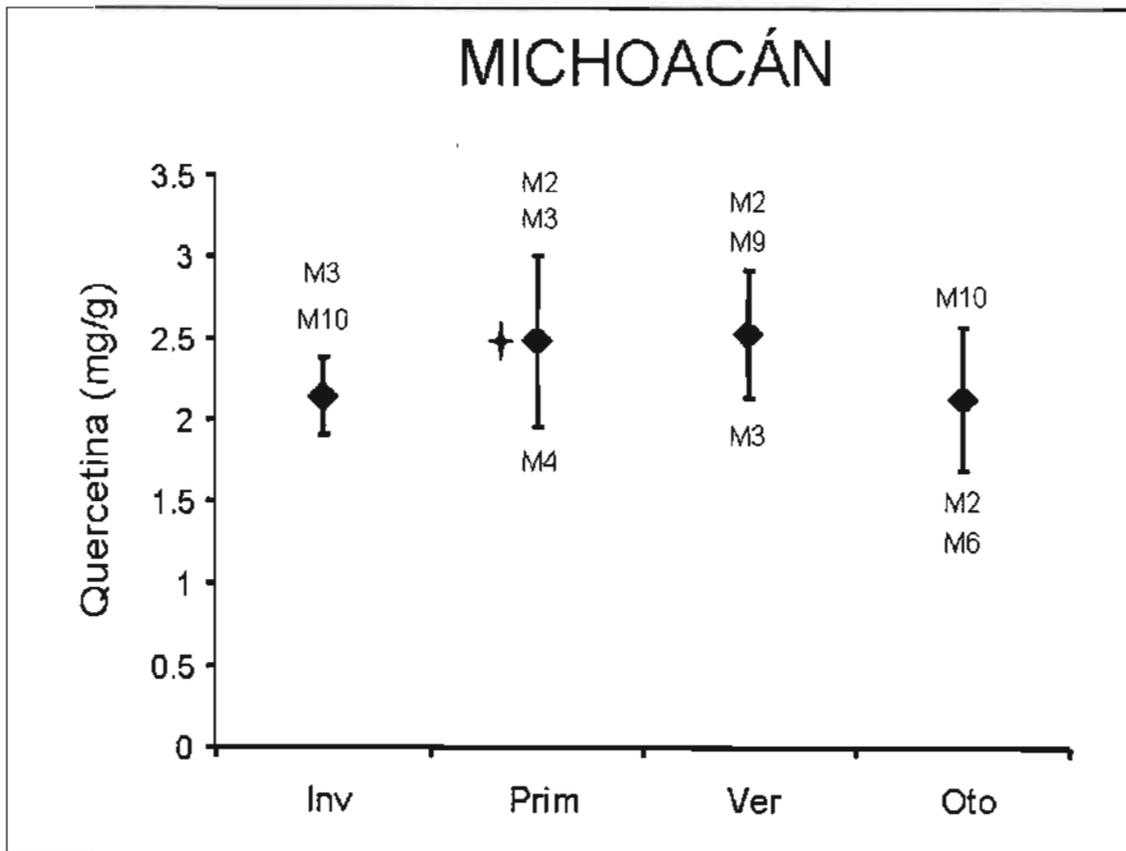
\*Metros sobre el nivel del mar

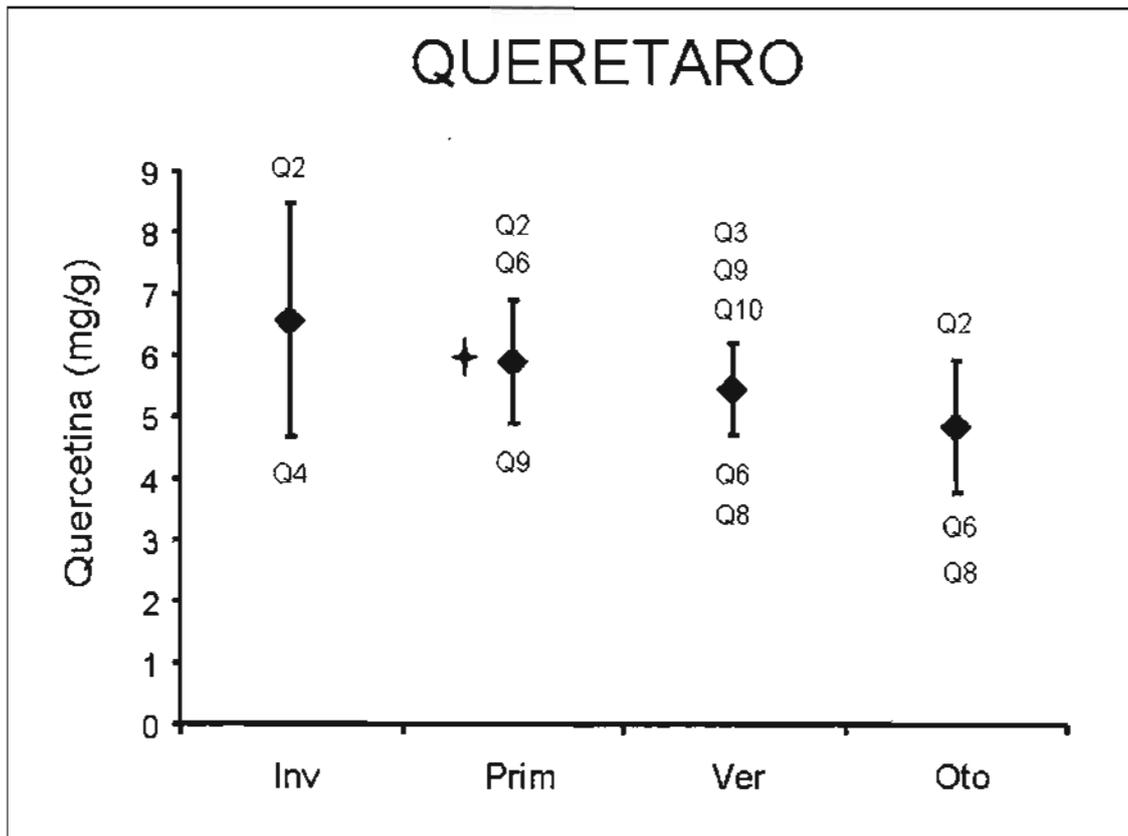
La tabla resume las diferencias altitudinarias entre las regiones geográficas donde se encuentran los cultivares de guayaba seleccionados en el estudio y las condiciones climáticas (temperatura) y pluviales que se presentaron en promedio durante el 2002.

ANEXO 3. Individuos con una mayor o menor acumulación de quercetina evaluados por la prueba de normalidad de Kolmogorov-Smirnov.









El anexo muestra a los individuos que salen de la normalidad en la acumulación de queratina o que están fuera del rango promedio  $\pm$  DS según la prueba estadística de Kolmogorov-Smirnov, en las regiones geográficas seleccionadas y en las diferentes estaciones del año. En Aguascalientes los individuos con mayor acumulación son A2, A3, A8, A9 y A10, predominando A9 en el año; los individuos con menor acumulación son A1, A3, A5 y A7, predominando A1 y A7 en el año. En el estado de México los individuos con mayor acumulación son EM2, EM3, EM5 y EM10; los individuos con menor acumulación son EM1, EM4, EM6, EM7 y EM9, sobresaliendo EM1 en invierno y primavera. En Michoacán los individuos con mayor acumulación son M2, M3, M9 y M10, sobresaliendo M2 y M3 en 2 épocas del año diferentes; los individuos con menor acumulación son M2, M3, M4 y M6. En Querétaro los individuos con mayor acumulación son Q2, Q3, Q6, Q9 y Q10, predominando Q2 en el año; los individuos con menor acumulación son Q4, Q6, Q8 y Q9, predominando Q6 y Q8 en verano y otoño.

ANEXO 4. Evaluación y selección de las muestras en base a su concentración de Quercetina

Estado	Clave (No.)	Concentración de Quercetina mg / g hoja				
		Observaciones*	Invierno	Primavera	Verano	Otoño
AGUASCALIENTES	A1 (I)	Baja Acum.	7.46	7.81	6.59	2.57
	A7 (II)	Baja Acum.	6.31	7.50	6.44	2.00
	A9 (III)	Alta Acum.	9.03	10.67	7.84	8.58
	A5 (IV)	X	6.31	8.82	8.11	5.35
	A8 (V)	Y	8.09	9.79	9.33	5.98
EDO. MEXICO	E1 (VI)	Baja Acum.	6.95	5.95	5.79	5.85
	E3 (VII)	Alta Acum.	9.14	9.00	8.23	8.58
	E5 (VIII)	Alta Acum.	9.45	12.62	6.68	4.16
	E10 (IX)	Alta Acum.	9.74	9.62	8.51	7.00
MICHUACÁN	M2 (X)	Alta Acum.	2.12	3.28	3.16	1.56
	M3 (XI)	Alta Acum.	2.45	3.05	1.84	2.05
	M4 (XII)	Baja Acum.	2.00	1.48	2.18	2.36
	M10 (XIII)	Alta Acum.	2.60	2.29	2.49	2.73
QUERETARO	Q2 (XIV)	Alta Acum.	11.61	7.00	5.46	6.16
	Q8 (XV)	Baja Acum.	5.92	6.03	4.43	3.01
	Q3 (XVI)	Z	6.61	5.04	6.52	5.91

Clave: Primera letra del estado y número de árbol. No: Número asignado a la muestra

\* Alta acumulación (+) =  $\geq$  Promedio + SD; Baja Acumulación (-) =  $\leq$  Promedio - SD; X, Y, Z: Muestras que presentaron polimorfismo con los oligonucleótidos 1, 2 y 3.

ANEXO 5. Diferencias polimórficas en los patrones de bandeo tipo RAPD y su correlación con la mayor acumulación de quercetina obtenida a lo largo del año

# muestra	árbol	RAPD 1	RAPD 2	RAPD 3	RAPD 4	RAPD 4 1	Quercetina (mg/g)
1	A1	CC	GG	TT	CC	AC	7.81
2	A2	CC	GG	TT		AA	10.50
3	A3	AC	GT	TT	TC	AC	10.63
4	A4	CC	GT	TT	TC	AC	9.47
5	A5	AC	GT	TT	TC	AC	8.82
6	A6	CC	GT	TT	TC	AC	9.70
7	A7	AC	GT	TT	TC	AC	7.50
8	A8	AC	GT	TT	TC	AC	9.79
9	A9	AC	GT	TT	TC	AC	10.67
10	A10	CC	GT	TT	TC	AC	8.75
11	EM1	CC	GG	TT	TC	AC	6.95
12	EM2	CC	GG	TT	TC	AC	9.19
13	EM3	CC	GG	TT	TC	AC	9.14
14	EM4	CC	GG	TT	TC	AC	9.31
15	EM5	CC	GG	TT		AA	12.62
16	EM6	CC	GG	TT	CC	AC	8.22
17	EM7	CC	GG	TT	CC	AC	8.91
18	EM8	CC	GG	TT	CC	AC	7.80
19	EM9	CC	GG	TT	TC	AC	8.27
20	EM10	CC	GG	TT	TC	AC	9.74
21	M1	AC	GT	TT	CC	AC	2.67
22	M2	AC	GT	TT		AA	3.28
23	M3	AA	GT	TT	CC	AC	3.05
24	M4	AC	GT	TT	CC	AC	2.36
25	M5	AC	GT	TT	TC	AC	2.67
26	M6	AC	GT	TT	TC	AC	2.66
27	M7	AC	GT	TT	TC	AC	2.46
28	M8	AC	GT	TT	CC	AC	2.39
29	M9	CC	GG	AT	CC	AC	3.07
30	M10	AC	GT	TT	CC	AC	2.73
31	Q1	CC	GG	TT	CC	AC	5.64
32	Q2	CC	GG	AT	CC	AC	11.61
33	Q3	CC	GG	TT	CC	AC	6.61
34	Q4	CC	GG	TT	CC	AC	5.44
35	Q5	CC	GG	TT	CC	AC	5.57
36	Q6	CC	GG	TT	CC	AC	7.54
37	Q7	CC	GG	TT	CC	AC	6.20
38	Q8	CC	GG	TT	CC	AC	6.03
39	Q9	CC	GG	TT	TC	AC	6.36
40	Q10	CC	GG	TT	TC	AC	6.93

La tabla muestra las diferencias polimórficas encontradas en los perfiles de bandeo tipo RAPD y la máxima acumulación de quercetina cuantificada en el año de las 40 muestras recolectadas. Las bases nucleotídicas sustituyen el comportamiento alélico homocigoto o heterocigoto de las bandas que comparten una región específica en el gel.

## ANEXO 6

Secuencias de los amplificadores aleatorios generados en *Psidium guajava* por la técnica RAPD con el oligonucleótido 1, en dos escrutinios independientes.

CLONA	SECUENCIA					pb
1.33	GGTGCGGGAA	AA-GGGGAAA	ATTCAGACTA	GGCAAAATCA	ATAAGATTTC	49
1.2A	*****	**A*****	*****	*****	*****	50
1.33	CGTATTGTTT	TTAAGATTTA	CAGTTCCGA	GCAATGGATA	CAATTCTGGA	99
1.2A	*****	*****	*****	*****	*****	100
1.33	TTGATGTCAA	TTTTCTCTTG	AATCCATGGT	CTTAATCCCC	TTCTAAGCTC	149
1.2A	*****	*****	*****	*****	*****	150
1.33	CTTGTTGTGG	GTACCTGATA	TTGTCAGACG	ACAATGGGCC	AAATTATGAC	199
1.2A	*****	*****	*****	*****	*****	200
1.33	TTGTGCTTTC	ATTGCGGCTT	TGCAGACTCA	AAAATAGTGA	CAAAATGCAC	249
1.2A	*****	*****	*****	*****	*****	250
1.33	ATCTTGCATT	GCAGTGACAG	TGCGACTGAA	GATTGGGTAA	GGCGGTTCTC	299
1.2A	*****	*****	*****	*****	*****	300
1.33	TTATGTTCCC	GCACCA				315
1.2A	*****	*****				316
1.1	GGTGCGGGAA	CATAAGAGGA	CCGOCCTACC	CAATCTTCAG	TCGCACTGTC	50
	ACTGCAATGC	AAGATGTGCA	TTTTGTCACT	ATTTTTGAGT	CTGCAAAGCC	100
	GCAATGAAAG	CACAAGTCAT	AATTTGGCCC	ATTGTCGTTT	GACAATATCA	150
	GGTACCCACA	ACAAGGAGCT	TAGAAGGGGA	TTAAGACCAT	GGATTCAAGA	200
	GAAAATTGAC	ATCAATCCAG	AATTTGATCC	ATTGCTCGGA	AACTGTAAT	250
	CTTAAGAACA	ATACGGAAAT	CTTATTGATT	TTGCCTAGTC	TGAATTTTCC	300
	CCTTTTCCC	GCACCA				316
1.12A	GGTGCGGGAA	GGGCAACTTC	TCCGAGCTCT	TCAAGTCCAT	CCAGGAGTAC	50
	GAGAAGACGC	TGGAATACGG	GCGAGTCGCA	GGGCCGGCCT	CCGOCCTGAAC	100
	TGGTATGCC	TTTGGCCATT	TGTTTCCTTC	AAGCGCGTTT	TCTAATGGAA	150
	GGATGGTATG	TCCGCTGTGT	ATTTGAGTGC	TGAGTATGAA	GGGATAAGCA	200
	CTCTTCGCCG	ATAAAGTTGG	ATGCAATTCC	TGCTTCTCTG	ATGAACAATA	250
	ATGGAAAGTC	COCTGGTTCA	CCCCCATGAA	CCAGCAAAGA	AGGATGAACT	300
	TTGTGAACAA	AOCGATAATG	ACGGTTCCTT	CGTTACTGAC	TGCGTCTTAG	350
	TTATGCGGTT	GGGCTTAACG	TCACAATTCC	ATGAGCATCA	ATAATCGCTC	400
	AAGTAACTCA	GCAGATTTCC	GCACC			425

1.1A	GGTGCGGGAA	CAGAACAACA	GGGAAAGACC	CGACCGACCT	GCCCAGGGCT	50
	CGAGGGCGAG	CITTTATTGTT	GAGAGAATGG	GGAGTGAATC	GAAAGGCTTC	100
	CGTTTTCGTT	CITGGTTTGG	GGGCTTTCGG	GCTCCTCTCG	ATCTTATTCG	150
	TAGTTGGGAA	GGGGGAAGGA	GTCTAAATCA	ATGGACATTA	ATGAACCATC	200
	ATTGATGGAC	GTTGCACATG	ACACGATCAA	TTCGACTCAA	GGTCTGGCGC	250
	TAATAGAAGT	GGCTTACTCT	CCTAGTAGCG	AAGGAAAGGG	GCAGGGCTTT	300
	TTTCGTAGTA	ATAGTGGACG	GGTCTCATGA	GATCTATGCC	GACCGTCGGA	350
	ATCAAATGAA	GGTCGAAGGA	CCATTCGATT	CATTAAGAGG	CATAGATCTA	400
	GGCCTCTTTG	TTTGGTCAAT	CACCAAAGGC	GGGGGGAACC	ACTATGGACG	450
	AGACCGCCCG	GCCTCGATTC	TTTTGCATAG	TCCGGGGGCC	GGCCGCAGCA	500
	GACCANCGGG	GCATAGCGGC	GCCCTCGGCT	GGCGTCATTT	CCCGCACCA	549
1.5A	GGTGCGGGAA	TGTTCCCTGA	TGCGTATGGA	GCTCCCTTTT	GGGTTCCTGG	50
	TGCTTCTTTC	AAATGGCGTA	AAGTTTTTGT	TTTAACTGAG	TGCTTTTTAT	100
	TTACAATGC	AGAGTAGATG	TCTCTGGGGG	ACGTCCTCTT	CGGAATGCGT	150
	TCCATGGTGT	AGTCACCTTG	CAGGTTGCGT	TTCAGTCTC	CTTGACATTC	200
	ATAGTAGCTT	TCCTTGAGGG	TTTCTAATTT	TTATTTGTGA	CCGATGCTAA	250
	TGGAATTTAC	AAGAATTCTT	CGTGTGGGCC	TAGTCTTAAC	GGGTGTGATA	300
	AATTATGCTA	GTAACCTCGT	TCTCTAGATC	AAGTTAAGTG	GTGAACAATT	350
	GAATAATTCC	GGTTTACCTT	TTTAGGATGT	GCAAGAAAAC	CTGCCTAGTA	400
	CCTGTACCAG	CCATCTGGAT	CCAACCTCGG	TACTCCTCTC	TTTATTTCTG	450
	TACTTACTAT	GGAAAATTC	ATCCTCTTTG	CAGATTAACC	CTACCTCTCT	500
	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CTCTCCCTCT	CTCTCAGTGC	550
	TTCTTCCC GC	ACCA				564
1.3A	GGTGCGGGAA	GAACAAAGAA	AGCATAGACA	TGCTAGATAG	GCAACAGTCT	50
	TGTAAGGGC	ATATGAAAAG	GAAAAAGGAA	TCAGATCATT	CTATGTGTTA	100
	ATGTATCCCA	CATAAGTCTA	CAGGGTTTAT	ATGCTCTTCA	AATACAGTCT	150
	COCTCCACCG	ATTATTTTAA	GCATTTGAAT	CCTTATTCGG	CTTGACCGTG	200
	AGATGGTGTG	TCAACATATG	GCTTCCACCA	ATTACCATGT	ATTGATAATT	250
	TCTCCACCTT	TATACGTCAA	TGAAATGCAG	ACTCCTAGAA	GCCAAAGAAA	300
	TATGGAACGA	GTAGCTTTTA	CACAGACAAG	TGAAAAACA	AAGTAAATCC	350
	CACCTTCTTG	CITATAAGAT	AGTTGAAAACA	AGATAAAAAT	TTACTTTCTA	400
	TTACAAATAC	CTCACCTATA	CAATCTATGA	TCTGTATATT	TACATAAITG	450
	ATTTATTTTT	TTACATCAAG	CAGCCTCAAG	GTAATCCCA	ATGCACCAAT	500
	GACTGAGCCA	GCTTCCCGAA	TTACTTTTCC	CATTGCCCCG	CTACGAGCAC	550
	AATCCACAA	GACTACCACA	AAATCACCAA	GTCAAAGGCA	GCGAGAATTA	600
	GCCAACTGGA	TTTAAGCAGC	GGTTTCAGCG	CTCAACAGAT	TAGGTTTAAAC	650
	ACTCAAGAGC	CCCGTCGCGA	GCAATTTTCT	TGCGATCTGA	CACTACCAAA	700
	TCACATCACA	TTTTCCCTTC	CCGCACCA			728

1.11A	GGTGCGGGAA	GAGAGATCCG	TTATCCAATT	GACAAAGGAG	TATATATATT	50
	ACTATAGCAA	AAGCAGGGTA	TTTTCTCAG	TTGTTTCATGC	AATGGGTATT	100
	CCCTTTAGTT	GTTTCATCCAA	TCAAACAAAC	TGAAATACAA	AGGCTGCATT	150
	CAGCAGGCTA	ATATTCAAGA	ATGGTCTTAA	ATTGGAAAGA	GAAAGTTCAA	200
	CCAAATATAT	CTAAATGAAT	AAATACTTGG	ATTCCTAAC	ATGATGCCAG	250
	ATCCAAAAC	CAGCAGAGAT	TGCATTCAAT	TCAGGAATCG	AGAATGCGCA	300
	TAATAGGTGG	AGACACCAGC	AGAGAAAGCA	GAAAATCAG	CAATGCTCAA	350
	TTTCAATCTT	TTACAGGATT	TTCTTTTAAAC	TCATTTTATT	TCATGTCTCC	400
	CCTGATACAA	ATGCCTTCCT	GTATTGCCAT	ATAACCCACC	ACCGTCCAAT	450
	GAACATTGTG	AGAAAGGGAG	ATTGATGCAA	CAAATGCCTC	CTCATATTGT	500
	TTACATTGTC	ATACCAACAT	GCTTCTCTAT	TCATGGAGAC	ACAGAGTGAC	550
	CATCCTTTAT	GCATGTTAAC	CTCATGAAAA	ATAAACAGGC	TTGGTTATTC	600
	TTTTCTACAC	CCATGCTAGG	CCATGCCAC	TAACATATCC	AACTCAATTT	650
	CATTCTTAT	CAGGCCAACA	ATCCAAATTC	GAGAAAAAGG	GGGAAAGGGG	700
	ATGTTTAGCT	CTTGCAAAAG	ACATGGAAGG	ACCCATAAAT	AGTCAGAAGC	750
	AACTCTTGCA	ACAAAAAATC	AGTATTAAGA	GGCCTTGATG	GGAGTATTGG	800
	AGAACAAGAA	AAATAGTTCA	ATGAAACACA	CGAAAAACAA	GACTTACTTA	850
	GGTAGTTTCT	ATGTGCATAT	TAGGGAGACT	ATTCAAATCA	CAGAAAAAGA	900
	TTTACAATAC	AAAGAACCAA	GTACAAATGG	ACGAGTATCC	AAATTTTGAA	950
	GTTTCAAGTG	TAAAGGTAAA	CATGGAAAAC	ATCTATCACC	TCTGAATTAT	1000
	TATAAGCACG	AAGAACAGTG	AAAGGTTGAG	GCTTTGCCCG	GATATCATAA	1050
	AAAATATCC	GACCGTTGCT	GGTCCCGCA	CCA		1083

2.2A	GTTTCGCTCC	AAAGCCCGCC	CAAAACCCAA	AACTAAATCA	GAACCACATT	50
	AAATTAAGAT	CCTGAATCTA	GAAGCTTTTA	ATTTTCCTCT	GACATTTTCT	100
	GCTCCTTCTT	TGTTTTACTT	ATTTATTTAT	TTATTTTAA	ATCCGCCTTC	150
	TACGTGGGTT	CTCAAATGCA	AAAATAATTC	TGCAGTCTTA	ACCTATGAAG	200
	GAGCGAAACA					210

2.1	GTTTCGCTCC	TTCATAGGTT	AAGACTGCAG	AATTATTTT	GCATTTGAGA	50
	ACCCACGTAG	AAGCGGATT	TAAAAATAAA	TAAATAAATA	AGTAAACAA	100
	GGGAGCAGAA	GAATGTCAGA	GGAAAATTA	AAGCTCCTAG	AATCCAGGAT	150
	CTACAATGGG	ATGTGAGAAT	GCGAAAAAAA	TTTGGGAGTT	GGGCCCTCTG	200
	AAGGACCGAA	CAGGC				215

2.2	GTTTCGCTCC	ATTAACATAA	ATTTAAATTC	AACTATATCT	TCCAGTAATG	50
	TGAAAAAGAT	AAITGCACGT	AATAATAATG	CCCTAATAAC	AACTACACAC	100
	NNNNGTACAT	CTGATAATAA	ATACCTTTTC	CTTGTGTAGC	TCAGGATGAA	150
	AATAAATCTA	GCTTTTTGTT	CAGTTGAAAG	TAATGCGGTT	GGCTCTAGTG	200
	TAAAGTGTTT	AAAGAAATGA	GAAAAGTATA	AAGGGAAAAC	ACATCTAAAT	250
	GATTAATTGG	TAGGCCTGTA	CTTATTTTCT	ATTACTTTGG	AGCGAAAC	298

2.3A	GTTTCGCTCC	AACTGTTTAA	GAGCTTATAC	AACCCATTAT	CAAGCTAGGC	50
	TTAGGACTGG	CGCATGGAAT	TTGATTAGCG	GAGAGGGTAC	ACGTTTGTGG	100
	AGCCACTTTT	GACTAATATA	AAAGTTTGAG	CTGTTGTTAG	GTGGAAGGAG	150
	ACCGCACGTG	TATAAAAAGT	ACATAAGACT	CATTATCAAG	CGACATGAGA	200
	TATTTCAACA	ACACGAGCAA	TGGAGAGTAA	TTATAGCTCA	GGACTCAGAC	250
	ATTGAGAAAC	AAGGAGGAGA	CTCTTATGGC	TCGATATAGT	TGAATGCGAT	300
	ACCGAATCAC	TGCCGTCCCG	AGGAGGAGCG	AAACA		335
2.34A	GTTTCGCTCC	TGGTACTACT	GGAGAAGGTT	GATCATGGAA	CTTCTGGTGA	50
	AATTATTGCT	GTGAATGAGA	TCAGGTATTT	CAGGGGGGTA	CTGGTACGTC	100
	GGCCACTGGG	AACGAATATC	TGCGGTAGTG	TACCACTAAT	GTAGTTGACC	150
	TGGTGGTAGT	CGCGAATAAT	TTCTCTTTAC	GAACATGTTG	ACGGTTTCTT	200
	GCTTGGGACG	GGAGAAATGA	TCTGCTAGGA	CATTGGATT	TCCTGGGAGA	250
	TACTTTACAT	CGAAAAGAGTA	TCGTGAAAAC	CATTCGGCCC	AGCGCATGGG	300
	CTAAGGATGA	GGGATTTGCT	TCTGTTGAAC	TTGAACATTT	GGGAAAAGGA	350
	GGACATGTCC	ATTTGACGCA	GAAAGGTTAC	GCCGATGAGA	TGGAATTCAA	400
	ACTTTTGTAT	GCCATACTTG	ACTGCCAGAA	TTTCTTGAA	CGTGGAATGG	450
	TAGTGGAGTT	CAGCTGGAGA	AAATCGTCCA	CTTGTATATC	CACATATGGA	500
	GCGAAACA	=				508
2.45A	GTTTCGCTCC	GACCAGATCA	AGTAGCAGAG	AGCCGCAAAG	GAGAATCTTA	50
	TGATGCATTG	GTGGAATTTT	TTACCACCAA	ATTTGTTGGT	GGCCCAACGG	100
	AGATTTTCTT	GCCAGTCTCT	GTTGCGCCAA	GAGAGACCAC	ACTTATCCGC	150
	CCAAAAAACC	GCAAGGGAAC	CCGTGACTTG	ACATTCAAAA	AACAAGTGGT	200
	CCACCGAGTC	GGGTCTGAAA	CGGCAAAAGG	AGCATAGTCC	GTTTGGGATT	250
	CCGTGATAGT	TCAAAAGACG	CACCTGAGTC	GGCAGCCGAT	TAAGGGTTGC	300
	AAGCCATAGA	AGGAAGGAGT	GGCGTGAAT	AATGTGGCAG	TCCTAGATAA	350
	AGTGATGCCA	AAAAACCGTG	CTGCCCTTGC	GCCTAATATC	GTCCCAAGCA	400
	GAAGCCACCG	AGAAGGAGCC	AGATGGGTGG	CAAAGCCAAA	TCAGACAATC	450
	CCTCTAGCC	GGGGCAATAA	TCGGAGGGAA	ATCGGACCAC	GTGTGAAGCC	500
	GATCCAGAAC	AGAATTGGCC	AGAAACTGAC	CATCAATGAA	GTCTCGAACC	550
	GAAGCCATCC	GAAAGACACC	CGAGAGATAG	ATCTCCCTAT	CCGAAAAGGC	600
	CTTATACAGA	GGGCCATTTG	CATGCCAAGG	GTCAAACCAA	AAGGAAGTTT	650
	CGETCCAAAG	CCCGCCAAA	ACCCAAAAC	AAATCAGAAC	CACATTAAAT	700
	TAAGATCCTG	AATCTAGAAG	CTTTAATTT	TCCTCTGACA	TTTTCTGCTC	750
	CTTCTTTGTT	TTACTTATTT	ATTTATTTAT	TTTTAAATCC	GCCTTCTACG	800
	TGGGTTCTCA	AATGCAAAAA	TAATTCTGCA	GTCTTAACCT	ATGAAGGAGC	850
	GAAACA					856

2.1A	GTTTCGCTCC	TGCTTGATTC	GCCGTGAAGC	TGTCCTCCT	CCCCAAGAGC	50
	ACATTCCAAT	CTGGACCTCC	TGACTGCAAT	TTCAGTAGCA	ATTGTGATTG	100
	GTGGAGGCAG	CTAAGTATTG	GAGACGGACC	CTACTGATAC	TTTGAATTGA	150
	TTCTCAATGC	TTCTTAGGCC	ACAACATGT	ATACCAGGAA	GACGGAAGCT	200
	TGAGCAGCGA	GGGCGAGAAT	GTCGGCGCAC	GATACAGTCC	CAGGACAAGA	250
	GCTCTCAATA	GCCGCTTTGA	TGTTGTCCAC	CACATCGAAA	CCCCTCGCGG	300
	AATTTGCATT	CGGACCCGCG	TCTTTCTCGC	TCTTAATGGT	GGCGCTGTTG	350
	TCCAGCAATA	GCGATCCATC	GCACCCCTTG	TAACAAATGG	TAAGCACAAC	400
	CTTTAGGTCA	CGTGACTAAA	GCTACCCGAA	AGTCGGATTG	AAAGTACGAT	450
	AAGAAACGAA	GTGATTATTT	AGGACTCGCT	TTTAATGGAA	AGATGACCTA	500
	GCTTAGGGAC	TCACGTCGAC	GAAGCAATCG	TGGAAATGGA	GACGGACGAG	550
	GCTGGCGGTG	ATGCGCGGGT	CGGATTGAAG	GGCCCGGTGG	ATCACGGATT	600
	GGGTAATCCT	GGAAACATTC	GGGCATGTCG	TGGCGTAAAA	GCTTGCCTC	650
	AACTGAGCGC	TCGATGGACG	GAGGAGAAAC	CACAATGCGG	CGAAAATGAT	700
	CATCGCTATC	AGCGAGAAGT	GAGCTGGAAC	ACGGACCATG	ACATGTAAAT	750
	TAACCCCGTT	CGGACCGATT	GAATCCCTCT	CAGAAACAGT	GGAATTTGAA	800
	TGCTAAAGCA	CAGTGTGGT	CCTATACGTA	TATATAGTAA	GGGTACGTCG	850
	AAGCTCCTAG	GGCTCAAAGT	CATCATGAGC	GCGTAAATC	ATTGAAAAAT	900
	TACTCTTCT	TACTCCATCG	TCTCAAGGCG	GAATTATTAA	TGTTGAATTA	950
	GATGCAGAGG	ATCGAACATA	TATAAACTT	TGGAGCGAAA	CA	992
327	GTAGACCCGT	CATGAACGTC	ATGAAAACGA	GTTCCATAAA	GCTAAAAATG	50
	AAAGACTCGG	AGCAGAGAAT	ACAAGGTATG	GAGAAGCTTT	TGACTACTGCC	100
	ACNTGCTCTA	ACTGTAGTGA	CCTCTTACC	CTCAACGAAA	TGCAGCATCT	150
	AAAGATAAAA	AAAAATGCC	GTTTGAAGG	TGAGTTAGGT	TAATCCCAG	200
	CAGAATCAAG	TTCTATTANA	GGAAATACG	AGAAGACCAA	TCTTATGGGG	250
	TCAGTATTAA	TATCTTTCGA	GAAAAGGAAG	ATGTATCACT	ACCAGTCAAG	300
	TAACGTTGAC	GACGAGCTTA	TTGCCATGAT	CAACTAGTTT	GACGTTGATG	350
	AAGCTNTAAA	GCTGTAATGT	TCATACAAAG	TTTGNTTTT	AAAGCCAGTG	400
	ATCACGGGTC	TAC				413
3.15A	GTAGACCCGT	TCTTGACTTT	TGCCCCGCA	CCGGGAGAAA	TGCGTGTGAA	50
	ATAAATCAGG	TAAGGACTCG	AGATATTCTA	GCTATCTATT	GGAGACAACT	100
	GTCATGCTG	TCATATGCC	CATAAGAAGA	AAAATTGAGC	ATCATCTTGA	150
	AATTGGTATG	AAACAAAAAT	GATATACTCA	TTTTCCCAAG	TTTGTGTTCA	200
	TTGTTATTTT	GCCCATCTG	CTATCAGAGT	AGACTGCATT	AGGTGGCCAA	250
	AGCTTTACAG	AGATCTAGGG	TTGCCTCCA	AGTATGATAA	ATTGCAAGAA	300
	TTTGGTTTCA	GGCTTTGGCA	ATGTTCTATC	CAACACTTGC	COCTGCAAGA	350
	AGGAAAATAA	AATGTCATTT	TCAGAATTGG	TGAAACTTGA	ATGGGGATAT	400
	TGCCCTTAG	ATTGAGTAGC	CITTGCTACA	CTTCCATCA	ACTGATATGG	450
	TATGATCGAG	GTTAAAGTAT	GAATTAATCT	TCCATTGACC	GATAACCTGA	500
	TTTATAAGCT	CGTACGGGGA	TCAGCCAGGT	CTTCTAGCCT	AACTATGAGC	550
	TACGGGTCTA	CA				582

4.1	AAGAGCCCGT	GGATACATCA	GTAAGTTACC	AACAACCTCG	TGAAAGCATC	50
	TGAAGATCAT	TCTGATTGTT	TTGAATCACA	GGCCATTAGT	ATTTATTTGC	100
	GTTAGAATGA	ACATGTGAGT	TGGGTTGGAA	AAGTCTAAA	TTTGGAGAAT	150
	TAATGTGATG	CGATGCAAAG	CAGTTAATAT	ATCTAAATTG	GTGTATAAAT	200
	TATTGGCTTA	AGTTTTTGAG	TGAATGTGGA	TTCAATTGGA	TATGCTGACG	250
	GGCTCTTA					258

4.39	AAGAGCCCGT	GGGGTTCTGG	GCGTGACAAG	TGATGAAAAG	GAGCATTGGC	50
	GAAGAAGCAC	AGCGTTCTC	AACGGAAATT	GGTGGCTCAC	GAAGGAGGCG	100
	ATGGTCCCTC	AGCTGAGCAT	GGGTGATTGG	GGTGGTCCAA	AGTATCTTCT	150
	ATACCTAGGA	AAGACAAGAG	CCCGTCTCGT	CGAAGTCCA	AAACCCGTTT	200
	GTCGTTTTAA	GGGGTGCATC	GTCCATGGAT	GCTAAATCAG	GTACGCCGTT	250
	TCAAATCTT	GGCTAATAGG	TTGTTAGAAC	ATGACTACGG	AGGTCCTAGG	300
	GTGCCGGAAT	CGCCGGCCAA	CAGTTGGTAT	CAGAGCCACC	ATAGCTCGGT	350
	TTAGCACAT	ACGGGCTCTT	A			371

4.3	AAGAGCCCGT	CTTCTCAGAT	TTAGCGCGT	ATCCAACCTG	AAATTGCTAG	50
	TTAGAGCTCA	TCCGGAGTCC	AAATCGATCT	CCATTTCTC	TATGGGACCG	100
	TAAGACCTTG	GGTTTCAATG	TAAGTTGGCT	AGAAATTTCC	TAAATTTTCG	150
	TAGATTAATG	CTGAACGAG	GCTGGTCATG	GGGAGTTAAG	CTCGACTGCT	200
	ACCGTTCGCC	ATCCGCCGAT	GCTCCGGCAC	CTCTGTACCC	CTCATTACCC	250
	TTGTTCTCAG	TCTTCTCTAG	GTATAGAAGA	TACTTTGGAC	CACCCCAATC	300
	ACCCATGCTT	AGCCGAGGGA	CCATCGCCTC	CTTCGTGAGC	CGGCAGTTTC	350
	CGTTGAGAAC	CGCTGTGCTT	CTTCGCCAAT	GCTCCTTTTC	ATCACTTGTC	400
	ACGCCAGAA	CCCCACGGGC	TCTTA			425

4.9	AAGAGCCCGT	GCGACGTAGT	GAAATGCTAA	TGACAATTGA	CTGATGGGTC	50
	CTAACACAG	CCTACCGTCG	TGTTCTGTGC	GTGTACATC	ACTGCTAATG	100
	TTAGCAGCTG	GCCACTTAAA	AACCAAGGT	TGAAAGATCC	AAAGAATATC	150
	AGCCTAGTCT	ACATAGAAAA	GTCAAGATAC	GGTAGCAATC	AATTTAGTCT	200
	TCACTTAGCG	TTCATATCTG	GATTTACAT	GAACCTAATT	ATGGCAGAGT	250
	TCCAACCAAGT	TTGCCAGCTT	AGAAACCTAC	TCGAGCTTAC	AAAACATCCA	300
	TCTAATCAGG	CATGCTCAAC	AGAAATTTGA	TGTTTCGCAA	ACGACACAAC	350
	AGTTCGATAG	GAATACGAAA	AATCATATGT	GGCACGTAAT	CACGAAGCAG	400
	TCGGCCATTT	TCAATTTACC	TACAAAGGGG	AACCCGGCAA	GAATCGGGCT	450
	GGTCACAGAT	CGAAGAGTGC	AATCTGAGAA	GCTGCCACAT	CCGCTTGAT	500
	CGCGAGCAAC	CTCCACCGAC	CCGCTTCTCG	CAGGTCCGCA	AATGCTTGAT	550
	CGAGAGCTGG	ATGCCGACGC	AGATGCGCAA	CITGCCGCAC	GGCTCCTTGT	600
	TCCTCACGGG	CTCTTA				616

5.8	AACGCGCAAC	GAAATTATCT	GCCCCATCCA	ACTAGTTCCA	AGCATCTCGC	50
	TAAATTAGAA	CTAATCTAAA	TCTAGCTTAA	AAAAGTGCCT	TATCCTCATT	100
	TATTCTATCT	CTATTCTTGT	CTTCAAGGCA	AATAGACAAG	AGTAGATTTT	150
	CTTTGGAAAT	CTCGAATGGA	CCAATTAAGC	ATGGGCTAGG	AAAGGTGTCT	200
	ATGTGCTCCA	CACATAGACT	AAGAAACAAT	TAATGACTAA	TGGGTTGCGC	250
	GTTA					254

5.5	AACGCGCAAC	TGAGCCTAAA	GTCTAGCACC	AAAGOCCAAA	TTTGTTTGCT	50
	GAGCCAAACC	GACTIONAACC	GTTTGAGCCT	ACCATGCCTG	GATCTAACAA	100
	GTCCATCCCA	GCCCAAAGTA	ACCAATTCOA	ATATGATTAG	AATTAACCT	150
	ATCTAAGCGA	TAACCAACAT	ACGTCAATGA	CACAAGCATC	AAGAAAATT	200
	TTTCCAATT	TCAATCAATT	CAGAACCAGT	TATGCACCAT	AGGCTAGTAT	250
	TCTGACACTA	CCGCATCAAC	AAGAGGAAAG	GAAACACACA	ATTGATAAAC	300
	AAAACCGTGA	ACTGTATCTA	AAATAGGAAA	AATGCAACGA	CATAACAACC	350
	ACGTACCTCT	ATTTGCCCTG	GTTGCGCGTT	A		381

5.39	AACGCGCAAC	AACAGAATGC	TCAAGGGTTG	TGAAGAATGT	GGCCTDATGT	50
	CCATGAGCGG	TAGTTGTTTA	AACAGCAGCA	GTGAATTAGC	TAAAGCGAAA	100
	AGGCAGGAAT	GGGCTCGAGA	TGAAACTTCT	CTTATGGTCA	GGGGCGGACG	150
	GCTAGTCCTT	GGCTTCGAGG	ATTTCCACCC	AGAATATCGG	AGGGAGGATC	200
	GGAGAAGAGG	GGGGAGAGTC	CCAGGGGTCG	GTGAAGAAGC	AGAGGTACGG	250
	AAGATTGCTT	TTAAAGTTCC	CAATAATGGA	TTGTTGGCCG	AGGAGTTGTT	300
	TTTGGTAGCT	CGTGAAGTAT	ATGACTCCAG	GTCTCGCCTA	TTCATTATG	350
	TGTCAAGAGT	TGCCAAAAAA	TACGCCTGTT	TATACACGTA	GCTTGTGTGG	400
	AGGAGTTCAT	GTTAGTAATC	CACCATGTTG	GGGGAAGCCG	TGTGAAGAGC	450
	ATAGCTAGAA	ATGTGGAGTC	ATTGAGGGAG	CTTATAAAGT	TTTCAGCGAG	500
	ATACCCGTGA	AGGATCTTGG	CATGTGGAGT	GCGATTATGA	TTGCTTGCCT	550
	TCAGCATGCA	CACATGGAGA	AGACATTCGA	GCTGTTTAAA	TAGATGGGGG	600
	GTCACGAGAT	GAAGCCAAAT	TTCATTACTT	TCCGGAGTAT	CCTTACGCT	650
	TGCAGTTGCG	CGTT				664

CAG.28	TCTAAGGCCT	TGCTAGCAGA	AGCACTTCTG	TTTCTAAGG	CCTTGCTAGC	50
	AGAAGCAGCA	GGAGCTTTTC	TAAGGCCTTG	CTAGCAGAAG	CAACTGTACA	100
	TTAGAATCGG	ACAAGAGTGA	TGGTCATAGC	TGTTTCA		137
SK5	ATTCGATTCT	AAGGCCTTGC	TAGCAGAAGC	AGCAGAAGCT	TTTCTAAGGC	50
	CTTGCTAGCA	GAAGCACTTC	TGTTTTCTAA	GGCCTTGCTA	GCAGAAGCAA	100
	CTGTACATTA	GAATCGGACA	AGAGTGATGG	TCATAGCTGT	TTCCAATCAC	150
	TAGTGAATTC					160
CAG.4	TCTAAGGCCT	TGCTAGCAGA	AGCACAACIT	AAAAAGAATC	TTTCTAAGG	50
	CCTTGCTAGC	AGAAGCATTI	CTGTTTTCTA	AGGCCTTGCT	AGCAGAAGCA	100
	CTTCTGTTTT	CTAAGGCCTT	GCTAGCAGAA	GCACAACITTA	AAAAGAATCT	150
	TTTCTAAGGC	CTTGCTAGCA	GAAGCAACTG	TACATTAGAA	TCGGACAATT	200
	NNTGATGGTC	ATAGCTGTTT	CCA			223
CAG.1	TCTAAGGCCT	TGCTAGCAGA	AGCACAGTGG	CTCATGCCTG	TAATCTCAGC	50
SK2	*****	*****	*****	*****	*****	50
CAG.1	ATTTTGGGAG	GCTGTGGCNN	AAGCATCACT	TGAAGCCAGA	AGTTTGNTGA	100
SK2	*****	***** AG	*****	*****	***** A..*	100
CAG.1	CCAGCCTGGG	CAACATAGCA	AGACCCATC	CCTCCACACA	CAAACACACA	150
SK2	*****	*****	*****	*****	*****	150
CAG.1	CACACACACA	CACACACACA	CACACACACA	CACACACACA	CACACAAACA	200
SK2	*****	*****	*****	*****	*****	200
CAG.1	AGCTCTTGCC	AGAATTAGAG	CTACAAATTG	COCTCAGGT	CCTAGAAGAT	250
SK2	*****	*****	*****	*****	*****	250
CAG.1	CAGTCCTTCA	ATTAGATTCA	GATTGAGATG	CTTCCTCTTT	TAAACAATGA	300
SK2	*****	*****	*****	*****	*****	300
CAG.1	TTCCCTTTCT	ATCATGCCCA	ATAAGAAAAC	AAATAAAAAT	TAAACAATAC	350
SK2	*****	*****	*****	*****	*****	350
CAG.1	TGCCTGTAAT	CTCAGCTACC	CAGGAGGCAG	AAGCAGAACT	GCTTCTGCTA	400
SK2	*****	*****	*****	*****	*****	400
CAG.1	GCAAGGCCTT	AGA				413
SK2	*****	***				398

CAG.38	TCTAAGGCCT	TGCTAGCAGA	AGCATTTCAG	AGAGAACAGG	GGTCAAAGTC	50
CAG.37	*****	*****	*****	*****	*****	50
CAG.38	CTCCCCAAGT	AAGGAACACA	GGCACAGCCA	CACTGCATAC	ACACACGTGC	100
CAG.37	***** A	*****	*****	*****	*****	100
CAG.38	ACACACACAG	GCGCACACAC	CCACTCACAC	ATGTGCACAC	TCACACCOCC	150
CAG.37	*****	*****	*****	*****	*****	150
CAG.38	CACACGTGTG	CACACCCACG	CACACACGCA	CACACCCACA	GTGCACACTC	200
CAG.37	*****	*****	*****	*****	*****	200
CAG.38	ACACCCACTC	ACACACGTGC	ACTCACACCC	ACGCACACAC	GTGTGTGCAC	250
CAG.37	*****	*****	*****	*****	*****	250
CAG.38	ACCCATACAC	ACCCACACAC	GCGTGCACAC	ACCCACTCAC	ACACGTGCCA	300
CAG.37	*****	*****	*****	*****	*****	300
CAG.38	AACACATGCG	TGCACACCOCA	TGCACACACA	CTTGCACACT	COCAGGTGCA	350
CAG.37	*****	*****	*****	*****	*****	350
CAG.38	CACACACCCC	CTTACACGTG	CACACTCAA	CCCCACACT	CACATGCATG	400
CAG.37	*****	*****	*****	*****	*****	400
CAG.38	CACAOCATG	TGCACACACC	CACACACAAC	CCACACGGTG	CACTCACACC	450
CAG.37	*****	*****	*****	*****	*****	450
CAG.38	CTCACACACG	TGCAATTCAC	ACGCATTCTA	CACACATGCA	CACACCCATG	500
CAG.37	*****	*****	*****	*****	*****	500
CAG.38	CATGTGCACT	CGCACCCATG	CACACTCACG	TGTGCTTTTG	CTAGCAAGGC	550
CAG.37	*****	*****	*****	***** C	***** G	550
CAG.38	CTTAGA					556
CAG.37	*****					555
CAG.20	TCTAAGGCCT	TGCTAGCAGA	AGCACTTCTG	TTTCTAAGG	CCTTGCTAGC	50
	AGAAGCACAC	GTGAGTGTGC	ATGGGTGCGA	GTGCACATGC	ATGGGTGTGT	100
	GATGTGTGT	AGAATGCGTG	TGAATTGCAC	GTGTGTGAGG	GTGTGAGTGC	150
	ACCGTGTGGG	TTGTGTGTGG	GTGTGTGCGC	ATGGGTGTGC	ATGCATGTGA	200
	GTGTGGGGGT	TTGAGTGTGC	ACGTGTGAGG	GGGTGTGTGT	GCACCTGGGA	250
	GTGTGCAAGT	GTGTGTGCAT	GGGTGTGCAC	GCATGTGTTT	GGCACGTGTG	300
	TGAGTGGGTG	TGTGCACGCG	TGTGTGGGTG	TGTATGGGTG	TGCACACACG	350
	TGTGTGCGTG	GGTGTGAGTG	CACGTGTGTG	AGTGGGTGTG	AGTGTGCACT	400
	GTGGGTGTGT	GCGTGTGTGC	GTGGGTGTGC	ACACGTGTGG	GGGGTGTGAG	450
	TGTGCACATG	TGTGAGTGGG	TGTGTGCGCC	TGTGTGTGTG	CACGTGTGTG	500
	TATNCAGTGT	GAGCTGTGCC	TGTGTTCCCT	CTTGGGNAGG	ACTTTGACCC	550
	CTGTTCTNCT	ACTGAAATGC	TACTGCTAG	CAAGGCCTTA	GA	592

CAG.7	TCTAAGGCCT	TGCTAGCAGA	AGCACTCTGC	CTCAGTGCTT	CTCTTCCTAC	50
	ACACTTCACC	TCCOCTTGCC	TCTCTGAATG	TAAGTGTCTC	TTCCACAGCA	100
	ATCTGTGCAA	AGACTAACAC	CAGAGTAACT	ATCTTACAGT	ATTGTAACAC	150
	TTAGCTAATT	TTTCGCCTC	CAACAATAAA	TTCTGATCTA	TTTGAGGAAG	200
	TGAACTCCAA	TACCATTCTC	AAGGAAGTCC	TTTTTTAGAA	ATATAGACTG	250
	AAACCTGACA	GAAAAGGGAT	TCAGTGTATA	TCCTAAAGGA	TGGCATAGGA	300
	GAGCTGAGAC	TTAGCAGGGT	ATACACTAGA	GTGGTCTCAT	GCTCTTCTTG	350
	CCAGCACACC	ACTAGGTTAG	CAGTTCTCAA	CTTTGGCTGT	AGTTGGTATA	400
	ACCTGGAAGC	TTTAAGATAT	TTTTGATGTC	TAGGTCTCAG	TTCTAGAAAT	450
	TTTGAATTCA	CTGGAATGGA	GTAGATTTGG	GTCATTAAGG	TTTTTAACTG	500
	TTCCAGGTG	ATACTGGGGA	TGATGAGAAC	TTAAGGATGA	GAATCAATGC	550
	ACAGCAAGAC	ACTGTGAAAT	GGGAATAGAT	GTTAGAAGCC	ACCATTGGCA	600
	CTAGGGGTCA	CAAATGGATT	GCCAATTTAA	GAATACCTTA	TTTTTCACAA	650
	TTCTCTTAAG	GAATAAGTTT	TCTTCCACCT	GAACTGCTTC	TGCTAGCAAG	700
	GCCTTAGA					708

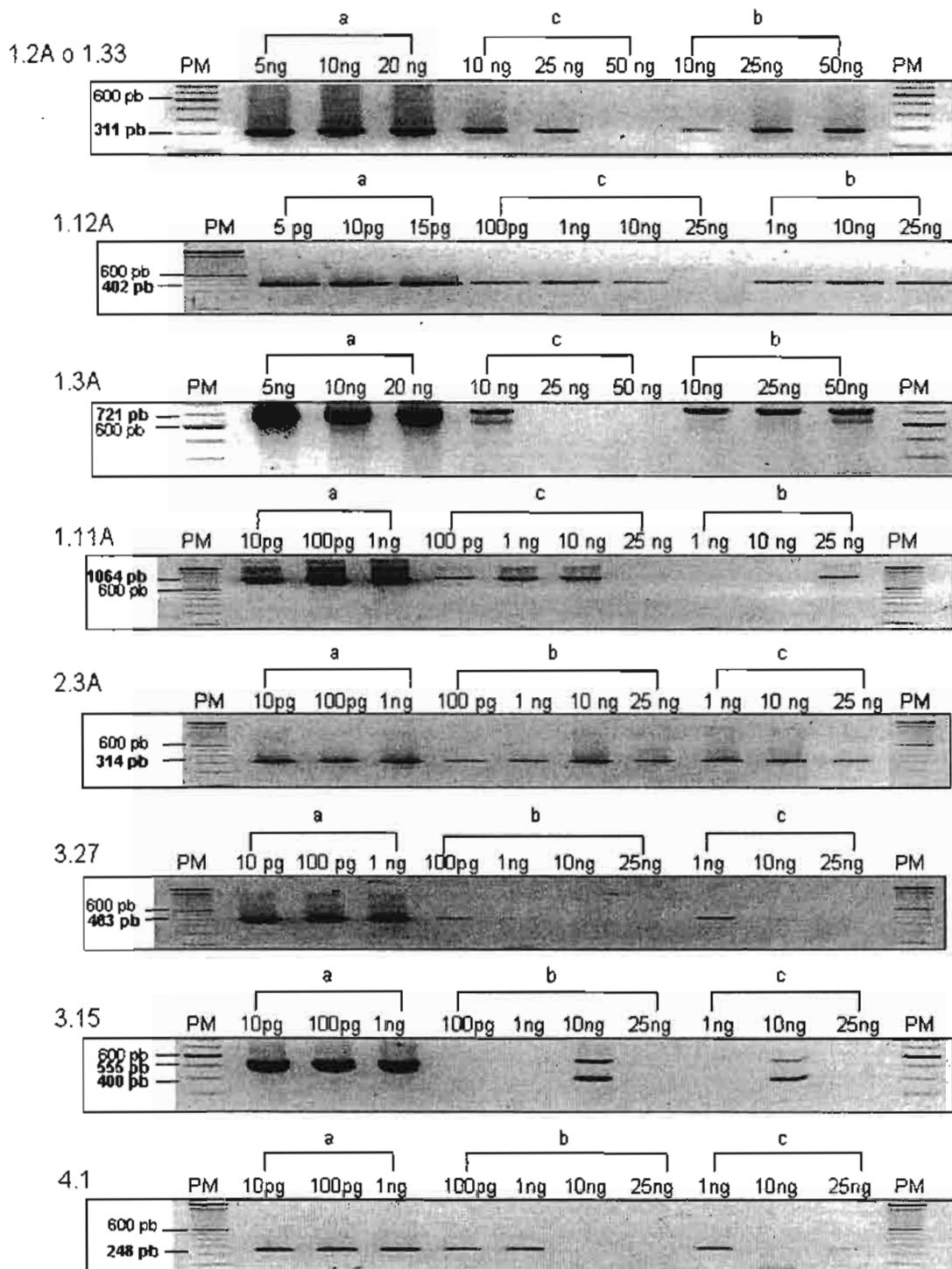
SK1	ATTTCCATGG	AGTGCCTGAG	TGTTTGAGTT	GACGAGTAGC	AATACCTAGG	50
	ATTGGACACA	GCCAATTTAT	ATCTCCTAAT	AGTTGNTGGA	AATCACTGAA	100
	GGTTTGTAAT	CTGTCCCTAT	GGAGGACTGC	TTTCTGAGGT	CGAACATTTG	150
	TTTCAGTAAC	AATAGTGCCT	AAGTATTGGT	ATGGGGAGGT	TGCTTCTACC	200
	TITTTGTTGAA	CTATTTTGAG	ATGTCCCOCT	GTTTTGCTTC	TCTCAACAAT	250
	TGGTGTAAATA	GCTCTTCCAT	TGGAGCAGCC	AAAAGAATTT	CATCCATATA	300
	ATGAATGATG	TAGGCAGTAG	GAAATATATT	ATGAGGCTCC	TTTAAGGACT	350
	GGCCTACAAA	ACGCTGAAAT	AGCGTAGGAC	CGGTGAGCAT	GCTTTGGGGT	400
	GAAACTTTGC	TTATCCTTCT	CATGTAAGGG	TATAGTAAGG	AAATAATCTT	450
	TGAGATCTAC	TACTGTAGGA	GGTGAATCAC	TTGGAATGGC	TGCCAGGAAT	500
	GGCAGACCTT	GCTGTAATGC	ACOCATTGAT	TAAATCTGTG	CATCTAATCG	550
	CTCTCAAATC	ATGTAACAAC	TTGCTGGCAA	ATCGGTCTGT	TTTTATAGTA	600
	AAATTTTGCT	GAAACCTTAC	ACTGTGTTG	ATACATATGT	ATATGTATAT	650
	GTGTGTATAT	ATCTCTATAT	GTGTGTATAT	ATATACATAT	ACACATGCAC	700
	ACACATATGT	ATGTGTGTGC	TTCTGCTAGC	AAGGCCTTAG	AATCACTAGT	750
	GAATTC					756

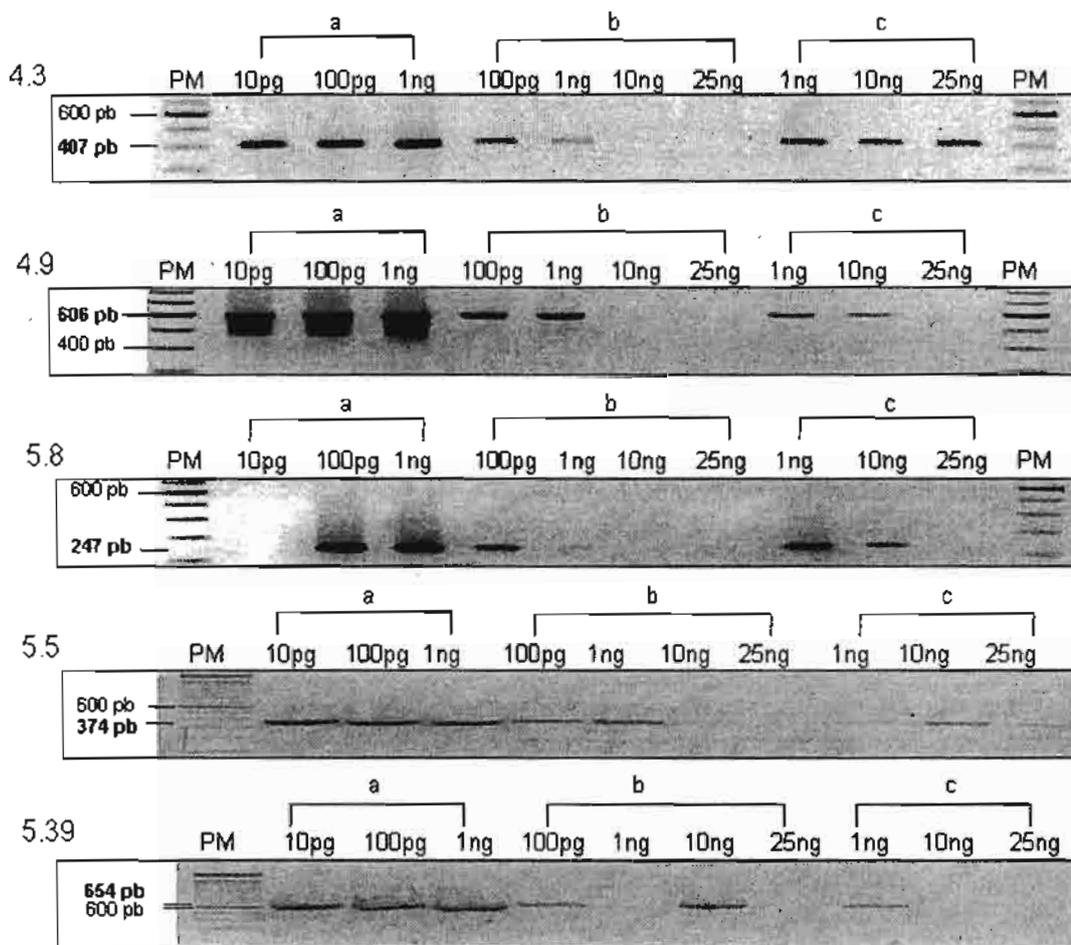
SK3	GAATTCACTA	GTGATTCTAA	GGCCTTGCTA	GCAGAAGCGG	TTCCTGGGAT	50
	CAGCCGCTTA	CCTTGAATG	CTGCAGCAGA	AGTATCTGCT	CGTCAAGCTG	100
	TTGTGCGTTG	CCTTCTATTT	CAGTCTGCCT	CTTTCTTTTT	TTCTTGAAAA	150
	TAAAGAGACA	GAGGCTGTAA	GAGTGGGTGG	GTGTTGGCTC	CAGAAATTTT	200
	AAGGCAGAAA	GCAGCAGCAG	CAGCAGCACA	CATCTGTATA	GACACTTCCA	250
	GAGACACCGG	GCACACACGC	AGTGTACCT	GCCCCACCGT	CTCTGGGTGT	300
	GTTGCAAGCA	GCAGCATTCT	CCTAACATGT	CTGCTGTAGG	CTCGAAGAGT	350
	GGGAGAAAAT	CTCCTCACTC	AGCCAGGCTG	CAAGAATTTA	ATCCAAAAGA	400
	AAAGACGGTT	TCTACACACA	CACACAGACA	CACACACACA	CACACACACA	450
	CACACACACA	CACACAGAGA	GAACCATGCT	CTGTGCAGTG	TTAATTTGGA	500
	CACAAAGAAT	AAAAATGTCT	CAATCCAAAA	GACAGAGGTT	CTTACAAACT	550
	GCAAATACCA	TAGAATAAGA	CAGTGGGAAA	TTCTGGGTGA	GATTTATTTG	600
	ATCAAATCA	CAGACATAAA	TAAGCAGACG	TTGTTTCAA	TTTTGTAGGA	650
	CAAATCAOCC	AGTTGGACAT	ACCAAACAAT	GCAAATGAAA	TGGGTTATAT	700
	TTGCTGCCAC	TAGCCTTGGT	GTAGCAITTC	ATGTGGGTGG	GGCAGCTGAG	750
	TTACGATAAA	TATGAAATCT	GGAATCTGTC	TGTCCGGGTT	CACTCCCAGC	800
	CCTACCCTC	ACTACCTGTC	TGACTTCTGA	CAAG		834
SK4	GCTTCAGCGG	CACGAACAGC	TGAAGCGCCA	ACAGGACCCA	AAGCAGAGGG	50
	ATCGTATGCA	ACAAGAGCGA	AATTGTGAAC	AGGGTGGAGA	AAAACCACT	100
	GCAATGAAAG	GTAAAAAGTG	GCACAAATCA	GACTATAAAG	GAAGAGAATT	150
	CATCATAACT	AACATCCACG	AATTTTACC	TCCCCCGGAA	TTTCAATTGG	200
	AAAAGCAGCA	AATGAAAGCA	TCACATCAGA	GACAGATATA	GCGACTGTGT	250
	TTTTATCAAC	TGCAACCAAT	CCCAAGCCTT	GAGAATGGTA	TACAATGACG	300
	CCAGTCCCAA	AGAAGTGCTG	AGAATGGACG	CCATCTAOCA	TACATGATGG	350
	TGGCACGTGA	ACCTGCAAGC	AAAAGCATGA	TAAGTCTGAA	TTAAGAGCC	400
	ACCACAAGCA	ATGGCTAAAC	ATCTCCTGAA	ACTACAAAA	TTAAGACTTG	450
	CTACTGGAGC	CCCTAAGACG	AGGATAAGAA	CTTATCCAAA	ATCAGATGGT	500
	CAAAGTCTGT	CTTCATCAA	TGGAGATAAA	TCCACACACA	CACACACACA	550
	CACACACACA	TTTGAAGAAG	CAOCCACACA	GTGTTGTCAC	CACTAGTTTA	600
	CTGATACAGT	TAAAGGTAGC	ACAATCAAAC	ACAAGTGGAA	GCAAGGAGAC	650
	GGTGATATCC	CAAAGCACA	TAAATCAAGG	TAGGGTTAAA	TTATTAAGG	700
	ACGCAGTAGG	GTGACCAATT	AATCAATAAA	CTTGGAAAAT	AATACCAATA	750
	AAATTTTTTG	AAATTACTTA	CCTCAAACAT	TACAAGAGTA	GGCTCTATCA	800
	CATGTTACGC	GTATGAAGCA	TTGGCTGATA	ATGCTGGCGG	ACTTTGATAA	850
	TCTGTCAAAA	CTGTACTGTC	CCCAITTTCTG	GTATCTTTCA	GCTCCAGTTC	900
	CCTATTATCA	AGCAATGAAT	AGTCGGAAGC	TGCCCATCA	GCTGATAAAT	950
	CTTCTCCAC	ACGTCGTTTC	TTCTTTCCAA	TATCTGATTC	ATCTTGAGCA	1000
	TGGGACCCCTT	CAGAAGAATG	TCCATCATTI	GTTTCCATGC	TAGCAACACC	1050
	ATCAATCAAC	TCTGTCTCC	TTTCTGATC	TACTTGTTC	ATTGTGACCT	1100
	CAGCAGTACA	ACCTGCTTGC	TTGGCTAGCA	CTGGATCACC	ATCGAGCGCT	1150
	TCTGCTAGCA	AGGCCTTAGA	ATCACTAGTG	AATTC		1184

ANEXO 7. Diseño de oligonucleótidos empleando el programa *Web Primer DNA* para la amplificación de regiones de ADN específicas para *P. guajava* L.

CLONA	SENTIDO	CONTRA SENTIDO	Tamaño
1.33 / 1.2A	CGG GAA AAG GGG AAA ATT CA	TGG TGC GGG AAC ATA AGA GAA	311 pb
1.1	GGA ACA TAA GAG GAC CGC CTT	CGG GAA AAA GGG GAA AAT TC	305 pb
1.12A	CAA CTT CTC CGA GCT CTT CAA	TCT GCT GAG TTA CTT GAG CGA	402 pb
1.5A	TGT TCC TTG ATG CGT ATG GA	GGG AAG AAG CAC TGA GAG AGA	548 pb
1.3A	GTG CGG GAA GAA CAA AGA AA	GGG AAG GGA AAA TGT GAT GTG	721 pb
1.11A	GGA AGA GAG ATC CGT TAT CCA	AGC AAC GGT CGG ATA GTT TT	1064 pb
2.3A	CGC TCC AAC TGT TTA AGA GCT	GGA CGG CAG TGA TTC GGT A	314 pb
3.27	CCG TCA TGA ACG TCA TGA AA	ACC CGT GAT CAC TGG CTT TAA	403 pb
3.15A	AGA CCC GTT CTT GAC TTT TGC	ACC CGT AGC TCA TAG TTA GGC	555 pb
4.1	AGC CCG TGG ATA CAT CAG TAA	CCG TCA GCA TAT CCA ATT GA	248 pb
4.38	TTC TGG GCG TGA CAA GTG AT	TGC TAA AAC CGA GCT ATG GTG	343 pb
4.3	AGC CCG TCT TCT CAG ATT TT	TTC TGG GCG TGA CAA GTG AT	407 pb
4.9	TGC GAC GTA GTG AAA TGC TAA	AAG AGC CCG TGA GGA ACA A	606 pb
5.8	CAA CGA AAT TAT CTG CCC CA	AAC GCG CAA CCC ATT AGT CAT	247 pb
5.5	AAC GCG CAA CTG AGC CTA AA	CAA CCA GGG CAA ATA GAG GTA	374 pb
5.39	CGC AAC AAC AGA ATG CTC AA	CAA CTG CAA GCG TAA AGG AT	654 pb
SK2	TGG CTC ATG CCT GTA ATC TCA	TAC CCA GGA GGC AGA AGC A	343 pb
CAG.1	TGG CTC ATG CCT GTA ATC TCA	TGC TTC TGC CTC CTG GGT A	358 pb
CAG.38 / CAG.37	TTT CAG AGA GAA CAG GGG TCA	ACA CGT GAG TGT GCA TGG GT	508 pb
CAG.20	TCT AAG GCC TTG CTA GCA GAA	GGT CAA AGT CCT CCC AAG AA	433 pb
SK3	TTTGGTATGTCCAACCTGGGTG	GGTGTGGCTCCAGAAATTT	497 pb
SK4	AGAATGGACGCCATCTACCAT	CGTATGAAGCATTGGCTGAT	509 pb

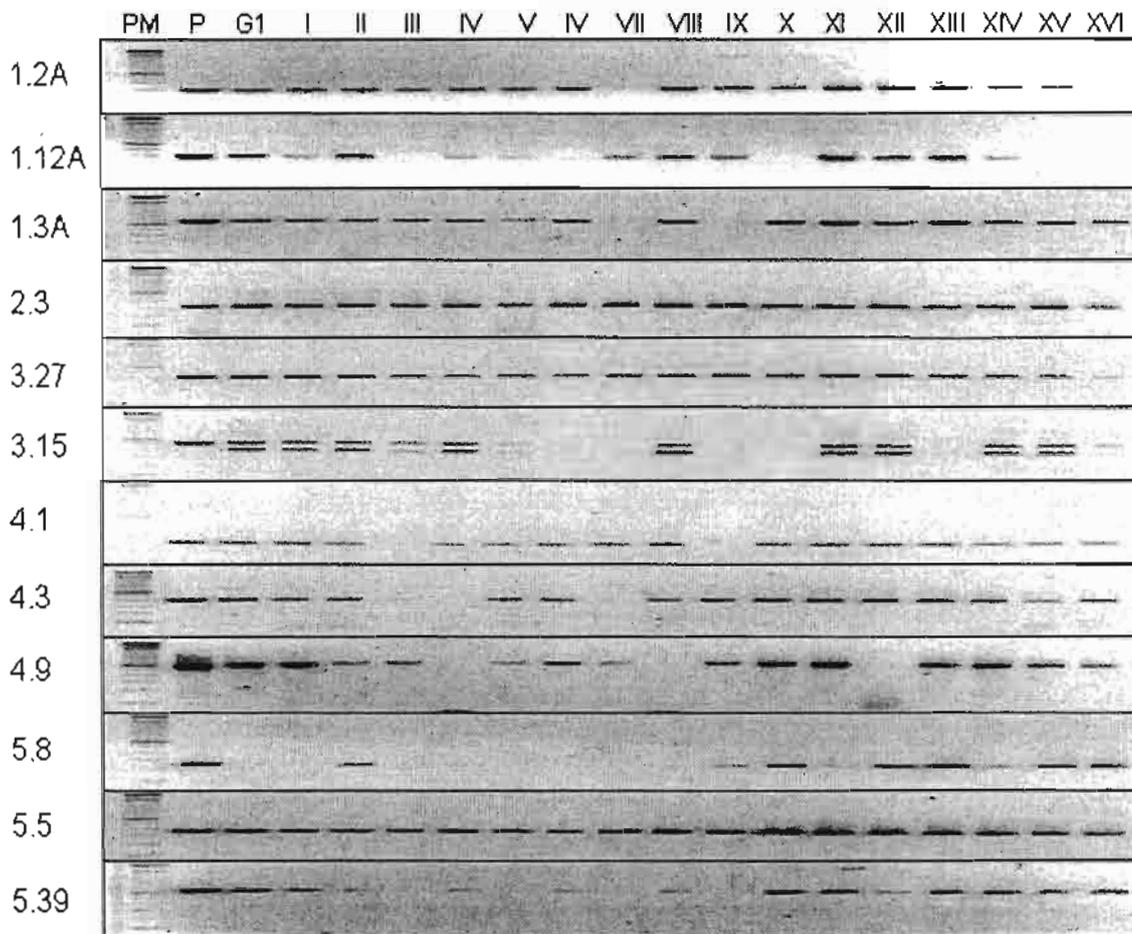
ANEXO 8. Evaluación de la concentración de ADN plasmídico y genómico como parámetro en la amplificación de segmentos específicos tipo RAPD de *P. guajava*.





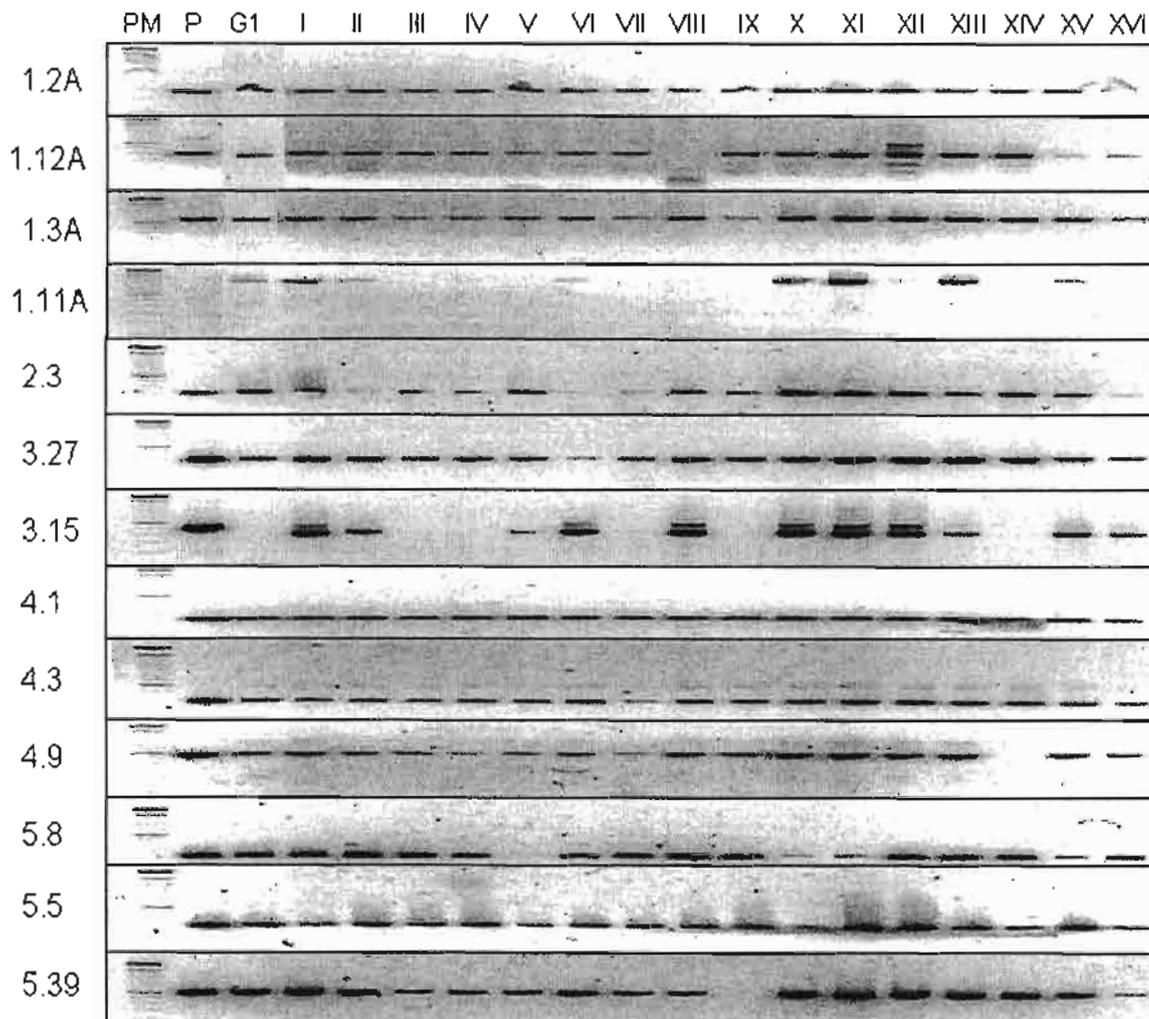
Amplificación de los segmentos 1.2A, 1.3A, 1.12A, 1.11A, 2.3, 3.27, 3.15, 4.1, 4.3, 4.9, 5.8, 5.5 y 5.39 empleando diferentes concentraciones de ADN plasmidico (a), ADN genómico de hoja fresca (b) y ADN genómico de hoja seca (c), empleando 1.0 mM de  $MgCl_2$ . PM: Marcador de 100pb (Invitrogen Life Technologies, USA).

ANEXO 9. Amplificados de segmentos específicos tipo RAPD de *P. guajava*, usando 10 ng de ADN genómico y 1.0 mM MgCl<sub>2</sub>.



PM: Marcador de 100pb (Invitrogen Life Technologies, USA); P: ADN plasmídico; I - XVI: Muestras seleccionadas (**ANEXO 4**); G1: Árbol ancestral de guayaba.

ANEXO 10. Amplificados de segmentos específicos tipo RAPD de *P. guajava*, usando 5 ng de ADN genómico y 1.5 mM MgCl<sub>2</sub>.



PM: Marcador de 100pb (Invitrogen Life Technologies, USA), P: ADN plasmídico; I - XVI: Muestras seleccionadas (**ANEXO 4**), G1: Árbol ancestral de guayaba.

# Aplicación de la biología molecular en el control de calidad de los fitomedicamentos

## Quality control of phytodrugs applying molecular biology tools

Iris A. Feria R.<sup>1,2</sup>, Marisol López L.<sup>3</sup>, Héctor Serrano<sup>4</sup>, Sandra Orozco, Xavier Lozoya<sup>5</sup>

<sup>1</sup> Laboratorio de Plantas Medicinales. Hospital de Especialidades. Centro Médico Nacional SXXI, IMSS, <sup>2</sup> Programa de Doctorado en Ciencias Biológicas, UAM-Xochimilco, <sup>3</sup> Departamento de Sistemas Biológicos, UAM-Xochimilco, <sup>4</sup> Departamento de Ciencias de la Salud, UAM-Iztapalapa, <sup>5</sup> Unidad de Investigaciones en Enfermedades Neurológicas, Hospital de Especialidades, Centro Médico Nacional SXXI, IMSS.

**RESUMEN:** en los últimos años las plantas medicinales se han utilizado en la industria de fitomedicamentos elaborados a partir de material vegetal o de sus principios activos. Actualmente, los procedimientos generales para su manufactura son reconocidos por la Secretaría de Salud u organismos internacionales como la OMS. Las técnicas actuales de control de calidad del material vegetal incluyen identificación taxonómica y huellas químicas dificultándose una adecuada selección sobre todo, si se cuenta solamente con partes de la planta, cuando la morfología es muy semejante entre los individuos o cuando los compuestos químicos de identificación son muy parecidos. Por ello, utilizar marcadores moleculares permite caracterizar regiones polimórficas del ADN para autenticar especies, variedades o alguna variación importante independientemente si se cuenta con el individuo completo o sólo una parte, contribuyendo a mejorar el control de calidad de los fitomedicamentos.

**ABSTRACT:** in recent years, medicinal plants have been used in the phytodrug industry directly from the plant drug or their active compounds. At this time, general manufacturing procedures are recognized both from national authorities like the National Health Secretaria and international organisms like the World Health Organization. Techniques for quality control from raw material include taxonomical identification and chemical fingerprinting showing some limitations as is the case when only parts of the plant are available, morphological aspects are very similar to other plants, or when chemical specific compounds are shared. The use of molecular markers allows the characterization of polymorphic DNA regions useful for species, variety or some individual characteristic authentication that do not depend on the availability of the whole organism. The use of these techniques could improve the quality control needed for phytodrug manufacture.

**Palabras claves:** marcador genético, control de calidad, droga vegetal **Key words:** genetic marker, quality control, plant drug.

### Correspondencia:

Iris A Feria Romero  
Laboratorio de Plantas Medicinales, UIEN,  
Hospital de Especialidades,  
Centro Médico Nacional Siglo XXI, IMSS.  
Avenida Cuauhtémoc 330, Col. Doctores  
México, D.F. CP 06720  
Teléfono/Fax (55) 56276900 ext. 21367  
e-mail: irisferi@yahoo.com.mx.

Fecha de recepción: 24 de marzo de 2006  
Fecha de aceptación: 13 de agosto de 2006

## Introducción

Actualmente, las plantas medicinales tienen gran importancia en el área de salud, captando la atención de instancias internacionales como la Organización Mundial de la Salud que señala que el 80% de la población de países en desarrollo resuelve sus problemas de salud mediante la medicina alternativa<sup>1</sup>. Esta práctica ha logrado incrementar considerablemente el número de remedios herbolarios y medicamentos elaborados a partir del material vegetal o de sus extractos.

La demanda de material vegetal ha inducido el cultivo de especies con importancia terapéutica en diversas zonas geográficas, propiciando una variabilidad en la concentración de los principios activos provocada por un medio ambiente diferente. Además, la explotación comercial y necesidad económica por parte de los proveedores, ha favorecido la confusión, sustitución y adulteración con otras especies botánicamente similares a la planta de interés.

Debido a lo anterior, los países y sus instancias gubernamentales han desarrollado documentos oficiales o farmacopeas herbolarias donde se recopila toda la información reglamentaria, los métodos para el control de calidad de la parte de la planta con uso medicinal o droga vegetal y sus monografías. Particularmente, las monografías descritas en la Farmacopea mexicana se enfocan en la descripción de la droga vegetal, ensayos de identidad, análisis químicos y conservación; variando su contenido de acuerdo al conocimiento que se tenga de la droga vegetal.<sup>2</sup>

Por otro lado, tanto la Organización Mundial de la Salud como diversas asociaciones científicas alrededor del mundo que se relacionan con la farmacognosia de las plantas medicinales y fitoterapia, han publicado monografías no oficiales con información técnica y científica.

A manera de ejemplo podemos mencionar algunos puntos que se consideran más sobresalientes en este tipo de monografías: 1) nombre científico y variedad, 2) nombre común, 3) origen botánico, 4) sinónimos, 5) parte utilizada de la planta, 6) principales componentes químicos, 7) ensayos macro y microscópicos, 8) acciones farmacológicas, 9) indicaciones, 10) interacciones con otros medicamentos, 11) contraindicaciones, 12) efectos secundarios, 13) posología, 14) presentaciones farmacéuticas comerciales simples y compuestas, 15) adulteraciones, 16) conservación, 17) bibliografía.<sup>3,4</sup>

Como se puede observar, los aspectos más sobresalientes en el control de calidad del material vegetal son botánicos, morfológicos y químicos, entre otros, los cuales proporcionan los caracteres de autenticación de la materia prima.

En una primera instancia, la información botánica y morfológica se obtiene de la identificación taxonómica proporcionada por

el ejemplar de interés colectado en campo. Una vez corroborada la especie, se realiza una serie de descripciones macro y microscópicas de la droga vegetal; el nivel de la descripción dependerá de la integridad del material vegetal y la similitud con sus adulterantes. La información microscópica proporciona una información más completa y detallada en comparación con la macroscópica, sobre todo en muestras donde la parte de la planta está incompleta. Los especialistas en esta técnica hacen uso de técnicas histológicas con la finalidad de describir las estructuras celulares propias de la especie<sup>5</sup>. En un trabajo realizado por Rivera-Arce y col.,<sup>6</sup> en hojas de *Psidium guajava*, se proporcionan los datos anatómicos y morfológicos de esta droga vegetal para su autenticación, dando pauta a un primer acercamiento al control de calidad en el desarrollo de un fitomedicamento a partir de esta especie.

El método químico por su parte permite la autenticación de la especie y adecuada selección del material vegetal al identificar el conjunto de compuestos químicos provenientes del extracto e identificados por un perfil cromatográfico. Las principales técnicas utilizadas para este propósito son cromatografía en capa fina (TLC, *Thin Layer Chromatography*), cromatografía de líquidos (HPLC, *High Performance Liquid Chromatography*) y la cromatografía unida a la espectrometría de masas (HPLC-MS, *High Performance Liquid Chromatography mass spectrometry*). La información obtenida se considera confiable debido a que los individuos de la misma especie y que sean cultivados en la misma zona geográfica presentan un perfil cromatográfico muy similar como consecuencia de crecer en un mismo nicho ecológico, lo que favorece una adecuada acumulación de principios activos<sup>7</sup>. En la descripción de este fenómeno, Yan y col.,<sup>8</sup> emplean la cromatografía de gases para la detección de compuestos volátiles presentes en el fruto de *Gardenia jasminoides*, viéndose afectados por la época de año y la región geográfica en que se cosecha. Este fruto es utilizado en la medicina herbolaria para el tratamiento de encefalitis, hepatitis, entre otros padecimientos. De igual forma, Lago y col.,<sup>9</sup> utilizan la cromatografía de gases con y sin espectrometría de masas para la identificación de aceites volátiles en hojas de *Guarea macrophylla* ssp. colectadas en diferentes épocas del año. El efecto farmacológico del extracto crudo y fracciones sesquiterpénicas se midió a través de la respuesta eléctrica de antenas aisladas de polillas, con el fin de determinar la mayor actividad en las muestras colectadas.

Una vez identificadas las huellas químicas por el perfil cromatográfico, estas son utilizadas para determinar la autenticidad, calidad, seguridad y eficacia de la materia prima antes de convertirse en un medicamento elaborado a partir de componentes de origen vegetal o fitomedicamento. En un trabajo realizado por dos Santos y col.,<sup>10</sup> la cuantificación de las huellas químicas de la cumarina y el ácido *o*-cumárico en hojas del complejo guaco (*Mikania glomerata* y *M. laevigata*) permitió establecer que el proceso de liofilización disminuía la concentración de estos compuestos en un 50 y 60% respectivamente en comparación con el extracto acuoso fresco, detectar la presencia de un nuevo com-

puesto químico como consecuencia de la liofilización, la ausencia en la actividad anti-inflamatoria y hemorragia en el pulmón del modelo murino; todos estos aspectos repercuten en elaboración de un fitomedicamento a partir del extracto liofilizado. Las hojas del guaco son utilizadas en la medicina tradicional brasileña en forma de extracto, jarabe o infusión para el tratamiento de asma y bronquitis.

De forma adicional, las huellas químicas también permiten la identificación de adulterantes considerada una práctica común entre los proveedores que desean obtener un mayor beneficio económico, en especial cuando escasea la planta o droga vegetal de interés. Esta identificación se realiza cuando se observa alguna alteración importante al comparar los perfiles cromatográficos entre el extracto del material vegetal de interés y sus posibles adulterantes. En este contexto, Schaneberg y col.,<sup>11</sup> se basaron en el perfil cromatográfico generado por los alcaloides de efedrina presentes en *Ephedra sinica*, para detectar especies del mismo género usados como adulterantes biológicos o la adición química como la metanfetamina. Esta planta utilizada como infusión en la medicina tradicional china como estimulante y antiasmático, actualmente tiene gran demanda por su efecto en la pérdida de apetito y pérdida de peso.

A pesar de los avances tecnológicos en la caracterización morfológica y química del material vegetal, estas técnicas presentan limitaciones cuando la muestra que se pretende analizar se compone de individuos de diferentes especies con una estrecha cercanía filogenética. En estos casos, los ejemplares poseen estructuras celulares o perfiles cromatográficos muy similares, distinguibles sólo por la presencia de diferentes partes de la planta. Por lo anterior, diferentes grupos de investigación han aplicado herramientas de la biología molecular con el fin de proporcionar una información más clara en la corroboración de especies e identificación de adulterantes.

En la tabla 1 se hace una comparación entre las metodologías descritas en esta revisión y sus técnicas correspondientes, el nivel de identificación, ventajas, desventajas, así como el costo relativo de cada una de ellas.

#### Corroboración taxonómica o autenticación del material vegetal

El ADN (ácido desoxirribonucleico) contiene toda la información genética capaz de identificar al individuo de una especie, ya que en esta molécula se encuentra almacenada la información necesaria para la expresión de características físicas o morfológicas (fenotipo) que los botánicos utilizan como claves para la identificación taxonómica. El ADN se encuentra en el núcleo y en algunos organelos de las células (mitocondria y cloroplasto) y es el mismo en cualquier célula de la planta, lo que permite la identificación de la especie independientemente de si se cuenta o no con el ejemplar de herbario completo o con la droga vegetal

de interés, además de que su estructura física de doble cadena súper enrollada le confiere una gran estabilidad a la exposición de factores ambientales.<sup>12</sup>

#### A) Identificación de la especie a partir de material vegetal fresco

Actualmente existen diferentes sistemas comerciales que permiten la extracción del ADN a partir de material fresco, sin embargo, no hay que descartar el uso de técnicas especiales que contemplan un alto contenido de polifenoles, polisacáridos o mucopolisacáridos presentes en el material vegetal y que interfieren con su pureza<sup>13,15</sup>. Una vez que se ha extraído el ADN, se corrobora la integridad del material genético con el objeto de garantizar la reproducibilidad de los resultados, ya que un fraccionamiento parcial o total del ADN modifica el patrón de expresión visual al momento de la separación electroforética. Posteriormente, el ADN es fraccionado bajo un procedimiento controlado dependiendo de la técnica seleccionada, que se basa principalmente en la utilización de enzimas de restricción (endonucleasas que cortan la doble cadena de ADN) y/o la reacción en cadena de la polimerasa (PCR, *polymerase chain reaction*)<sup>16</sup>.

1. RFLP (*Restriction Fragment Length Polymorphisms* o Polimorfismo en los Fragmentos de Restricción). Esta técnica se ha utilizado en plantas para estudios filogenéticos de diversidad genética e identificación de cultivos con el propósito de proteger a las diferentes variedades existentes utilizando ADN nuclear, mitocondrial y de cloroplasto con buenos resultados<sup>17,18</sup>. Este método se basa en la digestión del ADN total empleando diferentes enzimas de digestión. Los fragmentos obtenidos son separados por electroforesis, transferidos a una membrana y analizados por hibridación con sondas marcadas. Estas sondas se diseñan para detectar pérdidas, inserciones o sustituciones de secuencias o nucleótidos únicos, lo que puede visualizarse como la presencia o ausencia de bandas de ADN que evidencian la ganancia o pérdida de sitios reconocidos por la enzima de restricción. Este método permite estudiar más ampliamente al ADN, ya que la enzima de restricción sólo reconoce el sitio de corte y no la funcionalidad de la secuencia, con lo que pueden detectarse cambios genéticos puntuales o modificaciones en las bases, como sucede en la metilación. Los fragmentos generados son ilimitados y no se afectan por el medio ambiente, por el desarrollo de la planta o por efectos pleiotrópicos (condición por la cual un gen tiene efectos múltiples en la vida de un organismo), se heredan en forma mendeliana simple, se identifica al *locus* y sus variaciones o alelos, y la expresión de estos marcadores es codominante, por lo que se distinguen todas las formas alélicas del gen. La principal desventaja de la técnica es la selección de la(s) enzima(s) de restricción adecuada(s), el diseño de las sondas, la detección y análisis de los fragmentos digeridos.

2. RAPD (*Random Amplified Polymorphic DNA* o Fragmentos polimórficos de ADN amplificados al azar). Esta técnica se basa

Tabla 1. Comparación entre los métodos morfológicos, químicos y genéticos, y las técnicas utilizadas en el control de calidad de la materia prima

MÉTODO	NIVEL DE IDENTIFICACIÓN	VENTAJAS	DESVENTAJAS	COSTO
<b>MORFOLÓGICO</b>				
MACROSCÓPICO	género / especie	Accesibilidad, no requiere equipo especializado	Nivel de identificación bajo	\$
MICROSCÓPICO	especie	Accesibilidad, de baja complejidad técnica	Uso de solventes, no distingue fácilmente adulterantes	\$\$
<b>QUÍMICO</b>				
TLC	especie	Accesibilidad, técnicas simples	Uso de solventes, confusión en la identificación de compuestos	\$
HPLC	especie	Cuantificación e identificación de principios activos, época del año y región geográfica.	Equipo especializado, uso de solventes, métodos de extracción específicos	\$\$
HPLC-MS	especie	Cuantificación e identificación de principios activos, época del año y región geográfica.	Equipo especializado, uso de solventes, métodos de extracción específicos	\$\$\$\$
<b>GENÉTICO</b>				
<b>NUCLEAR</b>				
RFLP	especie	Relativa accesibilidad, uso de enzimas de restricción	Confusión en la identificación de fragmentos, requiere µg de DNA.	\$\$
RAPD	especie	Relativa accesibilidad, Técnica PCR	Reproducibilidad en el método sin estandarizar	\$\$
AFLP	especie	Mayor especificidad en comparación a RFLP y RAPD	Especificidad en la identificación de fragmentos, cantidades mayores de DNA por el uso de enzimas de restricción	\$\$\$
Región intergénica 5s-RNA	especie	Amplificación de región específica	Requiere una secuenciación inicial y el uso de enzimas de restricción.	\$\$\$
<b>PLASTIDOS</b>				
Región intergénica trnL/trnF (cloroplasto)	especie	Amplificación de región específica	Requiere secuenciación	\$\$\$\$
Región atpF/atpA (plástido)	variedad	Se identifican variedades	Requiere secuenciación	\$\$\$\$

en la reacción en cadena de la polimerasa, mejor conocida como PCR, que bajo condiciones *in vitro* imita al medio celular para permitir la replicación del ADN. Sin embargo, en la técnica RAPD se busca la amplificación de múltiples fragmentos y se logra mediante la hibridación aleatoria de una pequeña cadena de nucleótidos o cebador, clásicamente de 10 pares de bases (pb), con el ADN total a temperaturas relativamente bajas (36° C), lo que permite una interacción más laxa. De esta forma, los productos de amplificación serán polimórficos cuando se ha producido una pérdida, inserción o cambio de un solo nucleótido en la cadena de ADN molde. Al igual que en el caso de los RFLP, los fragmentos son separados por electroforesis y las bandas con diferente peso molecular representarán diferentes *loci* de herencia dominante.<sup>19,20</sup>

La obtención de un mayor número de fragmentos de diferentes partes del ADN se logra empleando diferentes combinaciones en la secuencia del fragmento cebador. En vegetales, las secuencias amplificadas indican su herencia dominante, su habilidad para detectar regiones de ADN altamente variables (5-10 *loci* por fragmento cebador), su potencialidad en la localización de genes, identificación de variedades, estudios de hibridación inter e intraespecífica y el estudio de la variación genética en poblaciones emparentadas.<sup>21,22</sup> Las principales ventajas de esta técnica son la rapidez en la obtención de resultados, el bajo costo, una menor inversión en equipo y la baja cantidad de ADN que se requiere. Sin embargo su principal desventaja es la inconsistencia de datos, ya que la reproducibilidad se afecta con pequeñas variaciones en las condiciones de amplificación. Esta desventaja se puede minimizar al estandarizar las condiciones de reacción y de amplificación, usando un control interno para asegurar la reproducibilidad de los productos de la amplificación y llevando un registro de las bandas nítidas y consistentes.<sup>23</sup> Mi y col.,<sup>24</sup> utilizaron esta técnica en la diferenciación de 6 especies de *Artemisa* con 18 cebadores aleatorios de 10 oligonucleótidos cada uno. El análisis polimórfico de los fragmentos generados permitió determinar la distancia filogenética entre las especies e identificar el oligonucleótido capaz de discriminarlas. Las plantas del género *Artemisa* son utilizadas en el tratamiento de una gran variedad de padecimientos que van desde la protección celular de las úlceras pépticas, protección del hígado, anti-malárico y efectos antitumorales.

2. AFLP (*Amplified Fragment Length Polymorphisms* o polimorfismo de fragmentos amplificados de ADN). Esta técnica combina la digestión del ADN empleando diferentes enzimas de restricción, la extensión de fragmentos de secuencia conocida o ligación de adaptadores que se usan como ancla de los nucleótidos del sitio de corte, la amplificación selectiva de uno de los fragmentos de restricción por medio de cebadores complementarios a los adaptadores y por último, la separación electroforética. El polimorfismo es detectado por la presencia o ausencia de fragmentos y son normalmente de herencia domi-

nante. A diferencia de las dos técnicas anteriores, la eficiencia en la detección de regiones polimórficas puede ser mayor, ya que es posible detectar una variación considerable del genoma empleando un par de cebadores. Esta técnica se ha empleado en el estudio de la diversidad genética de diferentes plantas, en la saturación de mapas genómicos y en la identificación de fragmentos ligados a genes de interés comercial.<sup>25,26</sup> Entre sus ventajas está la identificación de hasta 50 *loci* en poco tiempo y la reproducibilidad, ya que no depende de las condiciones de la reacción siempre y cuando se emplee un ADN de buena calidad en el proceso de digestión. Su principal desventaja es la interpretación y análisis de datos cuando no se cuenta con un analizador de geles.<sup>27</sup> Esta técnica se ha utilizado para la identificación de la *Radix Quinquefolii*, *Radix Astragali*, *Radix Notoginseng*, *Cortex Cinnamomum*, *Radix Isatidis*, *Radix Codonopsis* y *Radix Remana*, a partir de las cuales se desarrollan siete drogas vegetales comúnmente utilizadas en la medicina tradicional china. Los fragmentos de ADN obtenidos después de la digestión con EcoRI/MseI y la amplificación con 16 oligonucleótidos específicos fueron analizados para determinar el índice de similitud entre las especies seleccionadas.<sup>28</sup>

3. Región intergénica ARNr-5S (ácido ribonucleico 5S). En los eucariotes superiores, el gen ARNr-5S se compone de una región codificable de unidades repetidas (cistrón) de aproximadamente 120 pb, separadas por un espacio de aproximadamente 300 pb. El gen ARNr-5S es considerado altamente conservado, pero el espacio que separa las regiones es variable entre las diferentes especies. Esta característica permite utilizarlo como un método de identificación. La técnica consiste principalmente en la utilización de cebadores conservados que delimiten la región intergénica ARNr-5S de la planta de estudio para llevar a cabo su amplificación y secuenciación. Posteriormente, se pueden utilizar enzimas de restricción que relacionen el producto digerido con la identificación de la especie. En un estudio realizado a 25 individuos, en donde se incluían especies y variedades del género *Fritillaria*, la amplificación específica de la región intergénica ARNr-5S y su digestión con la enzima EcoRI permitió identificar a *Fritillaria cirrosa* de *F. pugiensis*, *F. anhuiensis* y *F. thunbergii* una vez que el fragmento de interés fue secuenciado en cada una de las especies.<sup>29</sup> Como dato interesante *Fritillaria cirrosa* es muy apreciada por sus propiedades antitusivas y expectorantes, por lo que su bulbo llega a ser hasta 100 veces más caro que el de *F. thunbergii*, aunque su toxicidad sea mucho mayor, de ahí la importancia en la identificación del material vegetal.

#### B) Identificación de la especie a partir de material vegetal seco o procesado

El conocimiento cada vez más extenso sobre las proteínas y genes involucrados en el funcionamiento de las plantas ha permitido identificar con mayor facilidad tanto secuencias codificables (exones) como no codificables (intrones y regiones

intergénicas) en un mayor número de especies<sup>30</sup>. Las regiones codificables son muy apreciadas en el campo de la biología, química, bioquímica y medicina ya que permiten dilucidar vías metabólicas y enfermedades relacionadas con la presencia o ausencia de una sustancia o proteína específica. En el caso de las regiones no codificables se desconocía o ignoraba su función, ahora se sabe que muchas de las regiones no codificables permiten el ordenamiento de exones, como sucede con los intrones, y logran una distribución adecuada de los genes en el cromosoma, como es el caso de las regiones intergénicas. Otro ejemplo son los telómeros, regiones de ADN no codificante altamente repetitivas presentes en los extremos de los cromosomas y cuya función principal es la estabilidad estructural de los cromosomas en las células con núcleo, la división celular y el tiempo de vida de las estirpes celulares<sup>31</sup>. Recientemente, se han localizado en regiones no codificantes segmentos cuya transcripción a ARN permite la regulación fina de la expresión genética: ARN interferente (ARNsi) y microRNA (ARNmi).<sup>32</sup>

En la práctica, la amplificación de regiones no codificantes ha permitido la identificación de especies cuando el ADN está muy fraccionado, como sucede en muestras secas de plantas. Esto se debe a tres razones principales: 1) el tamaño del producto amplificado no se ve afectado por la degradación parcial del ADN, 2) los oligonucleótidos flanqueantes son altamente conservados y 3) las regiones amplificadas presentan una variación franca entre secuencias de diferentes especies.

En un trabajo encaminado a la identificación de seis especies de *Panax*, se utilizó la región intergénica ribosomal ITS1-5.8S-ITS2 a partir de raíces secas<sup>33</sup>. Cabe destacar que mientras *P. ginseng* y *P. quinquefolius* L. son utilizadas como agentes tónicos, profilácticos y anti-envejecimiento; *P. notoginseng* es utilizado como hemostático en hemorragia; *P. japonicus* es utilizado como antitúxico leve y en la reducción de flemas; *P. trifolius* L. y *P. major* no presentan estudios documentados de sus propiedades terapéuticas. Una vez amplificada la región ribosomal ITS1-5.8S-ITS2 de ADN en cada una de las especies analizadas, se procedió a identificar las diferencias empleando enzimas de restricción específicas en sitios de corte conocidos. En este estudio, los fragmentos generados lograron identificar a cada una de las especies.

#### - Selección del material vegetal idóneo

Los individuos agrupados en una misma especie comparten características morfológicas semejantes. Sin embargo, en ocasiones suelen presentar variaciones tan evidentes, causadas en gran parte por el medio ambiente, que se clasifican como variedades. Hosokawa y col.,<sup>34</sup> hicieron una comparación de la región intergénica de los genes *atpF* y *atpA* del ADN plastídico obtenida por amplificación entre 3 variedades de la especie *Angelica acutiloba* cuya raíz es utilizada en la medicina tradicional

japonesa. El interés en la diferenciación de estas variedades se debe a que las raíces secas de *A. acutiloba* var. *acutiloba* es más eficaz como inmunopotenciador que la *A. acutiloba* var. *sugiyama*, por otro lado las diferencias morfológicas entre *A. acutiloba* var. *sugiyamae* y *A. acutiloba* var. *iwatensis* son muy similares dificultando su discriminación.

No obstante, existe otro tipo de variaciones cuyas manifestaciones físicas no son suficientes como para poder considerar al organismo como variedad. En este caso, se le puede considerar una variación individual que se refleja como la acumulación de metabolitos secundarios en plantas y forma parte del acervo propio del germoplasma.<sup>35</sup>

En la selección del material vegetal para la producción de fitomedicamentos se buscan las condiciones ambientales óptimas para que la droga vegetal presente la mayor concentración de principios activos. Es por ello que, en el diseño del fitomedicamento, la parte inicial resulta una labor ardua donde se contemplan diferentes etapas geográficas, épocas del año, estados fenológicos de la planta y, si es posible, variedades de la especie. Actualmente, la selección de estos individuos sólo es posible mediante la cuantificación química del metabolito secundario, sin la garantía de que esta propiedad se herede a su descendencia o se conserve durante el proceso de propagación, como sucede al aplicar herramientas biotecnológicas como la embriogénesis somática *in vitro*. La característica sobresaliente de este proceso, por el cual se forman embriones a partir de células vegetativas, es que los embriones somáticos o semillas artificiales presentan la potencialidad de producir duplicados de un genotipo específico, mejor conocidos como clones.<sup>36</sup>

Las técnicas de biología molecular, por su parte, identifican fragmentos del ADN conocidos como marcadores moleculares, que tienen una relación directa con la característica de interés, por ejemplo, la acumulación de metabolitos secundarios, entre individuos que contengan la propiedad deseada y los que carezcan de ella. Una vez identificado el marcador genético se puede seleccionar adecuadamente a los individuos en campo, a su descendencia o a aquellos que fueron propagados biotecnológicamente sin necesidad de cuantificar el compuesto químico.

#### - Identificación de adulterantes vegetales o contaminantes biológicos

De forma complementaria, la aplicación de técnicas como RFLP, RAPD, AFLP o la amplificación de regiones específicas del ADN con o sin la identificación de huellas químicas, han permitido autenticar la materia prima cuando se utilizan forma rutinaria para el control de calidad del producto herbolario. Tal es el caso de un trabajo realizado en *Angelica sinensis* y sus adulterantes más comunes: *A. acutiloba* y *A. gigas*. La raíz de *A. sinensis* se ha utilizado en la medicina tradicional china por

más de 2000 años en padecimientos relacionados con el sistema circulatorio; actualmente es empleada en el tratamiento de desórdenes menstruales, modulación del sistema inmunológico y como emoliente y laxante de la constipación crónica. Utilizando al ácido felúrico y Z-ligustilida como huellas químicas exclusivos para control de calidad, por su abundancia e independencia de la región geográfica en *A. sinensis*, el estudio parecía contar con una muestra de características adecuadas. El marcador genético empleado fue la región intergénica ARNr-5S empleando cebadores flanqueantes conservados, donde se observaron diferencias significativas en las secuencias de las 3 especies identificadas con enzimas de restricción únicas para cada una<sup>37</sup>.

También existen trabajos donde la sola aplicación de marcadores moleculares ha permitido la identificación de adulterantes, como es el caso de *Dendrobium officinale*. Los tallos de esta planta se utilizan ampliamente en la medicina tradicional china y tienen aplicaciones tan variadas como el tratamiento de padecimientos estomacales hasta la prolongación de la vida. Su costo llega a ser hasta de 3000 dólares el kilo, por lo que es muy cotizada en el mercado. Los adulterantes que se utilizan son tallos del mismo género: *D. devonianum*, *D. loddigessi*, *D. moniliforme* y *D. gratiosissimum*, presentando una apariencia y estructura similar que impide la identificación de las especies si se utilizan criterios morfológicos. En este caso, se empleó la amplificación de la región compuesta por los intrones ITS1 e ITS2 (ADNr ITS), la cual separa a los genes 18S, 5.8 y 26S del ARN ribosomal, para la diferenciación e identificación taxonómica de especies, ya que presenta una variación polimórfica clara entre individuos del mismo género<sup>38</sup>. De manera análoga, se ha podido discriminar a *Pinellia ternata* de sus adulterantes al amplificar el gen de la proteína MBL (*mannose binding lectin*) con cebadores específicos que contienen sitios de corte para las enzimas *Bam*HI y *Nco*I. Tanto el diseño de los cebadores como la selección de los sitios de corte se obtuvieron al secuenciar la región seleccionada en cada una de las especies de interés<sup>39</sup>. La Farmacopea china describe que los tubérculos de *Pinellia ternata* son utilizados como antitúxico y expectorante.

De forma complementaria, se han desarrollado marcadores genéticos que amplifican regiones específicas del cloroplasto como la del intrón *trnL* y su espacio intergénico (secuencia *trnL/trnF* del cloroplasto) para la identificación de especies. La principal justificación en la aplicación de esta región es la estabilidad en la transmisión de información genética de esta región debido a su herencia materna. Este marcador fue utilizado en la identificación del complejo rubarbo, el cual se compone de los rizomas y raíces de tres especies principales del género *Rheum*: *R. palmatum* L., *R. tanguticum* Maxim. ex Balf. y *R. officinale* Baill. Este grupo de plantas se utiliza en la medicina tradicional china por sus efectos laxantes, antiflogístico y hemostáticos, y sus adulterantes son generalmente *R. botaoense*, *R. compactum*, *R. undulatum* y *R. emodi*. La principal diferencia en el procedimiento para la obtención de las secuencias polimórficas, es la

utilización de cebadores flanqueantes que identifican la región *trnL/trnF* del cloroplasto a pesar de que se extrae ADN total. De igual forma, los resultados permitieron identificar a los individuos de las diferentes especies en estudio.<sup>40</sup>

## Conclusiones

Debido al auge de la medicina alternativa como una opción de la población para resolver sus problemas de salud, la explotación y comercialización de las plantas medicinales también se ha incrementado considerablemente.

Sin embargo, la mayor demanda en el uso de plantas medicinales o alguna de sus partes ha provocado su uso indiscriminado, una aplicación terapéutica inespecífica, el desconocimiento de las buenas prácticas de recolección y su adulteración propiciando la desconfianza entre la población. Debido a lo anterior se ha creado la necesidad de establecer parámetros de calidad cada vez más exigentes, por lo que se ha hecho necesario aplicar técnicas y conocimientos de diferentes áreas para el desarrollo de fitomedicamentos con el fin de garantizar la autenticación y recolección de la especie, estandarización del principio activo y eficacia terapéutica. Hasta el momento, sólo la identificación macro y microscópica del material vegetal y el análisis de perfiles cromatográficos son las técnicas aceptadas por las diferentes instancias gubernamentales.

Hasta el momento y a pesar de lo ampliamente fundamentadas y evaluadas que están, las técnicas de biología molecular no han sido contempladas como parámetros de control de calidad, ni tampoco son descritas en las farmacopeas o monografías de plantas medicinales, limitando entonces la incorporación de la información que brindaría esta herramienta útil en la autenticación y selección del material vegetal idóneo, en la identificación de adulterantes y contaminantes microbianos o para predecir la existencia de compuestos químicos.

Las técnicas de biología molecular, en especial los marcadores moleculares, son una alternativa en la identificación de especies, independientemente de si se cuenta con el individuo completo o solamente unas cuantas células. El desarrollo de estos marcadores aplicados a las plantas medicinales se enfoca a regiones que son polimórficas entre especies pero que pueden ser amplificadas gracias a lo conservado de sus extremos flanqueantes y evidenciadas gracias al corte específico de la región polimórfica. Así pues, existen ejemplos donde la amplificación de la región intergénica ARNr-5S, la región ADNr ITS o la secuencia *trnL/trnF* del cloroplasto, en combinación con el uso de enzimas de restricción, han permitido la autenticación de especies y sus adulterantes ya sea a partir de muestras frescas, secas o procesadas. Es de esperarse que estas herramientas puedan ser incorporadas una vez que sea mayor la cantidad de laboratorios en los cuales converjan el desarrollo de fitofármacos y la certeza científica aplicada a este campo.

## Bibliografía

- Romero-Cerecero O., Reyes-Morales H., Herrera-Arellano A., Lozoya-Legorreta X., Tortoriello J. 2004. Aceptación de los fitofármacos por médicos y pacientes en clínicas de atención primaria. *Revista Médica IMSS*, 42(2):125-130.
- Métodos generales de análisis de la Farmacopea Herbolaria y extrafarmacopea. 2001 En: *Farmacopea herbolaria de los Estados Unidos Mexicanos*. Secretaría de Salud. FEUM, México, pp. 27-69.
- WHO monographs on selected medicinal plants. 1999. En: *World Health Organization*, Geneva, vol 1 y 2, pp. 16-285 y 12-349.
- Vanaclocha B., Cañigueral S., Vila R., Risco E., Pérez F., Portillo A., Freixa B., Milo B., Alonso J. R., Rios J. L., Morales M. A., Bachiller L. I. 2003. Monografías de plantas medicinales. En: Vanaclocha B., Cañigueral S. (eds.), *Fitoterapia. Vademécum de prescripción*. 4ª ed. MASSON, S.A., Barcelona, pp. 81-511.
- Ruzin S. E. 1999. Infiltrating and embedding tissues. En: *Plant microtechnique and microscopy*, Oxford University Press, New York, pp. 61-85.
- Rivera-Arce E., Gatusso M., Lozoya X. 2003. Anatomical identity parameters of the crude drug *Psidium guajavae folium*. *Pharmaceutical Biology*, 42(7): 516-521.
- Drasar P., Moravcova J. 2004. Recent advances in analysis of Chinese medical plants and traditional medicines. *Journal of Chromatography B*, 812: 3-21.
- Yan S., Xin W., Luo G., Wang Y., Cheng Y. 2006. Chemical fingerprinting of *Gardenia jasminoides* fruit using direct sample introduction and gas chromatography with mass spectrometry detection. *Journal AOAC International*, 89(1): 40-45.
- Lago J. H., Soares M. G., Batista-Pereira L. G., Silva M. F., Correa A. G., Fernandes J. B., Vieira P. C., Roque N. F. 2006. Volatile oil from *Guarea macrophylla* ssp. *tuberculata*: seasonal variation and electroantennographic detection by *Hypsipyla grandella*. *Phytochemistry*, 67(6): 589-594.
- dos Santos S. C., Krueger C. L., Steil A. A., Kreuger M. R., Weber B. W., Wisniewski J. A. 2006. LC characterization of guaco medicinal extracts, *Mikania laevigata* and *M. glomerata*, and their effects on allergic pneumonitis. *Planta Medica*, 72: 679-684.
- Schaneberg B. T., Crockett S., Bedir E., Khan I. A. 2003. The role of chemical fingerprinting: aplicación to *Ephedra*. *Phytochemistry*, 62: 911-918.
- Singer M., Berg P. 1993. Las bases moleculares de la herencia: una visión general. En: *Genes y genomas*. Ediciones Omega, S.A., Barcelona, pp. 2-72.
- Friar E. A. 2005. Isolation of DNA from plants with large amounts of secondary metabolites. *Methods in enzymology*, 395: 3-14.
- Wulff E. G., Torres S., González V. E. 2002. Protocol for DNA extraction from potato tubers. *Plant Molecular Biology Reporter*, 20: 187a-187e.
- Tel-Zur N., Abbo S., Myslabodski D., Mizrahi Y. 1999. Modified CTAB procedure for DNA isolation from Epiphytic cacti of the genera *Hylocereus* and *Selenicereus* (Cactaceae). *Plant Molecular Biology Reporter*, 17: 249-254.
- Becerra V., Paredes C. M. 2000. Uso de marcadores bioquímicos y moleculares en estudios de diversidad genética. *Agricultura técnica*, 60(3): 270-281.
- Desplanque B., Viard F., Bernard J., Forcioli D., Saumitou-Laprade P., Cuguen J., Van Dijk H. 2000. The linkage disequilibrium between chloroplast DNA and mitochondrial DNA haplotypes in *Beta vulgaris* ssp. *maritima* (L.): the usefulness of both genomes for population genetic studies. *Molecular Ecology*, 9(2): 141-154.
- Vaillancourt R. E., Petty A., McKinnon G. E. 2004. Maternal inheritance of mitochondria in *Eucalyptus globulus*. *The Journal of heredity*, 95(4): 353-355.
- Williams J. G., Kubelik A. R., Livak K. J., Rafalski J. A., Tingey S. V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, 18(22): 6531-6535.
- Welsh J., McClelland M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research*, 18(24):7213-7218.
- Gan S., Shi J., Li M., Wu K., Wu J., Bai J. 2003. Moderate-density molecular maps of *Eucalyptus urophylla* S. T. Blake and *E. tereticornis* Smith genomes based on RAPD markers. *Genetica*, 118(1): 59-67.
- Amarger V., Mercier L. 1995. Molecular analysis of RAPD DNA based markers: their potential use for the detection of genetic variability in jojoba (*Simmondsia chinensis* L. Schneider). *Biochimie*, 77(12): 931-936.

23. Antoni J. 1997. Randomly amplified polymorphic DNA (RAPD) analysis. En: Caetano-Anollés G. y Gresshoff P.M. (eds.), *DNA Markers. Protocols, applications, and overviews*. WILEY-LISS, New York, pp. 75-83.
24. Mi Y. L., Eui J. D., Chae H. P., Young H. K., Eung S. K., Byong S. K., Seung-Eun O. 2006. Development of SCAR marker for discrimination of *Artemisia princeps* and *A. auyi* from other *Artemisia* herbs. *Biological & pharmaceutical bulletin*, 29(4): 629-633.
25. Nguyen T.T., Klueva N., Chamareck V., Aarti A., Magpantay G., Millena A. C., Pathan M. S., Nguyen H.T. 2004. Saturation mapping of QTL regions and identification of putative candidate genes for drought tolerance in rice. *Molecular genetics and genomics*, 272(1): 35-46.
26. Bove J., Lucas P., Godin B., Oge L., Jullien M., Grappin P. 2005. Gene expression analysis by cDNA-AFLP highlights a set of new signaling networks and translational control during seed dormancy breaking in *Nicotiana glauca*. *Plant molecular biology*, 57(4): 593-612.
27. Tohme J., González O., Beebe S., Duque M. 1996. AFLP analysis of gene pools of a wild bean core collection. *Crop Science*, 36:1375-1384.
28. Yuang M., Hong Y. 2003. Heterogeneity of chinese medical herbs in Singapore assessed by fluorescence AFLP analysis. *The American Journal of Chinese Medicine*, 31(5): 773-779.
29. Cai Z. H., Li P., Dong T.T.X., Tsim K. W. K. 1999. Molecular diversity of 5S-rRNA spacer domain in *Fritillaria* species revealed by PCR analysis. *Planta Medica*, 65: 360-364.
30. Lewin B. 2000. *From genes to genomes*. En: Genes VII. Oxford University Press, New York, pp. 37-79.
31. Hernández F. R. A. 1999. Telómeros y Telomerasas. *Revista Cubana Investigaciones Biomédicas*, 18(2): 121-129.
32. Duchaine T. F., Wohlschlegel J. A., Kennedy S., Bei Y., Conte D., Pang K., Brownell D. R., Harding S., Mitani S., Ruvkun G., Yates J. R., Mello C. C. 2006. Functional proteomics reveals the biochemical niche of *C. elegans* DCR-1 in multiple small-RNA-mediated pathways. *Cell*, 124(2): 343-354.
33. Ngan F., Shaw P., But P., Wang J. 1999. Molecular authentication of *Panax* species. *Phytochemistry*, 50: 787-791.
34. Good agricultural practices for medicinal plants. 2003. En: *WHO guidelines on good agricultural and collection practices (GACP) for medicinal plants*. World Health Organization, Geneva, pp. 7-16.
35. Hosokawa K., Hishida A., Nakamura I., Shibata T. 2006. The sequences of the spacer region between the *atpF* and *atpA* genes in the plastid genome allows discrimination among three varieties of medicinal *Angelica*. *Planta Medica*, 72: 570-571.
36. Gray D. J., Purohit A. 1991. Somatic embryogenesis and development of synthetic seed technology. *Critical Review in Plant Sciences*, 10(1): 33-61.
37. Zhao K. J., Dong T.T.X., Tu P. F., Song Z. H., Lo C. K., Tsim K. W. K. 2003. Molecular genetic and chemical assessment of radix *Angelica* (Danggui) in China. *Journal of Agricultural and Food Chemistry*, 51: 2576-2583.
38. Ding X., Xu L., Wang Z., Zhou K., Xu H., Wang Y. 2002. Authentication of stems of *Dendrobium officinale* by rDNA ITS region sequences. *Planta Medica*, 68: 191-192.
39. Lin J., Zhou X., Gao S., Wu W., Lui X., Sun X., Tang K. 2006. Authentication of *Pinellia ternata* and its adulterants based on PCR with specific primers. *Planta Medica*, Published online.
40. Yang M., Zhang D., Liu J., Zheng J. 2001. A molecular marker that is specific to medicinal rhubarb based on chloroplast trnL/trnF sequences. *Planta Medica*, 67: 784-786.

## Reconocimientos

A las valiosas sugerencias de los tres revisores anónimos que permitieron elevar la calidad de esta recopilación bibliográfica.

Dear Dr. Iris FERIA-ROMERO  
We are pleased to inform you that your paper

No. 7965

I. FERIA-ROMERO, H. ASTUDILLO-DE LA VEGA, M.A. CHAVEZ-SOTO, E.  
RIVERA-ARCE, M. LÓPEZ, H. SERRANO, X. LOZOYA

Molecular RAPD markers associated with quercetin accumulation in *Psidium*  
guajava L. individuals

has been accepted for publication in *BIOLOGIA PLANTARUM*.

Yours sincerely  
H. Kymrova - *Biologia Plantarum*

## BRIEF COMMUNICATION

### Molecular RAPD markers associated with quercetin accumulation in *Psidium guajava*

I. A. FERIA-ROMERO\*<sup>1</sup>, H. ASTUDILLO-DE LA VEGA\*\*, M. A. CHAVEZ-SOTO\*\*, E. RIVERA-ARCE\*\*, M. LÓPEZ\*\*\*, H. SERRANO\*\*\* and X. LOZOYA\*\*

\* Doctorado en Ciencias Biológicas, UAM, Mexico DF, Mexico

\*\* Centro Medico Nacional Siglo XXI, IMSS Mexico DF, Mexico

\*\*\* Universidad Autonoma Metropolitana, Mexico, D.F., Mexico

Running title: RAPD markers in *Psidium guajava*

#### Abstract

We used a RAPD amplification method to identify molecular markers associated with high quercetin accumulation in the leaves of *Psidium guajava* L. trees, selected from four different Mexican agronomic regions. We identified six polymorphic RAPD fragments of 620, 590, 370, 690, 480 and 460 bp among individuals of *P. guajava*. Genetic linkage disequilibrium analysis revealed that three RAPD profiles considered as DNA markers (620/590 bp, 370 bp and 480/460 bp) had a positive, direct association with quercetin content ( $P < 0.05$ ). These informative molecular markers can be used for selective identification of plants with the highest accumulation of flavonoids.

*Additional key words:* guava leaves, DNA profile, linkage disequilibrium, flavonoid accumulation, QTL.

#### Abbreviations:

**Acknowledgements:** The authors would like to thank Silvia Arzate-Trujillo and Akemi Sons-Matzumoto for technical assistance, as well as the people involved in the maintenance and processing of the *P. guajava* L. cultivars in the four sampled regions. IA Feria-Romero is under the Doctorado en Ciencias Biológicas, UAM. This work was supported by an IMSS-FOFOI FP-2003/039 grant.

<sup>1</sup> Corresponding author present address: M.Sc. Unidad de Investigación en Enfermedades Neurológicas, Centro Médico Nacional Siglo XXI, IMSS. Avenida Cuauhtémoc 330, Col. Doctores CP 06720. México, D.F., México; fax: (+525) 56 276900 21367, e-mail: [irisferi@yahoo.com.mx](mailto:irisferi@yahoo.com.mx).

## Introduction

*Psidium guajava* L. (*Mirtaceae*) originated in Mesoamerica is now widely distributed in many tropical and subtropical areas of the world. The fruit is extensively consumed either fresh or processed and the leaf infusion is used in traditional herbal medicine for gastrointestinal disorder treatment (Lutterodt 1989) due to the presence of specific flavonoids (quercetine glycosides) (Lozoya *et al.* 1994). Furthermore, the pharmaceutical importance of quercetin glycosides (Lozoya *et al.* 2002) and other flavonoids in guava leaves, physiologically these compounds take part in gene regulation and growth metabolism, and their accumulation is related to the geography, phenology of plant, and UV incidence (Havsteen 2002). For the implementation of distinguished individuals with high capacity in flavonoid accumulation, practical tools are needed as indicators which avoid non desirable material, resulting in low confidence results (Rivera-Arce *et al.* 2003). An alternative could be the chemical evaluation of a metabolic end product that is expensive and time consuming (Vargas-Alvarez *et al.* 2006). Beside the current knowledge of the genes related to flavonoid production, there are not conclusive data of a particular gene or group of genes related to quercetin glycosides accumulation (Hagsteen, 2002). From this perspective, indirect, non- related indicators or markers are needed. Random Amplified Polymorphic DNA (RAPD) markers have been used for the assessment of genetic relationship among cultivars, to estimate chemical and genetic diversity between varieties, and to evaluate the correlation with active compound in medicinal species (Mohapatra and Rout 2006, Padmesh *et al.* 2006, Narasimhan *et al.* 2006). RAPD markers have been used in guava characterization; Prakash *et al.* (2002) estimate the molecular diversity of 41 genotypes from different regions of India and Padilla-Ramírez *et al.* (2002) use them to evaluate the germplasm through fruit characteristics in a very narrow region of Mexico. To the best of our knowledge, there are no reports dealing with the study of genetic markers and flavonoid accumulation in *P. guajava* that could be used as an auxiliary or complementary method for quality control in plant drug production. The aim of this study was to find a group of DNA markers that could identify those *P. guajava* individuals with the highest quercetin content, using RAPD analysis in leaf samples from four different regions of Mexico with variable levels of guava fruit production.

Leaves collected from forty *P. guajava* individuals grown in four different Mexican regions were identified according to Rivera-Arce *et al.* (2003), air-dried in a 40 °C forced-ventilation oven and powdered. One g of powder was weighed and exhaustively extracted by refluxing in a Soxhlet apparatus with 25 cm<sup>3</sup> of methanol for 3 h; then, the solvent was evaporated *in vacuum* and the residue was recovered. The extracted quercetin glycosides were hydrolyzed by refluxing in 20 cm<sup>3</sup> of hydrochloric acid 0.5 M in methanol for 1 h; afterwards, methanol was added for a final volume of 25 cm<sup>3</sup> in a volumetric flask and a 1 cm<sup>3</sup> aliquot was diluted 1:10 with the same solvent. This solution was filtered through a 0.45 µm filter for HPLC analysis of the aglycone quercetin obtained.

Quercetin in a 0.01 cm<sup>3</sup> sample injection volume was analyzed by HPLC on a reverse-phase C<sub>18</sub> column of 4.6 x 150 mm (Varian; Palo Alto, CA, USA) in a Waters 2796 system (Waters; Milford, MA, USA), equipped with a Waters 996 PDA detector maintained at 35 °C and by using a 20-80 % increasing acetonitrile gradient in 10 mM phosphoric acid. The flow rate was kept constant at 0.6 cm<sup>3</sup> min<sup>-1</sup> for a total running time of 35 min. All the operation conditions, data acquisition and analysis were controlled by Millennium software (Waters; Milford, MA, USA). Quercetin peak was compared to a commercial standard one (Sigma-Aldrich Co.; St. Louis, MO, USA), based on retention time and quantified by interpolation on a calibration curve, as described by Vargas-Alvarez *et al.* (2006).

Young and disease-free leaves were washed in 70% ethanol for 2 min and sterile, distilled water for 2 min, in sterile containers. The clean leaves were stored at -70 °C until use. A 150 mg leaf sample was cut into small pieces and ground to a fine powder under liquid nitrogen with a mortar and pestle. DNA was extracted from the sample using the Qiagen DNeasy Plant Mini Kit (Qiagen; Hilden, Germany) according to the manufacturer's instructions. The extracted DNA was quantified and checked for purity and stored at 4 °C until use. The DNA solutions were adjusted to a final concentration of 15 µg cm<sup>-3</sup> prior to their use in the RAPD reactions.

Six 10-mer commercial primers (Amersham Biosciences; Piscataway, NJ, USA) were used for RAPD-PCR amplification. Fifteen ng of template DNA were amplified in a 0.025 cm<sup>3</sup> total PCR reaction including one RAPD Analysis Bead (Amersham Biosciences; Piscataway, NJ, USA) to guarantee the reproducibility in the samples and 25 pmol of each primer. DNA amplification was performed using a GeneAmp 9700 PCR thermocycler (PE Biosystems Inc.; Foster City, CA, USA) programmed for 1 step of 5 min at 94°C followed by 45 cycles of 1 min at 94°C, 1 min at 36°C and 2 min at 72°C and a final 10 min, 72°C extension. To assess that RAPD amplification was reproducible, at least two individual experiments were performed with no significant differences (data not shown).

Amplified RAPD fragments were separated on a 2% agarose gel stained with ethidium bromide, visualized under UV fluorescence, digitalized in a BioImaging System (UVP; Cambridge, UK) and evaluated with the LabWorks v1.1.27 software (UVP; Cambridge, UK).

One-way analysis of variance was used for statistical evaluation of total quercetin accumulation in the samples collected. Statistical evaluation of RAPD profiles and their correlation with quercetin accumulation was obtained by using the SNPAnalyzer ISTECH web service facility, available at [http://www.istech.info/istech/board/login\\_form.jsp](http://www.istech.info/istech/board/login_form.jsp). This software takes into account the frequency of a determined character or band and a specific chemical trait and returns a numerical value that is the statistical correlation for each pair of data. For this purpose, the banding patterns were classified according to polymorphic variations in amplified RAPD fragments, if they differed in at least two samples with the same primer. We consider that the RAPD profiles have an allelic behavior when they share some bands at the same position on the agarose gel. The homozygote showed a simple band, while heterozygote showed either one, two or three bands for the same primers used with different samples. An estimation of major allele frequencies was obtained by chi square statistical and their correlation with quercetin accumulation trait, as a genetic Linkage Disequilibrium (LD) was obtained according to Lewontin (1964, 1995). Additionally, each banding pattern was associated with the quercetin concentration as a phenotypic trait. The association of RAPD profiles generated from each sample with different primers with quercetin accumulation was obtained through the expectation-maximization (E-M) algorithm of the SNPAnalyzer software (Excoffier and Slatkin 1995).

## Results

Ten mature trees from each area were randomly selected and evaluated for their quercetin content during different seasons throughout the year. Significant differences were observed in quercetin accumulation throughout seasons in some states. Whereas Queretaro and Michoacan did not have a significant season variation ( $5.57 \pm 0.95 \text{ mg g}^{-1}$  and  $2.32 \pm 0.43 \text{ mg g}^{-1}$  respectively), cultivars from Aguascalientes showed the highest quercetin content of  $9.34 \pm 1.14 \text{ mg g}^{-1}$  followed by Estado de Mexico with  $8.51 \pm 1.94 \text{ mg g}^{-1}$  in spring; the lowest accumulations were in autumn with  $5.52 \pm 2.02 \text{ mg g}^{-1}$  and  $5.53 \pm 1.91 \text{ mg g}^{-1}$  respectively, showing a significant difference between seasons ( $P < 0.001$ ). Interestingly, from all the collected trees, one Estado de Mexico individual analyzed in spring was the highest producer, with a quercetin content of  $12.62 \text{ mg g}^{-1}$  (data not shown). Our results are in agreement with those of Vargas-Alvarez *et al.* (2006), who stated that May and July 2000 were the peak times for the flavonoids (myricetin, quercetin, luteolin, and kaempferol) accumulation in mature guava leaves in the sampled geographic regions included in this study.

RAPD analysis is a molecular biological tool that allows an easy evaluation of the genome constitution of different plant samples, and the reproducibility of the results depends on the standard method and the highest quality of reagents employed. The six decamer primers that were used generated a total of 91 RAPD fragments, resulting in different amplification fragment patterns (Table 1); sharing a 74 to 100% by all the analyzed individuals, depending on the primer used. The validation of the RAPD technique was implemented by considering fragments from 200 to 1000 base pairs (bp) in length (Welsh and McClelland 1990, Williams *et al.* 1990), and band pattern reproducibility was determined in at least two individual experiments, as indicated above. Also, the similarity obtained among all collected samples allowing the identification of polymorphic differences, was used as an indicator of reproducibility.

Among the 91 generated fragments, only six were polymorphic in at least one individual, corresponding to primers 1, 2, 3 and 4 (arrows in Fig. 1). An allelic behavior upon RAPD amplification in these primers, along with the correlation between RAPD profile and a specific Mexican state, was observed. Pattern 1b was present only in Michoacan, patterns 1c and 2b in Aguascalientes and Michoacan, pattern 3b in Queretaro and Michoacan and pattern 4a in Aguascalientes, Estado de Mexico and Michoacan states. The amplified fragments for primers 5 and 6 were shared by all individuals without polymorphic fragments.

In a preliminary correlation between band patterns and quercetin accumulation, we found two individuals containing  $\geq 10.5$  mg total quercetin per gram of dried leaves collected in spring from Aguascalientes and Estado de Mexico, sharing the same banding patterns. An statistical analysis in all RAPD patterns with SNPAnalyzer and EM algorithm allowed us to determine that 1a, 2a, 3a and 4c are the most informative combinations (MIC), due to the high frequencies of the 590 bp fragment (allele 1a) and 480 and 460 bp fragments (compound allele 4c); alleles 2a and 3a are the lacking of the 360 or 690 bp bands, respectively. This MIC refers to the correlation between DNA fragments and the quercetin accumulation. In order to explore the possible interaction among all these allele-fragments, we evaluate their linkage disequilibrium (LD) association. As it can be seen in table 1, the major LD for allele 4c was with 1a and 2a fragments, whereas for allele 4b, significant LD values were with the 3a allele, the lack of the 690 bp fragment.

With the purpose of identifying a genetic factor that was involved in the variation of quercetin accumulation, we suggested a DNA molecular marker analysis to identify genomic characteristics related to quercetin variation. Many authors have treated the RAPD technique as equivalent to single genetic alleles in such a way that when population genetics statistics are used, there are relationships to the observed phenotype (Becerra and Paredes 2000, Padilla-Ramírez *et al.* 2002). According to our analysis, the frequency of fragments resulting in MIC, correlates to quercetin content, a dominant trait

with a Lod score of 9.498 for a Quantitative Trait Loci (QTL) relationship between the 1a and 2a with 4c banding patterns ( $p < 0.05$ ), a significant high value that correlates with our phenotype trait.

In conclusion, we found clear and defined amplification fragments linked to quercetin glycoside accumulation in *P. guajava* individuals and demonstrated that, even when different individuals of the same region were analyzed, the presence of these amplified fragments correlated with the quercetin accumulation capability of a specific individual.

## References

- Becerra, V., Paredes, M.: Use of biochemical and molecular markers in genetic diversity studies. - *Agr. tec.* **60**: 270-281, 2000. [In Spanish]
- Excoffier, L., Slatkin, M.: Maximum likelihood estimation of molecular haplotype frequencies in a diploid population. - *Mol. Biol. Evol.* **12**: 921-927, 1995.
- Havsteen B.H.: The biochemistry and medical significance of the flavonoids. - *Pharmacol. therap.* **96**: 67-202, 2002.
- Lewontin, R.: The detection of linkage disequilibrium in molecular sequence data. - *Genetics.* **140**: 377-388, 1995.
- Lewontin, R.: The interaction of selection and linkage II. Optimum models. - *Genetics.* **30**: 757-782, 1964.
- Lozoya, X., Meckes, M., Abou-Zaid, M., Tortoriello, J., Nozzolillo, C., Thor, J.: Quercetin glycosides in *Psidium guajava* L. leaves and determination of a spasmolytic Principle. - *Arch. Med. Res.* **25**: 11-15, 1994.
- Lozoya, X., Reyes-Morales, H., Chavez-Soto, M.A., Martinez-Garcia, M.C., Soto-Gonzalez, Y., Doubova, S.V.: Intestinal anti-spasmodic effect of a phytodrug of *Psidium guajava* folia in the treatment of acute diarrheic disease. - *J. Ethnopharmacol.* **83**: 19-24, 2002.
- Lutterodt, G.D.: Inhibition of gastrointestinal release of acetylcholine by quercetin as a possible mode of action of *Psidium guajava* leaf extracts in the treatment of acute diarrhoeal disease. - *J. Ethnopharmacol.* **25**: 235-247, 1989.
- Mohapatra, A., Rout, G.R.: Optimization of primer screening for evaluation of genetic relationship in rose cultivars. - *Biol. Plant.* **50**: 295-299, 2006.
- Narasimhan, S., Padmesh, P., Nair, G.M.: Assessment of genetic diversity in *Coscinium fenestratum*. - *Biol. Plant.* **50**: 111-113, 2006.
- Padilla-Ramírez J.S., González-Gaona E., Esquivel-Villagrana F., Mercado-Silva E., Hernández-Delgado S., Mayek-Pérez N.: Characterization of outstanding guava germplasm from the Calvillo-Cañones region, México. *Rev. Fitotec. Mex.* **25**: 393-299, 2002. [In Spanish]

- Padmesh, P., Reji, J.V., Jinish Dhar, M.: Estimation of genetic diversity in varieties of *Mucuna pruriens* using RAPD. - Biol Plant. **50**: 367-372, 2006.
- Prakash, D.P., Narayanaswamy, P., Sondur, S.N.: Analysis of molecular diversity in guava using RAPD markers. - J. Hortic. Sci. Biotechnol. **77**: 287-293, 2002.
- Rivera-Arce, E., Gattuso, M., Lozoya, X.: Anatomical identity parameters of the crude drug *Psidium guajavae folium*. - pharmaceut. Biol. **41**: 516-521, 2003.
- Vargas-Alvarez, D., Soto-Hernández, M., González-Hernández, V.A., Engleman, E.M., Martínez-Garza, A.: Kinetics of accumulation and distribution of flavonoids in guava (*Psidium guajava* L.). - Agrociencia. **40**: 109-115, 2006. [In Spanish]
- Welsh, J., McClelland, M.: Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. **18**:7213-8, 1990.
- Williams, J.G., Kubelik, A.R., Livak, K.J., Rafalski, J.A., Tingey, S.V.: DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. - Nucleic Acids Res. **18**: 6531-6535, 1990.

Table 1. The sequences of primers used for RAPD analysis and a summary of the frequency and linkage disequilibriums of the banding patterns obtained from *P. guajana* L. individuals using RAPD polymorphic fragments [bp]. The directionality of each nucleotide sequence is 5' to 3'. Polymorphic fragment analysis is indicated between parentheses. The major pattern band obtained in RAPD analysis correlating with linkage disequilibrium association is indicated by an asterisk.

Primer sequence	Nucleotide	Number of polymorphic fragments	Total number of fragments	RAPD profiles from polymorphic fragments	Frequency of major band	Linkage disequilibrium association
1	GGTGCGGAA	2 (590, 610)	22	1a* (590) 1b (610) 1c (590/610)	0.82	4c
2	GTTTCGCTCC	1 (370)	16	2a* 2b (370)	0.79	4c
3	GTAGACCCGT	1 (690)	15	3a* 3b (690) 4a	0.98	4b
4	AAGAGCCCGT	2 (460, 480)	14	4b (460) 4c* (460/480)	0.69	
5	AACGGCAAC	0	12	5		
6	CCCGTCAGCA	0	12	6		

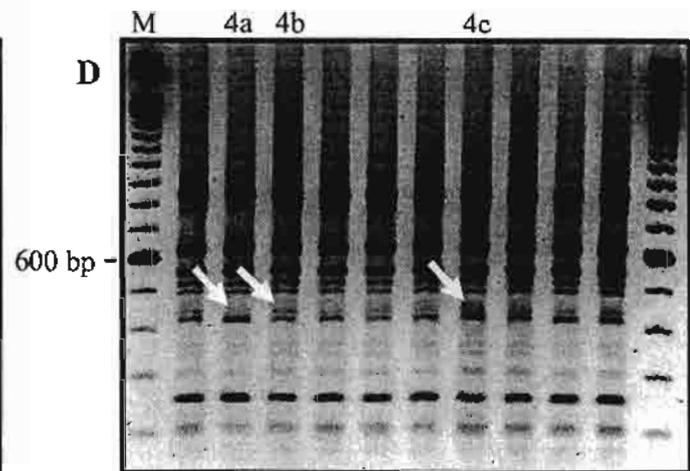
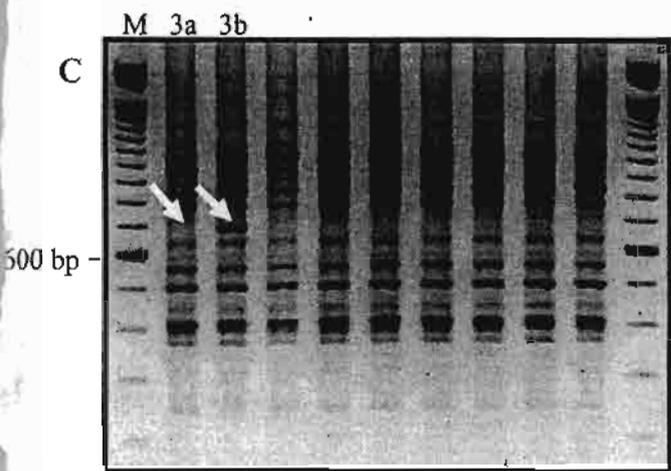
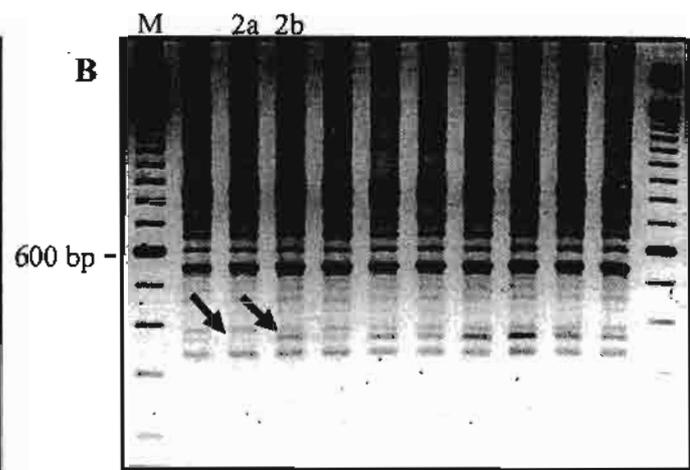
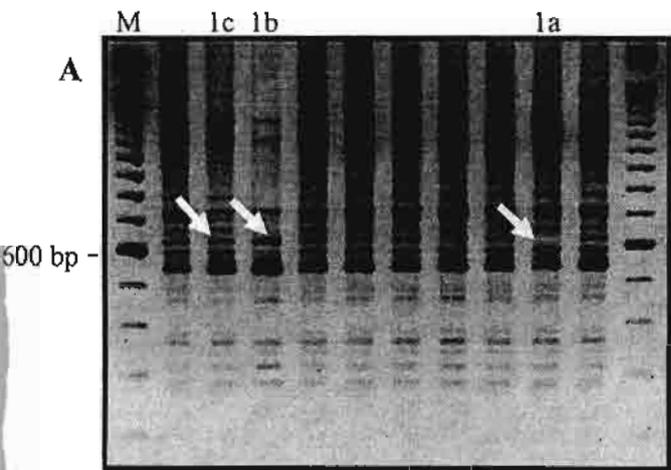


Fig. 1. RAPD profiles of *P. guajava* from different Mexican states.

Lane M: 100 bp ladder. **A:** 1a, 1b and 1c represent polymorphic RAPD profiles using primer 1 for Michoacan individuals. **B:** 2a and 2b represent polymorphic RAPD profiles using primer 2 for Aguascalientes individuals. **C:** 3a and 3b represent polymorphic RAPD profiles using primer 3 for Queretaro individuals. **D:** 4a, 4b and 4c represent polymorphic RAPD profiles using primer 4 for Aguascalientes individuals. Arrows indicate informative fragments with allelic behavior.