

Análisis bioinformático y predicción de genes en secuencias genómicas de *Clostridium* sp. IBUN22A

Bioinformatic analysis and gene prediction in genomic sequences from *Clostridium* sp. IBUN22A

José David Montoya Solano*, Zulma Rocío Suárez Moreno**,
Dolly Montoya Castaño***, Fabio Ancízar Aristizábal Gutiérrez****

RESUMEN

Este trabajo tuvo como propósito identificar secuencias de genes en clones obtenidos en una librería genómica de la cepa colombiana de *Clostridium* sp. IBUN 22A. Los insertos de nueve clones con tamaños superiores a 500 pb fueron secuenciados y analizados por medio de búsquedas de los insertos completos o de sus marcos abiertos de lectura (ORFs) traducidos en GenBank 141.0 y Uniprot 6.6 con BLAST 2.2.8. Se identificaron seis genes con alta similitud a genes de metabolismo básico (*housekeeping*) en diferentes especies de *Clostridium*. En el clon pBsIBUN22A-1 se localizó una secuencia de 851 pb con una identidad del 99.74% respecto al gen codificante de la enzima glicerol deshidratasa (*DhaB1*) de *Clostridium butyricum* (AY968605) involucrada en la producción de 1,3 Propanodiol (1,3-PD). La identificación de genes presentes en la cepa nativa *Clostridium* IBUN 22A abre la puerta a la investigación básica y a la ingeniería metabólica para hacer más rentable el proceso de producción de 1,3-PD junto al conocimiento de los genes presentes en la cepa nativa.

Palabras clave: análisis de secuencias, librería genómica, *Clostridium*, glicerol deshidratasa, 1,3-propanodiol.

ABSTRACT

This work was aimed at identifying gene sequences in recombinant clones obtained from a genomic library from the Colombian *Clostridium* sp. IBUN 22A native strain. Nine clones greater than 500 bp were sequenced and analysed using BLAST 2.2.8 to search for complete inserts or their translated open reading frames (ORFs) in GenBank 141.0 and Uniprot 6.6. Seven genes having high similarity to housekeeping genes from different *Clostridium* species were identified. An 851 bp sequence was located in the pBsIBUN22A-1 clone, having 99.74% identity with the gene encoding the *Clostridium butyricum* enzyme glycerol dehydratase (*DhaB1*) which is involved in 1,3-propanediol (1,3-PD) production. Identifying genes in the native IBUN 22A strain formed the starting point for basic research and metabolic engineering aimed at making 1,3-PD production more profitable and increasing knowledge of genes present in the native strain.

Key words: glycerol dehydratase, 1,3-propanediol, genomic library, *Clostridium*, sequence analysis.

Recibido: enero 22 de 2006

Aceptado: mayo 02 de 2006

* QFUN. Universidad Nacional de Colombia. Instituto de Biotecnología de la Universidad Nacional de Colombia.

** QFUN. MSc. Microbiología. Instituto de Biotecnología de la Universidad Nacional de Colombia.

*** QFUN. MSc. PhD. Instituto de Biotecnología de la Universidad Nacional de Colombia. Correo electrónico: dmontoyac@unal.edu.co

**** QFUN. PhD. Departamento de Farmacia, Facultad de Ciencias. Universidad Nacional de Colombia.

INTRODUCCIÓN

Durante los años 90 se aislaron 13 cepas del género *Clostridium* provenientes de diferentes muestras de suelos colombianos caracterizadas por producir solventes (acetona, butanol, etanol) en una concentración superior a la de cepas de referencia del mismo género, y con capacidad para degradar polisacáridos naturales que incluyen almidón y carboximetilcelulosa (CMC) (Montoya et ál., 2000). Posteriormente se realizaron análisis de taxonomía polifásica que incluyeron la secuenciación del gen 16S rARN, caracterización molecular por campo pulsado (PFGE), detección de polimorfismos de longitud de fragmentos de restricción amplificados (AFLP), hibridación ADN-ADN y pruebas bioquímicas y enzimáticas que sugirieron que las cepas mencionadas podrían constituir una especie nueva cercanamente relacionadas con *Clostridium butyricum* (Montoya et ál., 2001; Jaimes et ál., 2006; Suárez et ál., 2006). Adicionalmente se determinó que cinco de las cepas tenían la capacidad de producir 1,3-propanodiol (1,3 PD) a partir de glicerol en una concentración mayor a la de tres cepas de referencia del género *Clostridium* (Cárdenas y Pulido, 2004) mediante una ruta bioquímica propia de algunos clostridios (figura 1). La cepa *Clostridium* sp. IBUN 22A coincide con los subgrupos con ma-

yor capacidad celulolítica, solventogénica y de producción de 1,3-PD a partir de glicerol. Por esta razón se generó una librería genómica en *Escherichia coli* a partir de *Clostridium* IBUN 22A, destinada a la búsqueda de genes de interés en clones con insertos de tamaño superior a 500 pb.

A través de una estrategia de minería de datos se analizaron secuencias genómicas obtenidas con el propósito de identificar sectores codificantes de enzimas asociadas a procesos metabólicos de interés comercial. Uno de los resultados más relevantes fue la identificación de la secuencia parcial de uno de los genes involucrados en la síntesis de 1,3-PD. Así mismo se identificaron seis genes de mantenimiento celular con secuencias altamente similares a los genes equivalentes en otras especies del género *Clostridium*.

MATERIALES Y MÉTODOS

Cepas bacterianas y plásmidos. La cepa celulolítica *Clostridium* sp. IBUN 22A, aislada de suelos colombianos (Montoya et ál., 2000), fue usada para la construcción de una librería genómica en *E. coli* XL1 BLUE con el vector de clonación pBluescript® II KS+/-, como lo describe Vargas et ál. (2002). Adicionalmente, el clon PBS-25 con actividad celulasa fue subclonado en *Escherichia coli*

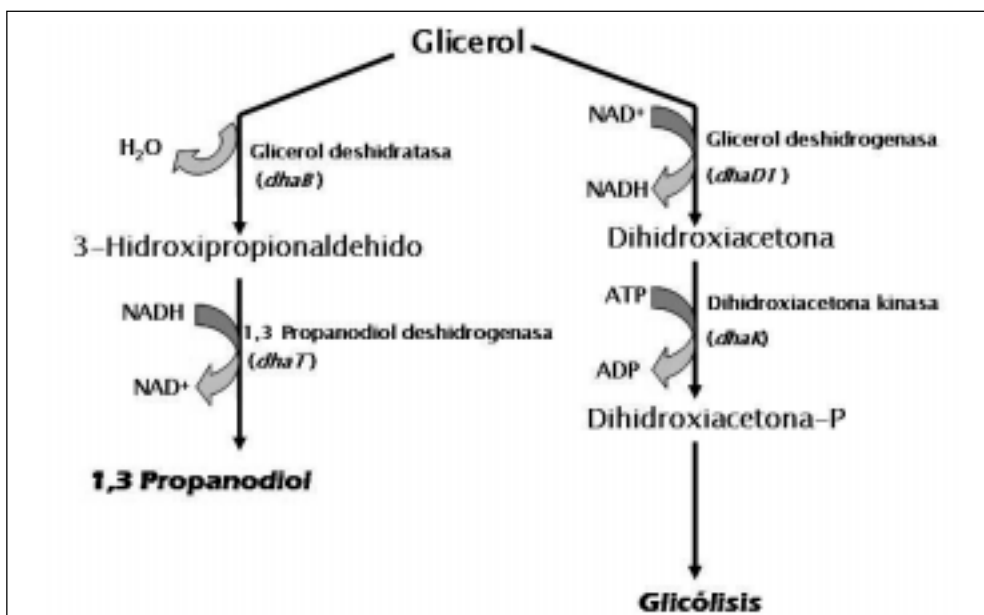


Figura 1. Ruta bioquímica de la asimilación de glicerol y producción de 1,3 Propanodiol en *Clostridium butyricum*. Se describen las enzimas implicadas en el proceso y sus genes identificados a la fecha (Raynaud et ál., 2003).

DH5á, obteniéndose así los clones PBSH-26 y PBSH-37 (Roncancio et ál, 2005).

T7 (T3, 5'-ATTAACCCTCACTAAAGGGA-3'; T7, 5'-TAATACGACTCACTATAGGG-3').

Secuenciación de clones representativos de la librería genómica y productos de subclonaje.

Todos los plásmidos fueron evaluados electroforéticamente y se seleccionaron para secuenciación aquellos cuyo tamaño de inserto fuera superior a 500 pb. Nueve plásmidos fueron seleccionados y extraídos con el kit Wizard® de Promega (tabla 1). La secuenciación se realizó por triplicado en las siguientes instituciones: servicio de secuenciación automatizada de la UNAM (México), el Laboratorio de Genómica e Expressao de UNICAMP (Brasil) y MACROGEN (Corea), usando los *primers* M13, F y R. (F, 5'-CCCAGTCACGACGTTGTAAAACGAC-3'; R, 5'-AGGAAACAGCTATGACCAT GATTAC-3'), T3 y

Análisis bioinformático de las secuencias de los insertos.

Las secuencias depuradas fueron comparadas contra la base de datos completa de Genbank 141.0 (Benson et ál., 2005), al igual que contra bases de datos específicas de *Clostridium acetobutylicum*, *C. botulinum* A, *C. difficile*, *C. perfringens* ATCC 13124 y str. 13, *C. tetani* E88 y *C. thermocellum* ATCC 27405 almacenadas en el *website* de NCBI (www.ncbi.nlm.nih.gov/), usando las aplicaciones BLASTN 2.2.8 y TBLASTX. Las secuencias traducidas en los seis marcos de lectura posibles fueron comparadas contra la base de datos de proteínas SwissProt 42.8 (Boeckmann et ál.2003) usando BLASTX 2.2.8 desde el portal de NCBI y contra la base de datos UniProt 6.6

Tabla 1. Clones secuenciados y registrados en GenBank

CLON	LONGITUD	GENES ENCONTRADOS	NÚMERO GENBANK
pBsIBUN22A-1	851	Contiene el extremo 5' del gen <i>dhaB1</i> que codifica para Glicerol Deshidratasa	AY968805
pBsIBUN22A-2	808	Contiene el gen <i>nifN</i> , que codifica para la proteína de biosíntesis del cofactor Fe -Mo de la nitrogenasa. El extremo t7 de la secuencia 16 de Corea no pudo ser alineado con esta secuencia, probablemente porque existe una región no secuenciada del inserto entre los dos extremos y por ello al extremo t7 se le denominó pBsIBUN22A-2-B.	DQ002548
pBsIBUN22A-3	228	Contiene el gen <i>tufA</i> que codifica para el Factor de Elongación TU.	DQ060835
011901-2	851	No se identificaron genes.	DQ223967
012901-20	1002	No se identificaron genes.	DQ223968
020801-30	716	Contiene el extremo 3' de un gen que codifica para Purina-nucleósido Fosforilasa, con secuencia conservada en <i>Clostridium</i> .	DQ225170
pBs25	591	El dominio C-terminal del gen Transcetolasa.	DQ228721
020803-16	342	Contiene el extremo 5' del dominio de unión a piridina (Transket_pyr) del gen Transcetolasa.	DQ228722
pBS25-B	665	Contiene el extremo 3' de un gen que codifica para Purina-nucleósido Fosforilasa.	DQ228723

(Leinonen et ál., 2004). Por último se extrajeron los marcos abiertos de lectura (ORFs) más largos usando el programa getorf y plotorf de EMBOSS (Centro Internacional de Bioinformática del IBUN, www.bioinf.ibun.unal.edu.co; Rice et ál., 2000) y se confirmaron con Artemis v5 (Rutherford et ál., 2000), ORFfinder de NCBI (Wheeler et ál., 2005). Los ORF más largos se compararon nuevamente contra UniProt usando BLASTP. En caso de hallarse una semejanza superior al 70% con genes o proteínas conocidos, sumada a la detección de señales de transcripción en posición -10 o -35, RBS (secuencias de unión a ribosomas) o palíndromos de terminación, se consideró que la secuencia correspondía con alta probabilidad a un gen. Las secuencias seleccionadas fueron registradas manualmente en GenBank usando la herramienta BankIt (en línea).

RESULTADOS

Secuenciación de clones representativos de librería genómica y productos de subclonaje. Se obtuvieron 25 secuencias entre 248 y 1546 pb. Se sometieron a GenBank nueve secuencias, en siete de las cuales se identificaron genes de *housekeeping* o de interés para las rutas metabólicas de *Clostridium* (tabla 1).

Predicción del gen *dhaB1* en el clon pBsIBUN22A-1. Se determinó que el inserto del clon pBsIBUN22A-1 tenía una longitud inicial de 851 pb. El análisis reveló que el ORF más largo estaba ubicado entre las posiciones 76 y 851 en sentido 5' 3' (figura 2), con una región de 386 pb, con identidad de 99.74% respecto a la región codificante del gen *dhaB1* de *C. butyricum* ($E=1e-70$) (figuras 3 y 4), con un sitio de unión a ribosoma localizado entre las posiciones 57 y 61, mientras que el resto de la secuencia exhibió una similaridad menor a una proteína transportadora de prolina y triptófano dependiente de sodio en *Clostridium tetani* (Brueggemann et ál., 2003).

El alineamiento de la traducción en el segundo marco de lectura del inserto de pBsIBUN 22A-1 con la secuencia aminoacídica de *DhaB1* en TBLASTX presenta una región de 152 aminoácidos (aa) con identidad perfecta y valor ($E=2e-77$). Este resultado es sobresaliente, dado que el gen *dhaB1* es el primero del operón 1,3-PD de *C.*

butyricum, y codifica para la enzima glicerol deshidratasa que permite la conversión de glicerol en 3-hidroxi propionaldehído en la ruta bioquímica de producción de 1,3 PD (Raynaud et ál., 2003). En este caso se identificó un sitio de unión al ribosoma (RBS) en las posiciones 57 a 61, hecho que permite predecir el gen con alta probabilidad y que sugiere la presencia del resto de los genes en la cepa nativa en atención a que se ha demostrado que tiene la capacidad de producir 1,3 propanodiol a partir de glicerol. Se requieren estudios posteriores para establecer la secuencia de los demás genes involucrados en la ruta de la cepa nativa, así como su organización física (tabla 1).

Detección de genes metabolismo básico o *housekeeping*. En el tercer marco abierto de lectura del clon pBsIBUN22A-2 se halló una región con 78% de identidad con el gen codificante de la proteína *NifN-B* de *Clostridium beijerinckii* (GenBank AAS91670) entre las posiciones 81 y 782, y con un sitio de unión a ribosoma entre la posición 68 y 72 (Chen et ál., 2001). La calidad de los alineamientos obtenidos ($E=3e-25$) y el hecho de que la mayor parte de los resultados de la búsqueda sean homogéneos en otras especies de *Clostridium* permiten predecir que esta secuencia codifica una proteína de 233 aa requerida para la biosíntesis del cofactor hierro-molibdeno (o hierro-vanadio) usado por nitrogenasas implicadas en la fijación de nitrógeno en bacterias (tabla 1).

En la secuencia del inserto del clon Gen22A-C100 entre la posición 20 y 127 se halló una región de 93% de identidad con parte del gen que codifica para el factor de elongación *Tu* de *Clostridium perfringens* str. 13 (Shimizu et ál., 2002), así como con el mismo gen de otras especies de Clostridios. El factor *Tu* promueve la unión del aminoacil-tRNA al sitio A del ribosoma durante la biosíntesis de proteínas. Adicionalmente, se ubicó un palíndromo de terminación entre las posiciones 155 y 174 (tabla 1).

En el extremo 5' del inserto del clon 020801-30, entre las posiciones 2 y 112, se identificó un sector que guarda un 85% de identidad con una región de la enzima purina-nucleósido fosforilasa de *Clostridium tetani* E88 (GenBank NP 782010) y con otras del género *Clostridium*. Esta enzima tiene la función de catalizar la reacción de formación

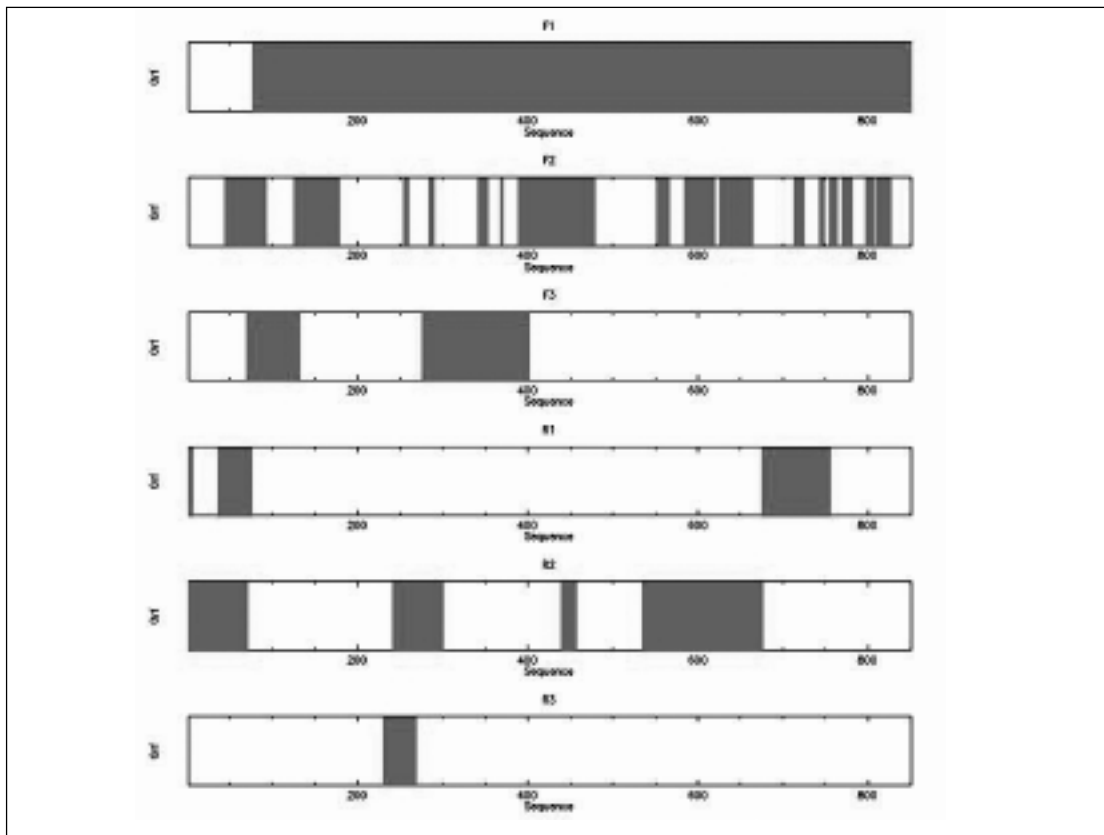


Figura 2. Marcos abiertos de lectura detectados en el análisis de la secuencia del clon pBSIBUN22A-1. Visualización obtenida por PLOTORF de Emboss (Rice et ál., 2000).

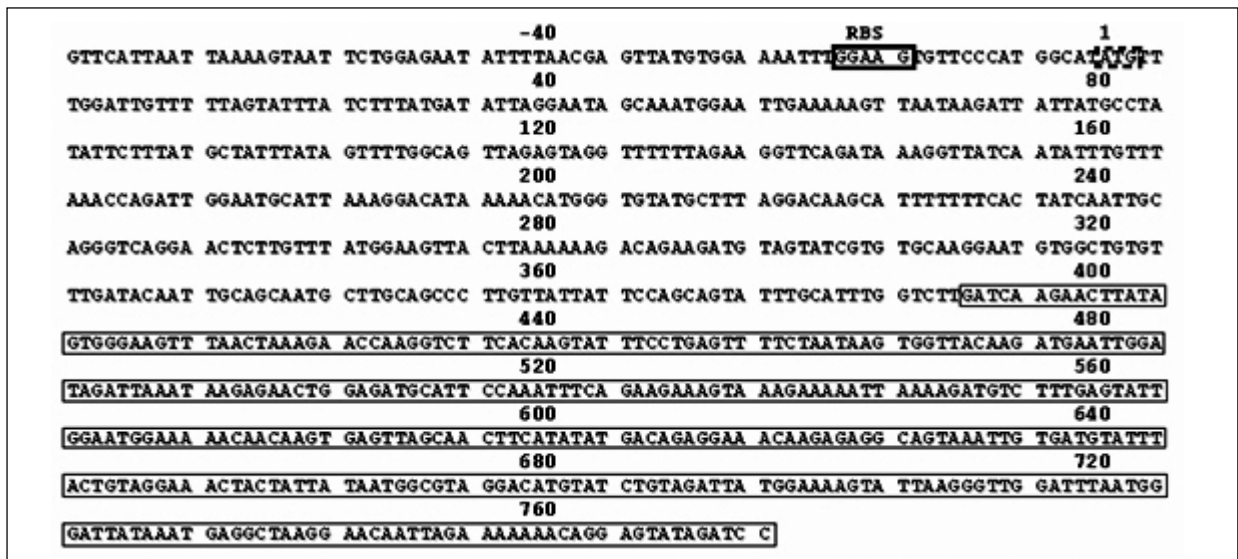


Figura 3. Secuencia del inserto del clon pBS IBUN 22A-1 registrada en GenBank (registro AY968605). El sitio de unión al ribosoma (RBS) y el codón de inicio (en líneas punteadas) han sido definidos a partir de secuencias consenso de estos elementos en diferentes genes de *C. butyricum*. La numeración positiva inicia en el primer nucleótido del codón de inicio. El rectángulo de línea continua delgada indica la región idéntica a *dhaB1* de *C. butyricum*.



Figura 4. Salida de Blastp para el ORF más largo identificado en la secuencia del clon pBsIBUN22A-1. En rojo la secuencia codificante de la proteína *DhaB1* y en azul con menor identidad la secuencia del transportador de prolina y triptofano dependiente de sodio.

de α -D-ribosa 1-fosfato en el metabolismo de las purinas. Este resultado sugiere que la secuencia obtenida corresponde a parte del gen codificante para la proteína mencionada, por lo que fue sometida a GenBank con el número DQ225170. Adicionalmente en la región 3' del inserto del clon pBS25B se detectó una región de 121 pb con similitud del 67% a la misma proteína (purin-nucleósido fosforilasa *DeoD*) de *Clostridium tetani* (tabla 1) (Brueggemann et ál., 2003).

En el inserto del clon pBS-25, en la región comprendida entre las posiciones 29 y 430, se encontró una región con una identidad del 69% ($E=6e-38$) con respecto al gen completo que codifica para el dominio carboxi terminal de la proteína transcetolasa de *Clostridium thermocellum* ATCC 27405 (GenBank ZP 00510628) y de 60% en el caso de la misma enzima en *Clostridium perfringens* str. 13 (GenBank: BAB80003) (Shimizu et ál., 2002). Se identificaron RBS entre la posición 17 y 22, y una señal -10 del promotor entre las posiciones 8 y 12, así como un palíndromo de terminación entre los sitios 458 y 482, hechos que sugieren que la secuencia completa sí corresponde al gen de una transcetolasa posiblemente involucrada en la vía de las pentosas-fosfato, por lo que fue sometida a GenBank con el número DQ228721. Adicionalmente en el inserto del clon 020803-16 se detectó que la traducción de uno de los ORF más largos presenta una similitud del 82% con el dominio de unión a piridina en transcetolasas de *Clostridium thermocellum* y *perfringens* (GenBank: ZP_00510628 y BAB80003), respectivamente. Confirman la predicción de esta proteína el hallazgo de una señal -10 del promotor entre las posiciones 21 y 26 (tabla 1) (Copeland et ál., 2005; Shimizu et ál., 2002).

En los insertos de los clones 011901-2 y 012901-20 (GenBank DQ223967 y DQ223968, res-

pectivamente) no se identificaron ORF representativos.

DISCUSIÓN

La identificación o reconocimiento de genes es uno de los problemas más importantes en la biología molecular, que ha incrementado su complejidad debido al advenimiento de un gran número de proyectos genoma. Las estrategias usualmente empleadas para resolver este problema se dividen en dos tipos, una de las cuales emplea la estadística de composición de las secuencias mientras que la otra recae en la búsqueda de similitud en bases de datos (Shibuya y Rigoutsos, 2002). En este trabajo se utilizó la segunda estrategia y las secuencias fueron contrastadas teniendo en cuenta las características estructurales de los genes.

En los insertos de los clones pBsIBUN22A-2, DQ060835, 020801-30, pBs25, 020803-16, pBS25-B, fue posible identificar genes de mantenimiento celular por similitud con genes descritos para especies del género *Clostridium*. La detección de tales genes de metabolismo permite la exploración genómica de las cepas nativas, aunque se requiere llevar a cabo pruebas bioquímicas y moleculares que demuestren la expresión y funcionalidad de estas enzimas en la cepa IBUN 22A.

La cepa colombiana *Clostridium* sp. IBUN 22A ha demostrado la capacidad de fermentar glicerol hasta 1,3-PD, y los últimos análisis taxonómicos la ubican en el mismo nodo de las especies productoras de 1,3-PD como *C. butyricum* (Suárez et ál., 2006). Cabe mencionar que Biebl y Spröer (2002) previamente demostraron que especies cercanamente relacionadas a la cepa nativa tienen la capacidad de metabolizar el glicerol y producir 1,3 propanodiol, como es el caso de algunas cepas

de *Clostridium diolis*. La producción de este solvente en las cepas nativas sugiere la existencia de un sistema genético que regule la producción de 1,3-PD, de manera similar a la descrita por Raynaud et ál., 2003, para *Clostridium butyricum* VPI1718 (2003).

La ruta de conversión de glicerol a 1,3-propanodiol involucra enzimas agrupadas en el regulón *dha* o en el operón 1,3-PD de acuerdo al organismo. Dos de ellas (glicerol deshidrogenasa y dihidroxiacetona kinasa) se encargan de convertir al glicerol en piruvato, produciendo así equivalentes NADH que son necesarios para la reducción del glicerol por medio de otras dos enzimas (glicerol deshidratasa y 1,3-PD deshidrogenasa) (Sun et ál., 2003) (figura 1). El paso limitante en este proceso es la conversión de glicerol en 3-hidroxi-propanaldehído catalizado por la glicerol deshidratasa. En este paso se produce un radical intermediario que suele involucrar a la coenzima B₁₂ como cofactor (Knietsch et ál., 2003). La glicerol deshidratasa *DhaB1* de *C. butyricum* es una excepción en dicha familia enzimática, ya que está compuesta por sólo una subunidad (gen *dhaB1*) y no requiere coenzima B₁₂ como cofactor (Nakamura y Whited, 2003). El hallazgo de una secuencia casi idéntica a una glicerol deshidratasa en IBUN 22A tiene particular relevancia dado que de su expresión depende toda la ruta de producción de 1,3-PD, siendo interesante averiguar en estudios posteriores si es o no dependiente de coenzima B₁₂, considerando que este es un costo adicional dentro del proceso industrial de 1,3-PD (Raynaud et ál., 2003, Sun et ál., 2003).

Se debe resaltar que la identificación de un gen del operón *dha* en IBUN 22A además de contribuir con la información sobre la cepa, abre las puertas para realizar mejoramiento de cualquiera de las cepas nativas aisladas junto con IBUN 22A, en las cuales se determinó productividad volumétrica y rendimiento de 1,3-PD igual o superior a ella (Cárdenas y Pulido, 2004). Actualmente se adelanta el proceso de secuencia de todo el operón que regula la producción de 1,3 PD en varias cepas nativas y se planea realizar la caracterización de las enzimas involucradas en la ruta para optimizar la producción de este metabolito. En el futuro una de estas cepas podría ser usada para la producción industrial del solvente a partir de

efluentes con alto contenido de glicerol como aquellos procedentes de las plantas de producción de biodiesel.

CONCLUSIONES

Los resultados de este trabajo demuestran la utilidad del análisis de secuencias de insertos en librerías genómicas para el hallazgo de genes con aplicación industrial y permite proyectar la investigación básica hacia la rentabilización del proceso de producción de 1,3-PD.

AGRADECIMIENTOS

El desarrollo de este proyecto fue posible gracias al respaldo económico de Colciencias y la Universidad Nacional de Colombia dentro del marco del proyecto "Producción de 1,3 Propanodiol a partir de glicerol generado del proceso de producción de biodiesel, empleando cepas nativas de *Clostridium* spp: estudio del operón, condiciones de producción por fermentación y su viabilidad económica" (1101-12-17848), desarrollado por el Instituto de Biotecnología. Algunas de las secuencias fueron obtenidas por el Dr. Gonçalo Guimarães Pereira del Laboratorio de Genómica e Expressao, Departamento de Genética e Evolução del Instituto de Biología de la Universidad de Campinas. Agradecemos la colaboración del Dr. Emiliano Barreto y Andrés Pinzón por su soporte en el análisis bioinformático.

BIBLIOGRAFÍA

- Altschul, S.; Madden, T.; Schaffer, A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids. Res.* 25: 3389-3402.
- Benson, D.A.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Wheeler, D.L. 2005. GenBank: *Nucleic. Acids. Res.* Update. 32 (database issue): D34-D38.
- Biebl, H.; Sproer, C. 2002. Taxonomy of the glycerol fermenting clostridia and description of *Clostridium diolis* sp. *Nov. Syst. Appl. Microbiol.* 25: 491-497.
- Boeckmann, B.; Bairoch, A.; Apweiler, R.; Blatter, M.C.; Estreicher, A.; Gasteiger, E.; Martin, M.J.; Michoud, K.; O'Donovan, C.; Phan, I.; Pilbout, S.; Schneider, M. 2003. The SWISS-PROT protein

- knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* 31: 365-370.
- Brueggemann, H.; Baeumer, S.; Fricke, W.F.; Wiezer, A.; Liesegang, H.; Decker, I. et ál. 2003. The genome sequence of *Clostridium tetani*, the causative agent of tetanus disease. *Proc. Natl. Acad. Sci. U.S.A.* 100: 1316-1321.
- Cárdenas, D.P.; Pulido, C. 2004. *Evaluación del potencial de las cepas nativas de Clostridium spp. para producir 1,3-Propanediol por medio de la fermentación de glicerol*. Trabajo de grado, Universidad Nacional de Colombia, Bogotá D.C., Colombia.
- Chen, J.S.; Toth, J.; Kasap, M. 2001. Nitrogen-fixation genes and nitrogenase activity in *Clostridium acetobutylicum* and *Clostridium beijerinckii*. *J. Ind. Microbiol. Biotechnol.* 27: 281-286.
- Copeland, A.; Lucas, S.; Lapidus, A.; Barry, K.; Detter, C.; Glavina, T.; Hammon, N.; Israni, S.; Pitluck, S.; Richardson, P. 2005. Sequencing of the draft genome and assembly of *Clostridium thermocellum* ATCC 27405. US DOE Joint Genome Institute (JGI-PGF). (Datos no publicados disponibles en GenBank acceso NZ_AABG 00000000).
- Jaimes, C.P.; Aristizábal, F.A.; Bernal, M.; Suárez, Z.R.; Montoya, D. 2006. AFLP fingerprinting of Colombian *Clostridium* spp. strains, multivariate data analysis and its taxonomical implications. *J. Microbiol. Methods*. (Disponible en <http://www.sciencedirect.com>).
- Knietsch, A.; Bowien, S.; Whited, G.; Gottschalk, G.; Daniel, R. 2003. Identification and characterization of coenzyme B12-dependent glycerol dehydratase- and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. *Appl. Environ. Microbiol.* 69: 3048-3060.
- Leinonen, R.; Diez, F.G.; Binns, D.; Fleischmann, W.; López, R.; Apweiler, R. 2004. UniProt Archive. *Bioinformatics* 20: 3236-3237.
- Montoya, D.; Arévalo, C.; Gonzáles, S.; Aristizábal, F.; Schwarz, W.H. 2001. New solvent-producing *Clostridium* sp. Strains, hydrolyzing a wide range of polysaccharides, are closely related to *Clostridium butyricum*. *J. Ind. Microbiol. Biotechnol.* 27: 329-335.
- Montoya, D.; Espitia, S.; Silva, E.; Schwarz, W.H. 2000. Isolation of mesophilic solvent-producing clostridia from Colombian sources: physiological characterization, solvent production and polysaccharide hydrolysis. *J. Biotechnol.* 79: 117-126.
- Nakamura, C.; Whited, G. 2003. Metabolic engineering for the microbial production of 1,3-propanediol. *Curr. Opin. Biotechnol.* 14: 454-459.
- Raynaud, C.; Sarçabal, P.; Meynial-Salles, I.; Croux, C.; Soucaille, P. 2003. Molecular characterization of the 1,3-propanediol (1,3-PD) operon of *Clostridium butyricum*. *Proc. Natl. Acad. Sci.* 100: 5010-5015.
- Rice, P.; Longden, I.; Bleasby, A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 16: 276-277.
- Roncancio, I.A.; Suárez, Z.R.; Montoya, D.; Aristizábal, F. 2006. Aislamiento de clones con actividad endo- β -1,4-glucanasa a partir de un segmento de ADN de 13kb de *Clostridium* sp IBUN22A. *Universitas Scientiarum.* 11(1): 29-40.
- Rutherford, K.; Parkhill, J.; Crook, J.; Horsnell, T.; Rice, P.; Rajandream, M.A.; Barrell, B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics.* 16: 944-945.
- Sun, J.; Van der Heuvel, J.; Soucaille, P.; Qu, Y.; Zeng, A. 2003. Comparative analysis of *dha* regulon and related genes for anaerobic glycerol metabolism in bacteria. *Biotechnol. Prog.* 19: 263-272.
- Shibuya, T.; Rigoutsos, I. 2002. Dictionary-driven prokaryotic gene finding. *Nucleic Acids Research.* 30: 2710-2725.
- Shimizu, T.; Ohtani, K.; Hirakawa, H.; Ohshima, K.; Yamashita, A.; Shiba, T. et ál., 2002. Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. *Proc. Natl. Acad. Sci. USA.* 99: 996-1001.
- Suárez, Z.R.; Cháves, B.; Aristizábal, F.; Montoya, D. 2006. Polyphasic approach to the taxonomy of thirteen *Clostridium* spp. strains by means of multivariate data analysis: strategy for describing a new species. *Proceedings of the 158th SGM Meeting*. Warwick-Inglaterra.
- Vargas, L.C.; Montoya, D.; Aristizábal, F. 2002. Clonación y expresión en *Escherichia coli* de genes de celulasas de *Clostridium* IBUN 22A. *Rev. Colombiana Biotecnol.* 4: 29-35.
- Wheeler, D.L.; Barret, T.; Benson, D.A.; Bryant, S. H.; Canese, K.; Church, D.M. et ál. 2005. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 33 (Database Issue): D39-D45.