

# Generación de la plataforma bioinformática para palma de aceite de Cenipalma

## Generating Cenipalma's Bioinformatics Platform for Oil Palm

### AUTORES

**Pedro Jesús Rocha Salavarieta. Ph.D.**

Investigador titular, Director de la División de Biotecnología, Cenipalma, Bogotá.  
pedro.rocha@cenipalma.org

**Gabriel Mutis Namur**

Estudiante Tesista, Universidad de los Andes, Bogotá.

**Silvia Restrepo Restrepo, Ph.D.**

Profesora Asociada, Facultad de Ciencias, Universidad de los Andes, Bogotá.

### Palabras CLAVE

*Elaeis guineensis*, *Elaeis oleifera*,  
Genómica.

*Elaeis guineensis*, *Elaeis oleifera*,  
Genomics.

Recibido: 28 mayo 2008  
Aceptado: 11 junio 2008

### Resumen

Desde enero de 2008, la División de Biotecnología de Cenipalma inició el desarrollo de nuevas líneas de investigación. Una de ellas es la bioinformática, la cual se constituye en un pilar fundamental de apoyo a la investigación de las diferentes Divisiones del Centro. En Colombia, el desarrollo de la bioinformática es incipiente. Sin embargo, el Laboratorio de Micología y Fitopatología de la Universidad de los Andes (LAMFU) es una de las entidades líderes en esta área. Por esta razón y en atención a la estrategia de interacción cercana entre Cenipalma y la Universidad de los Andes, se estableció un proyecto en colaboración con el objeto de generar una plataforma bioinformática para palma de aceite. Como consecuencia de las actividades realizadas en el marco de tal proyecto, en el presente artículo se describe la primera versión de la plataforma de bioinformática del gremio palmicultor colombiano, un producto tecnológico con aplicación para los sectores académico, agrícola e industrial.

### Summary

Since January 2008, Cenipalma's Biotechnology Division started to develop new research lines. One of them is the bioinformatics, which is a fundamental stone to support research of the different Divisions of Cenipalma. In Colombia, the development of bioinformatics is incipient. However, the Laboratory of Mycology and Phytopatology of the Universidad de los Andes (LAMFU) is one of the leader entities in this field. For this reason and attending to the strategy of close interaction between Cenipalma and Universidad de los Andes, a cooperative project was established with the aim of generating a bioinforma-



tics platform for oil palm. So, as consequence of the activities carried out in the frame of this project, here we describe the first version of the bioinformatics platform for the Colombian oil palm sector. Such platform is a technological product with potential application for academic, agro and industrial sectors.

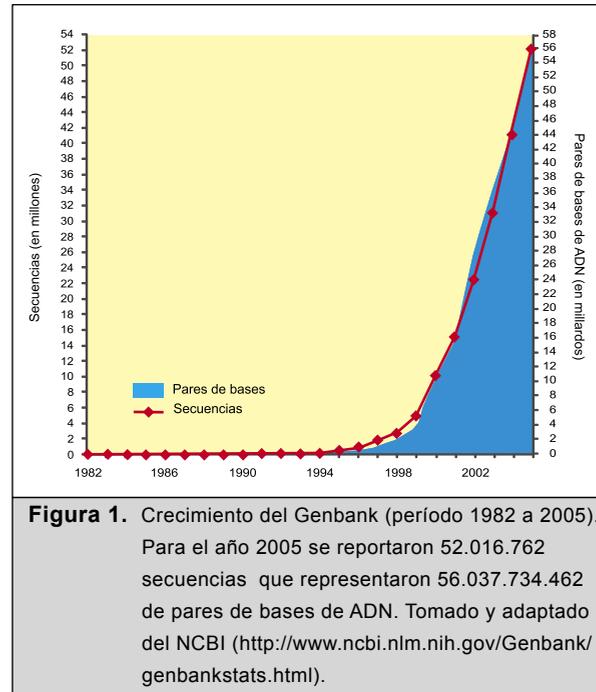


## Introducción

El gremio palmicultor colombiano ha considerado la biotecnología como un área de enorme importancia para su desarrollo tecnológico. En efecto, en la nueva estructura de Cenipalma se creó la División de Biotecnología (Fedepalma, 2007), en la cual se fortaleció el área de caracterización molecular y se establecieron nuevas líneas de trabajo, incluidas las de bioinformática, genómica y biotecnología industrial, entre otras (Rocha, 2008).

En el ejercicio prospectivo de la biotecnología en Colombia, realizado por Colciencias (Peña *et al.*, 2008), se encontró que la bioinformática es uno de los campos de la biología más relevante y prioritario para desarrollar en los próximos años. Ciertamente, el ejercicio prospectivo llega a esta conclusión, debido a que es necesario generar herramientas para el manejo sistemático de enormes cantidades de información (obtenidas por medio de disciplinas como la genómica y la proteómica, entre otras).

Existe una variada, enorme y creciente cantidad de información disponible en las bases de datos de secuencias de ADN y de proteínas (Figura 1; NCBI, 2008). Desde 1995, los genomas de más de 190 organismos han sido secuenciados (<http://www.genomenetwork.org>), incluidos aquel de *E. oleifera*, del híbrido OxG y de *E. guineensis* tipos Dura, Ténera y Pisífera, junto con un clon anormal (Budiman *et al.*, 2005). Así, para el año 2005, más de 400.000 secuencias de ADN nuclear habían sido generadas. Esta avalancha de información es el resultado de la implementación de nuevas técnicas de secuenciación de ADN y ha estado ligada al progreso de tecnologías basadas en computadores, ya que desempeñan un papel fundamental en el procesamiento de grandes cantidades de información (Rocha, 2005). Así, existen



numerosas plataformas bioinformáticas y bases de datos de utilidad para estudios en plantas. Con respecto al uso de la informática en la palma de aceite, se han reportado algunas iniciativas. Por ejemplo, PalmGenes del Mpob (<http://palmoilis.mpob.gov.my/palmgenes.html>), una base de datos que contiene 5610 secuencias de ADN, y Mimos (*Malaysian grid computing*), un sistema que busca generar aplicaciones de utilidad para atender desde los frentes propios del cultivo, hasta lo relacionado con investigación biotecnológica del mismo (Weng *et al.*, 2007).

Aunque los detalles técnicos de la bioinformática son relativamente complejos para quien no es especialista, ciertamente los resultados pueden tener un enorme impacto para el experto en biología (Rocha, 2007a y b). Por esta razón, y en atención a la tendencia de fortalecer esta área y generar herramientas amigables y



de manejo sencillo para dejar de lado la programación y enfocarse más en la explicación de fenómenos biológicos y la utilización racional y eficiente de enormes cantidades de información, con el presente trabajo se hace entrega a los investigadores de los distintos Programas y Divisiones de Cenipalma, a la comunidad académica colombiana y al sector palmicultor, de una herramienta con la que se permite el uso sistemático de gran cantidad de información, generada en diferentes laboratorios de distintos países y hasta el momento disponible de manera fragmentada.

## Bioinformática

En sentido amplio, la bioinformática se define como una ciencia dedicada al manejo y el análisis de información biológica (Peña *et al.*, 2008). En sentido estricto, el término bioinformática hace referencia a la unión de la manipulación computacional matemática y al análisis molecular de secuencias biológicas, es decir, a la aplicación de la ciencia computacional en la biología molecular (Attwood *et al.*, 2002).

La bioinformática considera aspectos tan importantes como (i) recopilación, organización y almacenamiento de información biológica (como secuencias de ADN, secuencias de aminoácidos, estructuras proteicas, etc.) en las bases de datos; (ii) desarrollo de herramientas eficientes de búsqueda y análisis de información en las bases de datos; (iii) generación y uso de aplicaciones computacionales (programas) para la realización de tareas de extracción, filtro y comparación de información que permitan interpretar los datos y otorgarles una explicación biológica (Attwood *et al.*, 2002; Luscombe *et al.*, 2001). Adicionalmente, el empleo rutinario de la bioinformática está velando por la calidad de la información biológica generada mediante el establecimiento de procedimientos rigurosos de curaduría.

### Manejo de bases de datos

Las bases de datos constituyen la esencia de la bioinformática. Este recurso computacional permite almacenar y revisar secuencias de ADN, ARN, proteínas e información de anotaciones asociadas, las cuales pueden variar en cantidad, calidad y variedad de su contenido (Attwood *et al.*, 2002). Aparte de secuencias, las principales bases de datos biológicas poseen

información taxonómica y de estructuras proteicas. Las bases de datos literarias son también incluidas, ya que generan un vínculo entre los datos moleculares disponibles y su respectiva interpretación. El acceso a la mayoría de este contenido se realiza mediante la red global, Internet. De esta forma, se permite no solo una interacción entre diferentes científicos y centros de investigación del mundo, sino acceder, en cualquier momento, a la información requerida (Arndt, 2007).

En la actualidad, bases de datos representativas incluyen Genbank, que pertenece al *National Center for Biotechnology Information* (NCBI, <http://www.ncbi.nlm.nih.gov>), *The European Molecular Biology Laboratory* (EMBL, <http://www.embl.de>), *DNA Databank of Japan* (DDBJ, <http://www.ddbj.nig.ac.jp/searches-e.html>) y SynaBASE (<http://www.synamatix.com>), empleada para la caracterización rápida y eficiente de genomas y transcriptomas de plantas (Anwar, 2007). Estas bases son de uso público al igual que muchas otras, sin embargo, existen entidades privadas que venden la suscripción para consultar sus páginas como lo hace la empresa Celera (<http://www.celera.com>).

Estrechamente ligado al manejo de bases de datos, existe el término “minería de datos”, el cual hace referencia a la búsqueda y descubrimiento de información valiosa, no obvia, ubicada en una gran colección de datos (Olafsson *et al.*, 2006). Los métodos utilizados son usualmente automatizados, pero se puede realizar manualmente el proceso, para lograr categorizar, organizar, agrupar y asociar un nuevo conjunto de información (Mutis, 2008).

### Análisis de secuencias

Uno de los retos recurrentes en la bioinformática es la comparación y análisis de secuencias nucleotídicas y de aminoácidos. Mediante la comparación lineal de secuencias (técnicamente conocido como proceso de alineamiento), en la que se introducen inserciones para lograr que posiciones equivalentes en secuencias adyacentes que se sitúen en el registro (Attwood *et al.*, 2002), se pueden establecer diferencias y similitudes. La enorme cantidad de alineamientos hace indispensable el uso de computadores y de algoritmos. Por ejemplo, algoritmos como los de Needleman-Wunsch y Smith-Waterman, los cuales emplean la programación dinámica donde se construye una solución a un problema, resolviendo una serie de subproblemas más

pequeños; BLAST y FASTA, que se basan en métodos de búsqueda local para encontrar emparejamientos cortos e idénticos (Attwood *et al.*, 2002), logran que se puedan realizar diferentes tipos de alineamientos que resultan en aplicaciones tales como la predicción de función y de estructura de familias proteicas y la elaboración de cebadores para amplificación de regiones de interés o predicción de genes, entre otras. Una herramienta de enorme importancia es el alineamiento múltiple, el cual ayuda a la predicción de motivos y dominios mediante el uso de programas como T-Coffee (<http://www.ebi.ac.uk/t-coffee>) y Clustal-W (<http://www.ebi.ac.uk/clustalw>).

Por otro lado, la identificación de genes mediante procesos computacionales y comparaciones de bases de datos, logra establecer características únicas de los genes y elementos involucrados en su expresión (Luscombe *et al.*, 2001).

### Otras herramientas bioinformáticas

Las herramientas anteriores son las principales en el campo de la bioinformática, sin embargo no son las únicas. La creación de árboles filogenéticos utiliza en gran medida elementos involucrados con la bioinformática, tales como un alto componente matemático y de procesamiento computacional. Por otra parte, la complejidad que exhiben las proteínas, no sólo se debe a su composición. La estructura es el aspecto fundamental para conocer las diferentes funciones que éstas pueden desempeñar. Al predecir estas estructuras se pueden dar a conocer las características únicas de cada familia proteica y, con base en esto, se pueden entender diferentes procesos moleculares vitales en los organismos vivos, como son la replicación del ADN, las rutas metabólicas de la fotosíntesis y la actividad enzimática celular, por sólo nombrar algunos. Diferentes programas, la mayoría disponibles en la red global, ayudan a realizar estos modelos de proteínas, por ejemplo, Geneious (<http://www.geneious.com>), el cual es uno de los más completos y sencillos de utilizar en la actualidad. Este tipo de programas permite acceder a información tanto general (compuestos encontrados en la estructura primaria), como detallada (distancia entre enlaces en una posición específica) (Arndt, 2007).

Finalmente, existen distintas herramientas que ayudan a la investigación de los diferentes sistemas biológicos.

Sin embargo, estas herramientas, sin estar acompañadas por un fuerte componente computacional, no son de utilidad. Entre éstas encontramos los microarreglos y la secuenciación de ADN, entre otros.

## Materiales y métodos

### Material Computacional y Ubicación Experimental

Se utilizó un computador Hewlett Packard modelo Mobile AMD Sempron, memoria 1,12 GB RAM, 1,79 GHz, procesador 3000+, con software Windows XP, editor de texto Context, y paquete de asistente Microsoft Office 2007. Adicionalmente, de la página (<http://appserv.softonic.com/>) se descargó de manera gratuita el paquete de instalación Appserv, el cual contiene los programas: Apache, MYSQL y PHP, necesarios para la programación de la base de datos. El proyecto se llevó a cabo conjuntamente entre los laboratorios de biocómputo de la Universidad de los Andes (Bogotá) y las instalaciones de Cenipalma (Bogotá).

### Minería de Datos

Se hizo una búsqueda manual de los datos en la fuente de bioinformática primaria, Genbank (<http://www.ncbi.nlm.nih.gov/Genbank/index.html>), para las especies *E. guineensis* y *E. oleifera*. Se rastreó toda la información disponible de secuencias nucleotídicas y de aminoácidos (con fecha de cierre al 15 de mayo de 2008). La forma en que se realizó la búsqueda fue escribiendo el nombre de la especie en el buscador de la base de datos llamado *Entrez* (Figura 2), el cual arrojó las secuencias disponibles. Se entró a los enlaces que contenían las secuencias de interés y ahí cada una de estas fue copiada y llevada individualmente a un archivo de Excel, donde la información fue organizada en tipo de secuencia, secuencia, descripción y código de acceso en Genbank (GI).

### Diseño del Sistema de Información

El diseño seleccionado para el sistema de información fue sugerido por el biólogo Andrés Pinzón (profesor de bioinformática de la Universidad de los Andes). El objetivo de dicho diseño fue crear una base de datos interrelacionada, donde un núcleo central de información relaciona una búsqueda con un dato específico ubicado en un grupo de información. Por



lo tanto, lo obtenido al realizar la minería de datos fue organizado en archivos de información de Excel. El primer archivo tenía la información general de las dos especies, *E. guineensis* y *E. oleifera*, en cuanto a nombre común, nombre científico y un código numérico asignado. El segundo albergaba todos los datos, involucrando secuencias recolectadas en la minería de datos. Finalmente, el último archivo almacenó una lista de patógenos asociados y las sintomatologías que éstos producen en la palma de aceite. Esta última información fue obtenida del compendio de plantas ornamentales de Elliot *et al.* (2004).

### Programación mediante PHP y MYSQL

Una vez organizada toda la información que almacenaría la base de datos, se realizó la estructura de la misma y la del Sitio Web que la contendría, mediante programación computacional empleando: (i) PHP, un lenguaje de programación multiplataforma, el cual es embebido directamente en un código base HTML que es transferido e interpretado por un servidor, para luego mostrarlo en una interfaz gráfica que el cliente pueda interpretar fácilmente, y (ii) MYSQL, sistema de administración de bases de datos relacionales, ideal al crear bases de datos con acceso de páginas web dinámicas, sistemas de transacciones en línea o cualquier otra aplicación que involucre almacenamiento y consulta de datos (Cobo *et al.*, 2005; Pavon, 2007). La elección de estas herramientas de

programación se debió a la sencillez de utilizar, sin previo conocimiento de programación, y la versatilidad para crear bases de datos que luego puedan ser modificadas sin mayores inconvenientes.

El trabajo de programación necesario para crear tanto la base de datos como el Sitio Web se dividió en dos bloques denominados Bloque Informático y Bloque Estructural.

#### *Bloque Informático*

Se utilizó la aplicación de MYSQL, denominada PHPAdmin, para poder tomar la información recolectada, de tal forma que fuera utilizada por una página web dinámica, elaborada con lenguaje PHP. Inicialmente se creó en PHPAdmin un nuevo registro de base de datos bajo el nombre *palma*, en el cual se hicieron tres tablas de información (especies, secuencias y patógenos) que albergaron los datos de los archivos antes mencionados. Posteriormente, se tomaron estos archivos y se guardaron nuevamente bajo un nuevo formato (CSV), el cual es compatible con MYSQL. Luego, al entrar en PHPAdmin, se utilizó la aplicación “importar” para guardar los nuevos archivos en su tabla correspondiente.

#### *Bloque Estructural*

Mediante programación por lenguaje PHP y HTML y, por medio del editor de texto Context, se crearon *scripts* (archivos con código de programación) con

todos los aspectos de estructura de nuestro Sitio Web, incluidos encabezados e información escrita en el sitio, presentación gráfica, diseño de tablas, imágenes ilustrativas, hipervínculos dentro de la página, sistema de acceso de usuario y un código oculto de enlace con PHPAdmin. Para este caso, se crearon 14 *scripts* (uno para la página inicial, dos para el acceso por usuario y contraseña, uno para la información general del sitio y diez para la base de datos).

### Implementación de la base de datos y sitio web

Una vez que todos los *scripts* fueron generados y la información necesaria ubicada en MYSQL, se procedió a pasar todo a la subcarpeta *www*, ubicada en la carpeta *Appserv*, del programa con el mismo nombre. Esto permite que la información sea tomada automáticamente por Apache, que es el servidor web gratuito instalado en el ordenador y puesto como una página web dinámica. Sin embargo, en este punto la única forma de tener acceso al Sitio Web era desde el computador servidor (Hewlett Packard modelo Mobile AMD Sempron). Para poder entrar desde cualquier computador era necesario trasladar toda la información a una subcarpeta *www* de un servidor que tuviera un dominio IP establecido. Un dominio IP es un código numérico único que tiene cada servidor, que es traducido a nombres y alberga sitios electrónicos. Algunos ejemplos familiares son Yahoo, Google o Hotmail (Aznar, 2005). El ordenador seleccionado fue el utilizado por Andrés Pinzón en la Universidad de los Andes, con dominio IP <http://bioinf.uniandes.edu.co>.

## Resultados y discusión

Una plataforma bioinformática se caracteriza por ser un conjunto de programas y bases de datos integrados a un lenguaje computacional que expresa gráficamente la información (Qiang *et al.*, 2005). Bajo esta premisa se desarrolló un primer modelo de una plataforma bioinformática para *E. guineensis* y *E. oleifera*, teniendo un Sitio Web con una base de datos moleculares, acompañado de información y herramientas bioinformáticas de interés.

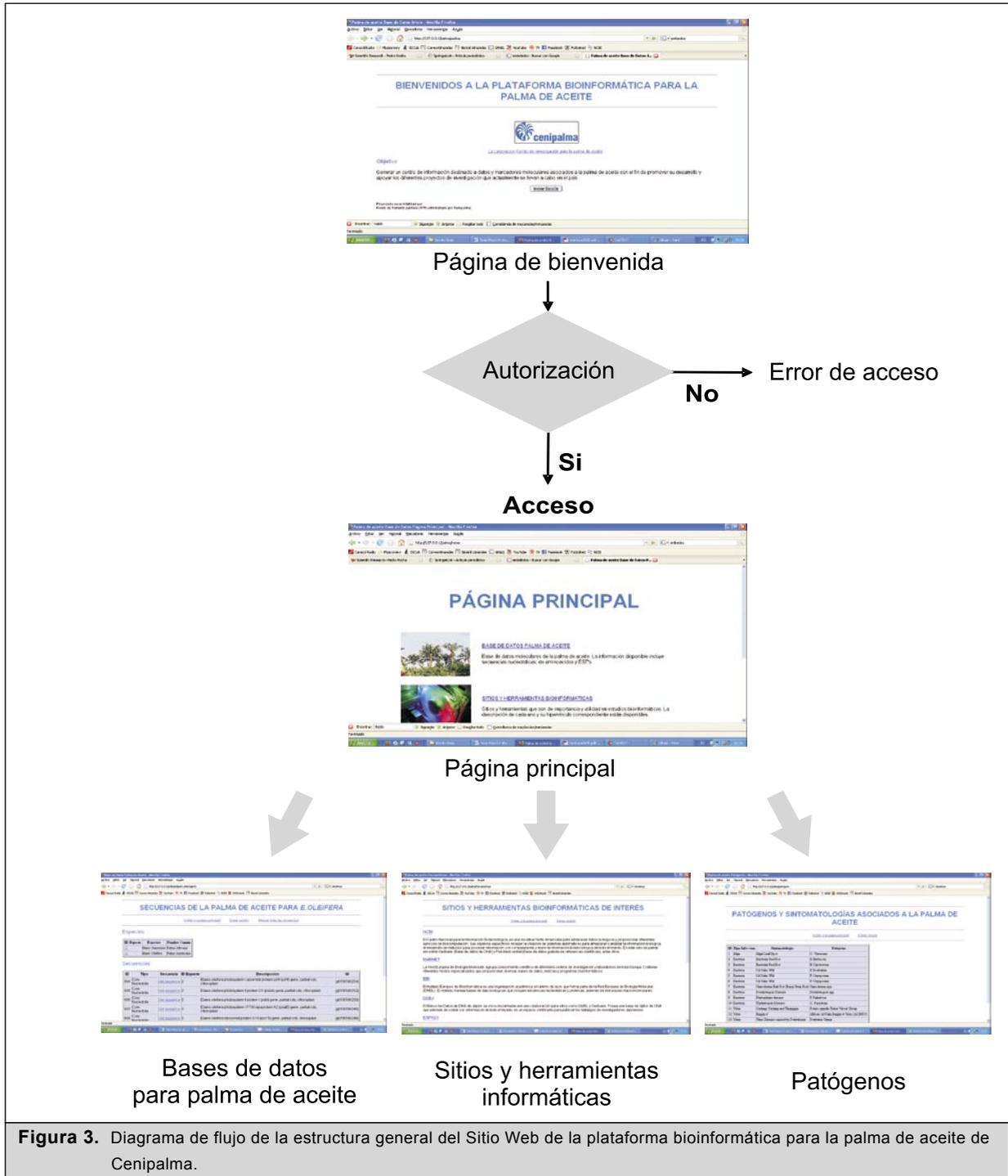
La plataforma creada puede ser consultada desde <http://bioinf.uniandes.edu.co> en el enlace *Palmasdb*. Al entrar en *Palmasdb* aparece una página de bienvenida

con el objetivo general y un botón para iniciar sesión, que al pulsarlo abre una página donde se requiere un nombre de usuario y una contraseña para autorizar el ingreso al sitio. Si se realiza incorrectamente esta acción, se indica que el usuario debe intentarlo de nuevo, de lo contrario se muestra un mensaje indicando que el acceso ha sido confirmado y queda habilitado un hipervínculo a la página principal. Al ingresar en esta página se muestran tres enlaces dinámicos (Figura 3): (i) la base de datos moleculares para la palma de aceite donde se encuentra toda la información de secuencias para las especies *E. guineensis* y *E. oleifera* y un buscador de secuencias por algoritmo (BLAST); (ii) la información de las principales herramientas y sitios esenciales en el campo de la bioinformática, y (iii) la lista de patógenos con sintomatologías asociadas reportadas en palma de aceite.

### Información biológica albergada

Con respecto a la información biológica contenida en la plataforma bioinformática generada, se tiene lo siguiente. En primera instancia, un total de 4.131 entradas están disponibles en la base de datos para la palma de aceite. Cada una de estas entradas está compuesta de un número de identificación interno (ID), el tipo de secuencia, la secuencia, la descripción, la especie a la cual pertenece (*E. guineensis* o *E. oleifera*) y un vínculo al identificador de la secuencia en la fuente original, en este caso el GI, utilizado en NCBI (Tabla 1).

Estos datos revelan que hay una mayor cantidad de información nucleotídica de *Core Nucleotide* y ESTs para *E. guineensis*, comparado con *E. oleifera*, aunque en secuencias de aminoácidos se tiene una cantidad de datos similares. La razón de este comportamiento es que al buscar la información molecular disponible para estas dos especies, la mayoría hace referencia a la palma africana (*E. guineensis*), pero al referirse específicamente a secuencias proteicas y sus dominios, se hablaba de la palma de aceite de manera global, ya que las pocas familias de proteínas estudiadas no varían aparentemente entre especies. En cuanto a las secuencias genómicas resultó una sorpresa: no el bajo porcentaje que ocupan en la base de datos (1,96%), sino que la mayoría pertenezca a *E. oleifera*, y que para *E. guineensis* no haya prácticamente ninguna. También cabe resaltar que el 76,74% de las secuen-



cias en la base de datos son del tipo ESTs, y al tener esta amplia cantidad se abre la posibilidad de tener múltiples marcadores moleculares para realizar mapeo genético. Este procedimiento ya ha tenido éxito en investigaciones moleculares con otras plantas como el arroz y el maíz (Cheah *et al.*, 1999).

Por otra parte, cabe resaltar algunas secuencias específicas que están en la base de datos, y que pueden mostrar su aplicación en el sector palmicultor. Por ejemplo, las proteicas para *E. guineensis*, refiriéndose a proteínas similares a la profilina (ID 486 a 489), las cuales tienen utilidad en estudios de manipulación

**Tabla 1.** Tipo y cantidad de secuencias que se encuentran en la base de datos para la palma de aceite, creada en este proyecto

Especies	Número de secuencias introducidas en la plataforma de Cenipalma			
	Nucleotídicas (Core)	Genómica	ESTs	Proteínas
<i>Elaeis guineensis</i> *	472	6	2.160	125
<i>Elaeis oleifera</i>	153	75	1.010	130

\* El tamaño del genoma reportado para esta especie es de  $3,4 \times 10^9$  pb o  $2C=3,76 \pm 0,09$  pg (Rival, 1997).

genética para aumentar o disminuir el crecimiento celular, no sólo en palma de aceite, sino en otras plantas como el modelo biológico *Arabidopsis thaliana* (Cheah *et al.*, 1999; Temnykh, 2000).

Otras secuencias de interés son las de *E. oleifera* (ID 757 a 831) que han sido utilizadas como sondas en investigaciones de cromosomas parentales del híbrido interespecífico (OxG) (Maria *et al.*, 1998).

En cuanto a la resistencia de patógenos hay secuencias (ID 462 a 466) de genes de resistencia (Genes R), que lo que hacen es codificar proteínas que confieren la capacidad para reconocer genes *avr* (avirulentos) y así detener el ataque infeccioso de un determinado patógeno. Aun cuando no ha sido estudiado propiamente en la palma de aceite, estudios como el de Betania *et al.* (2003) y Kover *et al.* (2002), en la interacción de *Arabidopsis thaliana* y *Pseudomonas syringae* pv. tomato, pueden ser tomados como modelo aplicable, debido a la similitud de los genes de resistencia entre familias de plantas.

También hay una gran cantidad de secuencias obtenidas de inflorescencias (ID 1036 a 1051, entre otras), lo cual se relaciona con la necesidad de encontrar las bases moleculares de la floración anormal en la propagación mediante técnicas de cultivo de tejidos.

Por último, existen múltiples secuencias nucleotídicas como las del gen completo de la plastoquinona oxidoreductasa (ID 633) y del gen de la ribulosa 1,5-bifosfato carboxilasa/oxigenasa (rubisco, ID 622), por citar unas pocas. Estos genes son de importancia, porque están involucradas en las rutas metabólicas de la fotosíntesis, tanto de la fase clara como del ciclo de Calvin. En consecuencia, por la eficiencia fotosintética desplegada por la palma de aceite, estos genes se constituyen en una fuente de enorme importancia para otras especies.

El resto de la información biológica en la plataforma bioinformática está ubicada en los enlaces restantes de patógenos/sintomatologías asociadas y sitios/herramientas bioinformáticas. El primer enlace alberga una lista de 13 patógenos y sintomatologías que han sido reportados en la palma de aceite. Sin embargo, las secuencias e información molecular relacionadas se incluirán en futuros proyectos. Un segundo enlace tiene una serie de vínculos a páginas útiles en estudios bioinformáticos, además de una pequeña descripción de las mismas.

## Conclusión

En conclusión, se logró desarrollar una plataforma bioinformática para la palma de aceite, con una buena cantidad (4.134 secuencias) y calidad de información, además de herramientas de utilidad para fomentar el uso de la bioinformática en las futuras investigaciones biotecnológicas relacionadas con el cultivo y lideradas por Cenipalma. El éxito logrado al desarrollar este primer modelo representa un paso importante en la implementación de la bioinformática en el sector palmicultor colombiano y complementa el sistema de información de Cenipalma (Martínez y Ruiz, 2005). Además, permite interactuar en redes unificadas con diferentes centros nacionales e internacionales (por ejemplo, Mpob, Cirad, etc.) dedicados a la investigación en diferentes áreas del cultivo de palma de aceite.

## Consideraciones finales

Durante varios años, Cenipalma obtuvo avances notables en lo relacionado con la utilización de marcadores moleculares para evaluar la diversidad genética de palma (Rocha y Rey, 2007a y b). Recientemente se ha incursionado en otras áreas de la biotecnología (Rocha, 2008), incluida la bioinformática. Con lo pre-



sentado aquí para palma de aceite, la bioinformática se convierte en una herramienta de bajo costo, que permite reducir la brecha tecnológica que se presenta, por ejemplo, con otros cultivos de oleaginosas y permite la comparación con información disponible de otras especies. Mediante la utilización de esta herramienta se ha concentrado la información actualmente disponible de palma de aceite y de organismos asociados, tales como aquellos involucrados a disturbios, patologías o enfermedades. Así, el gremio dispone de una herramienta valiosa que apoyará a

los programas de patología, fisiología, bioquímica, mejoramiento, etc.

## Agradecimientos

Los autores agradecen a Andrés Pinzón y Adriana Bernal (Profesores de la Universidad de los Andes) y al Comité de Publicaciones de Cenipalma, por sus valiosos comentarios en la crítica de este manuscrito. La investigación de Cenipalma es apoyada por el Fondo de Fomento Palmero (FFP) administrado por Fedepalma.



## Bibliografía

- Anwar, A. 2007. Elucidation of gene expression and genome structure of higher plants using novel high-speed genomics and bioinformatics technologies. In: Proceedings of the Pipoc 2007 International Palm Oil Conference (Agriculture, Biotechnology & Sustainability). Kuala Lumpur: 341.
- Arndt, T. 2007. Visual software tools for bioinformatics. *Journal of Visual Languages & Computing*. doi:10.1016/j.jvlc.2007.06.001.
- Aznar, A. 2005. La red internet. El modelo TCP/IP. España: Grupo Abanto Formación y Consultoría. pp. 4-7.
- Attwood, TK; Parry-Smith, DJ. 2002. Introducción a la Bioinformática. Prentice Hall, España. pp. 1-36, 209.
- Betania, F, Quirino; Bent, A. 2003. Deciphering host resistance and pathogen virulence: the *Arabidopsis/Pseudomonas* interaction as a model. *Molecular Plant Pathology*. 4 (6): 517-530.
- Budiman, MA; Singh, R; Low, ETL; Nunberg, A; Citek, R; Rohlfing, T; Bedell, JA; Lakey, ND; Martienssen, RA; Cheah, SC. 2005. Sequencing of the oil palm genespace. Proceedings of the Pipoc International Palm Oil Congress, Malaysia. pp. 628-639.
- Cheah, S; Rajinder, S; Maria, M. 1999. Generating molecular tools for oil palm improvement through the genomics approach. PORIM, Kuala Lumpur, Malasia. 61: 297-320.
- Cobo, A; Gómez, P; Pérez, D; Rocha R. 2005. PHP y MYSQL: Tecnología para el desarrollo de aplicaciones web. Díaz de Santos, España. pp. 99-100, 339-340.
- Elliot, ML; Broschat, TK; Uchida, JY; Simone, GW. 2004. Compendium of ornamental palm disease and disorders. The American Phytopathological Society Press, United States of America.
- Fedepalma. 2007. Reformada la estructura organizacional de Cenipalma. *El Palmicultor*. 423:6-7.
- Kover, PX; Wolf, JB; Kukel, BN; Cheverud, JM. 2005 Genetic architecture of *Arabidopsis thaliana* response to infection by *Pseudomonas syringae*. *Heredity* 94: 507-517
- Luscombe, N; Greenbaum D; Gerstein M. 2001. What is Bioinformatics: An introduction and overview. IMIA 2001 Yearbook. Recuperado el 22 de febrero de 2008, del Sitio Web de la Universidad de Yale <http://www.csb.yale.edu>.
- María, M; Clyde, MM; Cheah, SC. 1998. Chromosome painting in oil palm. PORIM, Kuala Lumpur, Malasia. 49.
- Martínez, R.; Ruiz, J. 2005. El sistema de información de Cenipalma. *Palmas* 26(4):61-76.
- Mutis, G. 2008. Desarrollo de una plataforma bioinformática de palma de aceite. Tesis de pregrado, Facultad de Ciencias, Universidad de los Andes, Bogotá.
- NCBI, 2008. Sitio Web oficial <http://www.ncbi.nlm.nih.gov>
- Olafsson, S; Li, X; Wu, S. 2006. Operations Research and Data Mining. *European Journal of operational research*. doi:10.1016/j.ejor.2006.09.023.
- Pavon, J. 2007. Creación de un portal con PHP y MYSQL. Tercera Ed. Alfaomega Grupo Editor. México. p. 1-3.
- Peña, M.; Del Portillo, P.; Castellanos, O.; Carrizosa, MS.; Peña, L. B. 2008. La biotecnología, motor de desarrollo para la Colombia de 2015. Colciencias, Primera edición. 264p.
- Qiang L; Pei H; Vasa C; Weizhong H; Yuan-Yuan L; Qing-Ming L; Yi-Ke G; Yi-Xue, L. 2006. KDE bioscience: platform for bioinformatics analysis workflows. *Journal of Biomedical Informatics*. 39(4): DOI 10.1016/j.jbi.200509.001.
- Rival, A; Beule, T; Barre, P; Hamon, S; Duval, Y; Noirot, M. (1997). Comparative flow cytometric estimation of nuclear DNA content in oil palm (*Elaeis guineensis* Jacq.) tissue cultures and seed-derived plants. *Plant Cell Reports*. doi:10.1007/s002990050339
- Rocha, P. J. 2005. Aportes de la biotecnología al cultivo de la palma de aceite en Pipoc 2005. *Palmas*. 26 (4): 53-59.
- Rocha, P. J. 2007a. Biotecnología en el cultivo de palma de aceite: Aspectos sobresalientes en Pipoc 2007. *Palmas*. 28(3): 47-55.
- Rocha, P. J. 2007b. La biotecnología como una herramienta para agilizar la selección de nuevos materiales en palma de aceite. *Palmas*. 28 (especial): tomo 1: 144-148.

- Rocha, P. J. 2008. División de biotecnología de cenipalma: Generador de herramientas para el sector palmicultor colombiano. *Ceniavances* 157:1-4.
- Rocha, P. J.; Rey, L. 2007a. Oil palm breeding program in Cenipalma, Colombia: A way to sustainability. Proceedings of the PIPOC 2007 International Oil Palm Congress (Agriculture, Biotechnology & Sustainability) Kuala Lumpur Convention Centre, Malaysia. 927-939.
- Rocha, P. J.; Rey, L. 2007b. Partial agronomical, biochemical and molecular characterisations of Colombian *Elaeis oleifera* germplasm bank. Proceedings of the PIPOC 2007 International Oil Palm Congress (Agriculture, Biotechnology & Sustainability) Kuala Lumpur Convention Centre, Malaysia. 1193-1198.
- Temnykh, S; Park, WD; Ayres, N; Cartinhour, S; Hauck, N; Lipovich, L; Cho, YG; Ishii, T; McCouch, SR. 2000. Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics*. Doi:10.1007/s001220051342
- Weng, KL; Maul, THB; Chen CL; Yuet, YL. 2007. Informatics in agriculture: issues and opportunities. In: Proceedings of the Pipoc 2007 International Palm Oil Conference (Agriculture, Biotechnology & Sustainability). Kuala Lumpur: 324-340.

## Pauta

## BANCO AGRARIO