

Bioinfo_eXtrema: un enfoque bioinformático para integrar información ambiental, bioquímica y genómica, enfocado en bioprospección y selección de consorcios de microorganismos con aplicaciones en biorremediación

Capdevielle, F. ⁽¹⁾, Ottati, C. ⁽²⁾, Lopretti, M. ^(2,3)

⁽¹⁾ Unidad de Biotecnología. Instituto Nacional de Investigación Agropecuaria, INIA - ⁽²⁾ Departamento de Bioprocesos y Biotecnología. Laboratorio Tecnológico del Uruguay, LATU - ⁽³⁾ Laboratorio de Bioquímica y Biotecnología. CIN, Facultad de Ciencias. Universidad de la República, Uruguay.

Contacto: fcapdevielle@inia.org.uy

Recibido: 28/06/2010 - Aprobado: 21/09/2010

Resumen

La identificación de componentes funcionales clave para diversos bioprocesos de interés industrial ha permitido seleccionar aislamientos adaptados a condiciones ambientales extremas en tres especies de hongos del género *Penicillium*. Dichos aislamientos fueron evaluados *in vitro* para caracterizar su potencial como componentes de un consorcio microbiano aplicable en biorremediación de efluentes industriales que contienen residuos lignocelulósicos.

Los resultados de la anotación de secuencias genómicas disponibles para una de las especies identificadas apuntan a la existencia de genes con alta similaridad respecto a los existentes en diversos hongos considerados como referencia en materia de degradación de lignina en ambientes naturales. Las anotaciones funcionales propuestas a partir de secuencias accesibles –identificadas a través de la base de datos Fungal Oxidative Lignin Enzymes– podrían contrastarse con los resultados experimentales para cepas creciendo en diferentes medios con lignina, representando ambientes industriales extremos. Mediante este trabajo se propone el ensamblado de Bioinfo_eXtrema como parte de un enfoque bioinformático centrado en la selección de consorcios de extremófilos para aplicaciones en biotecnología industrial, combinando diversas técnicas de minería de datos –integradas a través del Waikato Environment for Knowledge Analysis– para facilitar la integración de información molecular disponible e indicadores funcionales relevantes para aplicaciones en biorremediación.

Palabras clave: Extremozimas, lignina, minería de datos, *Penicillium*.

Abstract

The ability to identify microbe extremophiles with metabolic capabilities suited for specific bioprocesses opens the doors of extreme environments for development of new industrial products. Identification of key functional components from existing biodiversity supported selection of candidate isolates from three *Penicillium* fungal species. These candidates were evaluated *in vitro* to further characterize their potential as components of a microbial consortium for biorremediation of industrial effluents containing lignocellulosic residues. Results from annotation of available genomic sequences for one of these *Penicillium* species pointed to the existence of putative genes highly similar to those functionally identified in reference fungi for degradation of lignin in natural environments. Proposed functional annotations from available sequences were identified through a specialized database –Fungal Oxidative Lignin Enzymes (FOLy)– and could be contrasted straightforward with experimental results for strains growing in different media containing lignin, representing extreme industry-related environments. We propose the assembly of Bioinfo_eXtreme as an industrial-biotechnology-centered bioinformatics approach for consortia selection, combining diverse data mining techniques –components of the Waikato Environment for Knowledge Analysis (WEKA)–, to facilitate integration of available molecular information and relevant phenotypic indicators for biorremediation applications.

Keywords: Extremozymes, lignin, data mining, *Penicillium*.

Introducción

Por medio del mejoramiento en las técnicas de cultivo y/o detección de microorganismos se ha podido verificar que diversos ambientes con características “extremas” –caracterizados por condiciones restrictivas para el desarrollo de organismos vivos (alta o baja temperatura, acidez, salinidad, presión, etcétera)– presentan una importante biodiversidad de microorganismos “extremófilos” adaptados a dichas condiciones. La bioprospección de este tipo de microorganismos ha tenido un crecimiento sostenido en los últimos años en ambientes distribuidos en todo el mundo, considerando las aplicaciones alcanzadas en

diferentes industrias (diagnósticos, bioprocesos, entre otras) y servicios ambientales (remediación de ambientes contaminados), así como el potencial de uso industrial de las enzimas producidas por estos organismos (denominadas extremozimas).

El proyecto en el que se inscribe este trabajo apunta a establecer una plataforma biotecnológica como estrategia para impulsar el desarrollo productivo de microorganismos con características adaptativas “extremas” (tolerancia a factores físicos y/o químicos) con potencial para el desarrollo sustentable de nuevos productos y/o servicios biotecnológicos con alto valor industrial. A partir de muestras ambientales tamizadas a través de columnas que contienen “licor negro” –un residuo de la industria celulósica con alto contenido de

compuestos fenólicos— en condiciones de acidez (pH 4,5), se aislaron cepas de tres especies del género *Penicillium* (*Penicillium citrinum* Thom., *Penicillium variable* Sopp., y *Penicillium chrysogenum* Thom.) adaptadas a condiciones que representan posibles ambientes “extremos” vinculados con el tratamiento de residuos lignocelulósicos industriales.

En este trabajo se discute la importancia de incorporar enfoques bioinformáticos como soporte para el diseño de experimentos relevantes en el desarrollo de prototipos para nuevos productos biotecnológicos. El propósito es contribuir a la identificación de cepas de microorganismos extremófilos que contienen información genética para enzimas de interés biotecnológico industrial (extremozimas) con apoyo de procedimientos bioinformáticos. De esta manera se busca facilitar el acceso a información molecular y su integración, mediante técnicas de minería de datos, respecto a diversas categorías de descriptores fenotípicos y capacidades metabólicas estimadas en forma experimental.

Materiales y Métodos

Las cepas de *Penicillium* estudiadas (Figura 1) pueden crecer y desarrollarse en medios de cultivo con niveles elevados de compuestos fenólicos; ha sido evaluada su capacidad metabólica para modificar diferentes ligninas—organosolv, kraft y sulfonadas— y decolorar medios que contienen “licor negro”. Las enzimas potencialmente involucradas en la degradación de lignina fueron identificadas por medio del análisis comparativo de sus secuencias proteicas e integradas en una base de datos denominada Fungal Oxidative Lignin enzymes (FOLy), enfocada en la provisión de herramientas para la anotación funcional de los genomas de diversas especies de hongos que resultan de interés (Levasseur et al., 2008) como modelos de capacidades metabólicas aplicables en procesos de biorremediación.

Se utilizó el sistema WEKA (Waikato Environment for Knowledge Analysis, Witten y Frank, 2005) como soporte para la evaluación de diversas técnicas de minería de datos aplicables en la identificación de patrones diferenciales entre las secuencias de ADN correspondientes a diversas categorías funcionales definidas en FOLy. La utilización de información proveniente de secuencias anotadas con alto grado de confianza mediante integración de criterios estructurales y funcionales—accesibles en bases de datos *ad hoc* gestionadas por expertos del dominio funcional correspondiente— permitiría definir conjuntos de entrenamiento que faciliten la revisión continua de anotaciones realizadas en las primeras etapas de diferentes programas de investigación genómica. En nuestro análisis se consideraron dos categorías de enzimas con potencial ligninolítico definidas en FOLy a partir de referencias caracterizadas funcionalmente—con información bioquímica disponible—; una de ellas comprende enzimas con actividad oxidasa directa sobre lignina (Lignin Oxidasas: LOs) y otra agrupa enzimas con actividad oxidasa que actúan como auxiliares del proceso de degradación de lignina (Lignin Degrading Auxiliary enzymes: LDAs).

Asimismo, se utilizaron técnicas de minería de datos aplicadas a textos para integrar información funcional reportada en la bibliografía (National Center for Biotechnology Information) como apoyo en la identificación de posibles enzimas ligninolíticas en *Penicillium*. Las posibles asociaciones conceptuales entre identificadores taxonómicos y capacidades metabólicas fueron exploradas mediante dos herramientas del sistema PathText (Kemper et al. 2010): FACTA y KLEIO. FACTA (Tsuruoka et al. 2008) es un sistema de recuperación de información que permite reconocer asociaciones implícitas entre términos co-ocurrentes en una colección de artículos, mientras que KLEIO (Nobata et al. 2008) provee un rango de funciones para búsquedas semánticas y genera índices de conceptos para facilitar su identificación aun en casos donde existen diferencias apreciables en cuanto al uso de variantes ortográficas y acrónimos para los términos considerados.

Como referencia comparativa para los modelos utilizados en la anotación de genes para enzimas ligninolíticas se utilizó el genoma de

Phanerochaete chrysosporium (Martínez et al., 2004). La versión utilizada (2.0) integra la información genómica de esta especie (aproximadamente 35 millones de pares de bases) en más de 610.000 fragmentos ensamblados mediante el sistema “JAZZ assembler” (Joint Genome Institute), con 10.048 modelos de genes (BestModels v2.1) hipotéticos anotados.

A partir de 2008 se encuentra disponible en el National Center for Biotechnology Information una versión del genoma de la cepa Wisconsin 54-1255 de *P. chrysogenum* (32,2 millones de pares de bases secuenciadas) donde se han anotado los modelos génicos correspondientes a más de 12.000 secuencias codificantes para proteínas. Estos modelos fueron utilizados como base para evaluar combinaciones de procedimientos de minería de datos para contrastar las anotaciones accesibles vía FOLy para *P. chrysogenum*.

Los procedimientos de minería de datos utilizados incluyeron búsquedas por palabras clave extraídas de la anotación de secuencias genómicas de esta especie que presentaban niveles significativos de similitud (estimados mediante tBLASTn) con enzimas ligninolíticas de *P. chrysosporium*. Las proteínas hipotéticas de *P. chrysogenum* correspondientes a los modelos génicos identificados en FOLy, así como aquellas correspondientes a los modelos génicos identificados en forma *ad hoc* mediante uso de palabras clave fueron analizadas con un algoritmo de tipo “text mining” accesible vía web (<http://adios.tau.ac.il/DME/>), que permite identificar regiones de la secuencia con valor predictivo respecto a su posible funcionalidad enzimática (Weingart, Lavi y Horn, 2009).

Resultados

Los resultados experimentales obtenidos en el marco de este proyecto (Lopretti et al. 2010) indican que las tres cepas de *Penicillium* consideradas (Figura 1) pueden expresar capacidades metabólicas ligninolíticas en condiciones ambientales extremas (alta acidez y concentración de compuestos polifenólicos), reduciendo significativamente la pigmentación de los medios analizados y produciendo modificaciones de los compuestos fenólicos modelo y diversos compuestos metoxilados ensayados.

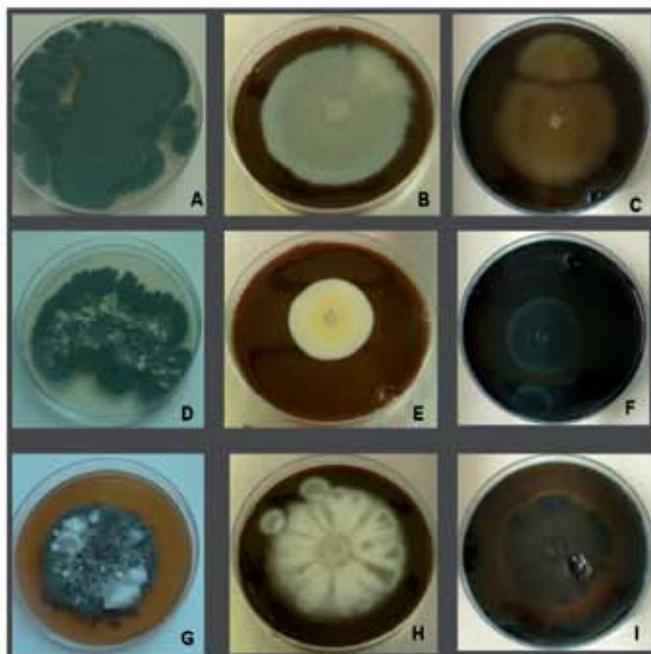


Figura 1. Crecimiento de cepas de *Penicillium* en diferentes medios: PDA (A, D, G); 100 mm glucosa + 2g/l NNaO_3 + 0,5% licor negro (B, E, H);, 1% FeCl_3 + 1% $\text{K}_3[\text{Fe}(\text{CN})_6]$. Las fotografías corresponden a *P. chrysogenum* (A, B, C), *P. variable* (D, E, F) y *P. citrinum* (G, H, I).

Utilizando como referencia palabras clave extraídas de la comparación –vía tBLASTn– entre secuencias de *P. chrysosporium* y *P. chrysogenum* se identificaron 13 proteínas hipotéticas anotadas para esta última especie como “multicopper oxidase”, que podrían corresponder a enzimas con actividad como lacasas. Aplicando el procedimiento de “text mining” –instrumentado por Weingart, Lavi y Horn (2009) siguiendo el método propuesto por Kunik et al. (2007)– se reconocieron péptidos específicos (Figura 3) en las secuencias de cinco de estas proteínas hipotéticas, apoyando su identificación como posibles enzimas con actividad de tipo lacasa (EC 1.10.3.2.).

En forma comparativa es posible indicar que dos de las tres secuencias anotadas por FOLy dentro de la categoría LO1 también fueron identificadas como lacasas mediante los procedimientos de minería de datos aplicados, apoyados tanto en el reconocimiento de similitudes con secuencias de lacasas anotadas en otros hongos ligninolíticos como de péptidos específicos asociados con diferentes clases de enzimas (“EC numbers”).

Discusión

Bioinfo_eXtrema I: selección de microorganismos para un consorcio aplicable en biorremediación de efluentes con alto contenido lignocelulósico.

Desde el punto de vista de los antecedentes en materia de degradación de lignina y compuestos fenólicos similares en condiciones “extremas”, Leitao (2009) revisó recientemente el potencial del género *Penicillium* para diversas aplicaciones en biorremediación, resaltando la importancia de incorporar procedimientos para caracterización funcional (a nivel bioquímico y molecular) que contribuyan a seleccionar cepas de alto potencial biotecnológico a partir de las etapas más tempranas de prospección y tamizado por adaptación a ambientes extremos. Los resultados experimentales sobre degradación de lignina obtenidos para las cepas de *Penicillium* adaptadas a condiciones ambientales extremas (acidez y presencia de contaminantes fenólicos) consideradas en este trabajo, así como la consistencia entre diversos métodos que respaldan la anotación de genes candidatos para enzimas ligninolíticas en *P. chrysogenum*, apuntan a la existencia de capacidades metabólicas valiosas para la selección de componentes de un consorcio fúngico aplicable en procesos de biorremediación.

Bioinfo_eXtrema II: integración de procedimientos para minería de datos en la interfase bioprospección / genómica / biotecnología.

Los resultados obtenidos en nuestro caso sugieren la conveniencia de utilizar un enfoque bioinformático apoyado por técnicas de minería de datos –en particular “text mining”– para integrar información sobre genes candidatos –establecida en las etapas realizadas o en proceso de anotación genómica para cada especie– que pueda contrastarse directamente con la determinación de actividades para enzimas específicas (lignin peroxidasas, manganeso peroxidasas y oxidasas en general) producidas por las cepas que mostraron crecimiento en los diferentes medios con lignina. Las actividades enzimáticas de tipo lacasa y peroxidasa que fueron determinadas experimentalmente para las tres cepas de *Penicillium* pueden ser interpretadas –con herramientas bioinformáticas y procedimientos de minería de datos– en términos de la posible base genómica existente en este género para metabolizar y utilizar compuestos lignocelulósicos.

La selección de microorganismos aplicables en procesos de biorremediación requiere generar, integrar y analizar múltiples fuentes de información para identificar combinaciones de capacidades metabólicas que resulten adecuadas para implementar diferentes tipos de bioprocesos conducentes a reducir y mitigar posibles daños ambientales asociados con actividades productivas. Dada la necesidad

de difundir los procedimientos evaluados y coordinar capacidades instrumentales y humanas para seleccionar cepas integrables en consorcios de biorremediación, se propone el desarrollo de un enfoque bioinformático –*Bioinfo_eXtrema*– como base de una plataforma biotecnológica interinstitucional articulada con la industria de base biotecnológica. El propósito es facilitar la integración de información fenotípica, genómica y de bioprocesos industriales para identificar funciones metabólicas y dominios estructurales aplicables en anotación funcional de microorganismos procedentes de ambientes naturales e industriales extremos, con aplicaciones potenciales en biorremediación.

Conclusiones

La selección de aislamientos de hongos candidatos para aplicaciones en biorremediación a partir de la biodiversidad existente en ambientes naturales e industriales extremos puede ser facilitada mediante la anotación genómica de componentes funcionales clave para diversos bioprocesos de interés productivo. Por medio de este trabajo se propone continuar avanzando en el ensamblado de una plataforma genómica-bioinformática-biotecnológica enfocada en el desarrollo productivo de microorganismos extremófilos con aplicaciones en biorremediación y biotecnología industrial. Esta iniciativa apunta a impulsar el diseño de nuevas aplicaciones biotecnológicas y bionegocios innovadores basados en la integración de múltiples fuentes de información (ambiental, genómica, bioquímica, industrial, etcétera), enfocadas en bioprocesos industriales de alto valor económico a nivel internacional, así como sostenibles en términos ambientales y de su contribución a un desarrollo productivo y social intensivo en conocimientos.

Referencias

- NATIONAL CENTRE FOR TEXT MINING; THE UNIVERSITY OF MANCHESTER. SCHOOL OF COMPUTER SCIENCE. *FACTA* + [En línea]. Versión 0.7. s.l.: National Centre for Text Mining School of Computer Science; The University of Manchester, 2009. [Consulta 27 de junio de 2010]. Disponible en: http://refine1-nactem.mc.man.ac.uk/facta_events/
- KEMPER, Brian; MATSUZAKI, Takuya; MATSUOKA, Yukiko; TSURUOKA, Yoshimasa; KITANO, Hiroaki; ANANIADOU, Sophia; TSUJII, Jun'ichi. PathText: a text mining integrator for biological pathway visualizations. En: *Bioinformatics*. 2010, 26(12):374-381.
- KUNIK Vered; MEROZ, Yasmine; SOLAN, Zach; SANDBANK, Ben; WEINGART, Uri; RUPPIN, Eytan; HORN, David. Functional representation of enzymes by specific peptides. En: *PLOS Comp Biol*. 2007, 3(8):1623-1632
- INRI; UNIVERSITES DE PROVENCE ET DE LA MEDITERRANEE; UMR. *FOLy Fungal Oxidative Lignin enzymes* [En línea]. Marseille: INRI; Universites de Provence et de la Mediterranee; UMR, 2010. [Consulta 27 de junio de 2010]. Disponible en: <http://folly.esil.univ-mrs.fr/>
- LEITAO, Ana Lucia. Potential of *Penicillium* species in the bioremediation field. En: *Int. J. Environ. Res. Public Health*. 2009, 6(4):1393-1417
- LEVASSEUR, Anthony; PIUMI, Francois; COUTIÑO, Pedro M.; RANCUREL, Corinne; ASTHER, Michèle; DELATTRE, Michel; HENRISSAT, Bernard; PONTAROTTI, Pierre; ASTHER, Marcel; RECORD, Eric. FOLy: An integrated database for the classification and functional annotation of fungal oxidoreductases potentially involved in the degradation of lignin and related aromatic compounds. En: *Fungal Genetics and Biology*. 2008, 45(5):638-645.
- LOPRETTI, Mary; OTTATI, Carolina; CAPDEVIELLE, Fabián; DAMBORIARENA, Agustín; SIBAJA, María. *Penicillium's* consortium: potential modifier of polyphenols for management and industrial use. En: EUROPEAN COMMISSION. *18th European Biomass Conference and Exhibition* (Lyon 3-7 de mayo de 2010). Lyon: European Commission, 2010.
- MARTINEZ, Diego; LARRONDO, Luis; PUTMAN, Nik; GELPKER, Maarten D.S.; HUANG, Katherine; CHAPMAN, Jarrod;

- HELFENBEIN, Kevin G.; RAMAIYA, Preethi; DETTER, J. Chris; LARIMER, Frank; COUTINHO, PEDRO M.; HENRISSAT, Bernard; BERKA, Randy; CULLEN, Dan; ROKHSAR, Daniel. Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. En: *Nature Biotechnology*. 2004, (22):1-6.
- NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. *Genomes & maps* [En línea]. Bethesda: NCBI, s.d. [Consulta 27 de junio de 2010]. Disponible en: <http://www.ncbi.nlm.nih.gov/guide/genomes-maps/>
 - NOBATA, Chikashi; COTTER, Philip; OKAZAKI, Naoaki; REA, Brian; SASAKI, Yutaka; TSURUOKA, Yoshimasa; ANANIADOU, Sophia. Kleio: a knowledge-enriched information retrieval system for biology. En: MYAENG, S.H. et al. *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. Singapore: ACM, 2008. pp. 787-788.
 - TSURUOKA, Yoshimasa; TSUJII, Jun'ichi; ANANIADOU, Sophia. FACTA: a text search engine for finding associated biomedical concepts. En: *Bioinformatics*. 2008, 24(21):2259-2260.
 - THE NATIONAL CENTRE FOR TEXT MINING. *KLEIO* [En línea]. Manchester: The National Centre for Text Mining, s.d. [Consulta 27 de junio de 2010]. Disponible en: <http://www.nactem.ac.uk/software/kleio/>
 - WEINGART, Uri; LAVI, Yair; HORN, David. Data mining of enzymes using specific peptides. En: *BMC Bioinformatics*. 2009, (10):446-456.
 - WEINGART, Uri; LAVI, Yair; HORN, David. *Data mining of enzymes. Peptide search* [En línea]. s.l.: Uri Weingar, s.d. [Consulta 27 de junio de 2010]. Disponible en: <http://adios.tau.ac.il/DME/>
 - UNIVERSITY OF CALIFORNIA. *Phanerochaete chrysosporium* [En línea]. Versión 2.0. California: University of California, s.d. [Consulta 27 de junio de 2010]. Disponible en: <http://genome.jgi-psf.org/Phchr1/Phchr1.home.html>
 - WITTEN, Ian H.; FRANK, Eibe. *Data mining: practical machine learning tools and techniques*. San Francisco: Morgan-Kaufmann, 2005.
-