

Diseño de cebadores degenerados para la caracterización de genes en *Musa*

Degenerated primers design for gene characterization in *Musa*

C. Giménez-Alvarado y M. Colmenares-Esqueda

Departamento de Biología, Facultad Experimental de Ciencias, Universidad del Zulia, Sector Grano de Oro, Edif. A1, piso 3, Laboratorio de Biotecnología Vegetal. Apto. 506, Maracaibo 4002, Venezuela.

Resumen

Existe gran cantidad de información de secuencias de ADN y proteínas con estudios funcionales en diferentes plantas modelo (*Arabidopsis thaliana*, *Medicago truncatula*) y de interés agrícola (*Oryza sativa*, *Solanum lycopersicum*). Esta información, disponible en las bases de datos en internet, ofrecen una excelente plataforma bioinformática para el diseño de cebadores degenerados que permitan la amplificación de genes homólogos de interés, en especies vegetales cuyos genomas son totalmente desconocidos. En esta investigación se planteó el uso de la estrategia CODEHOP para el diseño de cebadores y su validación como herramienta para la caracterización de genes homólogos relacionados con la floración en Plátano Hartón Enano. Con este software se diseñaron y sintetizaron 13 oligos degenerados, seis para el gen “*Flowering Locus T*” (*FT*) y siete para el gen “*CONSTANS*” (*CO*) de los cuales se probaron 36 posibles combinaciones para *FT* y 49 para *CO*. Con estas combinaciones se lograron identificar y secuenciar tres genes análogos a *FT* y uno para *CO* en DNA genómico de *Musa* AAB Plátano Hartón Enano, con lo que se demostró la aplicabilidad de esta estrategia para la caracterización de genes relacionados con la floración en *Musa* (AAB) cv. Plátano Hartón Enano.

Palabras clave: CODEHOP, genes de floración, genes homólogos, PCR, Plátano Hartón Enano.

Abstract

The huge amount of DNA and protein sequence information with functional studies from model (*Arabidopsis thaliana*, *Medicago truncatula*) and crop (*Oryza sativa*, *Solanum lycopersicum*) plants available in different databases offer an excellent bioinformatics platform to design degenerated primers for PCR amplification of homologous genes from crops plants with unknown genomes. This research shows the use of CODEHOP strategy for primer design and its validation as a tool for characterization of flowering time homolog genes of Dwarf Harton Plantain. With this software 13 degenerated primers were designed and synthesized, six for Flowering Locus T (*FT*) and seven for CONSTANS (*CO*) gene, which were tested in 36 possible combinations for *FT* and 49 for *CO*. With these combinations it was able to identify and sequence three genes analogous to *FT* and one for *CO* from genomic DNA of Dwarf Harton Plantain, showing the applicability of this strategy for characterization of flowering genes of *Musa* (AAB) cv. Dwarf Harton Plantain.

Key words: CODEHOP, Flowering genes, Homologous genes, PCR, Dwarf Harton Plantain.

Introducción

La secuencia de genomas completos y la gran cantidad de información sobre aspectos funcionales de secuencias genéticas, en diferentes plantas modelos (*Arabidopsis thaliana*, *Medicago truncatula*) y de interés agrícola (*Oryza sativa*, *Solanum lycopersicum*) ofrecen una excelente plataforma bioinformática para el diseño de cebadores, que permitan la amplificación de genes de interés en otras especies vegetales, cuyos genomas son totalmente desconocidos. Kanazin *et al.* (1996) reportaron por primera vez, el uso de cebadores degenerados para la caracterización de genes análogos de resistencia a enfermedades producidas por bacterias y hongos en plantas de soya, a partir de información de secuencias conocidas de estos genes en tabaco (gen *N*), *Arabidopsis* (*RPS2*) y linasa (*L6*), demostraron que estos

Introduction

The sequence of complete genomes and the big quantity of information about the functional aspects of genetic sequences in different model plants (*Arabidopsis thaliana*, *Medicago truncatula*) with agricultural interest (*Oryza sativa*, *Solanum lycopersicum*) offer an excellent bioinformatics platform for the design of primers that allow the amplification of genes with interest in other vegetable species, which genomes are totally unknown. Kanazin *et al.* (1996) reported for the first time the use of degenerated primers for the characterization of analogue genes of resistant to diseases produced by bacteria and fungi in soybean plants after the information of sequences known of these genes in tobacco (*N* gene), *Arabidopsis* (*RPS2*) and linasa (*L6*), showed that these resistant genes

genes de resistencia presentaron dominios de aminoácidos altamente conservados, inclusive entre especies de plantas muy distantes. Esta información podría utilizarse como herramienta para el diseño de cebadores para amplificar, clonar y secuenciar genes homólogos en diferentes especies.

Rose *et al.* (1998) desarrollaron un software para el diseño de cebadores llamado CODEHOP (según sus siglas en inglés, Consensus Degenerate Hybrid Oligonucleotide Primers) que ha sido utilizado desde entonces para la identificación y caracterización de genes ortólogos y parálogos en diferentes plantas (Drabešová *et al.*, 2014, Morant *et al.*, 2002), animales y especies bacterianas (Chakravorty y Vigoreaux, 2010). Con este software se diseñan cebadores altamente degenerados en la región 3', basándose en la información de secuencia de al menos tres o cuatro aminoácidos presentes en bloques conservados (<http://blocks.fhcrc.org/codehop.html>) que son contruidos mediante el alineamiento del mismo tipo de proteína en diferentes especies. El resto del cebador hacia 5' no es degenerado y se diseña en función de los nucleótidos más probables que se encuentran en las bases de datos, según su posición y secuencia entre los ADN's comparados.

En musáceas la estrategia para la caracterización de genes se ha valido de cebadores degenerados diseñados previamente para otras especies (Kanazin *et al.*, 1996). Backiyarani *et al.* (2013) utilizaron esta estrategia y secuenciaron genes de resistencia tipo NBS-LRR en *Musa* spp., y reportaron sus patrones de expresión durante la infección con *Pratylenchus coffeae*.

presented highly conserved amino acids, even between very distant plants. This information might be used as a tool for designing primers to amplify, clone and sequence the homologue genes in different species.

Rose *et al.* (1998) developed the software for designing primers called CODEHOP (named after the English acronym Consensus Degenerate Hybrid Oligonucleotide Primers), which has been used ever since for identifying and characterizing the orthologues and paralogues genes in different plants (Drabešová *et al.*, 2014, Morant *et al.*, 2002), animals and bacterial species (Chakravorty and Vigoreaux, 2010). With this software, highly degenerated primers are designed in the region 3', based on the sequence information taken from three to four amino acids present in conserved blocks (<http://blocks.fhcrc.org/codehop.html>) built after the alignment of the protein in different species. The rest of the primer until 5' is not degenerated and is designed in function of the nucleotides that are more likely to be in the database, according to the position and sequence between the DNAs compared.

In musaceae, the strategy for characterizing the genes has validated from degenerated primers previously designed for other species (Kanazin *et al.*, 1996). Backiyarani *et al.* (2013) used this strategy and sequence resistance genes type NBS-LRR in *Musa* spp., and reported their expression pattern during the infection with *Pratylenchus coffeae*.

Other strategies for the design of degenerated primers have been used for characterizing different gene

Otras estrategias para el diseño de cebadores degenerados se han utilizado para la caracterización de diferentes familias de genes en *Musa*. Chee *et al.* (2011), reportaron un estudio comparativo de genes argonautas en *Musa* spp. y demostraron la presencia del motivo conservado Ago-7 PIWI. Un año después Mlalazi *et al.* (2012), lograron publicar la secuencia de fitoeno sintasa y sus potenciales secuencias regulatorias. Sin embargo, ninguna de estas publicaciones describió la estrategia de diseño de los cebadores degenerados.

En este trabajo, el principal objetivo fue detallar paso a paso la estrategia de diseño de los cebadores degenerados mediante el software CODEHOP y demostrar su aplicabilidad en la caracterización de genes relacionados con la floración en musáceas.

Materiales y métodos

Secuenciación de genes análogos a *FT* y *CO* en ADN de *Musa* AAB “Plátano Hartón Enano”

Extracción del ADN genómico

Se realizó una extracción de ADN a pequeña escala, a partir de 100 mg de hojas frescas congeladas y trituradas en nitrógeno líquido según la metodología del Bromuro de Hexadeciltrimetilamonio (CTAB) (Doyle y Doyle, 1990), con las modificaciones incorporadas por Giménez *et al.* (2008).

Diseño de cebadores degenerados

Se realizó una búsqueda exhaustiva en bases de datos de secuencias de nucleótidos y proteínas de los genes CONSTANS (*CO*) y Flowering locus

families in *Musa*. Chee *et al.* (2011), reported a comparative research of argonauts genes in *Musa* spp. and showed the presence of the conserved Ago-7 PIWI. A year later, Mlalazi *et al.* (2012), published the sequence of phytoene synthase gene and its regulatory potential sequences. However, none of these publications described the design strategy of degenerated primers.

In this research, the main objective was to detail step by step the design strategy of degenerated primers using the software CODEHOP and to show its applicability in the characterization of flowering genes in musaceae.

Materials and methods

Sequencing of *FT* and *CO* genes analogue in *Musa* AAB “Dwarf Harton Plantain DNA”

DNA extraction was performed

Was realized at a small scale from 100 mg of fresh frozen leaves and crushed in liquid nitrogen following the hexadecyltrimethylammonium bromide (CTAB) methodology (Doyle and Doyle, 1990), with the modifications incorporated by Giménez *et al.* (2008).

Design of degenerated primers

An exhaustive search was conducted in databases of nucleotide and proteins sequences of genes like CONSTANS (*CO*) and Flowering locus T (*FT*), related to the process of floral transition in different plant species. The search for these genes that regulate flowering was carried out with

T (*FT*), relacionados con el proceso de transición floral en diferentes especies vegetales. La búsqueda de estos genes que regulan la floración, se realizó con el software disponible en la página Web del “National Center of Biotechnology Information” (NCBI, según sus siglas en inglés) (<http://www.ncbi.nlm.nih.gov/Database/index.html>).

Genes tipo *CO* y *FT* reportados para arabidopsis, arroz, maíz y cebada con estudios funcionales que los asociaran al control del tiempo de floración (Griffiths *et al.*, 2003; Peerapat, 2005) se usaron para generar grupos homólogos de secuencias usando el software “Basic Local Alignment Search Tool” (BLAST- <http://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Altschul *et al.*, 1997), contra bancos de secuencias de proteínas. Luego estos grupos de secuencias relacionados, se alinearon mediante el software ClustaW v 1.83 (Jeanmougin *et al.*, 1998). A continuación, este formato de secuencias alineadas se usó como entrada para encontrar bloques de secuencias consenso o conservadas mediante el programa “Blocks Multiple Alignment Processor” (http://blocks.fhcrc.org/blocks/process_blocks.html) (Henikoff *et al.*, 1999).

Estas regiones conservadas fueron utilizadas para el diseño de los cebadores degenerados con el software CODEHOP (Rose *et al.*, 1998) (figura 1, cuadro 1).

Condiciones para la PCR

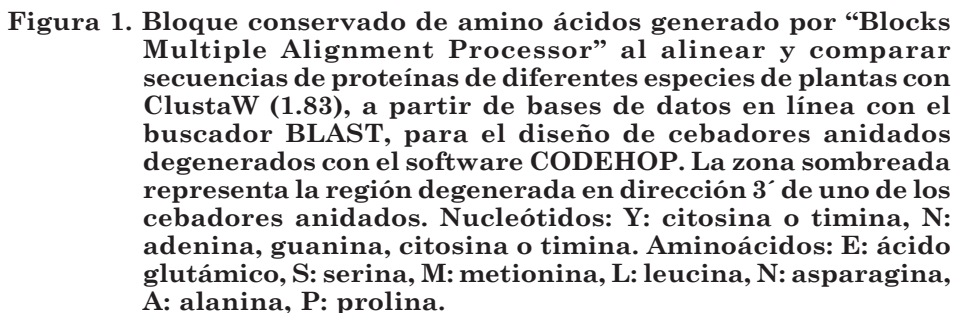
Los ciclos de temperatura necesarios para la reacción de PCR, se realizaron en un termociclador BioRAD modelo i-Cycler. La enzima utilizada fue la

the software available on the website of the “National Center of Biotechnology Information” (with the English acronym NCBI) (<http://www.ncbi.nlm.nih.gov/Database/index.html>).

Gene type *CO* and *FT* reported for arabidopsis, rice, corn, and barley with functional studies that associate them with the control of flowering time (Griffiths *et al.*, 2003; Peerapat, 2005) were used to generate sequence homologous groups using the software “Basic Local Alignment Search Tool” (BLAST - <http://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Altschul *et al.*, 1997), against banks of protein sequences. Then these related groups of sequences, were aligned using ClustaW software v 1.83 (Jeanmougin *et al.*, 1998). Subsequently, this aligned streaming format was used as input to find blocks of consensus or conserved sequences through the program “Blocks Multiple Alignment Processor” (http://blocks.fhcrc.org/blocks/process_blocks.html) (Henikoff *et al.*, 1999). These conserved regions were used for the design of the degenerate primers with the CODEHOP software (Rose *et al.*, 1998) (figure 1, table 1).

Conditions for the PCR

The temperature cycles required for PCR reaction were carried out in a thermal cycler, BioRAD model i-Cycler. The enzyme used was Taq polymerase *Platinum Invitrogene*, with an initial desnaturalization of 2 min at 94°C for achieving a “Hot start”, the rest of the cycle was carried out with a “touchdown” strategy according to Don *et al.* (1991), starting at 60°C as initial coupling temperature of the degenerated primers during 18 cycles,



Taq Polimerasa *Platinum Invitrogene*, con una desnaturalización inicial de 2 min a 94°C para lograr un “Hotstart”, el resto del ciclado se realizó con una estrategia “touchdown” según Donet *al.* (1991), comenzando en 60°C como temperatura inicial de acoplamiento de los cebadores degenerados durante 18 ciclos para luego descender la temperatura - 0,5°C por cada ciclo hasta alcanzar 54°C luego de 12 ciclos (figura 2). Seguidamente, una última extensión a 72°C durante 10 min para favorecer el clonaje de los fragmentos de PCR en el vector pCR®4-TOPO®.

then the temperature descended until - 0.5°C per cycle until reaching 54°C after 12 cycles (figure 2). Then, a final extension at 72°C for 10 min to favor the cloning of PCR fragments in the vector pCR®4-TOPO®.

After the PCR, the amplification products were resolved in 1.5% agarose gels in Tris Base (89 mM), Boric acid (89 mM), EDTA pH= 8 (2 mM) (TBE 1X) stained with an Ethidium Bromide solution (0.5 $\mu\text{g}\cdot\text{L}^{-1}$). Stained gels were observed in an UV light transilluminator at 392 nm. Subsequently, the PCR band was selected to sequence based on the expected size according to the degenerated primers design. Later,

Cuadro1. Cebadores degenerados diseñados sobre secuencias conservadas de los genes *Flowering Locus Ty* *CONSTANS* ensayados en *Musa* AAB Plátano Hartón Enano.

Table 1. Degenerate primers designed on conserved sequences of *Flowering Locus T* and *CONSTANS* genes essayed in *MUSA* AAB Dwarf Harton Plantain.

Código	Secuencia	Tm °C
FLT-1F	5'-CGGACCTTCTACACCCTGGTnatggtngayc-3'	62,3
FLT-2F	5'-CCTTCTACACCCTGGTGatggtngaycc-3'	62,4
FLT-3F	5'-CCTTCTACACCCTGGTGATGgtngayccnga-3'	62,4
FLT-1R	5'-GGACTCCCGCTGGcarttrwarwa-3'	60,5
FLT-2R	5'-GCCGGACTCCCGCtgrcarttrwa-3'	62,7
FLT-3R	5'-CGCAGCCGGACTCCckytgrcartt-3'	62,7
CO-2F	5'-AACCGGGTGGCCTCCmgncaygarmg-3'	60,4
CO-1F	5'-CGCCTACCTGTGCGCCwsntgygayrc-3'	62,7
CO-3F	5'-CGGGTGGCCTCCCGncaygarmng-3'	63,5
CO-4F	5'-TGGCCTCCCGGaygarmngt-3'	60,4
CO-1R	5'-CTTCTGGAACCTCCGGGTCtkyttytytc-3'	60,6
CO-3R	5'-CAGCACCCGGGCCctcnckntcywt-3'	62,0

geles de agarosa al 1,5% en Tris Base (89 mM), ácido Bórico (89 mM), EDTA pH= 8 (2 mM) (TBE 1X) teñidos con una solución de Bromuro de Etidio (0,5 $\mu\text{g.L}^{-1}$). Los geles teñidos se observaron en un transiluminador de luz UV a 392 nm. Posteriormente se seleccionó la banda de PCR candidata a

it was proceeded to purify from the agarose, cloning and transformation in chemically competent *E. coli* DH5 α -T1R (*Invitrogene*). The DNA recovery of agarose gels was done using the QG kit (QIAGEN) following the manual indications of the distributing company QUIAGEN www.qiagen.com/.

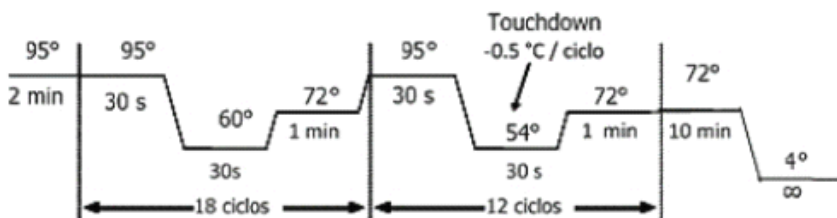


Figura 2. Perfil de termociclado utilizado para las amplificaciones de genes análogos de floración en *Musa* AAB Plátano Hartón Enano.

Figure 2. Thermal-cycle profile used for the amplifications of flowering analogues genes in *Musa* AAB Dwarf Harton Plantain.

secuenciar con base al tamaño esperado calculado según el diseño de los cebadores degenerados. Después se procedió a su purificación a partir de la agarosa, clonaje y transformación en *E. coli* DH5 α -T1R químicamente competente (*Invitrogene*).

La recuperación del ADN de los geles de agarosa se realizó mediante el uso del kit QG (QIAGEN), siguiendo las indicaciones del manual de uso de la compañía distribuidora QIAGEN www.qiagen.com/.

Procedimiento de ligación en vector pCR[®]4-TOPO[®] (*Invitrogene*)

El procedimiento de clonaje que se utilizó, se basó en la propiedad de la enzima Taq DNA polimerasa, la cual colocó después de cada fase de extensión, una adenina en los extremos de los fragmentos amplificados, lo que generó extremos cohesivos en el fragmento de PCR y el vector presentó extremos libres de timina en el sitio de ligación. El kit TOPO de *Invitrogene*, se utilizó para ligar el fragmento de ADN amplificado por PCR, donde la enzima ADN topoisomerasa permitió la ligación del fragmento de PCR al vector pCR[®]4-TOPO[®] en solo 5 min. El procedimiento para la ligación se realizó según las indicaciones del manual de uso de la compañía *Invitrogen* (<http://www.invitrogen.com/>). Una vez realizada la ligación se procedió a la transformación del vector pCR[®]4-TOPO[®] con el inserto de interés en bacterias *E. coli* DH-5 α .

Transformación de bacterias químicamente competentes

Para la transformación de bacterias químicamente competentes DH-5 α , se aplicó el protocolo de choque térmico en cloruro de calcio según la casa distribuidora *Invitrogene*. Después de

Ligation procedure in vector pCR[®]4-TOPO[®] (*Invitrogene*)

The cloning procedure used was based on the property of the Taq DNA polymerase enzyme, which placed after each extension phase an adenine in the ends of the amplified fragments, causing cohesive extremes in the PCR fragment, and the vector presented a tail ends of thymine for ligation. The TOPO kit of *Invitrogene* was used to ligate the DNA fragment amplified by PCR, where the DNA topoisomerase enzyme allowed the ligation of the PCR fragments to the vector pCR[®]4-TOPO[®] in only 5 min. The procedure for the ligation was performed according to the instructions of the manual of the *Invitrogen* company (<http://www.invitrogen.com/>). Once the ligation is done, it was proceeded to the transformation of the vector pCR[®]4-TOPO[®] with the ligated PCR fragment in *E. coli* DH-5 α bacteria.

Transformation of chemically competent bacteria

For the transformation of chemically competent bacteria DH-5 α , the thermal shock protocol in calcium chloride was applied according to the house distributor *Invitrogene*. After obtaining a good number of isolated colonies (PCR fragment library), it was proceeded to amplify the greatest number of isolates by colony PCR using the universal primers T3 and T7, which flanked the site of cloning vector pCR[®]4-TOPO[®] (*Invitrogene*).

PCR of bacterial colonies

The PCR of bacterial colonies consisted of a PCR reaction, using as templates the bacteria that adhered when picking a colony with the tip of

obtener un buen número de colonias aisladas (librería del fragmento de PCR), se procedió a la amplificación del inserto del mayor número de colonias aisladas mediante PCR en colonia, usando cebadores universales T3 y T7 que flanquearon el sitio de clonaje del vector pCR®4-TOPO® (*Invitrogene*).

PCR de colonias bacterianas

La PCR de colonias bacterianas, consistió en una reacción de PCR, usando como templado las bacterias que se adhirieron al picar una colonia con la punta de la micropipeta, la cual se replicó en una placa y se le asignó un código. Luego, se lavó la punta de la micropipeta en 50 µL de agua estéril, usándose esta suspensión como templado para la PCR. Para liberar el ADN de esta suspensión de bacterias, se calentó a 98°C por 10 min. Después se realizó una PCR usando los cebadores universales T3 y T7 que flanquearon el sitio de clonaje del vector pCR®4-TOPO® (*Invitrogene*), para asegurar que lo analizado fuera una sola secuencia y no una mezcla. Luego de obtener la amplificación del inserto, a partir del plásmido bacteriano por PCR, se procedió a un análisis con enzimas de restricción de corte frecuente.

Análisis de fragmentos de PCR amplificados a partir de clones bacterianos aislados

Precipitación de productos de PCR

Este procedimiento se empleó para concentrar y purificar los fragmentos de PCR, con la finalidad de que sus componentes no interfirieran con las reacciones de las enzimas de restricción.

A la reacción de PCR se le añadió 1/10 del volumen de acetato de sodio 3M a pH 4,6 y luego se añadió dos vo-

lumes de la micropipeta, which was replicated in a plate and later assigning a code. Then, the tip of the micropipette was washed in 50 µL of sterile water using this suspension as template for the PCR. To release the DNA of this bacteria suspension, it heated at 98°C for 10 min. Subsequently, a PCR was performed using the universal primers T3 and T7 that outflanked the cloning site of the vector pCR®4-TOPO® (*Invitrogene*) to ensure that the material analyzed was a single sequence and not a mixture. After getting the amplification of the insert on the basis of the bacterial plasmid by PCR, an analysis was carried out with restriction enzymes of frequent cut.

Analysis of amplified PCR fragments from isolated bacterial clones

Precipitation of PCR products

This procedure was used to concentrate and purify the PCR fragments, in order that the components would not interfere with the reactions of restriction enzymes.

Sodium acetate (3M, pH: 4.6) was added in 1/10 of the volume of PCR reaction; subsequently, two volumes of 95% ethanol were added and mixed in vortex until precipitating the DNA at -20°C for 30 min. Later, it centrifuged at 14.000 xg for 15 min and the supernatant was carefully aspirated with a micropipette, removing the liquid completely. Then, the pellet was washed with 250 µL of 70% ethanol and the ethanol was carefully aspirated with a micropipette and let it dry by evaporation. Finally the pellet was

lúmenes de etanol 95% y se mezcló en vórtex para seguidamente precipitar el ADN a -20°C durante 30 min. Posteriormente se centrifugó a 14.000 xg por 15 min y se aspiró cuidadosamente el sobrenadante con una micropipeta, removiendo el líquido completamente.

Posteriormente el precipitado se lavó con 250 μL de etanol 70% y se aspiró cuidadosamente el etanol con una micropipeta y se dejó secar por evaporación. Por último se resuspendió el precipitado en 20 μL de agua destilada estéril.

Luego que se purificaron los productos de PCR, se procedió a digerirlos con enzimas de restricción de alto corte, *HINF* I, *RSA* II, *DRA* I, que reconocieron cuatro pares de bases, según indicaciones del distribuidor. Los patrones de restricción obtenidos, se agruparon con la ayuda de un dendrograma que permitió seleccionar las secuencias según su patrón de restricción.

Secuenciación

Para la secuenciación se seleccionaron fragmentos de PCR de cada grupo del dendrograma y se amplificaron por PCR de colonias bacterianas, usando cebadores universales T3 y T7. Luego se purificaron con el kit QIAGEN para productos de PCR, descrito anteriormente y se procedió a su secuenciación en un secuenciador automático de la compañía "Eurofins Genomics" (<http://www.eurofinsgenomics.com/>). Obtenidos los resultados de la secuencia genética de los materiales amplificados por PCR, se contrastaron con los bancos de datos de nucleótidos y proteínas para verificar si lo clonado realmente fue un análogo de floración en *Musa*.

suspended in 20 μL of sterile distilled water. After purifying the PCR products, it was proceeded to digest them with high cut restriction enzymes, *HINF* I, *RSA* II, *DRA* I, that recognized four base pairs, according to the indications of the distributor. The restriction patterns obtained were grouped with the aid of a dendrogram that allowed selecting the sequences according to their restriction pattern.

Sequencing

For sequencing, the PCR fragments were selected from each group of the dendrogram and were amplified by PCR of bacterial colonies using universal primers T3 and T7. Then were purified with the QIAGEN kit for PCR products, as described above and proceeded to its sequencing in an automatic sequencer of the company "Eurofins Genomics" (<http://www.eurofinsgenomics.com/>). The results of the genetic sequence of the PCR-amplified materials were contrasted with the data banks of nucleotides and proteins to verify if the cloned was really a flowering analogue in *Musa*. After being cloned and sequenced, these results were contrasted in different databases and analyzed with the NCBI-blast2 program searching for homologous sequences. These analyzes were based on indexes calculated by algorithms that estimated the "*E*" probability (table 2 and 3) that a sequence seems to another by pure coincidence, which made it possible to infer whether the sequence might have some functional relationship to its counterpart already reported in the database.

Luego de clonados y secuenciados, estos resultados fueron contrastados en diferentes bases de datos y analizados con el programa NCBI-Blast2 en búsqueda de secuencias homólogas. Estos análisis se basaron en índices calculados por algoritmos que estimaron la probabilidad “E” (cuadro 2 y 3) de que una secuencia se pareciera a otra por pura coincidencia, lo que permitió inferir si la secuencia podría tener alguna relación funcional a su homóloga ya reportada en la base de datos.

Resultados y discusión

Diseño de cebadores degenerados

En total se diseñaron y sintetizaron 13 cebadores degenerados, seis para *FT* y siete para *CO*, de los cuales se probaron 36 posibles combinaciones para *FT* y 49 para *CO*. Con estas combinaciones se lograron secuenciar tres genes análogos a *FT* y uno para *CO* en DNA genómico de *Musa* AAB Plátano Hartón Enano.

Results and discussion

Design of degenerate primers

In total 13 degenerate primers were designed and synthesized, six for *FT* and seven for *CO*, out of which were tested 36 possible combinations for *FT* and 49 for *CO*. With these combinations were achieved three sequencing genes similar to *FT* and one for *CO* in genomic DNA of *Musa* AAB Dwarf Harton Plantain.

The combinations of degenerate primers that originated positive sequences for these genes were:

Gen *FT*: *FT-1F*: 5'-CGGACC TTCTACACCCTGGTnat ggtngayc-3'

FT-3R: 5'-CGCAGCCGGACTCC ckytgrcartt-3';

Gen *CO*: *CO-1F*: 5'-CGCCTAC CTGTGCGCCwsntgygayrc-3'

CO-2R: 5'-GATCCGCGGCCGG rrytngcrta-3'

Analysis of the sequences obtained by amplification using the degenerate primers designed by CODEHOP in *Musa* spp.

Cuadro 2. Análisis BLASTX de la secuencia MUSAFLT2 (DQ153047).

Table 2. BlastX analysis of the MUSAFLT2 sequence (DQ153047).

DB:ID	Descripción	a.a.	Índice	Ident%	Posit%	E
Q76CC3_POPNI	Flowering locus T.	174	351	90	91	6e-41
Q84XL0_SOLLC	SP3D.	177	350	86	91	7e-41
Q76BW3_POPNI	Flowering locus T.	174	350	90	91	7e-41
Q6R3R0_POPDE	Floweing locus T	174	350	90	91	7e-41
A8WES6_MAIZE	ZCN15.	177	341	87	90	9e-41
Q93WM7_ORYSI	Hd3a protein.	179	340	86	91	7e-31

Leyenda: Amino ácidos (a.a), índice de similitud basado en una matriz BLOSUM62 (Índice), Porcentaje de identidad (Ident%), Porcentaje de amino ácidos positivos (Posit%) y probabilidad de coincidencia por casualidad con la secuencia problema (E).

Cuadro 3. Análisis BLASTX de las secuencia MUSACO1 (DQ153049).**Table 3. BLASTX analysis of the *MUSACO1* sequence (DQ153049).**

DB:ID	Descripción	a.a.	Índice	Ident%	Posit%	E
Q7XQH7_ORYSJ	OSJNBa0067K08	333	551	49	58	9e-55
Q01IC4_ORYSA	OSIGBa0092E01	331	551	49	58	9e-55
Q8L4X3_HORVD	CONSTANS-like.	323	548	48	57	2e-54

Leyenda: Amino ácidos (a.a), índice de similitud basado en una matriz BLOSUM62 (Índice), Porcentaje de identidad de amino ácidos (Ident%), Porcentaje de amino ácidos positivos (Posit %) y probabilidad de coincidencia por casualidad con la secuencia problema (E).

Las combinaciones de cebadores degenerados que originaron secuencias positivas para estos genes fueron:

Gen *FT*: *FT-1F*: 5'-CGGACC TTCTACACCCTGGTn atggtn gayc-3'
FT-3R: 5'-CGCAGCCGGACTCC ckytgrcartt-3';

Gen *CO*: *CO-1F*: 5'-CGCCTACC TGTGCGCCwsn tgygayrc-3'

CO-2R: 5'-GATCCGCGGCCGGr yytngcrt-3'

Análisis de las secuencias obtenidas por amplificación utilizando los cebadores degenerados diseñados por CODEHOP en *Musa* spp.

Estas combinaciones de cebadores degenerados, permitieron la amplificación de bandas; sin embargo, solo para los cebadores de *CONSTANS* el tamaño teórico estimado de las bandas (~ 800 pb) coincidió con el amplificado.

Basado en el análisis BLAST, se encontró que los productos de PCR secuenciados eran análogos a genes de floración en otras especies de plantas, monocotiledóneas como arroz, avena y maíz y dicotiledóneas como *arabidopsis* entre otras (cuadro 2 y 3).

These combinations of degenerate primers allowed the amplification of bands; however, only for the *CONSTANS* primers the estimated theoretical size of the bands (~ 800 pb) agreed to the amplified. Based on the BLAST analysis, it was found that the PCR products that were sequenced were analogous to flowering genes in other plant species, monocotyledons such as rice, oats and corn and dicots as *arabidopsis* among others (table 2 and 3).

The theoretical sizes of the PCR products were the first indication in order to consider a band as a candidate to be sequenced. The combination of degenerate primers for the *FT* genes only amplified a single band by reaction (523 - 514 bp), which was always of a size greater than the expected (~300 bp). This phenomenon is due to the presence of a ~212 bp intron among the consensus regions used to design the degenerate primers (figure 3).

For the *CO* gen, the PCR amplified band corresponded to a partial exon with the expected theoretical size (figure 4). For the gene type *FT*, in plants such as *O. sativa* it has at least three introns (Peerapat,

Los tamaños teóricos de los productos de PCR fueron el primer indicio para considerar una banda como candidata a ser secuenciada. La combinación de cebadores degenerados para los genes *FT* solo amplificaron una sola banda por reacción (523 - 514 pb), la cual siempre fue de un tamaño superior al esperado (~ 300 pb). Este fenómeno se debió a la presencia de un intrón de ~ 212 pb entre las regiones consensos que fueron usadas para diseñar los cebadores degenerados (figura 3). Para el gen *CO*, la banda de PCR amplificada correspondió a un exón parcial con el tamaño teórico esperado (figura 4). Para el gen tipo *FT*, en plantas como *O. sativa* tiene por lo menos tres intrones (Peerapat, 2005; DQ157461). Mientras que para los tipo *CO*, estos tuvieron la particularidad que solamente se han reportado un intrón (Feng *et al.*, 2002; Q7XQH7). Por lo tanto, la probabilidad de conseguir un intrón entre dos motivos conservados fue muy baja, lo que determinó que el tamaño teórico estimado de las bandas fuera muy cercano al amplificado con los cebadores de generados.

Análogos a *FT* en *Musa* (AAB) cv. Plátano Hartón enano.

Para el gen tipo *FT*, se encontraron tres genes análogos en Plátano Hartón Enano, los cuales fueron registrados en la base de datos del GenBank del NCBI (<http://www.ncbi.nlm.nih.gov/Genbank/>), como *MUSAFLT1* (DQ153045), *MUSAFLT2* (DQ153047), *MUSAFLT3* (DQ153048).

Todos estos genes mostraron valores altamente significativos cuando se contrastaron en las bases de datos

2005; DQ157461). Meanwhile, for the type *CO*, these had the particularity that only one intron has been reported (Feng *et al.*, 2002; Q7XQH7). Therefore, the probability of getting an intron among two conserved motive was very low, which determined that the theoretical size of the bands estimated was very close to the amplified with the degenerate primers.

***FT* analogues in *Musa* (AAB) cv. Dwarf Harton Plantain**

For the gene type *FT*, three analogue genes were found in Dwarf Harton plantain which were recorded in the database of the NCBI GenBank (<http://www.ncbi.nlm.nih.gov/Genbank/>), as *MUSAFLT1* (DQ153045), *MUSAFLT2* (DQ153047), *MUSAFLT3* (DQ153048).

All these genes showed highly significant values when they were contrasted in the databases with the BLASTX program (table 2), showing identity values between 90 and 80% and indexes higher than 300, which was reflected in the low *E* probability that these sequences are similar to these genes by chance. These genes were homologous to genes like *Flowering locus T*, in *S. lycopersicum* (Q84XL0), *O. sativa* (Q93WM7), *Populus nigra* (Q76CC3), among others (table 2).

Possible structure of the *MUSAFLT1* and *MUSAFLT2* genes

An analysis of a greater amount of sequence information of genes *MUSAFLT1* (523 bp) and *MUSAFLT2* (514 bp) showed how both genes had the same organization of exons and introns, basically with variations at the level of the nucleotide sequence. The structure of these genes from 5' showed

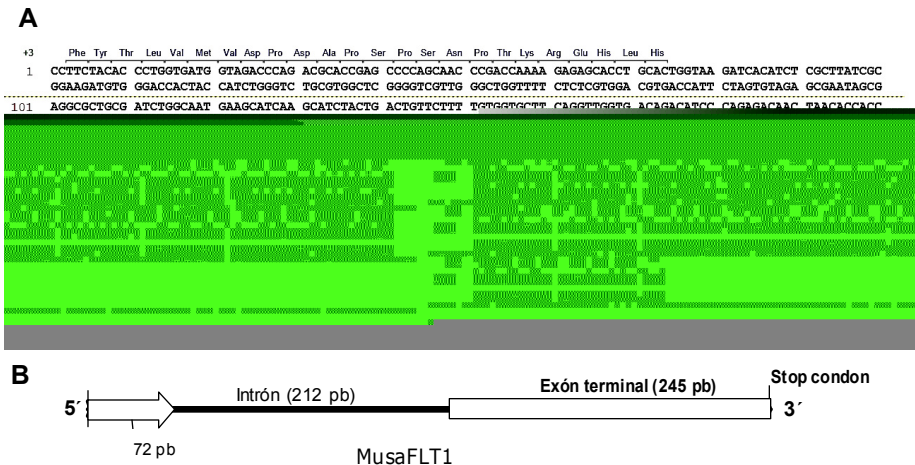


Figura 3. Esquema de la secuencia parcial del gen *MUSAFLT1* análogo a *FT*. A: Secuencia nucleotídica de exones e intrones y traducción de los marcos de lectura. B: Diagrama de la región secuenciada dividida en dos exones (72 y 245 pb) separados por un intrón de 212 pb.

Figure 3. Scheme of the partial sequence of *MUSAFLT1* gene analogue to *FT*. A: nucleus sequence of exons and introns and translation of the readings. B: Diagram of the sequence region divided into two exons (72 and 245 pb) separated by an intron of 212 pb.

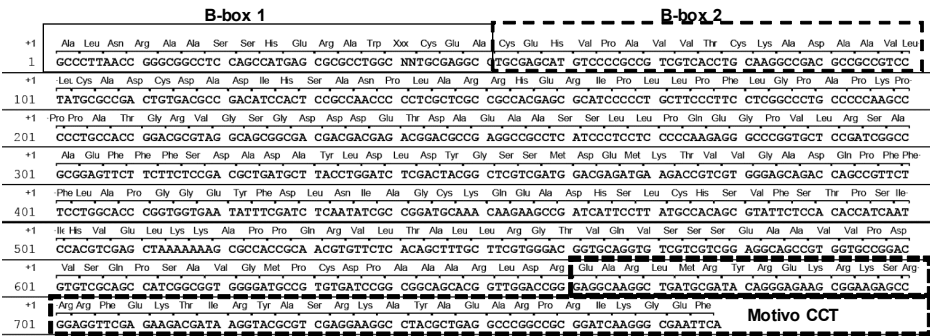


Figura 4. Esquema de la secuencia parcial del gen *MUSACO1* (778 pb)análogo a *CO*. Secuencia nucleotídica y traducción del marco de lectura. Dominios: B-box1, B-box 2 y CCT.

Figure 4. Scheme of the partial sequence of the *MUSACO1* gene (778 pb) analogue to *CO*. Nucleus sequence and translation of the readings. Domains: B-box 1, B-box 2 and CCT.

con el programa BLASTX (cuadro 2), mostrando valores de identidad entre el 90 y 80% e índices superiores a 300, lo que se reflejó en la baja probabilidad *E* de que estas secuencias se parecieran a estos genes por casualidad. Las secuencias análogas a estos genes fueron homologas a genes tipo *Flowering locus T*, en *S. lycopersicum* (Q84XL0), *O. sativa* (Q93WM7), *Populus nigra* (Q76CC3), entre otros (cuadro 2).

Posible estructura del gen *MUSAFLT1* y *MUSAFLT2*

Un análisis de una mayor cantidad de información de la secuencia de los genes *MUSAFLT1* (523 pb) y *MUSAFLT2* (514 pb) mostró como ambos genes tuvieron la misma organización de exones e intrones, básicamente con variaciones a nivel de la secuencia nucleotídica. La estructura de estos genes a partir de 5' mostró la secuencia incompleta de un exón intermedio (72 pb), seguida de un intrón (212 pb) y luego un exón terminal (245 pb) con un codón de parada TAG (figura 3).

El Gen *Hd3a* (DQ157461) (Peerapat, 2005) ortólogo de *FT* en arroz, presentó alta homología con *MUSAFLT2* (cuadro 2). Su secuencia genómica de 3.836 pb, se transcribe en un RNA mensajero de 540 pb, para una proteína de 180 aminoácidos (DQ157461, <http://www.ncbi.nlm.nih.gov/Genbank/>). El resto de la secuencia genómica está compuesta por al menos tres intrones y las regiones no traducidas UTRs (por sus siglas en inglés Untranslated Regions) en 5' y 3'. Al comparar la secuencia de *MUSAFLT2* con este gen *Hd3a*, se infiere que se debe secuenciar en dirección 5' para completar toda la región codante del gen e identificar la

the incomplete sequence of an intermediate exon (72 bp), followed by an intron (212 bp) and then a terminal exon (245 bp) with a stop codon TAG (figure 3).

The *Hd3a* gene (DQ157461) (Peerapat, 2005) *FT* ortholog in rice presented high homology with *MUSAFLT2* (table 2). Its genome sequence of 3.836 pb, is transcribed into a RNA messenger of 540 pb, for a 180-amino acid protein (DQ157461, <http://www.ncbi.nlm.nih.gov/Genbank/>). The rest of the genome sequence is composed of at least three introns and non-translated UTRs regions (for its acronym in English of Untranslated Regions) in 5' and 3'. By comparing the sequence of *MUSAFLT2* with this gene *Hd3a*, it can be inferred that it should be sequence in the 5' direction to complete the entire region and identify the intron structure of *MUSAFLT1* and *MUSAFLT2* genes.

CONSTANS analogue in *Musa* (AAB) cv. Dwarf Harton Plantain

For the *CONSTANS* gene (*CO*), only one homologue of this gene was sequenced in Dwarf Harton plantain, registering it in the database of the GenBank (NCBI) as: *MUSACO1* (DQ153049) (778 bp).

With the BLASTX analysis it was derived that the sequenced gene was a type *CO* gene, analogous to genes from other species such as *O. sativa* (Q01IC4) (Feng *et al.*, 2002), *Hordeum vulgare* (Q8L4X3) (Griffith *et al.*, 2003), etc. (table 3). The BLASTX index values were high, the highest rate corresponded to an *O. sativa* gene (Q7XQH7) (Feng *et al.*, 2002) with a

estructura de intrones de los genes *MUSAF1* y *MUSAF2*.

Análogo a *CONSTANS* en *Musa* (AAB) cv. Plátano Hartón Enano

Para el gen *CONSTANS* (*CO*), solo se secuenció un homólogo de este gen en Plátano Hartón Enano, registrándolo en la base de datos del GenBank (NCBI) como : *MUSACO1* (DQ153049) (778 pb).

Del análisis BLASTX se derivó que el gen secuenciado fue un gen tipo *CO*, análogo a genes de otras especies vegetales como *O. sativa* (*Q01IC4*) (Feng *et al.*, 2002), *Hordeum vulgare* (*Q8L4X3*) (Griffith *et al.*, 2003), etc. (cuadro 3). Los valores del índice BLASTX, fueron elevados, el mayor índice correspondió a un gen de *O. sativa* (*Q7XQH7*) (Feng *et al.*, 2002) con un valor de 551 (cuadro 3). Este gen se caracterizó por estar compuesto por dos exones separados por un único intrón y las regiones UTRs 5' y 3'.

En cebada (*Hordeum vulgare* L.) al estudiar la evolución de los genes tipo *CO* en comparación con *A. thaliana* y *O. sativa* utilizando un análisis Clustal W se mostró que estos podrían tener hasta tres intrones (Griffith *et al.*, 2003). Estos intrones se ubicaron entre los dominios B-box y CCT en la llamada región intermedia. En el caso descrito para el gen de arroz solo se encontró un intrón (Feng *et al.*, 2002, *Q7XQH7*) y para *Musa* AAB Plátano Hartón Enano no se detectó ningún intrón (figura 4), siendo toda la secuencia un marco de lectura terminal, sin codón de iniciación, evidenciándose que aún falta secuenciar hacia la región 5' del gen para identificar el motivo B-box1 completo y la región

value of 551 (table 3). This gene was characterized by being composed of two exons separated by a single intron and the UTRs regions 5' and 3'.

In barley (*Hordeum vulgare* L.) evolution studies of *CO* like genes in comparison with *A. thaliana* and *O. sativa* using a Clustal W analysis it was showed that these could have up to three introns (Griffith *et al.*, 2003). These introns were located between the domains B-box and CCT in the so-called middle region.

In the case described for the rice gene, only one intron was found (Feng *et al.*, 2002, *Q7XQH7*) and for *Musa* AAB Dwarf Harton Plantain none intron was detected (figure 4), being all the sequence a framework of terminal reading without starting codon, showing that it must be sequenced toward the 5' regionz of the gene to identify the complete B-box 1 and the promoting region of the gene. Additionally, to continue the sequence of the 3' region to characterize the possible UTRs sequences in this area.

Conclusions

This research showed the application of the CODEHOP strategy for sequencing genes or groups of genes in unknown genomes, taking advantage of the information published in the data banks of protein sequences in model plants with functional studies, or with completely sequence genomes for the design of degenerate primers that allowed sequencing genes or group of homologous genes with flowering (*MUSAF1*, *MUSAF2* and *MUSAF3* and *MUSACO1*) in *Musa* (AAB) cv. Dwarf Harton Plantain.

promotora del gen. Además, de continuar la secuenciación de la región 3' para caracterizar posibles secuencias UTRs en esta zona.

Conclusiones

Esta investigación mostró la aplicación de la estrategia CODEHOP para secuenciar genes o grupo de genes, en genomas desconocidos, aprovechando la información publicada en los bancos de datos de secuencias de proteínas en plantas modelos con estudios funcionales, o con genomas completamente secuenciados, para el diseño de cebadores degenerados que permitieron secuenciar genes o grupo de genes homólogos relacionados con la floración (*MUSAFLT1*, *MUSAFLT2* y *MUSAFLT3* y *MUSACO1*) en *Musa* (AAB) cv. Plátano Hartón Enano.

Agradecimientos

Este trabajo fue posible gracias al financiamiento de los Proyectos, FONACIT N° 2012001382 y CONDES CC-0334-13, que suministraron materiales y reactivos que permitieron la reapertura del laboratorio de genética molecular del Laboratorio de Biotecnología Vegetal de LUZ.

Literatura citada

- Altschul, S., T. Madden, A. Schäffer, J. Zhang, Z. Zhang, W. Miller y D. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acid Res.* 25(17):3389-3402.
- Backiyarani, S., S. Uma, G. Arunkumar, M. Saraswathi y P. Sundararaju. 2013. Cloning and characterization of NBS-

Acknowledgment

This research was done thanks to the economic support provided to the FONACIT projects N° 2012001382 and CONDES CC-0334-13, by supplying materials and reactive that allowed the re-opening of the molecular genetics laboratory of the Plant Biotechnology Laboratory of LUZ.

End of english version

- LRR resistance gene analogues of *Musa* spp. and their expression profiling studies against *Pratylenchus coffeae*. *African J Biotechnol.* 12 (27):4256-4268.
- Chakravorty S. y J. Vigoreaux. 2010. Amplification of orthologous genes using degenerate primers. *Methods Mol. Bio.* 634(3):175-185. En: *In Vitro Mutagenesis Protocols*. Bramey J. (Ed.).
- Chee, T., H. Pui, R. Othman y J. Harikrishna. 2011. Comparative analysis of Argonaute gene sequences in bananas (*Musa* sp.) shows conserved species-specific Ago-7 PIWI domains. *Genet Resour Crop Evol.* 58(5):713-725.
- Don, R., P. Cox, B. Wainwright, K. Baker, J. Mattick. 1991. Touchdown PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* 19(14):4008.
- Doyle, J. y J. Doyle. 1990. Isolation of DNA from small amounts of plants tissues. *BRL Focus* (12):3-15.
- Drabešová, J., J. Cháb, J. Kolář, K. Haškovcová y H. Štorchová. 2014. A dark-light transition triggers expression of the floral promoter CrFTL1 and downregulates CONSTANS-like genes in a short-day plant *Chenopodium rubrum*. *J. Exp. Bot.* 65(8):2137-2146.

- Ebrahim, S., T. Sharma, K. Usha y B. Singh. 2013. Resistance Gene Analogues in Mango against Mango Malformation. *ActaPhytopatholEntomolHungarica*. 48(1):39-52.
- Feng, Q., Y. Zhang, P. Hao, S. Wang, G. Fu, Y. Huang, Y. Li, J. Zhu, Y. Liu, X. Hu, P. Jia, Y. Zhang, Q. Zhao, K. Ying, S. Yu, Y. Tang, Q. Weng, L. Zhang, Y. Lu, J. Mu, Y. Lu, L. Zhang, Z. Yu, D. Fan, X. Liu, T. Lu, C. Li, Y. Wu, T. Sun, H. Lei, T. Li, H. Hu, J. Guan, M. Wu, R. Zhang, B. Zhou, Z. Chen, L. Chen, Z. Jin, R. Wang, H. Yin, Z. Cai, S. Ren, G. Lu, W. Gu, G. Zhu, Y. Tu, J. Jia, Y. Zhang, J. Chen, H. Kang, X. Chen, C. Shao, Y. Sun, Q. Hu, X. Zhang, W. Zhang, L. Wang, Ch. Ding, H. Sheng, J. Gu, S. Chen, L. Ni, F. Zhu, W. Chen, L. Lan, Y. Lai, Z. Cheng, M. Gu, J. Jiang, J. Li, G. Hong, Y. Xue, B. Han. 2002. Sequence and analysis of rice chromosome 4. *Nature*. 420 (6913):316-320.
- Giménez, C., E. de García y O. Haddad. 2008. Clonal micropropagation and analysis of stability of the resistance to black sigatoka disease in *Musa* somaclonal variant CIEN BTA-03. *Phyton* 77:65-79.
- Griffiths, S., R. Dunford, G. Coupland y A. David. 2003. The evolution of CONSTANS-Like gene families in Barley, Rice, and Arabidopsis. *Plant Physiol*. 131(4):1855-1867.
- Henikoff, J., S. Henikoff, y S. Pietrokovski. 1999. New features of the Blocks Database servers. *Nucl. Acids Res*. 27(1): 226-228.
- Jeanmougin, F., J. Thompson, M. Gouy, D. Higgins, T. Gibson. 1998. Multiple sequence alignment with Clustal X. *Trends Biochem. Sci*. 23(10):403-405.
- Kanazin, V., L. Marek y R. Shoemaker. 1996. Resistance gene analogs are conserved and clustered in soybean. *Proc. Natl. Acad. Sci. USA*, 93(21):11746-11750.
- Mlalazi, B., R. Welsch, P. Namanya, H. Khanna, R. Geijskes, M. Harrison, R. Harding, L. James. 2012. Isolation and functional characterization of banana phytoene synthase genes as potential cisgenes. *Planta*. 236(5):1585-1598.
- Peerapat, R. 2005. Detection of DNA methylation in *Hd1*, *Hd3*, and *Hd6* genes in rice KDML 105 (*Oryza sativa* L.). Peerapat Roongsattham. Bangkok: Kasetsart University.
- Rose, M., R. Emily, J. Schultz, G. Henikoff, S. Pietrokovski, M. Claire, Mc. Callum y S. Henikoff. 1998. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Research*. 26(7):1628-1635.