

Artículo Original

## Estrategias bioinformáticas y moleculares para aislar genes con potencial biotecnológico a partir de especies huérfanas de la ciencia

[Bioinformatics and molecular strategies to isolate genes with biotechnological potential from orphan species of science]

Juan C. Castro-Gómez<sup>1</sup> & Marianela Cobos-Ruiz<sup>1,2\*</sup>

<sup>1</sup>Unidad Especializada de Biotecnología, Centro de Investigaciones de Recursos Naturales, Universidad Nacional de la Amazonía Peruana. Psje. Los Paujiles S/N AAHH Nuevo San Lorenzo, San Juan Bautista, Iquitos-Perú.

<sup>2</sup>Laboratorio de Biotecnología y Bioenergética, Universidad Científica del Perú. Av. Abelardo Quiñones km 2,5, San Juan, Iquitos-Perú.

\*e-mail: [mcobos@ucp.edu.pe](mailto:mcobos@ucp.edu.pe)

### Resumen

La amazonía peruana es megabiodiversa y posee millones de genes con potencial biotecnológico que debemos descubrir. Para identificarlos, aislarlos y poder aprovecharlos tenemos que secuenciar los genomas y transcriptomas de las especies más importantes. Sin embargo, hasta concretar este tipo de proyectos, combinando estrategias bioinformáticas y moleculares podemos aislar genes con potencial biotecnológico. El objetivo fue aislar genes involucrados en la biosíntesis de vitamina C en *Myrciaria dubia*. De acuerdo a las relaciones filogenéticas de *M. dubia* con otras especies, se buscaron secuencias homólogas en el banco de genes del gen que codifica L-galactono-1,4-lactona deshidrogenasa (*L-GalLDH*). En base al alineamiento múltiple obtenido, se construyó un filograma consenso y de los clados formados se diseñaron cebadores degenerados. Los frutos fueron obtenidos de una colección de germoplasma y el ARN total se purificó, el ADNc se sintetizó y amplificó con cebadores degenerados. El amplicón sintetizado se clonó y secuenció con técnicas estándares. Con las aproximaciones empleadas se logró aislar, clonar y secuenciar un segmento de 921 pb del gen *L-GalLDH* de *M. dubia*. En conclusión, con las estrategias bioinformáticas y moleculares descritas se aisló el segmento de un gen que codifica la enzima L-galactono-1,4-lactona deshidrogenasa de la vía biosintética de vitamina C en *M. dubia*, especie de la que no se disponía de secuencias génicas. Con estas mismas aproximaciones será posible aislar genes de interés biotecnológico de esta y otras especies amazónicas que aún no cuentan con bases de datos de secuencias nucleotídicas.

**Palabras claves:** Biosíntesis de vitamina C, clonación de genes, *Myrciaria dubia*, vías metabólicas

### Abstract

The Peruvian Amazon is mega-biodiverse and has millions of genes with biotechnological potential yet to be discovered. To identify, isolate and use these genes we first must sequence the genomes and transcriptomes of the most important species. In the interim, however, we can combine bioinformatics and molecular strategies to isolate genes with biotechnological potential. The aim of the current study was to isolate genes involved in vitamin C biosynthesis of *Myrciaria dubia*. Using known phylogenetic relationships of *M. dubia*, we searched GenBank for homologous sequences of the gene that encodes L-galactono-1,4-lactone dehydrogenase (*L-GalLDH*). Based on multiple alignments, a consensus phylogram was constructed, and from clades degenerate primers were designed. The fruits were obtained from a germplasm collection and total RNA was purified, cDNA was synthesized and amplified with degenerate primers. The synthesized amplicon was cloned and sequenced by standard techniques. With these methods we have isolated, cloned and sequenced a 921 bp segment of the *L-GalLDH* gene of *M. dubia*. In conclusion, with the bioinformatics and molecular strategies described here, the gene encoding L-galactono-1,4-lactone dehydrogenase from the biosynthetic pathway of vitamin C in *M. dubia* was successfully sequenced even though gene sequences were not available. With this approach it will be possible to isolate genes of biotechnological interest in this and other Amazonian species in which nucleotide sequences databases are lacking.

**Keywords:** Vitamin C biosynthesis, gene cloning, *Myrciaria dubia*, metabolic pathways.

**Recibido:** 22 setiembre 2013

**Aceptado:** 10 noviembre 2013

**Este artículo puede ser citado como** JC Castro-Gómez & M Cobos-Ruiz. 2013. Estrategias bioinformáticas y moleculares para aislar genes con potencial biotecnológico a partir de especies huérfanas de la ciencia. **Cienc amaz (Iquitos)** 3(2), 84-95.

## INTRODUCCIÓN

La amazonía peruana se caracteriza por su megabiodiversidad, que hasta el momento ha sido aprovechada de manera muy limitada, principalmente por actividades extractivistas. Sin embargo, la mayor riqueza que albergan estas especies son los millones de genes y sus variantes alélicas, aún no descubiertos. Muchos de estos genes podrían ser aprovechados de múltiples formas, porque pueden ser patentados, codificar nuevas enzimas, ser empleados para la producción de sustancias de interés industrial y compuestos bioactivos en procesos biotecnológicos *in vitro* e *in vivo*, podrían servir para realizar el mejoramiento genético por intragénesis o transgénesis de diversas especies, entre otras aplicaciones. Pero no podremos aprovecharlos o evaluar su potencial biotecnológico mientras no sean identificados y aislados. Para acelerar el descubrimiento de estos genes debemos realizar proyectos para secuenciar el genoma y transcriptoma de las especies amazónicas de mayor importancia socioeconómica, las que en su gran mayoría son "huérfanas de la ciencia" por la escasez o ausencia de investigaciones científicas realizadas y publicadas. Actualmente, la obtención de información genómica y transcriptómica de cualquier especie son factibles porque disponemos de plataformas de secuenciamiento masivo (next-generation DNA sequencing) que han reducido drásticamente el costo y el tiempo para secuenciar genomas y transcriptomas completos. Estas nuevas plataformas son comercializadas por las compañías Roche 454, Illumina, Helicos BioSciences y Life Technologies (Ozsolak y Milos, 2011) y están siendo empleadas para descifrar los genomas y transcriptomas de diversas especies de animales y plantas (Sweetman *et al.*, 2012; Wang *et al.*, 2009). Pero hasta concretar proyectos de este tipo con las especies amazónicas, podemos emplear otras estrategias que combinen el uso de programas bioinformáticos y técnicas de biología molecular para aislar genes que codifican enzimas de vías metabólicas relevantes. Por ejemplo, *Myrciaria dubia* es un arbusto amazónico que desde hace 50 años

se ha demostrado que produce frutos con altas concentraciones de vitamina C (Bradfield y Roca, 1964; Imán *et al.*, 2011). Sin embargo, desconocíamos que vías metabólicas estaban involucradas en la biosíntesis de esta vitamina y no disponíamos de las secuencias génicas codantes de enzimas participantes de esta vía. Pero, empleando herramientas bioinformáticas y de biología molecular que describimos paso a paso en este artículo, nosotros hemos aislado con éxito todos los genes de la vía biosintética Smirnov-Wheeler, vía responsable de la producción de vitamina C en *M. dubia*, de los cuales, describimos el que codifica la enzima L-galactono-1,4-lactona deshidrogenasa. Por tanto, estas estrategias podrían ser empleadas para aislar y estudiar genes de interés en esta y otras especies de las que no disponemos de secuencias génicas.

## MATERIALES Y MÉTODOS

### Relaciones filogenéticas de la especie de interés

Es fundamental conocer este tipo de relaciones de filogenia molecular de la especie que estamos estudiando con otras especies de plantas. Esta información se obtiene del sitio web de Tree of Life Web Project (<http://tolweb.org/tree/phylogeny.html>) que muestra las relaciones filogenéticas de las especies en base a estudios moleculares (Maddison *et al.*, 2007) y también se puede obtener de Soltis y Soltis (2000), que muestran las relaciones entre los principales clados de 560 angiospermas en base al análisis filogenético combinado de secuencias del ADNr 18S, y de los genes *rbcl* y *atpB* (Figura. 1). El filograma muestra que *M. dubia* del orden Myrtales, tiene más similitud genética con el orden Geraniales, pero menos similitud genética con otros órdenes integrantes de Malvaceae (Crossosomatales, Picramniaceae, Huerteales, Brassicales, etc). Sin embargo, muestra mayor distancia genética con órdenes integrantes de Fabaceae (Fabales, Rosales, fagales, Cucurbitales, Malpighiales, etc). Como se muestra en el filograma, los Fabaceae y Malvaceae pertenecen a los Rósidos.

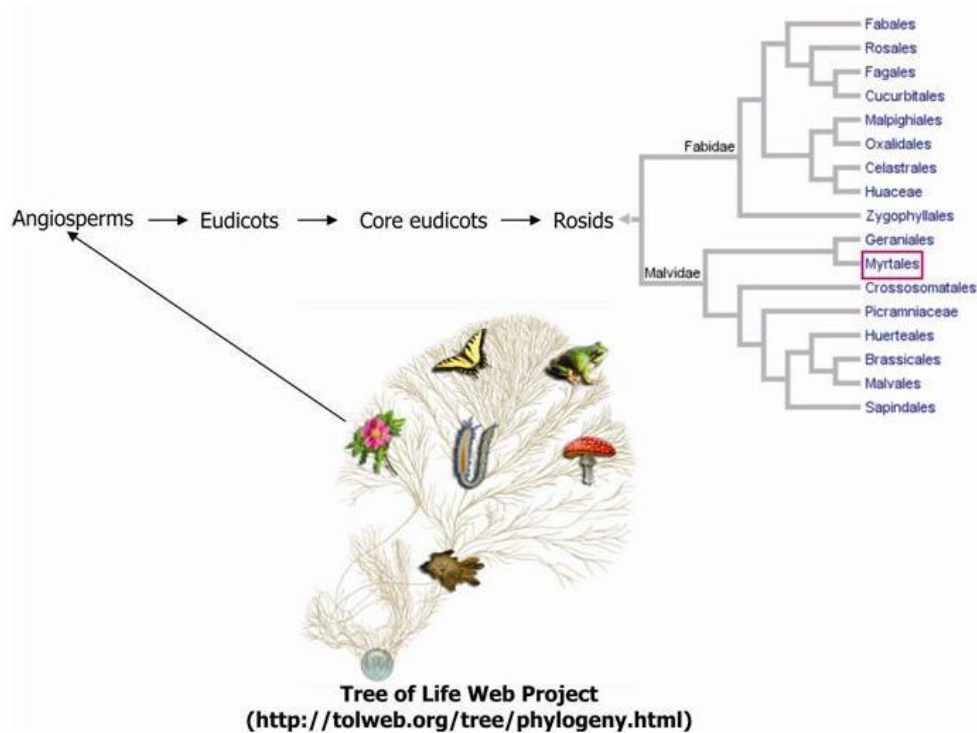


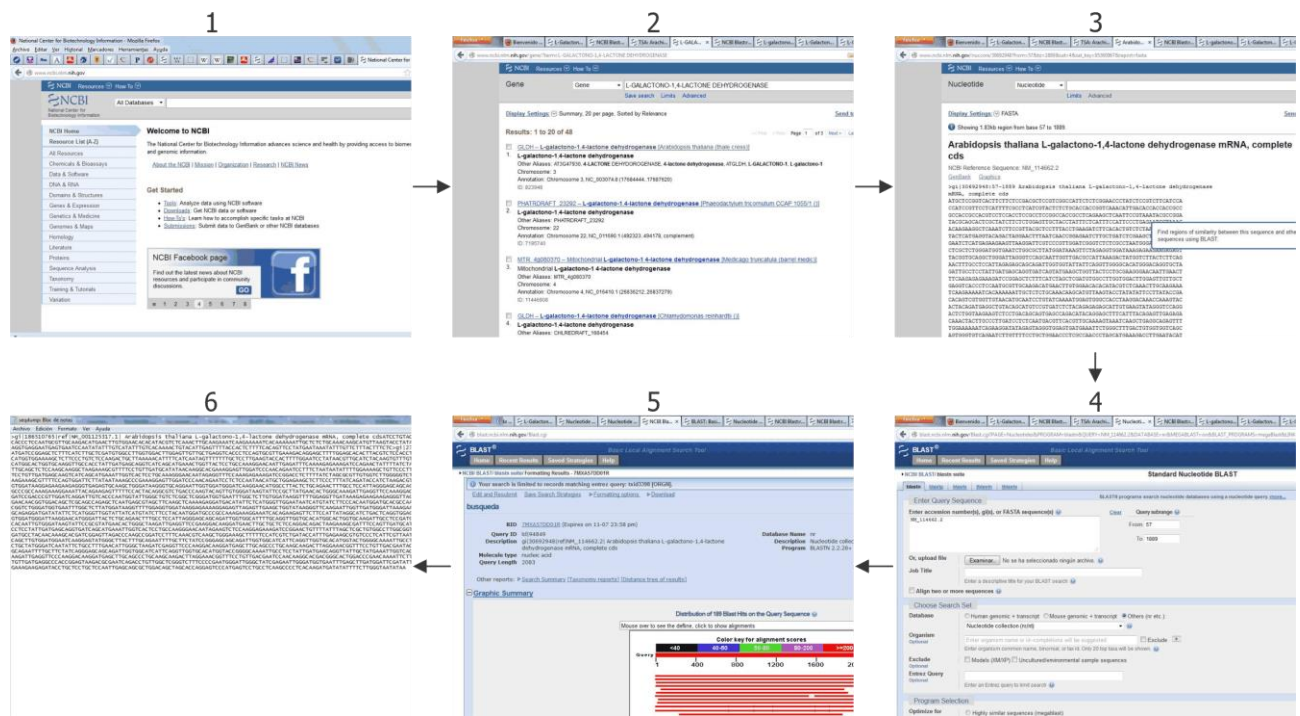
Figura 1. Relaciones filogenéticas moleculares entre rósidos

### Búsqueda de secuencias del gen de interés en el banco de genes

Conociendo las relaciones filogenéticas que tiene nuestra especie con otras especies de plantas, estamos listos para continuar con la siguiente etapa (Figura. 2). Para ello ingresamos al sitio web del NCBI:

<http://www.ncbi.nlm.nih.gov/> (1), del menú desplegable de bases de datos seleccionar **Gene** y en la ventana de búsqueda escribir **L-galactono-1,4-lactone dehydrogenase and Rosids** (el gen que nos interesa aislar y el grupo de plantas al que pertenece nuestra especie) y hacer click en **Search**, del listado que se muestra ingresar y buscar la secuencia completa del ARNm de una de las especies más cercanas, como *Arabidopsis thaliana* (2). Mostrar la secuencia del ARNm en formato FASTA y en la sección **Analyze this sequence** hacer click en **Run BLAST** (3). Para el análisis Blast se pueden emplear diversos filtros para buscar secuencias homólogas. Por ejemplo, en la sección **Choose search set** en bases de datos (**Database**), se pueden buscar en: Nucleotide Collection (nr/nt), Reference RNA sequence (refseq\_rna), Expressed sequences tag (est), Patent sequences (pat), Protein data bank

(pdb), Transcriptome Shotgun Assembly (TSA), etc. y en la opción organismo (**Organism**) seleccionar **flowering plants (taxid:3398), rosids (taxid:71275) ó malvids (taxid:91836)**. En la sección **Program Selection** se puede optimizar para **Highly similar sequences (megablast) ó More dissimilar sequences (discontiguos megablast)**. Luego de establecer los criterios de la búsqueda hacer click en **BLAST** (4). En la página de los resultados, del listado de secuencias encontradas en las bases de datos y mostradas en la sección **Sequences producing significant alignments**, seleccionar aquellas que presenten  $E_{value} \leq 0.0$  e Identidad (Ident)  $\geq 70\%$  y tengan de  $\geq 1500$  a  $\sim 2000$  nucleótidos. Para descargar y guardar las secuencias seleccionadas hacer click en **Download**, seleccionar **FASTA (complete sequence)** hacer click en **Continue** y guardar el archivo con las secuencias en formato fasta (5 y 6). Con el procedimiento descrito se han obtenido las secuencias completas de ARNm de nueve especies de plantas que cumplen con los criterios establecidos en la búsqueda (Tabla 1).



**Figura 2.** Pasos para la búsqueda y selección de secuencias homólogas del gen *L-GalLDH* en el banco de genes

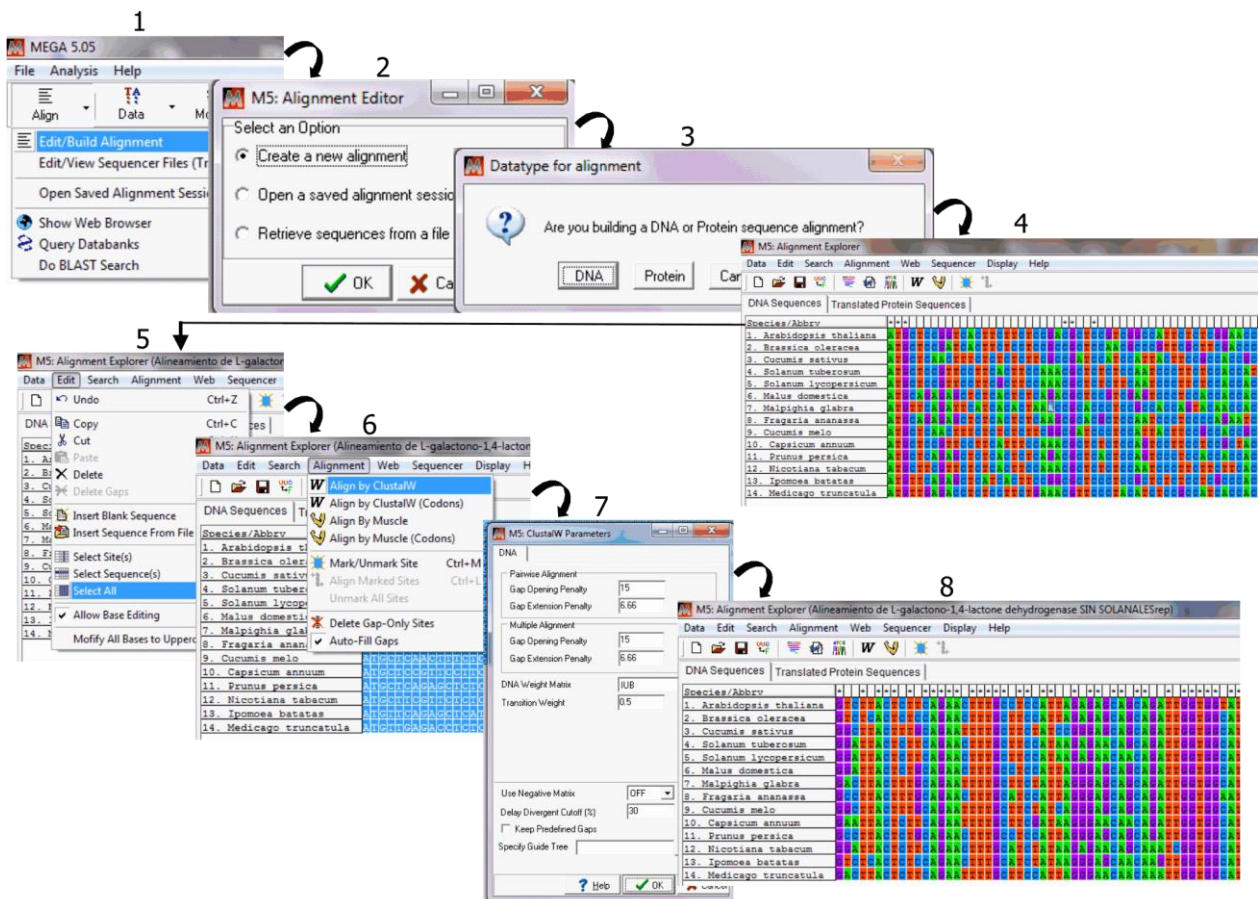
**Tabla 1.** Lista de especies de plantas con las secuencias de ARNm de *L-GalLDH* obtenidos del banco de genes

Especie	ARNm		Orden
	Código GenBank	Nº nucleótidos	
<i>Malpighia glabra</i>	EU683445.1	1779	Malpigiales
<i>Cucumis sativus</i>	HQ446099.1	1773	Cucurbitales
<i>Cucumis melo</i>	AF252339.2	1770	
<i>Fragaria x ananassa</i>	AY102631.1	1749	
<i>Malus domestica</i>	FJ752244.1	1800	Rosales
<i>Prunus persica</i>	AB457586.1	1766*	
<i>Medicago truncatula</i>	XM_003607586.1	1521	Fabales
<i>Arabidopsis thaliana</i>	NM_001125317.1	1539	Brassicales
<i>Brassica oleracea</i>	Z97060.1	1803	

### Alineamiento de secuencias

Después de obtener las secuencias se procede con su alineamiento múltiple. Un programa bastante amigable es el Mega 5.05 ó Mega 6 Beta 2 (Tamura *et al.*, 2013), que pueden ser descargados del sitio <http://www.megasoftware.net>. Para realizar el alineamiento, tal como describimos con imágenes en la Figura 3, en el menú del programa ingresamos a **Align** y seleccionamos **Edit/Build Alignment** (1), en la ventana del editor del alineamiento (**Alignment Editor**) que se muestra (2), seleccionamos la opción crear un nuevo alineamiento (**Create a new alignment**). A continuación se muestra la ventana del tipo de datos para el alineamiento (**Datatype for alignment**), seleccionamos **DNA** (3), seguidamente se muestra la ventana del explorador del alineamiento (**Alignment Explorer**). En esta ventana pegamos las 9 secuencias en formato fasta (4), las que previamente se seleccionaron y copiaron del bloc de notas. Luego, en el menú del explorador ingresamos a **Edit**, seleccionamos **Select All** (5), a continuación en el menú ingresamos a la opción alineamiento (**Alignment**), seleccionamos **Align by ClustalW** (6) y en la ventana de parámetros de Clustal W (**ClustalW Parameters**) hacemos click en **OK** (7) y a continuación las secuencias son alineadas, primero entre pares de secuencias y luego el alineamiento múltiple (8). Después de algunos segundos se muestra las secuencias alineadas en el explorador del alineamiento (**Alignment Explorer**). El archivo de las secuencias alineadas se guarda con un nombre apropiado y con la extensión **.mas**.

**Explorer**). En esta ventana pegamos las 9 secuencias en formato fasta (4), las que previamente se seleccionaron y copiaron del bloc de notas. Luego, en el menú del explorador ingresamos a **Edit**, seleccionamos **Select All** (5), a continuación en el menú ingresamos a la opción alineamiento (**Alignment**), seleccionamos **Align by ClustalW** (6) y en la ventana de parámetros de Clustal W (**ClustalW Parameters**) hacemos click en **OK** (7) y a continuación las secuencias son alineadas, primero entre pares de secuencias y luego el alineamiento múltiple (8). Después de algunos segundos se muestra las secuencias alineadas en el explorador del alineamiento (**Alignment Explorer**). El archivo de las secuencias alineadas se guarda con un nombre apropiado y con la extensión **.mas**.

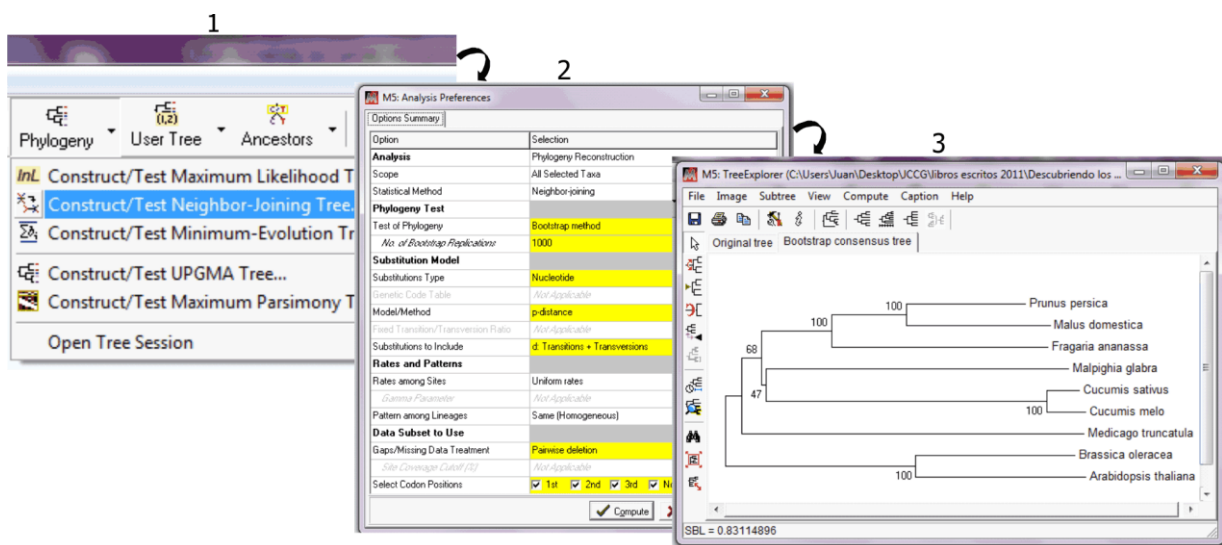


**Figura 3.** Pasos para realizar el alineamiento múltiple de las secuencias nucleotídicas con el programa Mega 5.05

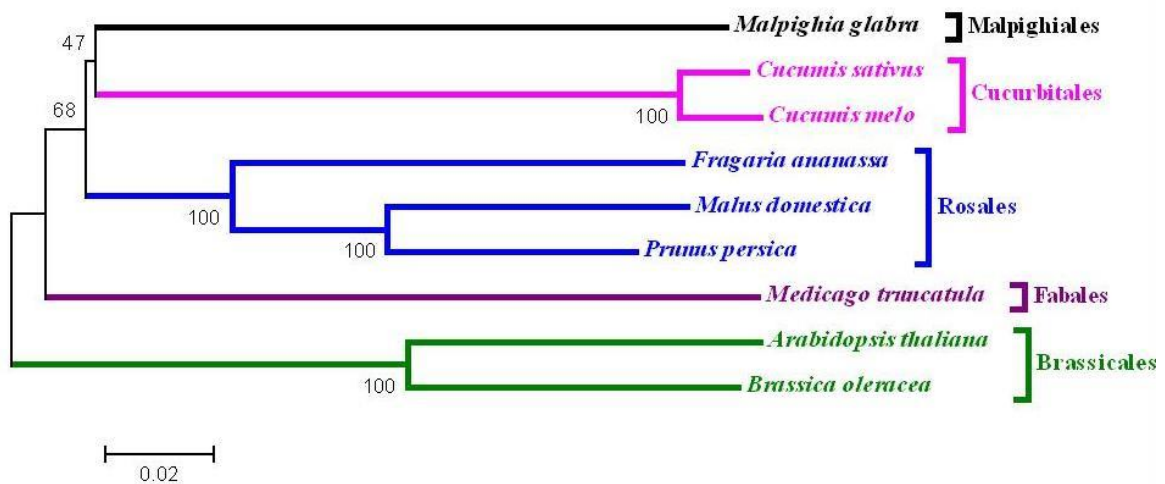
**Análisis filogenético**

Después de alinear las secuencias se construye el árbol filogenético (Figura 4). Para ello, en el menú ingresamos a **Phylogeny** y seleccionamos **Construct/Test Neighbor-Joining Tree** (1), inmediatamente en la ventana de **Analysis Preferences** (2), ponemos en **Phylogeny Test: Bootstrap method**, en **No. of Bootstrap replications: 1000**, en **Substitutions Type: Nucleotide**, en **Model/Method: p-distance**, en **Substitutions to include: d: Transitions + Transversions**, entre otros que se muestran en las imágenes, luego realizamos el

cómputo. Finalmente, el árbol filogenético consenso con los respectivos valores del análisis bootstrap se muestran en la ventana del **TreeExplorer** (3), en el que podemos editar el filograma y obtenemos la Figura 5. En el filograma observamos dos cladogramas, uno que incluye a los miembros de Fabidae (Malpighiales, Cucurbitales, Rosales y Fabales) y el otro al de los Malvidae (Brassicales). Es importante tener en cuenta este aspecto, porque el siguiente paso es diseñar varios pares de cebadores degenerados, en base al alineamiento obtenido de cada grupo (Fabidae y malvidae).



**Figura 4.** Pasos para realizar el análisis filogenético de las secuencias nucleotídicas alineadas con el programa Mega 5-05



**Figura 5.** Relaciones filogenéticas en base a las secuencias del ARNm de L-galactono-1,4-lactona deshidrogenasa, estimado con el test de Neighbor-Joining y un bootstrap de 1000 réplicas con el programa Mega 5.05

### Diseño de cebadores degenerados

Para diseñar los cebadores degenerados ingresamos al sitio web del programa SC Primer

(<http://scprimer.cpmc.columbia.edu/SCPrimerApp.cgi>). Esta aplicación web permite el diseño de cebadores degenerados a partir de alineamientos múltiples en formato fasta o clustalw. El algoritmo intenta minimizar el número de cebadores necesarios para amplificar todas las secuencias del alineamiento. La aplicación web requiere un registro simple por parte del usuario, después del cual podrá cargar el archivo que contiene las secuencias alineadas, monitorear el progreso del diseño y almacenar los resultados con todos los pares de cebadores degenerados.

### Colecta y transporte del material botánico

Los frutos verdes y maduros de *M. dubia* fueron obtenidos de la Colección Nacional de Germoplasma, Campo Experimental El Dorado, Estación Experimental San Roque-Loreto, Instituto Nacional de Innovación Agraria (INIA), ubicado en el Km 25½ de la carretera Iquitos-Nauta (03°57'17" LS, 73°24'55" LO). Las muestras botánicas fueron transportadas en refrigeración a la UEB-CIRNA-UNAP, donde fueron almacenadas a -20°C, de las que posteriormente se purificó el ARN total.

### Purificación de ARN total y síntesis de ADN complementario

El ARN total se purificó a partir de la pulpa y cáscara de frutos verdes y maduros según Castro *et al.* (2013) y la síntesis del ADN complementario (ADNc) se realizó a partir de 1,5 µg del ARN total empleando el Kit GeneAmp®-RNA PCR Core (Applied Biosystems). En un microtubo de 0,2 mL se añadió los siguientes componentes: PCR Buffer II 1X, MgCl<sub>2</sub> 5 mM, dNTPs 1 mM de cada uno, RNase inhibitor 1 U/µL, Oligo d(T)16 2,5 µM, MuLV Reverse Transcriptase 2,5 U/µL. La reacción se efectuó en un termociclador Eppendorf a 25°C por 10 minutos, 42°C por 60 minutos, 99°C por 5 minutos y 5°C por 5 minutos. Los productos obtenidos se almacenaron a -20°C.

### Amplificación del gen

Para la reacción en cadena de la polimerasa (PCR) se usó 10X PCR Buffer 1X, MgCl<sub>2</sub> 3 mM, dNTPs Mix 0,8 mM, cebadores 0,5 µM de cada uno, Amplitaq Gold 0,025 U/µL, ADNc (dilución 1:5) 2 µL y agua csp 20 µL. Las condiciones de amplificación en un termociclador Eppendorf ep gradient fueron: un ciclo de 95°C por 5 minutos, 40 ciclos de 95°C por 30 segundos, 60°C por 45 segundos, 72°C por 1 minuto y una extensión final de 72°C por 5 minutos. Los amplicones obtenidos fueron resueltos por electroforesis (con buffer TBE 0,5X a 85 voltios por 60 minutos) en gel de agarosa al 2% con 75 µg/mL de bromuro de etidio y los geles fueron fotografiados con el sistema Bio Doc Analyze.

### Clonación y secuenciamiento del gen

La clonación del gen se realizó de acuerdo al Kit de Clonación TOPO®TA (Invitrogen) y las colonias de *Escherichia coli* con inserto fueron seleccionadas por PCR y cultivaron en caldo LB por 24 horas. Luego, los plásmidos recombinantes fueron purificados según Sambrook *et al.* (1989). Quinientos nanogramos de plásmidos puros fueron empleados para secuenciar el inserto con el kit Big Dye Terminator V3,1 (Applied Biosystems) en un termociclador Eppendorf ep gradient. Los productos del secuenciamiento fueron purificados por precipitación con etanol/EDTA, desnaturalizados y resueltos en un analizador genético 3130XL. Las secuencias obtenidas fueron analizadas con el programa Sequencing Analysis v5,2,0 (Applied Biosystems). La identidad de la secuencia se verificó con análisis BLAST.

### RESULTADOS

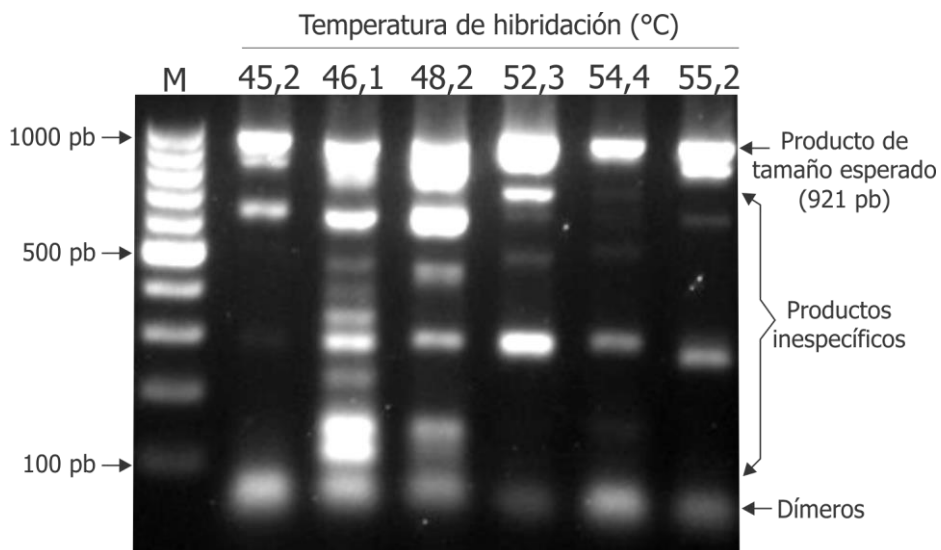
De los más de 50 pares de cebadores degenerados obtenidos con el programa SC Primer, sólo fueron seleccionados cinco pares para amplificar segmentos > 350 pb del gen de interés (Tabla 2). Los cinco pares de cebadores fueron sintetizados y empleados para amplificar regiones del gen (ADNc) bajo diferentes condiciones de PCR (temperatura de hibridación, [MgCl<sub>2</sub>], [cebadores], etc). De los cuales, el segundo par nos permitió amplificar un producto del tamaño esperado (921 pb). Tal como se muestra en la Figura 6, el producto fue amplificado en las seis temperaturas de hibridación evaluadas (45,2 a

55,2°C). Este amplicón fue purificado a partir del gel, ligado a un plásmido y clonado en *Escherichia coli*. Del cultivo bacteriano, se

purificó el plásmido recombinante y secuenció en ambas direcciones el segmento del gen clonado.

**Tabla 2.** Secuencia de algunos cebadores degenerados seleccionados para aislar el gen *L-GalLDH* de *M. dubia*.

Cebador Forward Tm (°C)	Cebador Reverse Tm (°C)	Tamaño esperado del producto (pb)
5'-TVGGNTGGRSTGATGAAATY-3' 59	5'-AYCTCRATYTTNGCCARTGYT-3' 62	391
5'-GGTCACGCCTGCCAAGGGTA-3' 68	5'-GGAACCTCAATCTTAGCCCAATGT-3' 64	921
5'-GGTGGACTTGGAGTTGTTGC-3' 61	5'-CGTGCTTTGTTGTATGCATCC-3' 61	938
5'-TGGCATTGTTTCAGGTGGGAG-3' 64	5'-CTGTAGTGGAAGAACTCCTCAGT-3' 57	948
5'-CTGAGGCTGAGTTCTGGA-3' 55	5'-CCTTGTCTTCGGAACCT-3' 58	444



**Figura 6.** Electroforesis en gel de agarosa al 2% de amplicones obtenidos con el segundo par de cebadores probados a diferentes temperaturas de hibridación durante el PCR. M: marcador de peso molecular de 100 a 1000 pb



Con la secuencia del segmento del gen de 921 pb, obtenida de los frutos de *M. dubia*, se buscó con un análisis BLAST las secuencias homólogas en los bancos de genes. Los resultados de este análisis nos muestran que el segmento del gen codifica la enzima L-galactono-1,4-lactona deshidrogenasa. Esto se deduce porque presenta una gran homología

con secuencias de este gen reportadas para varias especies de plantas como *Populus trichocarpa*, *Citrus unshiu*, *Vitis vinifera*, *Ricinus communis*, *Prunus persica*, *Rosa roxburghii*, *Malus x domestica* con las que presenta identidades máximas de 78 a 81% y E-values = 0,0 (Figura 7)

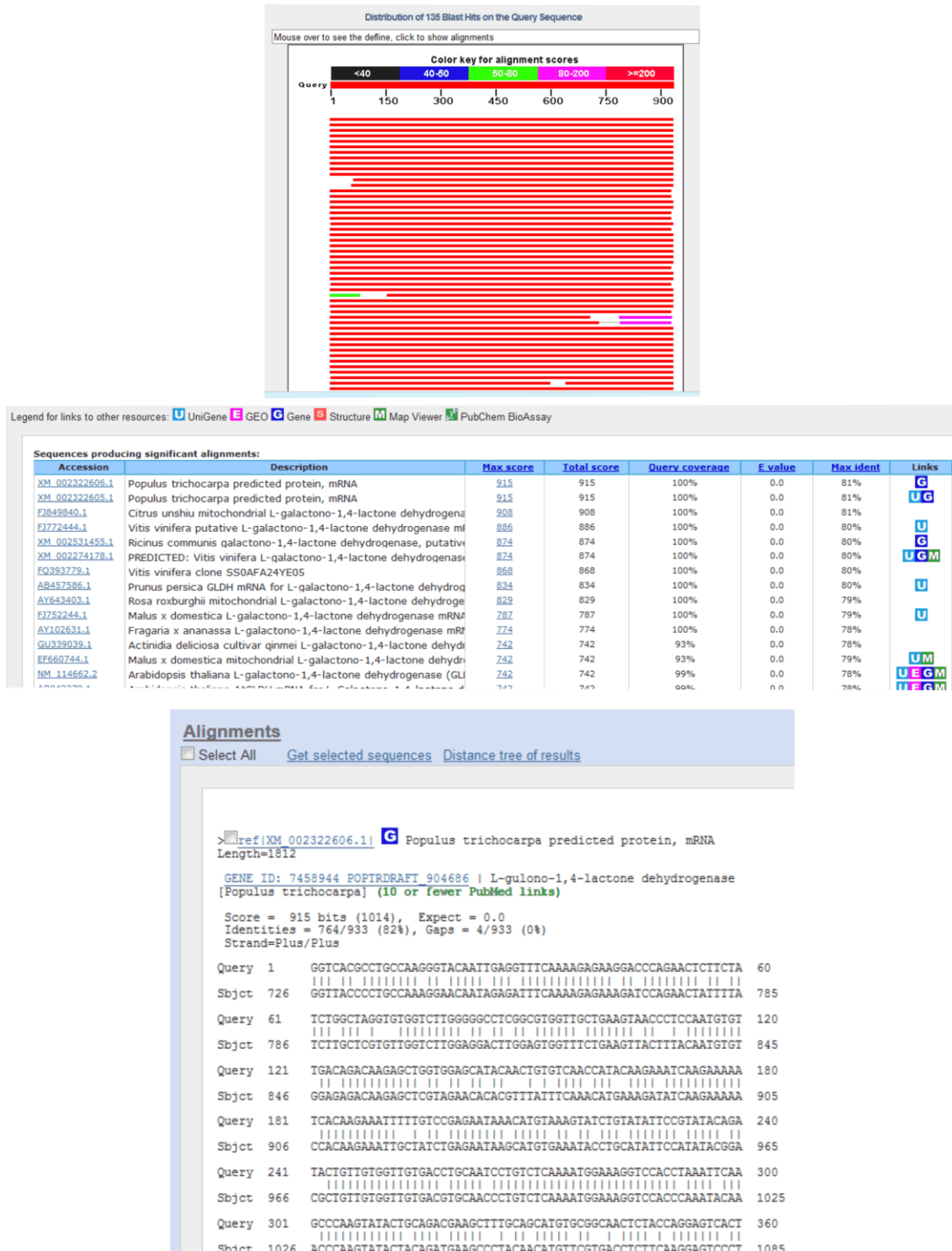
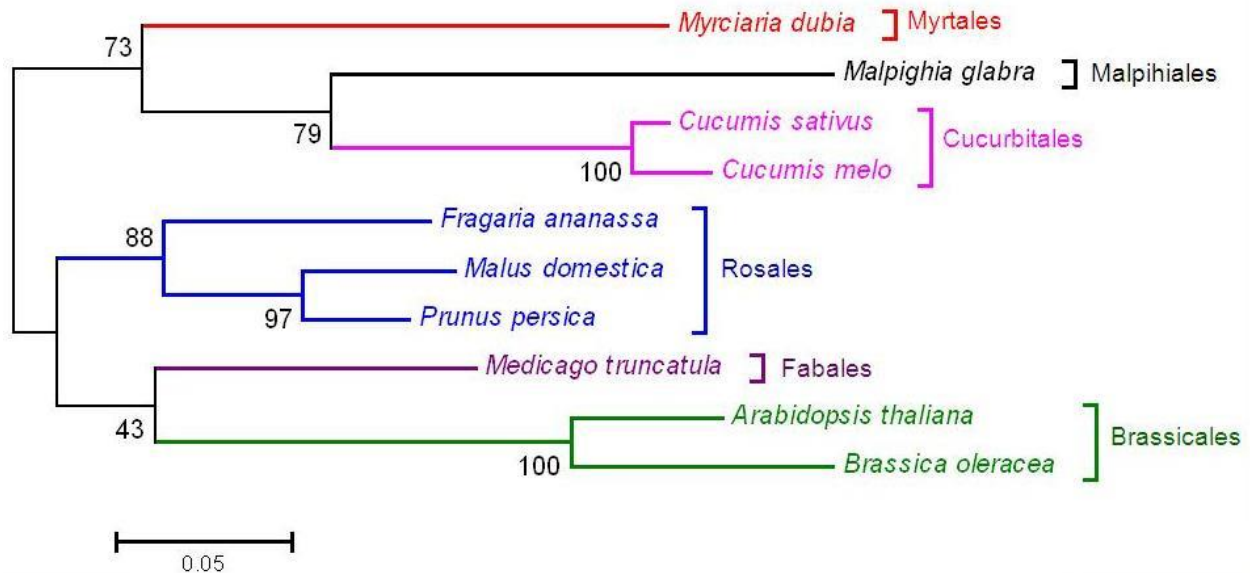


Figura 7. Resultados del análisis BLAST de la secuencia del gen aislado de frutos de *M. dubia* indican que se trata del gen que codifica la enzima L-galactono-1,4-lactona deshidrogenasa

Asimismo, el análisis filogenético de las secuencias del gen *L-GalLDH* de *M. dubia* y las secuencias de las otras especies empleadas para diseñar los cebadores degenerados, muestra que el gen de *M. dubia* tiene mayor

similitud genética con los genes homólogos de los Malpighiales y Cucurbitales. Sin embargo, muestra mayor distancia genética con los genes homólogos pertenecientes a los Fabales, Rosales y Brassicales (Figura 8).



**Figura 8.** Relaciones filogenéticas de las secuencias del ARNm de L-galactono-1,4-lactona deshidrogenasa de *M. dubia* con otras especies de Rósidos, estimado con el test de Máxima Probabilidad y un bootstrap de 1000 réplicas con el programa Mega 5.05.

## DISCUSIÓN

Con las estrategias bioinformáticas y moleculares descritas se aisló con éxito uno de los genes que codifica la enzima L-galactono-1,4-lactona deshidrogenasa. Esta enzima cataliza la última reacción (conversión de L-galactono-1,4-lactona en ácido L-ascórbico) de la vía metabólica para la biosíntesis de vitamina C (vía Smirnoff-Wheeler) en *M. dubia*. Con las estrategias mencionadas, también hemos aislado segmentos de otros genes de esta vía biosintética (Egoávil y Torres, 2012), considerada la vía más importante para la biosíntesis de vitamina C en plantas (Wheeler *et al.*, 1998). En base a la secuencia de estos segmentos génicos de *M. dubia* será factible obtener los genes completos empleando la técnica de amplificación rápida de extremos de ADNc (RACE 5'/3') tal como han sido reportados previamente (Badejo *et al.*, 2007; Dallmeier y Neyts, 2013). Otras investigaciones reportan varios programas bioinformáticos, que emplean algoritmos

diferentes al SC Primer, para diseñar cebadores degenerados en base a secuencias alineadas de nucleótidos (Jarman, 2004; Najafabadi *et al.*, 2008a, 2008b). Alternativamente, existen programas que nos permiten diseñar cebadores degenerados a partir de secuencias altamente conservadas de proteínas, es decir, estos programas identifican en los alineamientos protéicos múltiples las secuencias aminoacídicas consenso y diseñan los oligonucleótidos que tengan alta probabilidad de hibridar con las respectivas secuencias nucleotídicas de estas regiones (Boyce *et al.*, 2009; Rose *et al.*, 1998, 2003).

Estrategias similares podrían emplearse para aislar, a partir de especies de la biodiversidad amazónica, otros genes de interés biotecnológico para las industrias químicas, alimentarias y farmacéuticas. Pues nuestro país es uno de los centros de megabiodiversidad del planeta, pero aún desconocemos el número total de genes de

importancia biotecnológica, los que nos permitirían el tan ansiado desarrollo sostenible, pero en base a nuestras minas de genes. Muchos de estos genes controlan las vías metabólicas biosintéticas de principios activos, que actualmente se usan como fármacos o son precursores para la síntesis química de nuevos fármacos. En la mayoría de los casos, estos principios activos son purificados a partir de diversas partes de plantas (hojas, raíces, cortezas, etc) colectadas de los bosques amazónicos, causando impactos negativos en los ecosistemas y la biodiversidad. Sin embargo, aislando los genes que controlan la biosíntesis de estos principios activos, podríamos desarrollar sistemas de producción biotecnológica *in vivo* e *in vitro* para obtener cantidades suficientes que satisfagan la creciente demanda de estos productos. Estos sistemas productivos no tendrían impactos negativos en los ecosistemas y daría un valor agregado significativo a nuestros productos.

Pero actualmente, también disponemos de otras alternativas para realizar la bioprospección de genes de interés biotecnológico de la biodiversidad amazónica. Una de las alternativas más interesantes es el secuenciamiento del transcriptoma (Sweetman *et al.*, 2012). Millones de secuencias génicas obtenidas son ensambladas *de novo* y anotados por comparación con secuencias homólogas en bases de datos (Haas y Zody, 2009; Wang, *et al.*, 2009). Con estas estrategias podemos disponer de secuencias parciales y completas de todos los genes que se expresan en las células, tejidos u órganos de cualquier especie. Aunque en nuestro país no disponemos de estas tecnologías de secuenciamiento masivo, es posible ejecutar investigaciones de este nivel, porque varias instituciones extranjeras brindan servicios de secuenciamiento con estas tecnologías, a muy bajo costo. Debido a estas facilidades, actualmente estamos secuenciando el transcriptoma de frutos verdes y maduros de *M. dubia*. Pero, para contar con un catálogo de todos los genes y establecer todas las vías metabólicas de la especie, necesitamos secuenciar los transcriptomas en diferentes etapas de crecimiento y desarrollo de la especie tales como: en germinación, floración,

fructificación, callogénesis, embriogénesis, bajo diversas condiciones bióticas (ataque por fitopatógenos) y abióticas (sequías, inundaciones, etc) adversas. También, podríamos identificar todos los genes expresados diferencialmente, comparando entre plantas con características fenotípicas variables, como aquellas que producen frutos con alto y bajo contenido de vitamina C y antocianinas, aquellas que difieren en la cantidad y tamaño de frutos y las que son precoces o se retardan en la primera fructificación.

## CONCLUSIONES

Con las estrategias bioinformáticas y moleculares descritas se aisló el segmento de un gen que codifica la enzima L-galactono-1,4-lactona deshidrogenasa de la vía biosintética de vitamina C en *M. dubia*, especie de la que no se disponía de secuencias génicas. Con estas mismas aproximaciones será posible aislar genes de interés biotecnológico de esta y otras especies amazónicas de las que aún no disponemos de bases de datos de secuencias nucleotídicas de libre acceso.

## AGRADECIMIENTOS

A la Universidad Nacional de la Amazonía Peruana (UNAP) y al Consejo Nacional de Ciencia y Tecnología (CONCYTEC) por el financiamiento de la investigación. Asimismo, al Instituto Nacional de Innovación Agraria por las facilidades de acceso a la colección de germoplasma de *Myrciaria dubia*.

## REFERENCIAS

- Badejo AA, Jeong ST, Goto-Yamamoto N and Esaka M. 2007. Cloning and expression of GDP-D-mannose pyrophosphorylase gene and ascorbic acid content of acerola (*Malpighia glabra* L.) fruit at ripening stages. *Plant Physiol. Biochem.* 45, 665–672.
- Boyce R, Chilana P and Rose TM. 2009. iCODEHOP: a new interactive program for designing Consensus-DEgenerate Hybrid Oligonucleotide Primers from multiply aligned protein sequences. *Nucleic Acids Res.* 37, W222–228.

- Bradfield RB and Roca A. 1964. Camu-camu - a fruit high in ascorbic acid. *J. Am. Diet. Assoc.* *44*, 28–30.
- Dallmeier K and Neyts J. 2013. Simple and inexpensive three-step rapid amplification of cDNA 5' ends using 5' phosphorylated primers. *Anal. Biochem.* *434*, 1–3.
- Egoávil A and Torres J. 2012. Expresión de genes que codifican enzimas: GDP-L-Galactosa Fosforilasa, L-Galactosa Deshidrogenasa y L-Galactono-1,4-Lactona Deshidrogenasa de la ruta biosintética de Vitamina C en *Myrciaria dubia* (kunth) McVaugh "camu camu." Titulación de Biólogo. Universidad Nacional de la Amazonía Peruana.
- Castro JC, Egoavil A, Torres J, Ramirez R, Cobos M and Imán SA. 2013. Isolation of high-quality total RNA from leaves of *Myrciaria dubia* "CAMU CAMU." *Prep. Biochem. Biotechnol.* *43*, 527–538.
- Haas BJ and Zody M. 2009. Advancing RNA-Seq analysis. *Nat. Biotechnol.* *28*, 421–423.
- Imán S, Bravo L, Sotero V and Oliva C. 2011. Contenido de vitamina C en frutos de camu camu *Myrciaria dubia* (H.B.K) Mc Vaugh, en cuatro estados de maduración, procedentes de la Colección de Germoplasma del INIA Loreto, Perú. *Sci. Agropecu.* *2*, 123–130.
- Jarman SN. 2004. Amplicon: software for designing PCR primers on aligned DNA sequences. *Bioinforma. Oxf. Engl.* *20*, 1644–1645.
- Maddison DR, Schulz K-S and Maddison WP. 2007. The Tree of Life Web Project. *Zootaxa* *1668*, 19–40.
- Najafabadi HS, Saberi A, Torabi N and Chamankhah M. 2008a. MAD-DPD: designing highly degenerate primers with maximum amplification specificity. *BioTechniques* *44*, 519–520, 522, 524–526.
- Najafabadi HS, Torabi N and Chamankhah M. 2008b. Designing multiple degenerate primers via consecutive pairwise alignments. *BMC Bioinformatics* *9*, 55.
- Ozsolak F and Milos PM. 2011. RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* *12*, 87–98.
- Rose TM, Schultz ER, Henikoff JG, Pietrokovski S, McCallum CM and Henikoff S. 1998. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res.* *26*, 1628–1635.
- Rose TM, Henikoff JG and Henikoff S. 2003. CODEHOP (CONsensus-DEgenerate Hybrid Oligonucleotide Primer) PCR primer design. *Nucleic Acids Res.* *31*, 3763–3766.
- Sambrook J, Fritsch EF and Maniatis T. 1989. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press).
- Soltis ED and Soltis PS. 2000. Contributions of plant molecular systematics to studies of molecular evolution. *Plant Mol. Biol.* *42*, 45–75.
- Sweetman C, Wong DC, Ford CM and Drew DP. 2012. Transcriptome analysis at four developmental stages of grape berry (*Vitis vinifera* cv. Shiraz) provides insights into regulated and coordinated gene expression. *BMC Genomics* *13*, 691.
- Tamura K, Stecher G, Peterson D, Filipski A and Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.*
- Wang Z, Gerstein M and Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* *10*, 57–63.
- Wheeler GL, Jones MA and Smirnoff N. 1998. The biosynthetic pathway of vitamin C in higher plants. *Nature* *393*, 365–369.