

## Identificación y secuenciación de fragmentos polimórficos entre ecotipos de cocotero resistentes y susceptibles a la enfermedad del amarillamiento letal

### Identification and sequencing of polymorphic fragments between coconut ecotypes resistant and susceptible to lethal yellowing disease

Carlos A. Puch-Hau<sup>1,2</sup>, Carlos Oropeza Salín<sup>1</sup>, Adriana Quiroz Moreno<sup>1</sup>, Iván Córdova-Lara<sup>1</sup>, Verónica Limones-Briones<sup>1</sup>, Felipe Sanchez-Teller<sup>1</sup>, Luis Sáenz-Carbonell<sup>1\*</sup>.

1Unidad de Biotecnología, Centro de Investigación Científica de Yucatán, Calle 43 No. 130. Colonia Chuburna de Hidalgo, 97200 Mérida, YUC, México. 2Unidad Mérida, Departamento de Recursos del Mar, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV-IPN), Km. 6 Antigua Carretera a Progreso, Cordemex, 97310 Mérida, Yucatán, Mexico.

Autor de correspondencia: [vyca@cicy.mx](mailto:vyca@cicy.mx)

#### RESUMEN

La selección de germoplasma de cocotero resistente a patógenos como el fitoplasma que causa la enfermedad del amarillamiento letal (AL), es una prioridad dentro de los programas de mejoramiento genético. Por lo tanto, la de biomarcadores moleculares asociados a genes de resistencia (genes *R*) representa una herramienta esencial. En este estudio, usamos la técnica AFLP-RGC para identificar bandas polimórficas entre palmas de cocotero resistente y susceptible a la enfermedad AL. Se utilizaron un total de 24 combinaciones de cebadores a partir de las cuales se identificaron 17 fragmentos polimórficos. De los 17 fragmentos, nueve fueron de los ecotipos resistentes (AFLP-RGC 1-02, -06, -08, -09, -10, -11, -12, -13 y -17) y ocho de los susceptibles (AFLP-RGC 1-01, -03, -04, -05, -07, -14, -15 y -16). Todos los fragmentos fueron clonados y secuenciados. El análisis bioinformático reveló que nueve fragmentos de ADN (AFLP-RGC 1-01, -02, -03, -04, -06, -07, -12, -14 y -17) presentaron identidad con accesiones de GenBank, como la proteína de resistencia RPM1 de *Prunus mume* y el Receptor-tipo protein cinasa de *Elaeis guineensis* (AFLP-RGC 1-01 y -17, respectivamente). Las secuencias reportadas en este estudio constituyen una importante fuente de recursos genéticos para la identificación de potenciales marcadores moleculares asociados con la resistencia a patógenos en este cultivo.

**Palabras clave:** marcadores moleculares, patógenos, cocotero, genes de resistencia.

#### INTRODUCCIÓN

La palma de coco (*Cocos nucifera* L.) es un cultivo perenne que contribuye significativamente a la seguridad alimentaria y la generación de empleos e ingresos a nivel mundial. En México hace 40 años existía una superficie sembrada con este cultivo de aproximadamente 180,000 Ha, actualmente esta área ha disminuido significativamente a 124,000 Ha (FAOSTAT, 2020). Entre los factores responsables de la reducción de la superficie en el cultivo de la palma de coco

#### ABSTRACT

The selection of coconut germplasm resistant to pathogens such as the phytoplasma that causes lethal yellowing (LY) disease is a priority within the breeding programs. Therefore, the arrangement of molecular biomarkers associated with resistance genes (*R* genes) represents a very useful tool. In this study, we using the AFLP-RGC technique for identified polymorphism bands between coconut palm resistant and susceptible to LY disease. A total of 24 primers combinations were used from which 17 fragment polymorphism were identified. Of the 17 polymorphic fragments, 9 were of resistant (AFLP-RGC 1-02, -06, -08, -09, -10, -11, -12, -13 and -17) and 8 of the susceptible ecotypes (AFLP-RGC 1-01, -03, -04, -05, -07, -14, -15 and -16). All fragments were cloned and sequenced. The bioinformatics analysis revealed that nine DNA fragments (AFLP-RGC 1-01, -02, -03, -04, -06, -07, -12, -14 and -17) presented identity with accessions of GenBank, such as the resistance protein RPM1 of *Prunus mume* and Receptor-Like Protein Kinase-like of *Elaeis guineensis* (AFLP-RGC 1-01 and -17, respectively). The sequences reported on this study represents an important source of genetic resources for the identification of potential molecular markers associated with pathogen resistance in this crop.

**Keywords:** molecular markers, pathogens, breeding coconut palms, resistance genes

#### INTRODUCTION

Coconut palm (*Cocos nucifera* L.) is a perennial crop that contributes significantly to food security and the generation of employments and income worldwide. In Mexico 40 years ago there was a planted area of approximately 180,000 Ha, currently, this area has decreased significantly to 124,000 Ha (FAOSTAT, 2020). Among the factors responsible for the losses in coconut palm crop are biotic factors such as diseases caused by different pathogens such as viruses, viroids, protozoa, oomycetes, nematodes and phytoplasmas

se encuentran los factores bióticos como enfermedades causadas por virus, viroides, protozoos, oomicetos, nematodos y fitoplasmas (Puch-Hau, *et al.*, 2021).

Los fitoplasmas que afectan a las palmas de cocotero son responsables de la devastadora enfermedad conocida como el amarillamiento letal (AL) en las Américas (Harrison y Oropeza, 2008). En México, debido a esta enfermedad, la superficie del cocotero ha disminuido en un 40% a lo largo de la costa del Golfo de México (Córdova, *et al.*, 2019). Por esta razón, actualmente en nuestro grupo de trabajo se lleva a cabo el desarrollo de programas de mejoramiento para obtener palmas resistentes a AL. Uno de los principales objetivos de la investigación en fitomejoramiento genético es el aislamiento y caracterización de genes que pueden conferir resistencia (genes *R*) a una amplia gama de enfermedades causadas por virus, bacterias, hongos, nematodos e insectos plaga (Kourelis y van der Hoorn, 2018). La familia de genes *R* más grande contiene dominios de sitios de unión de nucleótidos (NBS) y repeticiones ricas en leucina (LRR). El dominio NBS está involucrado en la señalización e incluye varios motivos altamente conservados y estrictamente ordenados, como los motivos P-loop, kinase-2 y GLPL (Puch-Hau, *et al.*, 2015). El motivo LRR generalmente está involucrado en las interacciones proteína-proteína y la unión de ligandos con moléculas derivadas de patógenos, lo que sugiere que este dominio puede desempeñar un papel fundamental en la definición de la especificidad de reconocimiento de patógenos (Ravensdale, *et al.*, 2012). Sin embargo, hay muchos genes con funciones desconocidas que contienen estos dominios que se han identificado en diferentes especies de plantas y se ha demostrado que están relacionados filogenéticamente con genes *R*. Estas secuencias se denominan candidatos a genes de resistencia (RGC) (Pflieger, *et al.*, 2001) y representan una fuente importante para el descubrimiento de nuevos genes *R*.

Los RGC pueden utilizarse como herramienta para desarrollar nuevos marcadores vinculados a la resistencia a patógenos. Por ejemplo, un proceso modificado del AFLP (Polimorfismos en la Longitud de Fragmentos Amplificados) propuesto por Hayes y Saghai-Marrof (2000) y Egea-Gilabert, *et al.*, (2003) se basa en el uso de un cebador específico de la técnica AFLP y un cebador específico o degenerado del dominio NBS de los genes *R*. AFLPRGC (polimorfismo de longitud de fragmento amplificado acoplado a genes candidatos de resistencia) se ha empleado en diferentes cultivos como soya (Hayes y Saghai-Marrof, 2000), pimienta (Egea-Gilabert, *et al.*, 2003), albaricoque (Soriano, *et al.*, 2005), algodón (Zhang, *et al.*, 2007) y chícharo (Patil, *et al.*, 2014), para buscar polimorfismos que al validarse puedan usarse como marcadores moleculares para el mejoramiento genético de los cultivos.

(Puch-Hau, *et al.*, 2021).

Phytoplasmas infected coconut palms are responsible for the devastating disease known as lethal yellowing (LY) in the Americas (Harrison and Oropeza, 2008). In Mexico, due to this disease, the surface of the coconut palm has decreased along the coast of the Gulf Mexico (Córdova, *et al.*, 2019). Therefore, the developing of breeding programs to obtain resistance palms to LY is currently carried out in our group. One of the main objective of research on plant breeding was the isolation and characterization of the resistance genes (*R* genes) that conferring resistance to a wide range of disease caused by viruses, bacteria, fungi, nematodes and insects pests (Kourelis and van der Hoorn, 2018). The largest *R* gene family contains domains of nucleotide binding sites (NBS) and Leucine-rich repeats (LRR). The NBS domain is involved in signaling and includes several highly conserved and strictly ordered motifs such as P-loop, kinase-2 and GLPL motifs (Puch-Hau, *et al.*, 2015). The LRR motif is typically involved in protein-protein interactions and ligand binding with pathogen-derived molecules, suggesting that this domain may play a pivotal role in defining pathogen recognition specificity (Ravensdale, *et al.*, 2012). However, there are many genes with unknown functions, containing these domains that have been identified in different plant species and have been shown to phylogenetically related to known *R* genes and are referred as resistance gene candidates (RGC) (Pflieger, *et al.*, 2001) and could be an important source of novel *R* gene.

The RGC can be used as tool to develop novel markers linked to pathogen resistance. For example, a modified process of the AFLP (amplified fragment length polymorphism) proposed by Hayes and Saghai-Marrof (2000) and Egea-Gilabert, *et al.*, (2003) is based on the use of a specific primer of the technique AFLP and a specific or degenerate primer from the NBS domain of *R* genes. AFLP-RGC (amplified fragment length polymorphism coupled to resistance gene candidates) have been employed in different crops such as soybean (Hayes and Saghai-Marrof, 2000), pepper (Egea-Gilabert, *et al.*, 2003), apricot (Soriano, *et al.*, 2005) cotton In coconut palms, there is currently a RGC database that represents an important source for the discovery of new resistance genes in coconut (Puch-Hau, *et al.*, 2015; 2016), however, due to the slow growth of this crop, the lack of protocols for genetic manipulation (transformation and gene editing) and the difficulty in establishing controlled transmission systems for pathogens, it is difficult to advance in this way. In this sense, the implementation of a breeding program assisted by molecular markers could represent a faster and more viable option. Therefore, the objective of this work was the implementation of the AFLP-RGC technique

Esta técnica surge como una herramienta para identificar regiones genómicas que puedan contener genes de resistencia a patógenos. En cocotero, actualmente existe una base de secuencias de RGC que representa una fuente importante para el descubrimiento de nuevos genes *R* (Puch-Hau, *et al.*, 2015; 2016); sin embargo, debido al lento crecimiento de este cultivo, la falta de protocolos para la manipulación genética (transformación y edición de genes) y la dificultad para establecer sistemas de transmisión controlados de patógenos, dificulta avanzar en este camino. En este sentido, la implementación de un programa de mejoramiento genético asistido por marcadores moleculares podría representar una opción más rápida y viable. Por lo tanto, el objetivo de este trabajo fue implementar la técnica AFLP-RGC para la identificación, clonación y secuenciación de fragmentos polimórficos entre ecotipos de coco resistentes y susceptibles a la enfermedad del amarillamiento letal para identificar potenciales marcadores que puedan ser utilizados en la selección de germoplasma con características de interés, como la resistencia a plagas y patógenos.

**MATERIALES Y MÉTODOS**

**Material vegetal y aislamiento del ADN**

En este estudio se utilizaron dos ecotipos de cocotero. Los ecotipos analizados fueron el Enano Malayo Amarillo (EMA) resistente al AL, ubicado en Telchac Puerto, México (21° 21' N, 89° 16' W) y el Alto del Atlántico del Oeste (AAO) susceptible al AL ubicado en la (21° 22'N, 89° 00'W) (Zizumbo-Villareal, *et al.*, 2008). Para cada ecotipo se muestrearon cinco palmas (Tabla 1). Los ácidos nucleicos (ADN) se extrajeron a partir de 100 mg de hojas jóvenes siguiendo el protocolo descrito por Harrison, *et al.*,

for the identification, cloning and sequencing of polymorphic fragments between coconut ecotypes resistant and susceptible to the disease of lethal yellowing, in order to identify potential markers that can be used for the selection of germplasm with characteristics of interest, such as resistance to pests and pathogens

**MATERIALS AND METHODS**

**Plant material and DNA isolation**

Two ecotypes of coconut palm were utilized in this study. The ecotypes included were LY-resistant Malayan Yellow Dwarf (MYD), located in Telchac Puerto, México (21° 21' N, 89° 16' W) and the LY-susceptible West Atlantic Tall (WAT) located at the experimental planting of “Caridad del cobre”, San Crisanto, Yucatán, México (21° 22'N, 89° 00'W) (Zizumbo-Villareal, *et al.*, 2008). For each ecotype five coconut palms were sampled (Table 1). DNA was extracted from 100 mg of young leaves following the protocol described by Harrison, *et al.*, (1994). DNA quantification was performed using the Nanodrop 1000 (Thermo Scientific, USA).

**Tabla 1.** Descripción de los genotipos de cocotero empleados en este estudio.  
**Table 1.** Description of coconut palm genotypes used on this study.

<b>Number of samples</b>	<b>Ecotype</b>	<b>Localization</b>
5	MYD Resistant	Telchac Puerto, Yucatán, México
5	WAT Susceptible	San Crisanto, Yucatán, México

MYD: Malayan Yellow Dwarf, WAT: West Atlantic Tall

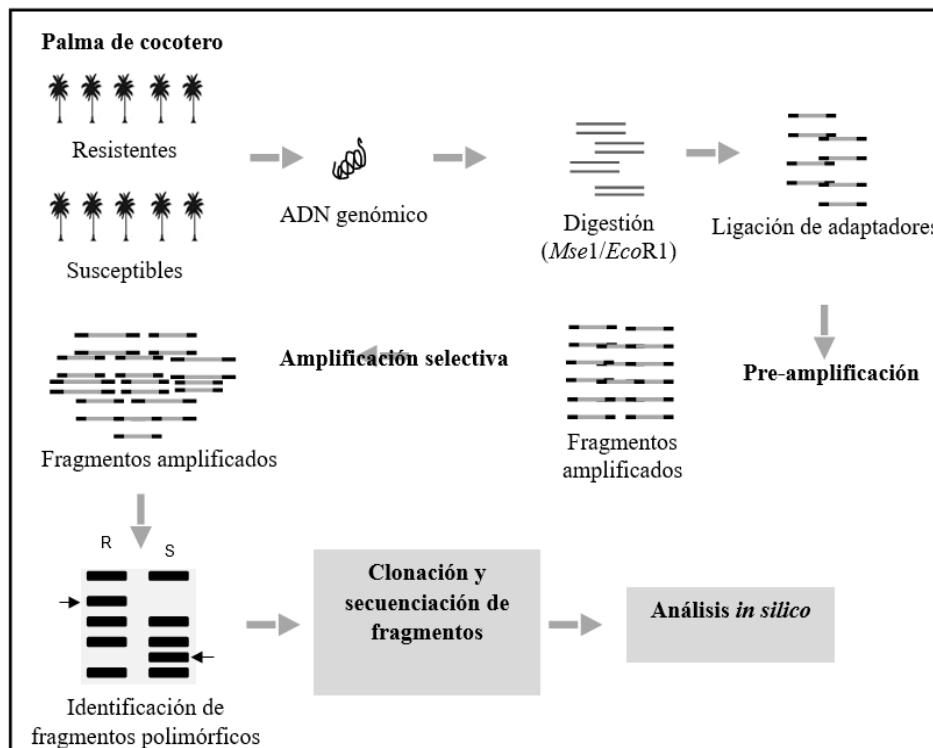
**Procedimiento AFLP-RGC**

La técnica de AFLP-RGC se llevó a cabo según lo descrito por Hayes y Saghai-Marrof (2000) y Egea-Gilabert, *et al.*, (2003) (Figura 1). Se realizó la digestión del ADN (100 ng) con las enzimas *EcoRI* y *MseI*; posteriormente se ligaron los

**AFLP-RGC procedure.** AFLP-RGC analysis was carried out as described by Hayes and Saghai-Marrof (2000) and Egea-Gilabert, *et al.*, (2003) (Figure 1). DNA digestion (100 ng) and adaptor ligations we carried out following the standard AFLP protocol (Vos *et al.*, 1995) with some minor

adaptadores siguiendo el protocolo estándar AFLP (Vos *et al.*, 1995) con modificaciones menores. Este ADN molde se preamplificó utilizando cebadores homólogos al adaptador *EcoRI* o *MseI* con un nucleótido selectivo (*EcoRI* + A y *MseI*+C). La preamplificación se realizó en un volumen total de 20  $\mu\text{l}$  que contenía 2  $\mu\text{l}$  de ADN digerido, 1 pmol de cada cebador, 0.2 mM de cada dNTP, 1.5 mM de  $\text{MgCl}_2$  y 1 U de Taq ADN polimerasa (Invitrogen). Las PCR se llevaron a cabo en un termociclador GeneAmp PCR System 9700 (Perkin-Elmer) utilizando las siguientes condiciones de amplificación: un ciclo de desnaturalización a 94 °C durante 2 min, seguido de 40 ciclos de 94 °C durante 30 s, 50 °C durante 30 s y 72°C por 1 min, y una elongación final a 72°C por 5 min. Los productos de preamplificación fueron diluidos (20 veces) y 2  $\mu\text{l}$  fueron usados como molde para la amplificación selectiva. Como se muestra en la Tabla 2, la amplificación selectiva se realizó con dos cebadores, uno complementario al adaptador *EcoRI* o *MseI* con dos y tres nucleótidos selectivos y otro con un cebador RGC degenerado correspondiente a cuatro motivos de aminoácidos conservados entre las proteínas de resistencia conocidas (P-loop, RNBS-D, GLPL). Para este proceso se emplearon los protocolos de PCR previamente descritos por Vos, *et al.*, (1995). A cada producto de la amplificación selectiva se le añadieron 3  $\mu\text{l}$  de tampón de carga (95% formamida, 0,01 M NaOH, 0,05% xilenocianol, 0,05% azul de bromofenol). Posteriormente, las muestras se desnaturalizaron mediante incubación a 95 °C durante 3 min, se colocaron directamente en hielo y se corrieron en un gel de poliacrilamida desnaturalizante al 6 % que se tiñó con nitrato de plata (Sambrook, *et al.*, 1989). Los tamaños moleculares de los fragmentos amplificados se estimaron utilizando estándares de ADN de 10 pb (Invitrogen).

modifications. This template DNA was the pre-amplified using primers homologous to the *EcoRI* or *MseI* adaptor with a selective nucleotide (*EcoRI* + A and *MseI*+C). Pre-amplification was performed in a total volume of 40  $\mu\text{l}$  containing 2  $\mu\text{l}$  of the template DNA, 1 pmol of each primer, 0.2 mM of each dNTPs, 1.5 mM  $\text{MgCl}_2$  and 1 U of *Taq* DNA polymerase (Invitrogen). PCR were carried out in a GeneAmp PCR System 9700 thermal cycler (Perkin-Elmer) using the following cycling conditions: a denaturation step at 94°C for 2 min, followed by 40 cycles of 94°C for 30 s, 50°C for 30 s and 72°C for 1 min, and final elongation step at 72°C for 5 min. Pre-amplification products were further diluted (20 fold) and 2  $\mu\text{l}$  was used as template for selective amplification. As show in Table 2, selective amplification was performed with two primers, one complementary to the *EcoRI* or *MseI* adapter with two and three selective nucleotides and another with degenerate RGC primer corresponding to four amino acid motifs conserved among the known resistance proteins (P-loop, RNBS-D, GLPL), respectively using the touchdown PCR protocols described by Vos, *et al.*, (1995). Selective amplification products were dried, and 3  $\mu\text{l}$  of loading buffer (95% formamide, 0.01 M NaOH, 0.05% xylene cyanol, 0.05% bromophenol blue) was added to each one. The samples were then denaturing by incubation at 95°C for 3 min, placed directly on ice and run on a 6% denaturing polyacrylamide gel that was stained with silver nitrate (Sambrook, *et al.*, 1989). Molecular sizes of the amplified fragments were estimated using a 10-bp ladder (Invitrogen).



**Figura 1.** Estrategia experimental para la búsqueda de fragmentos polimórficos entre ecotipos resistentes y susceptibles a la enfermedad del amarillamiento letal mediante la técnica AFLP-RGC.

**Figure 1.** Experimental strategy for the search of polymorphic fragments between ecotypes resistant and susceptible to lethal yellowing disease using the AFLP-RGC technique.

**Tabla 2.** Secuencia de nucleótidos de iniciadores usados para los AFLP-RGC.

**Table 2.** Sequence details of primers used for RGC-anchored amplified fragment length polymorphism analysis.

Iniciador	Secuencia (5'-3')	
	Iniciadores AFLP	
Pre-amplificación	<i>EcoR1</i> + A	GACTGCGTACAATTCA
	<i>Mse1</i> + C	GATGAGTCCTGAGTAAC
Amplificación selectiva	<i>EcoR1</i> + AA	GACTGCGTACAATTCAA
	<i>EcoR1</i> + AAG	GACTGCGTACAATTCAAG
	<i>EcoR1</i> + AAC	GACTGCGTACAATTCAAC
	<i>Mse1</i> + CAT	GATGAGTCCTGAGTAACAT
	<i>Mse1</i> + CTG	GATGAGTCCTGAGTAACTG
	Iniciadores degenerados RGC	
Amplificación selectiva	PIB	GGIATGCCIGGGIIIIGGIAARACIAC
	RNBS-D	GGRAAIARISHRCARTAVIRAARC
	P3A	AIITYIRIIRYIAGIGGYAAICC
	PIA	GGIATGCCIGGGIIIIGGIAARACIAC

Código de degeneración: I = inosine; R = A/G; Y = C/T; M = A/C; K = G/T; W = A/T; S=C/G; B = C/G/T; D = A/G/T; H = A/C/T; V = A/C/G; N = A/C/G/T. RGC: Candidatos a Genes de Resistencia.



### Desarrollo de la técnica AFLP-RGC

Para desarrollar la técnica de AFLP-RGC, evaluamos un total de 24 combinaciones de cebadores. Las amplificaciones selectivas de AFLP se realizaron combinando los cebadores degenerados RGC y cinco cebadores de AFLP (*EcoR1* + AA, *EcoR1* + AAG, *EcoR1* + AAC, *Mse1* + CAT y *Mse1* + CTG) seleccionados en base al alto número de polimorfismos obtenidos por Soriano *et al.*, (2005) en albaricoque (*Prunus armeniaca*).

### Clonación y secuenciación de los fragmentos polimórficos AFLP-RGC

Los fragmentos polimórficos de AFLP-RGC se cortaron del gel de acrilamida y se purificaron con un kit de extracción de gel QIAEX II (Qiagen, Valencia, CA, EE. UU.). El ADN se reamplificó en un volumen total de 25 µl que contenía 1.5 mM de MgCl<sub>2</sub>, 0.2 mM de cada dNTP, 50 ng de cada cebador y 1 U de ADN Taq polimerasa (Invitrogen). La PCR se realizó en un termociclador GeneAmp PCR System 97200 (Perkin-Elmer) siguiendo los protocolos de PCR descritos por Vos, *et al.*, (1995). Los productos de PCR se separaron por electroforesis en geles de agarosa al 1 % (Promega, Madison, Wisconsin, EE. UU.) teñidos con bromuro de etidio, visualizados con un transiluminador UV y fotografiados con un GelDoc 2000 (Bio-Rad, Hercules, CA, EE. UU.). Los productos de PCR del tamaño esperado se extrajeron del gel de agarosa y se purificaron usando un kit de purificación de PCR QIAquick (Qiagen, Valencia, CA, EE. UU.). El ADN obtenido se clonó en el plásmido pGEM-T Easy vector (Promega, Madison, Wisconsin, EE. UU.) y se transformó en células competentes de *Escherichia coli* JM109 según las instrucciones del fabricante (Promega, EE. UU.). Los clones de plásmidos recombinantes se extrajeron con el kit QIAprep Spin Miniprep (Qiagen, Valencia, CA, EE. UU.) y se cuantificaron con el espectrofotómetro NANODROP 2000c (Thermo SCIENTIFIC). El ADN del plásmido recombinante se secuenció por triplicado en Davis Sequencing Company (CA, EE. UU.).

### Análisis in silico de secuencias polimórficas

Los fragmentos polimórficos secuenciados se sometieron al programa VectorScreen para identificar y eliminar las secuencias correspondientes al vector de clonación. Posteriormente, se utilizó el programa BLASTN para identificar los porcentajes de identidad de los fragmentos con las accesiones del GenBank ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). Los marcos abiertos de lectura (ORF) de las secuencias fueron predichos por el traductor fr33 (<http://www.fr33.net/translator.php>).

**Development of AFLP-RGC.** To develop AFLP-RGC, we evaluated a total of 24 primer combinations. The AFLP selective amplifications were performed combining the degenerate primers RGC and five AFLP primers (*EcoR1* + AA, *EcoR1* + AAG, *EcoR1* + AAC, *Mse1* + CAT and *Mse1* + CTG) selected on the basis of the high number of polymorphisms obtained by Soriano *et al.*, (2005) in apricot (*Prunus armeniaca*).

### Cloning and sequencing of polymorphic AFLP-RGC

The AFLP-RGC polymorphic bands were excised from the acrylamide gel and purified using a QIAEX II Gel Extraction Kit (Qiagen, Valencia, CA, USA). The DNA was reamplified in a total volume of 25 µl containing 1.5 mM MgCl<sub>2</sub>, 0.2 mM of each dNTP, 50 ng of each primer and 1 U of Taq DNA polymerase (Invitrogen). The PCR was performed in a GeneAmp PCR System 97200 Thermal Cycler (Perkin-Elmer) following the touchdown PCR protocols described by Vos, *et al.*, (1995). PCR products were separated by electrophoresis on a 1% agarose gels (Promega, Madison, Wisconsin, USA) stained with ethidium bromide, visualized by UV transilluminator and photographed with a GelDoc 2000 (Bio-Rad, Hercules, CA, USA). PCR products of expected size were excised from the agarose gel and purified using a QIAquick PCR purification kit (Qiagen, Valencia, CA, USA). The obtained DNA was cloned into the plasmid pGEM-T Easy vector (Promega, Madison, Wisconsin, USA) and transformed into *Escherichia coli* JM109 competent cells according the manufacturer's instructions (Promega, USA). Recombinant plasmid clones were extracted using the QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA, USA) and quantified using the NANODROP 2000c Spectrophotometer (Thermo SCIENTIFIC). The recombinant plasmid DNA was sequenced by triplicate in Davis sequencing Company (CA, USA).

### In silico analysis of polymorphic sequences

The sequenced polymorphic fragments were subjected to the VectorScreen program to identify and eliminate the sequences corresponding to the cloning vector. Subsequently, the BLASTN program was used to identify the identity percentages of the fragments with the GenBank accessions ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). The open reading frames (ORFs) of the sequences were predict by fr33 translator (<http://www.fr33.net/translator.php>).

## RESULTADOS

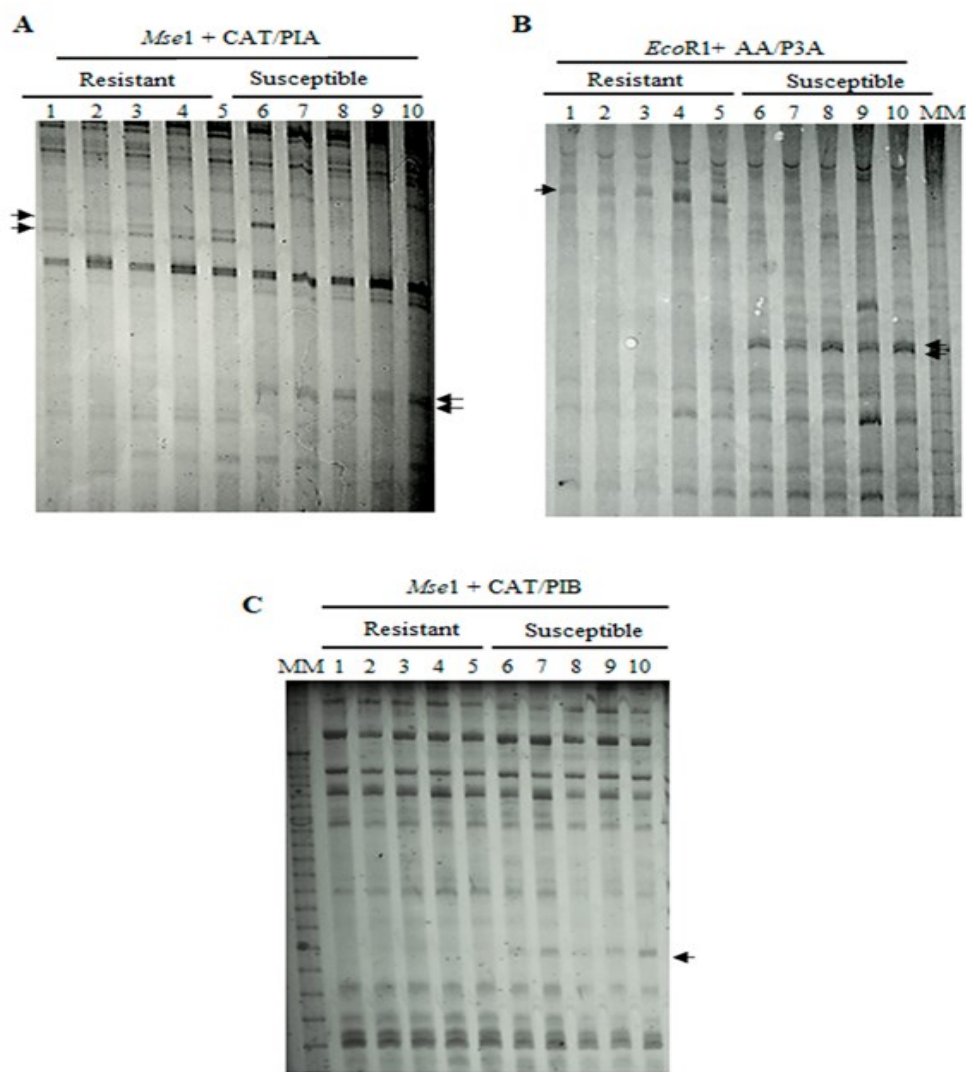
### Identificación de fragmentos polimórficos

De las 24 combinaciones de cebadores utilizadas, cuatro generaron un total de 17 bandas polimórficas; se identificaron nueve en ecotipos resistentes y ocho en susceptibles. El mayor número de bandas polimórficas se obtuvo con la combinación de cebadores *Mse1*-CAT/PIA (10 bandas polimórficas) (Figura 1A) y el menor número con las combinaciones de cebadores *Mse1*-CTG/P3A y *Mse1*/PIB (Figura 1B y C).

## RESULTS

### Identification of polymorphic fragments

Of the 24 primer combinations used, four generated a total 17 polymorphic bands; nine were identified in resistant ecotypes and eight in susceptible. The largest number of polymorphic bands were obtained with the primer combination *Mse1*-CAT/PIA (10 polymorphic bands) (Figure 1A) and the lowest number with primer combinations *Mse1*-CTG/P3A and *Mse1*/PIB (Figure 1B and C).



**Figura 2.** Perfiles de AFLP-RGC entre ecotipos de cocotero resistentes y susceptibles al AL en geles de poliacrilamida al 6% utilizando tres combinaciones de iniciadores. a) combinación *Mse1* + CAT/PIA b) Combinación *EcoR1*+ AA/P3A and c) combinación *Mse1* + CAT/PIB. Carriles: 1 al 5 individuo de ecotipos resistentes y carriles 6 al 10 individuo de ecotipos susceptibles al AL. Las flechas señalan las bandas polimórficas.

**Figure 2.** AFLP-RGC profiles between resistant and susceptible ecotypes on 6% polyacrylamide gels using three primers combinations. a) *Mse1* + CAT/PIA combination. b) *EcoR1*+ AA/P3A combination and c) *Mse1* + CAT/PIB combination. Lanes 1 to 5 resistant and lanes 6 to 10 susceptible ecotype. The arrows indicate the polymorphic bands.

### **Análisis in silico de secuencias polimórficas**

Los fragmentos polimórficos fueron secuenciados y mostraron que contenían los cebadores específicos AFLP y RGC; por lo tanto, no fueron detectadas bandas inespecíficas. El tamaño de los fragmentos obtenidos oscilaba entre 114 y 1026 pb. Las búsquedas de similitud utilizando el programa BLASTN revelaron que nueve marcadores compartían identidad con las accesiones del GenBank (Tabla 3). De estos nueve marcadores, AFLP-RGC 1-01 mostró homología con la proteína de resistencia a enfermedades tipo RPM1 de *Prunus mume* (acceso XP\_008235888.1) y AFLP-RGC 1-17 con el Locus 10 asociado con la resistencia a la enfermedad de la roya de la hoja (Proteína-Cinasa tipo receptor) en *Elaeis guineensis*. AFLP-RGC 1-02 mostró homología con una molécula chaperona de clase DnaJ de *Phoenix dactylifera* (acceso XP\_008786524.1) (Tabla 3), mientras que AFLP-RGC 1-03 y AFLP-RGA 1-07 con secuencias no caracterizadas de esta misma especie (Tabla 3).

Por otra parte, AFLP-RGC 1-04 presentó identidad con una supuesta poliproteína de *Oryza sativa*, cuya función aún no ha sido caracterizado (acceso AAT07561.1) mientras que AFLP-RGC 1-06, AFLP-RGA 1-12 y AFLP-RGA 1-14 con el gen DEF1 alterado, la subunidad 8 del complejo de partículas de proteína de transporte y la peptidil-prolil isomerasa de *E. guineensis*, respectivamente.

### **In silico analysis of polymorphic sequences.**

These fragments were sequenced and shown to contain both the AFLP and the specific RGC primers, therefore unspecific bands were not detected. The size fragments obtained ranging in size from 1026 to 114 bp. Similarity searches using the BLASTN program showed that nine markers shared identity with accessions of GenBank (Table 3). Of the nine markers, AFLP-RGC 1-01 showed homology with disease resistance protein RPM1-like of *Prunus mume* (accession XP\_008235888.1) and AFLP-RGA 1-17 with Leaf Rust 10 Disease-Resistance Locus (Receptor-Like Protein Kinase-like) of *Elaeis guineensis*. AFLP-RGC 1-02 showed significant homology to a DnaJ-class molecule chaperone of *Phoenix dactylifera* (accession XP\_008786524.1) (Table 4) while AFLP-RGC 1-03 and AFLP-RGA 1-07 with uncharacterized sequences of *P. dactylifera* (Table 3).

On the other hand, AFLP-RGC 1-04 presented identity with a putative polyprotein of *Oryza sativa* (accession AAT07561.1) while AFLP-RGC 1-06, AFLP-RGA 1-12 and AFLP-RGA 1-14 with disrupted DEF1 gene, trafficking protein particle complex subunit 8 and peptidyl-prolyl isomerase of *E. guineensis*, respectively.



**Table 3.** Polymorphic sequences isolated on this study, showing the size of amplicon, palm ecotype and the GeneBank accessions that presented the highest identity.

**Tabla 3.** Secuencias polimórficas aisladas en este estudio, mostrando el ecotipo, tamaño del amplicón y número de accesión del GenBank con las que presentaron mayor identidad.

Name/ GenBank Accession	Ecotype	Size (bp)	Highest similarity with GenBank accessions	E-value	Identity (%)	ORF
AFLP-RGC 1-01/ KY115216	S	311	Disease resistance protein RPM1-like ( <i>Prunus mume</i> )	1 <sup>e-12</sup>	40	Yes
AFLP-RGC 1-02/ KY115219	R	133	DnaJ-class molecular chaperone ( <i>Phoenix dactylifera</i> )	0.33	94	Yes
AFLP-RGC 1-03/ KY115217	S	237	Uncharacterized protein LOC103697661 ( <i>Phoenix dactylifera</i> )	4 <sup>e-29</sup>	57	No
AFLP-RGC 1-04/ KY115218	S	160	Putative polyprotein ( <i>Oryza sativa</i> )	9 <sup>e-10</sup>	63	Yes
AFLP-RGC 1-05/ KY230416	S	792	NS	-	-	No
AFLP-RGC 1-06/ KY225990	R	433	Disrupted DEF1 gene ( <i>Elaeis guineensis</i> )	6e-36	72	No
AFLP-RGC 1-07/ KY230417	S	155	Uncharacterized LOC120105813 ( <i>Phoenix dactylifera</i> )	1e-19	96	No
AFLP-RGC 1-08/ KY225991	R	192	NS	-	-	No
AFLP-RGC 1-09/ KY225992	R	133	NS	-	-	No
AFLP-RGC 1-10/ KY225993	R	136	NS	-	-	No
AFLP-RGC 1-11/ KY225994	R	176	NS	-	-	No
AFLP-RGC 1-12/ KY225995	R	192	Trafficking protein particle complex subunit 8 LOC105059818 ( <i>Elaeis guineensis</i> )	2e-11	76	No

AFLP-RGC 1-13/ KY225996	R	130	NS	-	-	No
AFLP-RGC 1-14/ KY230418	S	141	70 kDa peptidyl-prolyl isomerase LOC105038880 ( <i>Elaeis guineensis</i> )	3e-13	79.05	No
AFLP-RGC 1-15/ KY230419	S	142	NS	-	-	No
AFLP-RGC 1-16/ KY230420	S	142	NS	-	-	No
AFLP-RGC 1-17/ KY225997	R	114	Leaf Rust 10 Disease-Resistance Locus Receptor-Like Protein Kinase-like 1.1 LOC105048792 ( <i>Elaeis guineensis</i> )	8e-26	91	No

R: resistant ecotype; S: susceptible ecotype; NS: No Similarity; ORF: Open Reading Frame. E value represent the probability of the alignment occurring by chance

## DISCUSIÓN

En la palma de cocotero, mediante la técnica de AFLP-RGC se identificaron 17 fragmentos polimórficos en los ecotipos EMA y AAO; nueve se identificaron en ecotipos resistentes (EMA) y ocho en ecotipos susceptibles (AAO). Los niveles de polimorfismo obtenidos en palma de coco fueron relativamente bajos con respecto a otras especies como *G. hirsutum* y *G. barbadense* (12 fragmentos polimórficos por combinación de cebadores) (Zhang, et al., 2007) y *Prunus armeniaca* (en promedio 4.5 fragmentos polimórficos por combinación de cebadores) (Soriano, et al., 2005). Sin embargo, en otras especies, como *Glycine max* (2 polimorfismos por combinación) (Hayes y Saghai Marrof, 2000) y *Capsicum annum* (en promedio 0.36 polimorfismos por combinación de primers) (Egea-Gilabert, 2003) los niveles de polimorfismos fueron similares a los obtenidos en este estudio cuyo promedio fue de 0.7 fragmentos polimórficos por combinación. La secuenciación y análisis de los fragmentos polimórficos amplificados permitió identificar que los marcadores AFLP-RGC 1-01, -02, -03, -04, -06, -07, -12, 14 y -17) presentaban identidad con algunas accesiones del GenBank. AFLP-RGC 1-01 presentó identidad con la proteína R de *Prunus mume* homóloga a RPM1 de *Arabidopsis thaliana* involucrada en la resistencia al patógeno bacteriano *Pseudomonas syringae* (Grant, et al., 1995) lo que podría explicar su presencia en palmas susceptibles al AL. AFLP-RGC 1-02 presentó identidad con la proteína DnaJ de *Phoenix dactylifera*, también denominada HSP40 (proteína de choque térmico).

## DISCUSSION

In coconut palm, using the technique of AFLP-RGC we were able to identify 17 polymorphic fragments between MYD and WAT. Of the 17 polymorphisms, nine were identified in resistant ecotypes and eight in susceptible ecotypes. To identify these polymorphisms, we evaluated a total 24 combinations of primers obtaining on average 0.7 polymorphic fragments by combination. The levels of polymorphism obtained in coconut palm were relatively low, with respect to other species such as *G. hirsutum* and *G. barbadense* (12 polymorphic fragments by combination of primers) (Zhang, et al., 2007) and *Prunus armeniaca* (on average, 4.5 polymorphic fragments by combination of primers) (Soriano, et al., 2005), however, in other species, such as *Glycine max* (2 polymorphisms by combination) (Hayes and Saghai Marrof, 2000) and *Capsicum annum* (on average 0.36 polymorphisms by combination of primers) (Egea-Gilabert, 2003) the levels of polymorphisms were similar to those obtained in this study.

The sequencing of the polymorphic fragments allowed us to identify that the markers AFLP-RGC 1-01, -02, -03, -04, -06, -07, -12, 14 and -17) presented identity with GenBank Accession. AFLP-RGC 1-01 presented identity with the R protein of *Prunus mume* homologous to RPM1 of *Arabidopsis thaliana* that confers resistance to the bacterial pathogen *Pseudomonas syringae* (Grant, et al., 1995) that could explain its presence in susceptible palms to LY. Interestingly, AFLP-RGC 1-02 presented identity with the DnaJ protein of *Phoenix dactylifera*, also termed HSP40

Recientemente, se demostró que HSP40 funciona en la inmunidad de las plantas, su sobreexpresión provoca la muerte celular similar a una respuesta hipersensible (HR) mientras que, su silenciamiento aumenta la susceptibilidad al virus del mosaico de la soya (Liu y Whitham, 2013) en plantas resistentes. AFLP-RGC 1-06 no presentó identidad con genes involucrados en mecanismos de defensa, sino con el gen *DEF1* de *E. guineensis* asociado con la variante somaclonal “mantled” de la palma aceitera (Ong-Abdullah, *et al.*, 2015). Por otro lado, AFLP 1-17 presentó identidad con el locus 10 asociado con la resistencia a la enfermedad de la roya de la hoja (Proteína-Cinasa tipo receptor) de *E. guineensis*. La roya de la hoja es causada por el hongo *Puccinia hordei* que parasita una amplia gama de plantas, incluidos helechos y coníferas, y la mayoría de las familias de angiospermas dicotiledóneas y monocotiledóneas (Park, *et al.*, 2015). En palma de coco, hasta donde sabemos *P. hordei* no representa una amenaza fitosanitaria; sin embargo, esta secuencia podría estar asociada a la resistencia a otros patógenos de tipo fúngico que afectan a este cultivo como *Phytophthora palmivora*.

Las secuencias (AFLP 1-05, -08, -09, -10, -11, -13, -15 y -16) no presentaron identidad con ninguna de las accesiones de GenBank y se encuentran en regiones no codificantes (no presentaron ORF). La diversidad de este tipo de secuencias caracterizadas en palma de coco posiblemente es debido a la combinación de un iniciador para RGC más un cebador de la técnica de AFLP que permiten amplificar intrones y secuencias genómicas diferentes de los RGC (Zhang, *et al.*, 2007). A pesar de esto, es posible que estos fragmentos estén ligados a genes *R* u otros genes de interés agronómico.

## CONCLUSIÓN

En el presente estudio se aislaron e identificaron 17 marcadores moleculares basados en las técnicas AFLP-RGC que podrían aplicarse para la selección de germoplasma de cocotero resistente al fitoplasma causante de la enfermedad del amarillamiento letal.

## AGRADECIMIENTOS

El financiamiento parcial de la investigación fue proporcionado por CONACyT, México (Grant No. CB 129717) y Common fund for Commodities, Stadhouderskade 55, 1072 AB. Amsterdam (FIG00/22). Carlos Puch-Hau también agradece a CONACyT por la beca otorgada (No. 47245).

(heat shock protein). Recently, HSP40 was demonstrated to function in plant immunity, its overexpression of HSP40 causes HR-like cell death and silencing of HSP40 enhances susceptibility to Soybean mosaic virus in Soybean (Liu and Whitham, 2013). AFLP-RGC 1-06 did not present identity with genes involved in defense mechanisms, but also with the *DEF1* gene of *E. guineensis*, which is associated with the mantled somaclonal variant of oil palm (Ong-Abdullah, *et al.*, 2015). On the other hand, AFLP 1-17 presented identity with Leaf Rust 10 Disease-Resistance Locus Receptor-Like Protein Kinase-like of *E. guineensis*. Leaf Rust is caused by the pathogenic fungus *Puccinia hordei* that parasitize a wide range of plants, including ferns and conifers, and most families of dicotyledon and monocotyledon angiosperms (Park, *et al.*, 2015). In coconut palm, as far as we know, *Puccinia hordei* does not represent a phytosanitary threat, however, this sequence could be associated with resistance to other fungal-type pathogens that affect this crop such as *Phytophthora palmivora*.

The other sequences (AFLP 1-05, -08, -09, -10, -11, -13, -15 and -16) did not present identity with any of the accessions of GenBank and are found in non-coding regions (they did not present ORF). The diversity of this type of sequences characterized in coconut palm, possibly due to the combination of a primer for RGC plus an AFLP primer that allow amplifying variable genomic sequences including introns and non-RGC genomic sequences (Zhang, *et al.*, 2007), despite this, it is possible that these fragments are linked to *R* genes or other genes of agronomic interest.

## CONCLUSIONS

On this study were isolated and identified molecular markers based on AFLP-RGC technique that could be applied for the selection of germplasm resistance to the phytoplasma that causes lethal yellowing disease in coconut palms.

## ACKNOWLEDGMENTS

Partial funding of the research reported here was provided by CONACyT, Mexico (Grant No. CB 129717) and Common fund for Commodities, Stadhouderskade 55, 1072 AB. Amsterdam (FIG00/22). Carlos Puch-Hau also thanks CONACyT for the scholarship awarded (No. 47245).

## REFERENCES

- Córdova, I., Oropeza, C., Harrison, N., Ku-Rodríguez, S., Puch-Hau, C., Narvaez, M., and Sáenz, L. 2019. Simultaneous Detection of Coconut Lethal Yellowing Phytoplasmas (Group 16SrIV) by Real-Time PCR Assays Using 16Sr- and GroEL-Based TaqMan Probes. *Journal of Plant Pathology*. 101: 609-619. <https://doi.org/10.1007/s42161-019-00249-y>
- Egea-Gilabert, C., Dickinson, M., Bilotti, G., Candela M.E. 2003. Isolation of Resistance Gene Analogs in Pepper Using Modified AFLPs. *Biologia Plantarum*. 47: 27-32. <https://doi.org/10.1023/A:1027368528704>
- FAOSTAT 2022. Recovery 12 october 2022. <https://www.fao.org/faostat/en/#home>
- Grant, M.R., Godiard, L., Straube, E., Ashfield, T., Lewald, J., Sattler, A., Innes, R.W. and Dangl, J.L. 1995. Structure of the Arabidopsis RPM1 gene enabling dual specificity disease resistance. *Science*. 269: 843-846. <https://doi.org/10.1126/science.7638602>
- Harrison, N.A., Richardson, P.A., Kramer, J.B. and Tsai, J.H. 1994. Detection of the mycoplasma-like organism associated with lethal yellowing disease of palms in Florida by polymerase chain reaction. *Plant Pathology*. 43: 998-1008. <https://doi.org/10.1111/j.1365-3059.1994.tb01649.x>
- Harrison, N.A. and Oropeza, C. 2008. Coconut lethal yellowing. In: Harrison NA, Rao GP, Marcone C (eds) *Characterization, diagnosis and management of phytoplasmas*. Studium Press LLC, USA, pp 219-248.
- Hayes, A. and Saghai Maroof, M. 2000. Targeted resistance gene mapping in soybean using modified AFLPs. *Theor Appl Genet*. 100: 1279-1283. <https://doi.org/10.1007/s001220051435>
- Kourelis, J. and van der Hoorn, R.A.L. 2018. Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function. *The Plant Cell*. 30: 285-299. <https://doi.org/10.1105/tpc.17.00579>
- Liu, J. Z. and Whitham, S.A. 2013. Overexpression of a soybean nuclear localized type-III DnaJ domain-containing HSP40 reveals its roles in cell death and disease resistance. *The Plant journal: for cell and molecular biology*. 74: 110-121. <https://doi.org/10.1111/tpj.12108>
- Ong-Abdullah, M., Ordway, J., Jiang, N. *et al.* 2015. Loss of *Karma* transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature*. 525: 533-537. <https://doi.org/10.1038/nature15365>
- Park, R.F., Golegaonkar, P.G., Derevnina, L., Sandhu, K.S., Karaoglu, H., Elmansour, H.M., Dracatos, P.M. and Singh, D. 2015. Leaf rust of cultivated barley: pathology and control. *Annual review of phytopathology*. 53: 565-589. <https://doi.org/10.1146/annurev-phyto-080614-120324>
- Patil, P.G., Datta, S., Ag, bawa, I.O., Singh, I.P., Soren, K.R., Das, A., Choudhary, A.K. Chaturvedi, S.K. 2014. Using AFLP-RGA markers to assess genetic diversity among pigeonpea (*Cajanus cajan*) genotypes in relation to major diseases. *Acta Botanica Brasílica*. 28: 198-205.
- Pflieger, S., Lefebvre, V. and Causse, M. 2001. The candidate gene approach in plant genetics: a review. *Molecular Breeding*. 7, 275-291. <https://doi.org/10.1023/A:1011605013259>
- Puch-Hau, C., Oropeza-Salín C., Peraza-Echeverría S., Gongora-Paredes, M., Córdova-Lara, I., Narvaez-Cab, M., Zizumbo-Villareal, D., & Sáenz-Carbonell, L. 2015. Molecular cloning and characterization of disease-resistance gene candidates of the nucleotide binding site (NBS) type from *Cocos nucifera* L. *Physiological and Molecular Plant Pathology*. 89: 87-96. <https://doi.org/10.1016/j.pmpp.2015.01.002>
- Puch-Hau, C., Oropeza C., Góngora-Paredes, M., Cordova, I., Tun-Suárez, J. and Sáenz, L. 2016. New insights into the evolutionary history of resistance gene candidates in coconut palms and their expression profiles in palms affected by lethal yellowing disease. *Genes Genom*. 38: 793-807. <https://doi.org/10.1007/s13258-016-0422-6>
- Puch-Hau, C., Oropeza-Salín, C., Peraza-Echeverría, S., Córdova-Lara, I., & Sáenz-Carbonell, L. 2021. Resistance Gene Candidates (RGCs) in Coconut Palm: A Molecular Platform for the Genetic Improvement of Resistance to Pathogens. M. K. Rajesh et al. (eds.), *The Coconut Genome, Compendium of Plant Genomes*, [https://doi.org/10.1007/978-3-030-76649-8\\_12](https://doi.org/10.1007/978-3-030-76649-8_12). Springer Nature Switzerland AG 2021. Chapter12. ISBN 978-3-030-76648-1 ISBN 978-3-030-76649-8 (eBook). Pp 165-173.
- Ravensdale, M., Bernoux, M., Ve, T., Kobe, B., Thrall, P.H., Ellis, J.G. and Dodds, P.N. 2012. Intramolecular interaction influences binding of the Flax L5 and L6 resistance proteins to their AvrL567 ligands. *PLoS Pathog*. 8: e1003004. <https://doi.org/10.1371/journal.ppat.1003004>

- Sambrook, J., Fritsch, E.F. and Maniatis, T. 1989. *Molecular cloning: a laboratory manual*, 2nd edn. Cold Spring Harbor Laboratory Press, New York.
- Soriano, J.M., Vilanova, S., Romero, C., Llácer, G. and Badenes, M.L. 2005. Characterization and mapping of NBS-LRR resistance gene analogs in apricot (*Prunus armeniaca* L.). *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, 110: 980–989. <https://doi.org/10.1007/s00122-005-1920-0>
- Vos, P., Hogers, V., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M., Frijters, J., Pelman, J., Kuiper, M. and Zabeau, M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids. Res.* 23: 4407–4414. <https://doi.org/10.1093/nar/23.21.4407>
- Zhang, J., Yuan, Y., Niu, C., Hinchliffe, D. J., Lu Y., Yu S., Percy G.R., Ulloa M., and Cantrell R.G. 2007. AFLP-RGA markers in comparison with RGA and AFLP in cultivated tetraploid cotton. *Crop Science*. 47: 180-187. <https://doi.org/10.2135/cropsci2006.04.0249>
- Zizumbo-Villareal, D., Colunga-García, Marín, P., Fernández-Barrera, M., Torres-Hernández, N. & Oropeza, C. 2008. Mortality of Mexican coconut germplasm due to lethal yellowing. *Plant Genetic Resources Newsletter, FAO-Bioversity*. 156: 23-33.