

# CAPÍTULO IX

## ERA POST-GENÓMICA PARA EL DESARROLLO DE NUEVOS BIOCATALIZADORES

MARÍA EUGENIA GUAZZARONI, LUCÍA FERNÁNDEZ ARROJO,  
NIEVES LÓPEZ-CORTÉS, MANUEL FERRER

### RESUMEN

En la actualidad, la biología molecular moderna experimenta un aumento hacia enfoques "-ómicas», y, en particular, la genómica y la (meta) genómica como herramientas emergentes para el estudio de genes que codifican las enzimas a nivel de microorganismos individuales y comunidades imposibles de cultivar y, microorganismos sin explorar hasta ahora. Durante los últimos años, hemos visto enormes avances en nuestra capacidad para aislar microbios y analizar los genomas individuales, y la información proporcionada por estas secuencias es bastante sorprendente en lo que respecta a descifrar nuevas enzimas con nuevas actividades. Los métodos moleculares, en particular, (meta) genómica, se aplican ampliamente para caracterizar nuevos biocatalizadores. La (meta) genómica es la aplicación de modernas técnicas de la genómica al estudio de las comunidades de organismos microbianos directamente en sus ambientes naturales, evitando la necesidad de aislamiento y cultivo en el laboratorio de las especies individuales. Desde entonces, la (meta) genómica ha revolucionado la biotecnología desviando el interés lejos de cepas microbianas individuales hacia el 99% de las especies microbianas que actualmente se estima no se puede cultivar. Analizaremos las diferentes estrategias para aislar nuevas enzimas y como la aplicación directa de la (meta) genómica sobre las comunidades microbianas y sus enzimas revierte en una amplia gama de aplicaciones biomédicas, industriales y ambientales.

**Palabras clave:** Biocatalizador; Biodiversidad; Enzima; (meta-) genoma.

## ABSTRACT

### *Post-genomic era for the development of new biocatalysts*

Currently, modern molecular biology experiences a rise of '-omics' approaches, and in particular, genomics and (meta-) genomics as emerging tools to study gene coding enzymes at the level of single microorganisms and communities of uncultured and, so far, unexplored microorganisms. During the last few years, we have seen enormous strides in our abilities to isolate microbes and analyse single genomes, and the information that has poured out of these sequences is quite astonishing with respect to decipher new enzymes with new activities. Molecular approaches, in particular (meta-) genomics, are widely applied to characterize new biocatalysts. (Meta-) genomics is the application of modern genomics techniques to the study of communities of microbial organisms directly in their natural environments, bypassing the need for isolation and lab cultivation of individual species. Since then, (meta-) genomics has revolutionized biotechnology by shifting focus away from single microbial isolates towards the estimated 99% of microbial species that cannot currently be cultivated. We discuss here the different approaches to isolate new enzymes and how (meta-) genomics can direct application of microbial communities and their enzymes in a broad range of biomedical, industrial and environmental applications.

**Keywords:** Biocatalyst; Biodiversity; Enzyme; (meta-) genomics.

## 1. INTRODUCCIÓN

Los biocatalizadores son la respuesta de la Naturaleza a muchos retos industriales y medioambientales. Este artículo detalla los últimos avances para la identificación y mejora de biocatalizadores, con un enfoque mayoritario en la aplicación de técnicas independientes del cultivo microbiano.

Lo que los circuitos integrados son para los ordenadores, la biotecnología lo es para las industrias química, médica y medioambiental. En la primera línea de esta revolución están las enzimas microbianas o biocatalizadores – proteínas que juegan un papel clave en las actividades metabólicas de los microorganismos, confiriéndoles la habilidad de sobrevivir en un rango de condiciones ambientales muy diversas. De hecho, las enzimas cumplen una función vital en los seres vivos, y aunque su estudio se remonta a poco más de cien años, desde hace miles de años hasta la fecha se han utilizado en la preparación de productos de interés.

La comunidad científica ha logrado grandes avances en el aislamiento y expresión de genes que codifican enzimas que facilitan reacciones químicas de una forma sostenible (1, 2); en algunos casos los biocatalizadores pueden reemplazar compuestos químicos agresivos en numerosos procesos industriales (3). Esta capacidad ha promovido una iniciativa de la comunidad científica dedicada al descubrimiento y desarrollo de enzimas.

Dado que las enzimas presentan propiedades catalíticas específicas, solo una pequeña cantidad de las mismas se requiere para llevar a cabo una conversión deseada, y los rendimientos hacia el producto a sintetizar son, a menudo, superiores a aquellos obtenidos por rutas químicas. Derivadas de fuentes renovables, son totalmente biodegradables; cualquier exceso de biomasa puede transferirse a campos agrícolas como fertilizante. Además, las enzimas trabajan bajo condiciones suaves de reacción (a menudo, temperatura ambiente y pH neutro), lo que las convierte en una alternativa eficiente energéticamente.

Desafortunadamente, muchos de los procesos industriales en los que las enzimas pueden ofrecer importantes ventajas no se realizan en condiciones suaves. Es por ello, que el descubrimiento y desarrollo de enzimas que trabajen de una forma óptima a un rango de temperaturas y pH específico y que no pierdan o alteren su actividad en presencia de compuestos químicos adicionales, es de gran interés (4). También, si una nueva e inusual enzima es descubierta, se puede diseñar un esquema de producción eficiente y económicamente viable que lleve dicha enzima al mercado.

Las enzimas se obtienen de fuentes naturales, tales como hongos o bacterias, y posteriormente pueden ser genética- o molecularmente mejoradas o evolucionadas en el laboratorio para conferir nuevas propiedades específicas. El nuevo material genético se incorpora a un huésped recombinante, tal como el hongo *Aspergillus*, las levaduras *Pichia* o *Saccharomyces* o las bacterias *Escherichia*, *Bacillus* o *Pseudomonas*, que pueden ser fácilmente escalables mediante fermentación. Las razones para usar un huésped recombinante son entre otras, garantizar la producción segura (algunos microorganismos nativos pueden producir toxinas o pueden ser patógenos), conseguir una alta pureza de la enzima producida y un menor número y cantidad de sub-productos no deseados y favorecer la economía del proceso.

## 2. BIOTECNOLOGÍA Y BIOCATALIZADORES

En el siglo XXI se tiene que hacer frente a algunos problemas clave producidos fundamentalmente por el ser humano (5). Prácticas insostenibles en la utilización de los recursos naturales y una industrialización masiva han ocasionado un aumento de la contaminación y una degradación del medioambiente, que está provocando un calentamiento global del Planeta y cambios climáticos extremos. En paralelo, existen problemas sanitarios derivados no sólo del incremento y envejecimiento de la población, sino también de los estilos de vida sedentarios y del resurgir de epidemias, entre otros. Por lo tanto, hay una necesidad urgente de nuevas tecnologías, productos y procesos que por un lado, aporten mejoras en las condiciones de vida (por ejemplo, síntesis de compuestos bioactivos para nuevos fármacos) y, por otro, cumplan criterios de sostenibilidad (Química Verde). La Biotecnología, y dentro de ésta la Biocatálisis, ofrecen múltiples soluciones técnicas en salud, medioambiente y en el desarrollo de prácticas sostenibles, mediante el empleo de biocatalizadores, proteínas globulares complejas que están plegadas formando un surco o bolsillo en el que encajan la molécula o las moléculas reactivas – el sustrato – y donde tienen lugar las reacciones (6).

Los biocatalizadores contribuyen a avances médicos, asistiendo no solo a la síntesis de nuevos fármacos de prevención y cura de enfermedades como penicilinas y cefalosporinas semisintéticas, inmunomoduladores y moléculas

ópticamente activas, sino también a la producción de nuevos biomateriales biocompatibles (7, 8). Éstos constituyen una herramienta útil en la síntesis de alimentos nutricionales (como vitaminas y triglicéridos estructurados) y funcionales (por ejemplo, prebióticos, flavonoides, esteroides y ácidos poliinsaturados) que pueden ayudar a prevenir algunas enfermedades (9). En agricultura, se están desarrollando nuevas estrategias para la prevención de enfermedades de plantas mediante la síntesis, asistida por biocatalizadores, de biopesticidas (10). Los biocatalizadores ofrecen también la posibilidad de optimizar el consumo de materias primas y de energía, así como de reducir la contaminación medioambiental mediante técnicas como la biorremediación, biodesulfuración y síntesis de biodiesel, bioetanol o hidrógeno (5). En conclusión, los biocatalizadores tienen y tendrán un enorme potencial en una sociedad industrializada y que apuesta por procesos altamente selectivos. Es por ello que el desarrollo de nuevas y mejores enzimas es un campo en auge que proporciona nuevas expectativas para químicos y biólogos. Para los químicos, nuevas moléculas combinatorias, así como nuevas reacciones químicas con una enorme diversidad estructural y funcional, que puede llevarnos a imaginar el ensamblaje de cualquier compuesto. Y para biólogos, el descubrimiento de reacciones bioquímicas inesperadas, así como de estructuras únicas.

### **3. EN BUSCA DEL BIOCATALIZADOR IDEAL**

Las ribozimas (naturales y combinatoriales y deoxiribozimas sintéticas) han demostrado ser una herramienta útil con gran potencial en la biotecnología basada en ácidos nucleicos y presentan buenas expectativas para el desarrollo de terapias moleculares, biosensores y nuevos materiales (11). Sin embargo, aunque presentan nuevas actividades (y mecanismos de reacción), su aplicación práctica en biotecnología es limitada porque en la mayoría de los casos el sustrato de las ribozimas es ARN (12).

El diseño y obtención de anticuerpos catalíticos constituye también una alternativa en biotecnología. Sin embargo, presentan eficacias catalíticas  $10^4$ - $10^5$  veces inferiores a las de las enzimas, haciendo de estas últimas la elección primera para el futuro desarrollo de la biotecnología (13). El descubrimiento de nuevas y mejores enzimas, así como la ingeniería racional del catalizador para potenciar una

determinada aplicación, son claves para un futuro sostenible (3). No obstante, hay numerosos aspectos que deben ser considerados y que han estimulado diferentes áreas de investigación.

Así, muchos procesos requieren enzimas estables y activas en disolventes orgánicos o en sistemas con bajo contenido en agua (reacciones en sistemas bi- o polifásicos para facilitar la separación de productos) (14). Sin embargo, la baja estabilidad y la disminución de las velocidades de reacción limitan la aplicación extendida de los biocatalizadores en medios no acuosos. Si bien la utilización de líquidos iónicos, fluidos supercríticos y medios micelares se ha planteado como alternativa, la inactivación de las enzimas por falta de control del pH, la posible baja estabilidad en presencia de detergentes y los cambios estructurales inducidos por los cambios de presión, son inconvenientes que limitan la aplicación de los mismos. Es por ello que se necesitan nuevos biocatalizadores activos y estables en medios casi anhidros. Quizás, será importante analizar, mediante métodos computacionales y de dinámica molecular *in silico*, los efectos indeseados del disolvente sobre la enzima (15).

La aplicación práctica de las enzimas requiere también el desarrollo de estrategias para aumentar no solo su estabilidad y reutilización sino también ampliar su especificidad de sustrato (promiscuidad catalítica) (8, 16). Para ello existen en la actualidad técnicas de ingeniería enzimática (inmovilización, modificación química, diseño de enzimas (semi) sintéticas) y genética (evolución dirigida, mutagénesis dirigida, barajado de ADN, mutagénesis al azar y permutación circular) que permiten el diseño de novedosos procesos enzimáticos partiendo de biocatalizadores conocidos (3). Pese a ello, existe una necesidad creciente de desarrollar nuevos biocatalizadores (17).

#### **4. TÉCNICAS PARA EL DESCUBRIMIENTO DE BIOCATALIZADORES**

El primer paso para el descubrimiento de una enzima es determinar las condiciones de operación industrial de la misma, incluida su estabilidad en la disolución industrial y la identidad de los constituyentes que pueden inhibir su actividad. Por tanto, un mayor conocimiento del proceso puede ayudar en el proceso

de selección del mejor método de tamizado para el desarrollo de nuevos biocatalizadores (18). Preguntas importantes a plantearse en el proceso son, por ejemplo: ¿Es la primera vez que se emplean enzimas para la aplicación en particular para la cual se busca? ¿Es necesario diseñar una nueva enzima para llevar a cabo esta reacción, o existen accesibles enzimas con capacidades similares? y ¿La nueva enzima incrementará el rendimiento de un proceso existente? En la actualidad se pueden aplicar una variedad de métodos para el descubrimiento de enzimas, dependiendo del objetivo final:

- Tamizado clásico de microorganismos usando una colección establecida de cepas microbianas.
- Un acceso integrado que involucre tamizados moleculares de microorganismos aislados y muestras complejas medioambientales; los tamizados moleculares emplean herramientas genómicas, proteómicas y metabolómicas, acopladas con análisis bioinformáticos y computacionales.
- Evolución molecular mediante ingeniería iterativa de proteínas.

## 5. TAMIZADO CLÁSICO DE COLECCIONES DE MICROORGANISMOS

El tamizado clásico se basa en el cultivo de una amplia diversidad de microorganismos que se encuentran en las colecciones privadas o públicas de microorganismos. La diversidad viene reflejada no solo por la taxonomía, sino también por la variedad fisiológica, bioquímica y ecológica de los microorganismos. Aunque el tamizado es un juego de números (el éxito en la búsqueda dependerá del número de organismos a rastrear), nos permite identificar genes de interés si se sabe cómo y dónde buscar (19, 20). De esta forma, el criterio que un biocatalizador debe reunir ha de estar reflejado en la librería de microorganismos a rastrear. Por ejemplo, enzimas para uso en detergentes deben mostrar rendimiento óptimo a valores de pH altos y a temperaturas entre 20 y 50°C. Por otra parte, enzimas para alimentación deben trabajar a pH neutro y 37°C (por ejemplo, la temperatura del tracto digestivo). Además, las enzimas de alimentación deben mostrar una estabilidad extrema a pH bajos, dado que tiene que actuar en las condiciones drásticas que existen en el estómago. Las enzimas secretadas, aquellas que se

producen extracelularmente por microorganismos, tienen la mayor probabilidad de actuar en condiciones extremas muy diversas (4). Al contrario que las enzimas intracelulares, que son estabilizadas en el citoplasma, las enzimas secretadas por los microorganismos han sido optimizadas para trabajar en ambientes muy diversos por selección natural, dado que han evolucionado durante miles de años. Ejemplos de fuentes microbianas para aislar enzimas extracelulares son los lagos salados con valores de pH muy altos, o fuentes ácidas con valores muy bajos de pH, áreas geotermales con temperaturas cercanas a ebullición, o regiones polares (21).

El desafío del tamizado de microorganismos es la diversidad infinita de formas de vida (19, 22). En este punto hay que mencionar que se han identificado hasta la fecha cerca de 7500 especies microbianas que están representadas en varios centenares de miles de microorganismos, los cuales están accesibles en laboratorios y colecciones privadas y públicas (23, 24). Sin embargo, una vez que el método de tamizado ha sido diseñado, los microorganismos (independientemente de su número) pueden ser sometidos a un tamizado inicial para encontrar el mayor número posible de dianas positivas - microorganismos o enzimas que encajen con las condiciones de aplicación. Esto a menudo resulta en un medio de cultivo que contiene cientos o miles de enzimas (Figura 1, panel izquierdo). En un tamizado posterior, se aplican ensayos más restrictivos a los medios de cultivo.



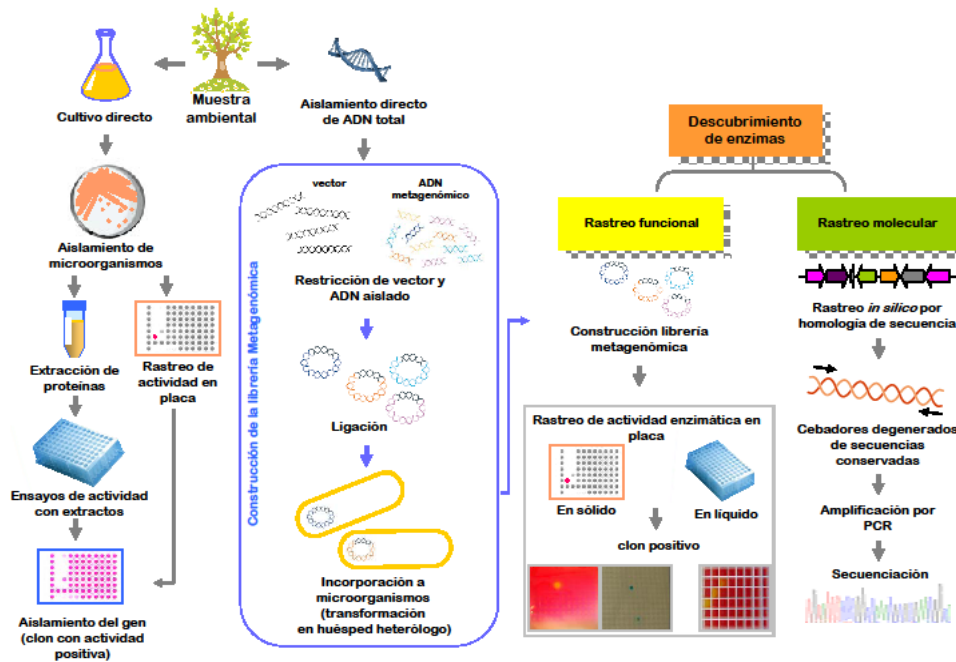


Figura 1.- Técnicas de tamizado que permiten acceder a nuevas enzimas de interés biotecnológico presentes en la biodiversidad microbiana.

El tamizado de una librería de microorganismos puede realizarse en placas de agar sólido o en micro-placas, dependiendo del ensayo a realizar. Los ensayos líquidos en formato de placas multipocillo ofrecen un amplio rango de posibilidades, especialmente porque el tamizado es independiente del crecimiento del microorganismo correspondiente (25-29). Normalmente, una librería se transfiere desde una placa de agar sólido a una placa multipocillo mediante el empleo de sistemas robotizados (30). Estas librerías se incuban para crecer una placa madre y parte de este cultivo se transfiere (preferiblemente mediante sistemas robotizados) a una placa de tamizado para finalmente identificar el organismo activo en una librería hecha de cientos o miles de ellos (29). Los organismos individuales que se encuentren activos en el tamizado son posteriormente analizados, y las enzimas de interés directamente purificadas. Alternativamente, el ADN microbiano del organismo en cuestión puede ser aislado y el gen que codifica para la enzima deseada se aísla de la mezcla y se inserta en un huésped adecuado. La librería (una secuencia de nucleótidos en un gen) se rastrea nuevamente usando el mismo ensayo que se aplicó en el tamizado original.

Los tamizados funcionales, basados en la identificación de microorganismos dentro de una colección que son capaces de transformar un sustrato deseado mediante el análisis de caldos de cultivo, son a menudo complementados con técnicas de tamizado molecular que permiten identificar genes de interés por similitud con secuencias de genes que codifican enzimas conocidas (31, 32). En otras palabras, la información de secuencias de ADN de un conjunto de enzimas relacionadas se usa para rastrear y clonar genes adicionales de otros microorganismos. Un alineamiento de las secuencias de nucleótidos o de las secuencias correspondientes de amino ácidos, seguido de la identificación de regiones conservadas evolutivamente, hace posible generar reacciones degeneradas en cadena de polimerasa (PCR) que contengan las secuencias de ADN correspondientes a regiones homólogas. En este caso, simplemente hay que aislar el ADN del organismo o de los organismos a investigar.

La reacción en cadena de polimerasa (PCR) es una técnica ampliamente usada en biotecnología para amplificar trazas de ADN (33). El primer paso para la reacción de PCR es sintetizar “oligo-nucleótidos” que son diseñados para complementar las regiones homólogas del gen de interés. Mediante el solapamiento de oligonucleótidos específicos al ADN molde y realizando múltiples ciclos de replicación dirigidos por los oligonucleótidos, la secuencia de ADN deseada puede ser amplificada millones de veces en varias horas. La limitación de este método es que se generan solo variantes de enzimas conocidas, pero no totalmente nuevas (34). La ventaja de este método es que no se basa en la expresión del gen en cuestión (tanto en el organismo donador como el receptor) y no requiere ensayos funcionales para detectar la actividad de la enzima. Este método se puede usar para rastrear librerías medioambientales, dado que no es un prerrequisito disponer de organismos vivos (este punto se discutirá posteriormente). Usando este método, comenzando con solo media docena de genes homólogos se pueden crear cientos de biocatalizadores de interés industrial (35).

Las técnicas de tamizado descritas anteriormente pueden ayudar a identificar candidatos de biocatalizadores que actúan de forma correcta sobre la aplicación en cuestión. Sin embargo, si la actuación de la enzima no es aún correcta, se deben

considerar otras alternativas: iniciar nuevos ciclos de tamizado basados en los resultados obtenidos en etapas anteriores o realizar programas intensos de optimización enzimática, por ejemplo, evolución molecular que involucra técnicas de ingeniería de proteínas, tales como mutagénesis al azar o recombinación homóloga *in vitro* de genes mutantes seleccionados procedentes de la fragmentación de ADN al azar (34).

Se han descrito métodos de tamizado alternativo en los cuales la librería de genes procedente de un organismo en cuestión se rastrea funcionalmente mediante el empleo de ensayos bioquímicos (36). Esto puede realizarse mediante inmuno-tamizado, que se basa en crear anticuerpos contra todas las proteínas secretadas en un cultivo libre de un microbio dado. Entonces, se genera una librería de genes del mismo organismo, usando un sistema de expresión que dirige la expresión de proteínas. Un vector es una molécula de ADN que replica en una célula huésped. Si una pequeña pieza de ADN foráneo se incorpora a un vector, éste replicará en la célula huésped. La librería de expresión se rastrea con los anticuerpos para selectivamente identificar clones que expresen las proteínas de interés (36). La secuenciación posterior de los clones identificados y la comparación de las mismas en bases de datos permite asignar actividades proteicas. La ventaja de este método es la rápida producción de todo un conjunto de clones de expresión sin la necesidad de usar métodos de tamizado funcional que solo permite la detección de la actividad enzimática deseada.

## **6. BIOCATALIZADORES Y CULTIVO DIRECTO DE MICROORGANISMOS**

Pese a que se dispone en estos momentos de colecciones importantes de microorganismos, se sabe que el Planeta es una fuente inagotable de nuevos organismos. Como manifestara el investigador estadounidense Stephen Jay Gould (1941-2002): "...en la Tierra no vivimos en la Era del Hombre, vivimos hoy, y siempre, en la Era de los microorganismos". Los microorganismos son, y siempre han sido, la forma de vida predominante en la Tierra. Pese a que fueron descubiertos hace unos 300 años por van Leeuwenhoek y que la mayor parte de la investigación microbiológica se ha producido en los últimos 100 años, iniciada con Louis Pasteur y Robert Koch, líderes en el desarrollo de la disciplina, se cree que

fueron los primeros habitantes del planeta, hace 3.500 millones de años. Esto significa que son 2.000 millones de años más antiguos que los organismos eucariotas, y tienen una tasa de reproducción mucho más rápida. Los procariotas pueden vivir, y sus enzimas trabajar, en ambientes muy diversos (4). La piel humana, por ejemplo, alberga a unos 100 mil microbios por  $\text{cm}^2$ , hay microorganismos que viven en volcanes a  $250^\circ\text{C}$ , otros en ambientes ácidos (pH entre 1 y 2), otros a profundidades entre 5 y 10 Km. de la superficie terrestre soportando presiones enormes (ver ejemplos en ref. 37). Son fundamentales para el planeta porque fabrican y mantienen la composición de la atmósfera y del suelo, convierten elementos clave como carbón, nitrógeno y azufre en formas accesibles para otros seres vivos, y contribuyen a formar la mayor biomasa del planeta. Como consecuencia de esas interacciones, se producen (macro) moléculas con actividades biológicas variadas, que juegan un papel importante en su supervivencia.

Todo esto convierte a los microorganismos en el foco de importantes proyectos de investigación. Incluso, los científicos destacan la importancia del estudio no sólo de los microorganismos como individuos sino de las comunidades de microbios, las cuales desarrollan tareas complejas esenciales para la salud de los seres humanos y la sustentación del planeta (37). El objetivo principal de este tipo de investigaciones, en lo que al tema de desarrollo de biocatalizadores se refiere, se centra en obtener una colección de microorganismos capaces de metabolizar un determinado sustrato y determinar posteriormente las enzimas responsables de tal transformación. Para ello se proponen diferentes programas de aislamiento y selección de microorganismos a partir del proceso de transformación de diferentes moléculas (38) (Figura 1, panel izquierdo).

Para lograr esta premisa mediante el aislamiento y selección de los microorganismos más adecuados, los científicos siembran los organismos en un medio de cultivo junto con los nutrientes básicos y, como única fuente de carbono, los sustratos sobre los que se pretende actuar (39). A continuación, se determina la concentración inicial de estos elementos y tras varios días de incubación se mide la concentración final. De esta forma, los expertos pueden comprobar la capacidad real

de los microorganismos presentes en una muestra para transformar los compuestos objeto de análisis. Este procedimiento se aplica por ejemplo para la selección de organismos capaces de eliminar metales pesados o para la búsqueda de organismos capaces de degradar los compuestos (ligno-) celulósicos para la identificación de microorganismos con exo-enzimas que intervengan en su degradación (40). Una vez identificado el organismo u organismos portadores de un biocatalizador de interés, se puede aplicar cualquiera de las técnicas descritas en el punto anterior para la localización y expresión del gen y la producción del biocatalizador.

El estudio de los microorganismos y sus funciones ha sido una fuente inagotable para el desarrollo de la biotecnología y su industria (3, 37). Diferentes nichos microbianos han sido la fuente de no solo moléculas con aplicaciones terapéuticas como compuestos antimicrobianos, anticancerígenos, antidepresores, sino también de organismos cultivables con aplicaciones industriales en diversos campos. A pesar de todas las herramientas de que disponen actualmente los microbiólogos y de todo el conocimiento generado desde la aplicación de técnicas de biología molecular, nos puede parecer que la complejidad de interrogantes que tiene ante sí la biotecnología es considerable. Así, hoy en día sabemos que toda la información generada en estos aproximadamente veinticinco años de aplicación de la biología molecular a la microbiología es tan sólo una pequeña parte de la que se está obteniendo desde hace unos pocos años con la aplicación de técnicas de secuenciación y análisis de genomas. No obstante, hoy es bien conocido que un simple gramo de suelo puede contener entre 2000 y 8.300.000 diferentes genomas (37), donde únicamente alrededor del 1% de éstos son accesibles con técnicas dependientes de cultivo. Es más, se estima que sólo un 10% de las especies han sido descritas, y que el 90% restante representa un enorme reservorio biológico y genético (41-44). Las limitaciones para conocer, y por lo tanto acceder a, la gran diversidad son la dependencia de medios de cultivo y la falta de conocimiento de las interacciones microbianas que nos imposibilita el cultivo de todos los microorganismos presentes en una muestra.

Es claro entonces que existe una gran cantidad de información con un potencial de aplicaciones biotecnológicas dentro de la porción no cultivable de microorganismos. Es por esto que aproximaciones como la (meta-) genómica representan una oportunidad para dilucidar más ampliamente el potencial biotecnológico inmerso en las comunidades microbianas (45). La (meta-) genómica supone la extracción, clonación, secuenciación y análisis del genoma de una comunidad microbiana, lo cual permite el estudio de una gran variedad de genes y sus productos, además de un conjunto de operones que codifican rutas de degradación o de biosíntesis. En este contexto, la (meta-) genómica es un campo nuevo de investigación que se ha desarrollado en la última década, con el objeto de entender la diversidad de diferentes nichos ecológicos conformados por microorganismos cultivables y no cultivables. Su principal objetivo ha estado centrado en la comprensión de los procesos ecológicos relacionados con la actividad microbiana en el planeta, tales como su participación en los ciclos biogeoquímicos, la diversidad ambiental, la evolución de poblaciones, la respuesta del ecosistema a los cambios ambientales y la influencia de los microorganismos en otros seres vivos y en el ambiente, entre otros (37). Así mismo, este campo se ha desarrollado gracias a la necesidad de identificar nuevos biocatalizadores con aplicación industrial, que cumplan funciones únicas, o bajo ciertas condiciones que sean de interés para su aplicación industrial (Beloqui et al., 2008).

## **7. BIOCATALIZADORES Y (META-) GENÓMICA**

Del estudio del ADN de comunidades de microorganismos, se ocupa la (meta-) genómica, un campo nuevo en el que se busca obtener y analizar secuencias del genoma de los diferentes microorganismos que componen una comunidad (46). Idealmente, los métodos independientes del cultivo permiten a los microbiólogos explotar el potencial biológico de comunidades microbianas completas. La (meta-) genómica, o clonación desde microorganismos no cultivables (aquellos que no crecen normalmente bajo condiciones estándar de laboratorio), es una estrategia alternativa para el aislamiento de biocatalizadores (17, 47, 48). La estrategia ha proporcionado evidencias de que las fuentes genéticas contienen millones de genes no caracterizados hasta la fecha (42, 47-49). La particularidad de esta técnica

radica en el hecho de que no se basa en el cultivo de microorganismos para obtener el material genético, sino en la extracción de ADN o ARN mensajero (mARN) de todo tipo de organismos presentes en una muestra ambiental (Figura 1). Las células de estos microorganismos se rompen mediante métodos químicos (condiciones alcalinas) o físicos (sonicación) y el material genético posteriormente aislado.

Una vez que el ADN está libre, éste es separado del resto de la muestra usando centrifugación diferencial, afinidad o precipitación (34). Enzimas de restricción se emplean posteriormente para fragmentar el ADN (ambas cadenas) en una secuencia particular de fragmentos más pequeños. Estos fragmentos son combinados con vectores – pequeñas unidades de ADN que pueden ser insertados en células para replicar y producir las proteínas codificadas en el ADN. Los vectores también contienen un marcador selectivo que proporciona al organismo modelo una ventaja de crecimiento que normalmente no tiene, por ejemplo, una resistencia a un antibiótico particular.

Posteriormente, los vectores que contienen el ADN (meta-) genómico se introducen en un organismo modelo, típicamente *E. coli*, que crece en un medio selectivo, de tal forma que solo las células que contiene el vector sobreviven, formando una colonia. La muestra de células que contiene todo el ADN (meta-) genómico en vectores se denomina librería (meta-) genómica y es la que típicamente se emplea para la búsqueda de nuevas formas de genes (des) conocidos. El paso final en el procedimiento es el análisis de ADN de la librería (meta-) genómica. Esta librería de genes puede ser rastreada empleando tanto métodos de PCR basados en homología de secuencias de ADN como ensayos funcionales, como se ha comentado en el punto anterior (Figura 1, panel derecho).

Sin embargo, la complejidad de una librería de ADN (meta-) genómico plantea retos técnicos. Estos incluyen: la calidad del ADN medioambiental, normalización de la librería para evitar el tamizado repetitivo de genes muy abundantes y ampliamente distribuidos; el tamaño de la librería (meta-) genómica, que a menudo requiere equipos de tamizado de alto rendimiento; la aplicabilidad de la técnica para organismos eucariotas, que indudablemente se encuentran en el medio-ambiente; la aprobación del producto y la producción, la cual, en muchos países requiere una

fuente microbiana conocida y definida (por ejemplo, una bacteria o un hongo); y el potencial del gen o de los genes detectados para ser expresado correctamente en un organismo huésped productor (46).

La construcción y tamizado de librerías (meta-) genómicas está todavía en su infancia, y la eficiencia de tamizado es baja en comparación con el coste y la complejidad técnica (48) del método en algunos casos. Solamente el tiempo nos dirá si el valor de esta tecnología residirá en estudios ecológicos o se extenderá a aplicaciones industriales, para el desarrollo de nuevos biocatalizadores y nuevos productos y procesos alrededor de los mismos. Independientemente de ello, la iniciativa de concretar proyectos metagenómicos requiere de la colaboración entre múltiples centros de investigación en todo el mundo, de la mejora de las tecnologías de secuenciación de genes (50) y el aprovechamiento de las herramientas más actualizadas de bioinformática (43, 51-56). El avance tecnológico es un requisito para la secuenciación de (meta-) genomas ya que los secuenciadores que se venían utilizando, incluso con el proyecto Genoma Humano (<http://nihroadmap.nih.gov/hmp/>), no resultan totalmente poderosos para secuenciar (meta-) genomas en corto tiempo (57).

## **8. HISTORIA DE LA (META-) GENÓMICA EN RELACIÓN A OTRAS “-ÓMICAS”**

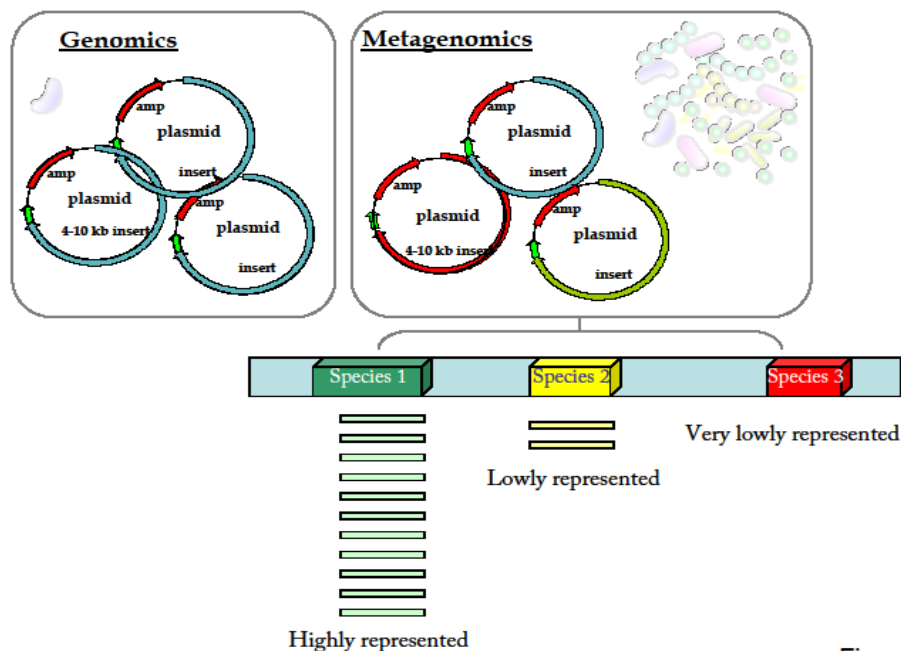
El primer término “-ómico” introducido fue “genómica” en 1998. Se estima que una bacteria produce intracelularmente unas 750-1.500 proteínas y excreta una media de 150-180 (58), mientras que el genoma más grande de un hongo filamentoso se estima que produce hasta 3.000 proteínas y excreta 300-400. Asumiendo que el tamaño medio de una bacteria es de 4 megabases (MB) (59) y que para un hongo filamentoso es de 30-40 MB (60), solo el 2%-5% de los marcos de lectura abierta (ORF) en un genoma completo es de interés primario en lo que a la presencia de biocatalizadores de interés se refiere (44, 50, 61, 62).

Como resultado de la baja proporción de enzimas identificadas en los análisis genómicos, se desarrollaron métodos alternativos para rastrear selectivamente nuevos biocatalizadores (34, 40). Así, un año después de la acuñación del término



genómico, el concepto “(meta-) genómica” empezó a usarse, definido como la aplicación de técnicas genómicas modernas al estudio de comunidades de microorganismos directamente en su medio natural, evitando la necesidad de aislar y crecer las especies individuales (37). Esta técnica tiene sus orígenes en la recuperación independiente de cultivo de genes 16S rARN, iniciada por Pace y colaboradores hace más de dos décadas (63). Desde entonces, la (meta-) genómica ha revolucionado la microbiología de tal forma que se ha pasado de investigar aislados microbianos individuales hacia el estudio estimado del 99% de especies microbianas que se estima no pueden ser cultivados (37). Mientras que el genoma representa el material genético (ADN) de un organismo individual, el (meta-) genoma representa el ADN de una comunidad entera de organismos (Figura 2). Sin embargo, el nivel de complejidad es extremadamente alto en comparación con el estudio de genomas individuales. Una de las razones principales son que cada comunidad microbiana puede contener 10, 100, 1.000, 10.000 organismos y, lo que es peor, un número desconocido de diferentes especies desconocidas. Además, puede existir una amplia variación en la abundancia de cada especie: algunas especies pueden representar más del 10% de la comunidad mientras que otras pueden solo representar menos del 0,1%.

En la actualidad, se encuentran disponibles varios miles de genomas y (meta-) genomas para investigación (<http://www.genomeonline.org>) (64). Por ejemplo, en 2007, la expedición “Global Ocean Survey” publicó el análisis de 41 (meta-) genomas (65), y en octubre de 2008,



**Figura 2.- Representación esquemática de las diferencias entre genómica y (meta-) genómica.**

Mientras la primera analiza el material genético de un organismo individual, la (meta-) genómica lo hace del conjunto de una mezcla organismos que se encuentran en diferentes proporciones.

El número de (meta-) genomas accesibles en las bases de datos públicas superó los 1300 (42). Hemos entrado en la era de los denominados “proyectos (meta-) genómicos” que incluyen, por ejemplo, el proyecto para la creación de la Enciclopedia Genómica de bacteria y Arqueas (GEBA) (<http://www.jgi.doe.gov/programs/GEBA/>) y el proyecto del Microbioma Humano (<http://nihroadmap.nih.gov/hmp/>), con otros proyectos visionarios en el horizonte.

La “proteómica” (el estudio de la producción y función de las proteínas totales de una célula siguió a la (meta-) genómica algo más tarde, en 1997. Las investigaciones en el área de proteómica crecieron rápidamente sobrepasando el número de publicaciones científicas relacionadas con la genómica en 2009, alcanzando un plató de 3000 publicaciones al año (37). Aunque las técnicas de aislamiento de proteínas se han aplicado tradicionalmente a los cultivos de

microorganismos puros, los últimos avances en la proteómica de alto rendimiento han proporcionado nuevas herramientas par obtener una visión integrada de mezclas complejas de proteínas. En este contexto, se introdujo el término (meta-) proteogenómica o (meta-) proteómica en 2004 para definir el estudio de mezclas complejas de proteínas en una muestra ambiental (66, 67). Técnicas como la electroforesis en dos dimensiones (2-DE) acoplada a la espectrometría de masas puede ser utilizada para separar e identificar cientos de proteínas de comunidades microbianas con más de 50 especies y es posible incluso identificar proteínas de una mezcla de bacterias con genomas sin secuenciar (68, 69). La (meta-) proteómica se ha empleado para identificar nuevos biocatalizadores (70), por ejemplo, involucrados en la degradación de biomasa (71, 72). A diferencia de las técnicas genómicas que identifica genes inducidos en un microorganismo frente a señales específicas, el proteoma analiza todos los genes que se expresan como proteínas.

El término “transcriptómica” siguió en 1999 y se utiliza para definir el seguimiento de los cambios globales en la expresión de genes (ARN). Por el contrario que la genómica y proteómica, los estudios de transcriptómica no se hicieron populares inicialmente debido a los problemas técnicos inherentes a la técnica, fundamentalmente que los microarrays (usados para detectar los niveles de los transcritos de ARN) son específicos de cada organismo, y por lo tanto, la secuencia del genoma del organismo en cuestión debe ser conocido (37). En estos momentos, el número de estudios se está incrementando hasta un número de 200 publicaciones al año (73-75), debido principalmente a la introducción de los microarrays de ADN ambiental que permiten medir el nivel de expresión de miles de genes al mismo tiempo, lo que se denomina (meta-) transcriptoma. También la introducción de las técnicas de secuenciación masiva de ARN, que serán discutidas posteriormente a lo largo de este capítulo, que han sufrido un auge en los últimos 5 años, han contribuido a la expansión de estas técnicas: la secuenciación directa del (meta-) transcriptoma permite estudiar el análisis de expresión génica de una comunidad sin conocimiento previo de su secuencia de ADN, el (meta-) genoma (76, 77).

Un micro-array de ADN contiene miles de segmentos de ADN adheridos a una superficie sólida, como por ejemplo vidrio, plástico o chips de silicona. La tecnología es un método popular para medir simultáneamente los niveles de expresión de un amplio número de genes (78) y puede ser también aplicada a programas de tamizado de enzimas para detectar genes inducidos bajo condiciones fisiológicas específicas (79). Por ejemplo, cuando un microorganismo se hace crecer en un sustrato específico, los genes involucrados en la utilización de este sustrato se inducen. Esta expresión diferencial permite a los científicos detectar genes que puedan degradar este sustrato. Posteriormente estos genes se expresan y testan en aplicaciones industriales (80). Los micro-arrays de ADN genómico aparecen como un método viable para identificar nuevas enzimas involucradas en la degradación de sustratos complejos, como nuevas peroxidasas que se expresan cuando el huésped se crece en pulpa tratada térmicamente (79).

Pese a que la identificación y cuantificación de metabolitos se desarrolló mucho antes usando técnicas de cromatografía de capa fina o cromatografía líquida o de gases para estudiar los extractos de plantas medicinales o aceites esenciales, no fue hasta 1998, que se introdujo el término “metaboloma” en un artículo relacionado con el genoma de la levadura *S. cerevisiae* (81). Dos años más tarde el término “metabolómica” se empezó a usar en analogía con las otras técnicas “-ómicas” desarrolladas hasta el momento. El número de publicaciones relacionadas con la metabolómica (o metabonómica) muestra un menor crecimiento que la genómica o proteómica, con aproximadamente unas 600 publicaciones en 2008 (37). Esta baja producción de debe al hecho de que no se han definido métodos de análisis definidos para el análisis de metabolomas, a diferencia que la genómica o proteómica. Avances en el campo del análisis de metabolitos incluyen mejoras en técnicas cromatográficas, incremento de la resolución en espectrometría de masas y resonancia magnética nuclear, diseño de técnicas bidimensionales y la identificación directa de metabolitos en extractos crudos, cuya presencia o ausencia está directamente ligada a la presencia o ausencia de la enzima correspondiente (82).

## 9. COMO SE TRABAJA EN (META-) GENÓMICA

La (meta-) genómica incluye una amplia variedad de técnicas y estrategias de trabajo. La mayoría de éstas constan de algunos primeros pasos comunes que se detallan a continuación.

En primer lugar los investigadores obtienen una muestra de un ambiente particular (suelo, agua de mar, la boca de un ser humano, etc.) y se realiza una extracción de ADN (o ARN) de todos los microbios presentes en la muestra. La mayoría de los proyectos de (meta-) genómica se enfocan en los microbios que tienen pequeñas cantidades de ADN, como bacterias o Arqueas (identificadas como una división mayor de organismos en 1977; pueden vivir en ambientes extremos y por lo tanto son portadores de biocatalizadores estables y activos en condiciones extremas) (4).

Una vez que el ADN es extraído se induce una replicación del mismo en un microorganismo para crear una “biblioteca” la cual contiene porciones de genomas de todos los microorganismos de la muestra o se lo secuencian directamente para estudios comparativos o para búsquedas de genes en particular (46), evitando la necesidad de crear una biblioteca. La secuencia no contiene genomas separados de cada especie, sino que consiste en una mezcla de millones de fragmentos de ADN al azar de todos los microorganismos muestreados de una comunidad.

Posteriormente, el tratamiento de la biblioteca de genomas ((meta-) genoma), depende del objetivo de búsqueda. En proyectos de (meta-) genómica basado en secuencias, los investigadores se focalizan en encontrar la secuencia genética completa -el patrón de bases nitrogenadas (A, C, G, T) en las cadenas de ADN – de los microorganismos descubiertos en la muestra. La secuencia puede entonces ser analizada de diferentes formas. Por ejemplo, los investigadores pueden usar la secuencia de una comunidad para determinar el genoma completo de especies microbianas individuales. Alternativamente, los investigadores podrían usar la secuencia para analizar el genoma de una comunidad como un todo, lo cual ofrece información sobre la ecología de la población y la evolución o la búsqueda de genes que codifican enzimas de interés.

Los proyectos de (meta-) genómica basados en la función, exploran los productos que producen los microorganismos en una comunidad natural. De esta forma es posible extraer directamente ADN a partir de una comunidad microbiana e identificar nuevas proteínas y metabolitos. A través de estos métodos los científicos pueden identificar funciones desconocidas hasta hoy (debido a que previamente el estudio se basaba sólo en microorganismos cultivados en laboratorio) (83-86).

## 10. BÚSQUEDA DE BIOCATALIZADORES MEDIANTE ANÁLISIS COMPUTACIONAL DE (META-) GENOMAS

Los marcos de lectura abierta (ORF) se encuentran normalmente a lo largo de la cadena de ADN cuando se trata de localizar un gen que codifica una proteína (86). Un ORF es un conjunto de tripletes de nucleótidos, cada uno de los cuales codifica un amino ácido que se encuentra entre un codón de iniciación que consta de tres nucleótidos (normalmente ATG), y un codón de terminación (normalmente TAA, TAG o TGA). Sin embargo, esta secuencia no contiene una señal de parada de la traslación antes de crear una proteína completa. Cuando se identifica un nuevo gen y su secuencia de ADN se descifra, no está claro cual es su secuencia proteica correspondiente. Esto se debe a que en ausencia de información adicional, la secuencia de ADN puede ser leída en seis marcos de lectura posible (tres por cada cadena, correspondiente a las tres posiciones de iniciación para el primer codón). En este sentido, un reto primario de la (meta-) genómica es establecer una conexión entre una secuencia génica y una función y una estructura (87, 88). El desarrollo y la automatización de las tecnologías de secuenciación, así como los avances en bioinformática, ha permitido secuenciar comunidades enteras en periodos de tiempo muy cortos y a un precio razonable, generando giga-bases de nuevas secuencias (89, 90). De esta forma se dispone de una colección anotada de secuencias de nucleótidos y amino ácidos accesibles en las bases de datos: más de 190 billones de bases están depositadas actualmente y este número se duplica cada 18 meses. Por una parte, esto abre una oportunidad sin precedentes para la búsqueda de nuevos biocatalizadores en esta “caja negra” de nuevas secuencias que pueden obtenerse mediante la aplicación de diferentes métodos de secuenciación (**Figura**

3); por otra, el alto número y variabilidad de secuencias levanta una serie de problemas de procesado y lleva a las técnicas bioinformáticas al límite.

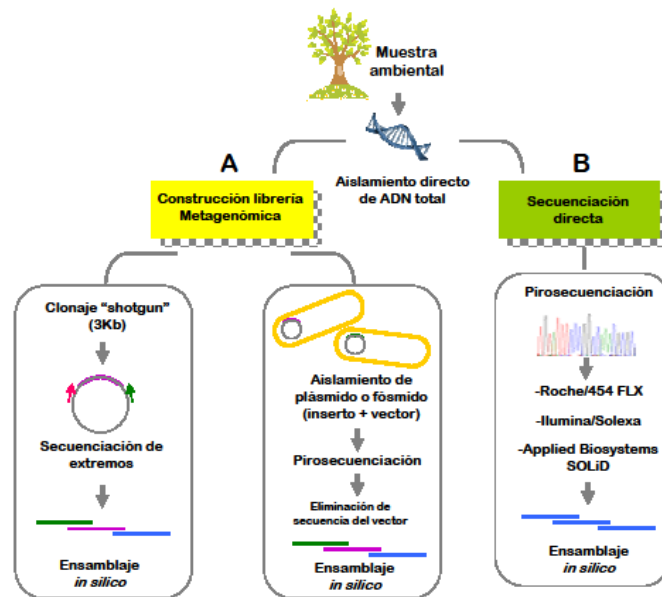


Figura 3.- Técnicas de secuenciación del ADN (meta-) genómico. (A) Dos enfoques diferentes para secuenciar clones de una librería (meta-) genómica. (B) Secuenciación de ADN aislado directamente de una muestra medioambiental.

En la actualidad las técnicas empleadas para el análisis de datos (meta-) genómicos se basan en las herramientas computacionales inicialmente diseñadas para el análisis de genomas microbianos (50). La diferencia radica en que mientras que actualmente los análisis computacionales generalmente dedican días o semanas por ordenador para el análisis completo de un genoma sencillo por comparación con bases de datos de proteínas no redundantes (una colección de proteínas de todos los organismos secuenciados hasta la fecha), el análisis de un (meta-) genoma requiere varios años si se utiliza un solo ordenador (91). Es por ello que uno de los principales cuellos de botella es la falta de algoritmos y de software para el procesamiento de un gran número de datos. Esto es de gran importancia dado que la cantidad y distribución de especies y, por tanto, de biocatalizadores

afecta el muestreo y que normalmente se sacrifica un muestreo extenso obtenido por piro-secuenciación por otro de más precisión para describir los microorganismos y biocatalizadores más abundante en una muestra (52).

Para resolver esto, la mayoría de estudios de (meta-) genómica se han centrado en comunidades con bajo número de miembros como minas ácidas que contienen menos de 10 especies (68). Otros ecosistemas actualmente investigados incluyen el análisis de comunidades microbianas y biocatalizadores en intestino de termitas para la búsqueda de biocatalizadores útiles en la producción de energía (37). La biorremediación es también un aspecto de interés desde el punto de vista de desarrollo de biocatalizadores dado que involucra la exposición a una mezcla de moléculas con estructuras químicas muy diversa a una red intrincada de múltiples especies en un escenario contaminado (92, 93).

El enfoque principal de la biología computacional es descifrar y entender la biología de los (meta-) genomas. Éstos contienen la información necesaria para el desarrollo y evolución de organismos vivos. Aunque esta información está incluida en un código de cuatro letras, ésta representa diferentes niveles de complejidad. En la actualidad el uso combinado de métodos experimentales y computacionales tiene como objetivo responder a cuestiones moleculares, evolutivas y aplicadas específicas mediante el uso de técnicas bioinformáticas, una disciplina multidisciplinar que aúna la biología, las matemáticas y ciencias de la computación (53). Existen en la actualidad una gran variedad de herramientas bioinformáticas y páginas webs accesibles para aplicaciones en (meta-) genómica. Éstas incluyen un amplio rango de herramientas para el ensamblado y asignación de (meta-) secuencias (91) pero también para la anotación y análisis automático de fragmentos de secuencias y la clasificación filogenética y funcional (37, 84-86) de millones de fragmentos de secuencias (meta-) genómicas (94-96). Estos análisis se facilitan gracias a la creación de un amplio número de bases de datos públicas que recogen la mayoría de los proyectos de (meta-) genómica.

Ejemplos de herramientas web accesibles y públicas para anotación y análisis automático son IMG/M (98), CAMERA (99), MG-RAST (42, 100), RAMMCAP (55), WebCARMA (85) y Sort-ITEMS (86). Además, se han desarrollado una serie de



herramientas (por ejemplo, TETRA) específicamente diseñadas para el análisis de (meta-) genomas basados en la composición de oligonucleótidos de diferente tamaño o motivos (101-103).

El reto en el ensamblaje de secuencias (meta-) genómicas deriva del pequeño tamaño de las mismas y la alta complejidad de especies que la forman. Pese a ello, de forma individual las pequeñas lecturas pueden ser comparadas con otras en las bases de datos de genes (o proteínas) conocidas para identificar secuencias homólogas. Este método tiene menos especificidad en la identificación de secuencias homólogas. Se han diseñado también herramientas para mejorar la predicción de genes (104). Una de ellas es el programa llamado “Metagenomic ORFome Assembly (MetaORFA)” (105), que consta de tres etapas. Primero, cada lectura se anota con un marco de lectura abierta (ORF) putativo que codifica una proteína. A continuación, cada ORF se ensambla en una colección de péptidos usando el método de ensamblaje EULER. Finalmente, los péptidos ensamblados (denominado ORFoma) se usan para realizar un escrutinio con bases de datos.

Recientemente, se ha desarrollado un programa que permite el análisis computacional de un alto número de secuencias, denominado MEGAN (90). En éste, antes de realizar el análisis las secuencias de ADN se comparan con secuencias conocidas depositadas en las bases de datos usando BLAST (105). MEGAN se usa entonces para explorar el contenido taxonómico de cada conjunto de datos, empleando la taxonomía de NCBI para resumir y ordenar los resultados. El software permite diseccionar los resultados sin necesidad de ensamblar o buscar marcadores filogenéticos específicos.

Los métodos de (meta-) genómica comparativa se usan para descubrir las adaptaciones de microorganismos a factores ambientales muy diversos y revelar la existencia de genes y biocatalizadores específicos (37, 57). Sin embargo, muchos de los parámetros extraídos de estos datos (por ejemplo, características genómicas, funcionales y filogenéticas) son interdependientes e influenciadas por aspectos técnicos tales como la preparación de la muestra y el tratamiento de datos, entre otras, que pueden producir una interpretación errónea de los resultados. Para evitar esto, se han propuesto iniciativas para disponer de estándares mínimos para la

obtención, tratamiento y análisis (meta-) genómicos, iniciativa que se ha denominado MINIMESS (106).

Pese al desarrollo de métodos bioinformáticos, uno de los problemas más grandes es la calidad de las anotaciones (meta-) genómicas (107, 108) y el alto número de marcos de lectura que se encuentran anotados como “proteína hipotética” o “proteína hipotética conservada” (61, 62, 106). Como ejemplo, el análisis de las secuencias obtenidas de una muestra de una planta de tratamiento de aguas produjo una cantidad considerable de marcos de lectura que codificaban proteínas hipotéticas conservadas (21%) y a otro tanto no se pudieron asignar actividades (o funciones) concretas (25%) (109, 110, 111). Además, las mutaciones naturales en una secuencia de amino ácidos de una proteína codificada en un (meta-) gen puede producir una proteína muy similar a otras descritas en las bases de datos, pero con una bioquímica muy diferente. Por lo tanto, en un escenario (meta-) genómico es extremadamente difícil extraer información bioquímica de una secuencia de ADN. En este contexto, se están diseñando nuevas bases de datos como MetaBioME (112), específicamente diseñadas para facilitar la identificación computacional basada en homología de nuevas enzimas de genomas y (meta-) genomas.

## **11. (META-) GENÓMICA FUNCIONAL: BÚSQUEDA DIRIGIDA DE BIOCATALIZADORES**

Mientras que los métodos de búsqueda basados en secuencias de ADN se basan en comparación con datos de secuencias homólogas, que son obtenidas de secuencias depositadas en las bases de datos, la (meta-) genómica funcional se basa en el tamizado directo de una librería de clones de ADN que muestran un fenotipo específico, es decir, los genes son reconocidos por su actividad y no por su secuencia (113, 114). El poder de esta aproximación radica en que no se requiere que el gen de interés sea reconocido por su secuencia, asegurando que esta técnica tiene el potencial de identificar directamente nuevas clases de enzimas con funciones no solo conocidas sino totalmente novedosas (115, 116). Otra ventaja de este método es que mientras que los métodos basados en análisis de secuencias pueden resultar en una anotación incorrecta de secuencias con baja similitud a

enzimas caracterizadas bioquímicamente, los resultados de la (meta-) genómica funcional son inequívocos (117).

La estrategia básica de la (meta-) genómica funcional se muestra en la Figura 1. Se toma en primer lugar una muestra ambiental y entonces el ADN total es extraído de la muestra. Este proceso de extracción es de vital importancia si se quiere disponer de una muestra representativa (118). En la mayoría de los casos se prefiere trabajar con ADN de gran calidad y que no esté fragmentado, preferiblemente con un tamaño de entre 30 y 100 mil pares de bases. Dos métodos que se usan de forma rutinaria son aquellos que incluyen la lisis directa de la muestra mediante ciclos de congelación y descongelación (119) o un fraccionamiento indirecto donde los microorganismos se separan y purifican del resto de componentes orgánicos o inorgánicos de la muestra previamente a la lisis (120). La lisis directa es mucho más rápida y puede proporcionar un mayor rendimiento en la extracción de ADN (121). Sin embargo, este método proporciona todo el material genómico sin discriminar el origen bacteriano, arquea o eucariota. Por lo tanto existe una gran probabilidad de la presencia de información redundante en los posteriores tamizados por actividad de las bibliotecas de ADN creadas, lo que requiere tiempo, es costoso y disminuye la eficiencia del tamizado. Además, la co-extracción de compuestos indeseados (por ejemplo, ácidos húmicos del suelo) puede resultar en procesos de inhibición en etapas posteriores de clonación del ADN resultante. Para evitar este problema se requieren etapas de electroforesis para purificar el ADN, lo que disminuye el rendimiento en la extracción del mismo. Por el contrario, los métodos de fraccionamiento indirecto, requieren más tiempo de procesamiento pero evitan los problemas asociados con la interferencia de contaminantes e incrementan la eficiencia de los tamizados. La sobre-representación de material genómico de organismos mayoritarios puede ser evitada mediante el uso de gradientes de cloruro de cesio capaces de separar ADN en función del tamaño y su contenido GC (34).

En algunos casos, por ejemplo ambientes bajos en nutrientes (por ejemplo, áreas del Ártico sometidas a de-glaciación), la cantidad de ADN extraída es muy limitada. En estos casos se pueden utilizar métodos de amplificación genómica total

(WGA: Whole Genome Amplification). Otro método es la amplificación con ADN polimerasa  $\phi$ 29 (Multiple Displacement Amplification, MDA) que es capaz de incrementar la cantidad inicial de ADN a una temperatura constante. Aunque estos métodos permiten amplificar al azar el ADN genómico (MDA proporcionando productos más largos), éstos se basan en el uso de métodos de PCR proporcionando una polarización de los resultados. Es por ello que en cada caso se requiere una optimización de los procesos y etapas de extracción de ADN de suficiente cantidad, pureza y calidad.

El ADN aislado se usa posteriormente para generar una librería (meta-) genómica usando un vector apropiado (114). En estos momentos se dispone de una gran variedad de vectores para la construcción de librerías (meta-) genómicas (121). Para la inserción de pequeños fragmentos (de 0.5 a 5 mil pares de bases) se utilizan plásmidos, mientras que para la inserción de fragmentos de hasta 35 mil pares de bases se utilizan fósmidos (por ejemplo, pCCFOS) o cósmicos (por ejemplo, pLAFR3) y en el caso de fragmentos de gran tamaño, cromosomas bacterianos artificiales (BAC).

La librería se transfiere posteriormente a un huésped y los clones individuales se someten a un tamizado para identificar la presencia de una enzima específica. Cuando este tamizado se realiza de forma robotizada el método es extremadamente eficiente para aislar nuevas enzimas de organismos que por métodos convencionales son inaccesibles. La Tabla I ilustra ejemplos de biocatalizadores aislados usando herramientas de (meta-) genómica funcional, que incluyen esterasas, lipasas, quitinasas, amilasas y amidasas, entre las más estudiadas.

Como se aprecia en la Figura 4 existe un desfase entre las enzimas que se requieren en biotransformaciones industriales y las aisladas por técnicas (meta-) genómicas.

*Tabla I.- Listado de biocatalizadores con potencial interés en biotransformaciones desarrollados por técnicas (meta-) genómicas en los últimos ocho años*

Muestra	Biodiversidad enzimática	Referencias
Suelo	53 alcohol oxidoreductasas	<b>130</b>
Suelo	3 glicerol dehidratasa y 2 diol dehidratasas	<b>131</b>
Sedimento marino	4 amidasas, 1 peroxidasa y 1 dehidrogenasa	<b>40</b>
Suelo	8 lipasas	<b>132</b>
Microbiota de ratón	Una 1,3-1,4- $\beta$ -glucanasa, tres 1,3(4)- $\beta$ -glucanasa, 1 $\beta$ -glucosidasa, a $\alpha$ -glucosidasa, 1 $\alpha$ -amilasa	<b>133</b>
Agua de lago	11 lipasas	<b>134</b>
Agua de río	1 metano y 1 amonio monooxigenasa	<b>135</b>
Mar de Sargaso	11 endo- $\beta$ -1,4-glucanasas, 2 aminopeptidasas	<b>136</b>
Rumen bovino	12 esterasas, 9 endo-beta-1,4-glucanasas y una ciclodextrinasa	<b>137-140</b>
Fosa marina	5 esterasas	<b>141</b>
Sedimento de lago	1 citocromo P450 monooxigenasa	<b>142</b>
Sedimento marino	1 lipasa	<b>143</b>
Suelo	1 esterasa	<b>144</b>
Rumen bovino	1 lacasa	<b>117</b>
Planta de tratamiento de aguas	43 extradiol dioxigenasas (EDO) capaces de transformar 2,3-dihidroxibifenol	<b>145</b>
Bioreactor de minería	Oxidasa (hidracina)	<b>146</b>
Planta de tratamiento de aguas	1 lipasa (activa a baja temperatura)	<b>147</b>

<b>Bacterioplancton marino</b>	110 ciclasas de isoprenoides	<b>148</b>
<b>Sedimento marino</b>	1 esterasa (alcalina) con actividad 3-oxoadipato enol-lactonasa	<b>149</b>
<b>Tubo digestivo conejo</b>	4 endo- $\beta$ -1,4-glucanasas de las familias 3 y 5 capaces de hidrolizar enlaces beta 1,3-, 1,4- y 1-6	<b>150</b>
<b>Fuente termal</b>	1 esterasa (hipertermofílica)	<b>151</b>
<b>Sedimento marino</b>	1 metalo-proteasa capaz de hidrolizar azocaseína y fibrina	<b>152</b>
<b>Lago hipersalino, alcalino</b>	2 quitinasas (familias 18 y 20)	<b>153</b>
<b>Suelo y compost</b>	12 esterases, una 1,2-dioxigenasa (hogentisato) y 1 amidasa	<b>154</b>
<b>Suelo contaminado</b>	1 decarboxilasa capaz de catalizar la decarboxilación de L-cisteína a cisteamina	<b>155</b>
<b>Suelo forestal</b>	4 lipasas (familia HSL)	<b>156</b>
<b>Bioreactores de minería</b>	3 sulfuro oxigenasas-reductasas	<b>157</b>
<b>Suelo contaminado con petróleo</b>	1 naftaleno oxidoreductasa capaz de catalizar la conversión de naftaleno a salicilato	<b>158</b>
<b>Suelo</b>	1 estireno monooxigenasa que cataliza la epoxidación de estireno y derivados para formar los (S)-epóxidos correspondientes	<b>159</b>
<b>Sedimento marino</b>	1 esterasa	<b>160</b>
<b>Sedimento marino</b>	2 alcano hidroxilasas, 9 dehidrogenasas, 1 proteasa, 2 oxidoreductasas, 1 amidasa, 1 dioxigenasa, 1 aril-sulfatasa, 1 beta-lactamasa y 2 hidrolasas	<b>161</b>
<b>Fuente termal (Tailandia)</b>	1 fosfolipasa y 1 esterasa (termoestables) activas a pH alcalinos	<b>162</b>
<b>Mar de China</b>	2 esterases	<b>163</b>

<b>Sedimento de estuario</b>	4-clorobenzoato dehalogenasa	<b>164</b>
<b>Compost</b>	1 poli(DL-lactico) depolimerasa	<b>165</b>
<b>Suelo</b>	2 esterasa (piretroides y piretrinas)	<b>166</b>
<b>Suelo</b>	4 nuevas 4-hidroxifenilpiruvato dioxigenasas (HPPD) que convierten 4-hidroxifenilpiruvato en homogentisato.	<b>167</b>
<b>Suelo</b>	1 celulasa	<b>168</b>
<b>Suelo</b>	1 xilanasas	<b>169</b>
<b>Suelo forestal</b>	6 lipasas, cuatro de ellas de la familia 4 y dos de la familia 2	<b>170</b>
<b>Suelo</b>	1 N-acilhomoserina lactonasa (NAHL)	<b>171</b>
<b>Agua del Río Ganges</b>	1 esterasa y 1 dehidrogenasa	<b>172</b>
<b>Suelo</b>	2 lipasas	<b>173</b>
<b>Suelo</b>	2 metalo-proteasas	<b>174</b>
<b>Sedimento de río contaminado</b>	1 bifenil-dioxigenasa	<b>124</b>
<b>Planta de tratamiento de aguas</b>	38 extradiol dioxigenasas	<b>175</b>
<b>Compost</b>	3 endoglucanasas (familia 5)	<b>176</b>
<b>Suelo</b>	2 extradiol 1,2-dioxigenasa	<b>177</b>
<b>Suelo antártico</b>	1 esterasa halotolerante alcalina (nueva familia)	<b>178</b>
<b>Rumen de búfalo</b>	14 celulosas (ácidas)	<b>179</b>
<b>Suelo</b>	1 lisina racemasa	<b>180</b>
<b>Suelo contaminado</b>	8 extradiol dioxigenasa	<b>181</b>
<b>Rumen bovino</b>	3,5,6-tricloro-2-piridinol hidrolasa	<b>182</b>
<b>Planta de tratamiento de aguas</b>	1 esterasa estable en presencia de detergentes y metales	<b>183</b>

Suelo de manglar	1 oxidasa multicobre	<b>184</b>
Suelo antártico	1 lipasa	<b>185</b>
Sedimento marino	15 lipasas	<b>186</b>
Mar de Sargaso	1 hidrogenasa	<b>187</b>
Sedimento marino	14 PAH dioxigenasas	<b>188</b>
Reactor enriquecido en biomasa	1 lipasa (alcalina y termoestable)	<b>189</b>
Rumen de vaca	2 lipasas	<b>190</b>
Muestras varias	6 agarasas, 55 amilasas, 27 celulasas, 11 quitinasas, 1 ciclodextrinasa y 17 esterasas	<b>191</b>
Sedimento marino	4-hidroxifenilpiruvato dioxigenasa	<b>192</b>
Sedimento marino	1 lipasa (activa a baja temperatura)	<b>193</b>
Suelo	64 quitinasas	<b>194</b>
Suelo	1 amilasa	<b>195</b>
Compost	1 esterasa (familia VIII) activa en presencia de 5-30% disolventes	<b>196</b>
Intestino gusano de tierra	<b>1 feruloil esterasa</b>	<b>197</b>

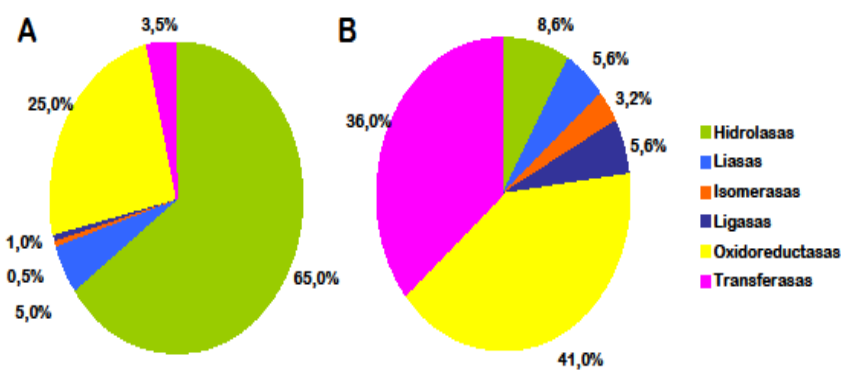


Figura 4.- Distribución de enzimas empleadas en biotransformaciones industriales (A) y enzimas presentes en (meta-) genomas bacterianos (B).



En particular, hidrolasas seguidas de oxido-reductasas y en menor medida liasas son las enzimas más utilizadas a escala industrial, mientras que los (meta-)genomas contienen un mayor número de oxido-reductasas y transferasas en comparación con otras actividades de interés (37).

Hasta la fecha, los datos bibliográficos indican que se han identificado y caracterizado aproximadamente 1.400 enzimas de microorganismos no cultivables mediante la aplicación de técnicas (meta-) genómicas, con un 10% de las mismas aisladas en el último año (Tabla I). En este sentido, pese al enorme potencial de esta técnica para el desarrollo de nuevos biocatalizadores existe una limitación en lo que se refiere al número de enzimas activas, es decir, el acceso a nuevos biocatalizadores mediante análisis (meta-) genómicos está sobreestimado (106). Así, se ha estimado que solo el 0,000001% de las enzimas presentes en (meta-)genomas microbianos pueden ser identificadas por métodos de tamizado convencional. Por ejemplo, en una muestra que contiene 1.000.000 de microorganismos y un total de aprox. 2.000.000.000 de genes, el 20% de los cuales codifica proteínas con actividad enzimática, solo 400 de ellas pueden ser identificadas, lo que limita el poder de las técnicas basadas en el tamizado por actividad (114). Sin embargo, estimaciones actuales indican que pese a los grandes esfuerzos en proyectos de mega secuenciación la identificación de nuevas familias de proteínas es lineal, lo que indica que los microorganismos continuarán siendo una fuente de nuevos biocatalizadores en el futuro (122). Como muchos de estos genes son enteramente novedosos, sus actividades no pueden inferirse por comparación con proteínas conocidas en bases de datos, y por lo tanto se han de desarrollar métodos funcionales para la identificación y futura caracterización de los mismos (117). Esto implica el desarrollo de métodos de tamizado de alto rendimiento (29).

La aplicación de métodos de tamizado por actividad y no por homología (32), accesibles a gran escala mediante el uso de sustratos cromogénicos y fluorogénicos, permiten analizar decenas de miles de clones en un solo ensayo (**Figura 1**), independientemente de la secuencia de los mismos. Las metodologías de tamizado y selección deben cumplir una serie de requisitos: (i) deben, si es

posible, seleccionar directamente la propiedad de interés – se selecciona lo que se busca, es una de las reglas principales en cualquier programa de búsqueda de enzimas; (ii) el ensayo debe ser lo más sensible posible en un rango adecuado y dinámico; (iii) el método debe ser aplicable en formato de alto rendimiento.

Se han diseñado numerosos ensayos para detectar actividades enzimáticas en medio de agar sólido (empleando colonias) o en líquido (empleando lisados celulares) mediante la producción de un fluoróforo o un cromóforo (25-27). Los ensayos en medio agar-sólido permiten rastrear unas 10.000 colonias al día, pero están muy limitados por la baja sensibilidad. Así, los productos de reacción difunden de la colonia y solo aquellas enzimas muy activas se detectan. Ejemplos clásicos de métodos de tamizado colorimétrico consisten en la suplementación de los medios de cultivo con tributirina para detectar actividad lipasa mediante la presencia de halos correspondientes a la hidrólisis del sustrato. El rango de aplicación de los ensayos en medio líquido es mucho más amplio, pero su aplicación en formato de alto rendimiento es todavía muy limitado: en ausencia de sofisticados sistemas robóticos que son inaccesibles a laboratorios académicos, solo se pueden rastrear de media los extractos de 1.000 a 10.000 clones (30). Esto convierte a ambos sistemas en métodos de bajo-medio rendimiento.

Para generar una librería (meta-) genómica efectiva, se necesitan métodos alternativos que permitan enriquecer la muestra en genomas individuales (114). El gen que codifica un biocatalizador de interés puede representar en muchos casos una fracción muy pequeña, en algunos casos menos del 0,01%, del total representado en la muestra y por lo tanto la baja proporción del gen en la librería juega un papel crítico en los programas de búsqueda de biocatalizadores (38). Por esta razón, se han desarrollado métodos de enriquecimiento alternativo (123), dependientes o independientes de la secuencia de ADN, como por ejemplo el método basado en el enriquecimiento del cultivo mediante el uso de medios selectivos o el marcaje isotópico (SIP: Stable Isotope Probing). Sin embargo, el inconveniente principal de esta técnica es el peligro de enriquecer la muestra en organismos de crecimiento rápido que no utilizan los nutrientes suplementados, la baja eficiencia en el marcaje y la posibilidad de un marcaje indirecto de

microorganismos que no utilizan el sustrato marcado directamente pero si utilizan intermediarios del mismo (34, 124).

El tamizado clásico funcional acoplado con técnicas innovadoras tales como la expresión inducida de genes por sustrato (SIGEX) (125, 126), la pre-amplificación inversa por PCR (PAI-PCR) (127) y el barajado de ADN (meta-) genómico (129), pueden incrementar las posibilidades de los tamizados. De entre ellos, el método de mayor relevancia para la identificación de biocatalizadores es el SIGEX, que se desarrolló inicialmente para detectar clones que expresan genes catabólicos inducidos por la presencia de un sustrato usando una proteína verde fluorescente (GFP) como gen reportero (125, 126). Los clones que contienen genes expresados constitutivamente se separan inicialmente mediante Fluorescence Activated Cell Sorting (FACS) y a continuación los genes deseados se inducen mediante la incorporación del sustrato de interés seguido de una separación por FACS. Utilizando este sistema, que permite analizar hasta 150 mil clones con un tamaño medio de inserto de 7 mil pares de bases en 17 días, se han identificado genes involucrados en la degradación de benzoato y catecol en una muestra obtenida de agua superficial. Pese a que el sistema puede ser aplicado en un formato de alto rendimiento, una de las mayores limitaciones es que solo se pueden detectar genes que son inducidos por ciertos sustratos, hecho éste que ocurre en un limitado número de actividades (menos del 0,2% total).

Otros métodos de tamizado, se basan en el tamizado fenotípico donde se introduce una ruta catabólica en un huésped para monitorizar posteriormente el crecimiento de las colonias frente a un sustrato inaccesible a dicha bacteria. Una modificación de este método es el uso de cepas modificadas en las que una actividad particular está ausente y la re-introducción de dicha actividad mediante la clonación de fragmentos de ADN (meta-) genómico permitiría identificar colonias que contengan el gen de interés. Este sistema permite rastrear la actividad mediante el método sencillo de crecimiento o no crecimiento, que se puede acoplar con espectrometría de masas para identificar diferencias en los productos de reacción de miles de clones (129).

## 12. BIOCATALIZADORES DESARROLLADOS POR (META-) GENÓMICA

Como se muestra en la Tabla 1, las actividades enzimáticas más estudiadas e identificadas por métodos (meta-) genómicos son hidrolasas que incluyen esterases, lipasas, glicosil hidrolasas (celulasas, glucosidasas y xilanasas, por citar las más estudiadas) y lactonasas, seguidas de oxidoreductasas, incluyendo mono- y dioxigenasas. Estas dos superfamilias de enzimas constituyen los dos tipos de biocatalizadores más empleados en la industria (3).

Además de estas enzimas, Xu et al. (161) han identificado una serie de alcanos hidroxilasas de sedimentos marinos profundos capaces de realizar oxigenaciones terminales. Un ejemplo que muestra la versatilidad de los (meta-) genomas microbianos como factorias enzimáticas es la habilidad que presenta la bacteria *E. coli* para transformar una gran variedad de compuestos aromáticos (desde catecol a 2,3-dihidroxibifenol) cuando expresa 38 extradiol dioxigenasas aisladas de una planta de tratamiento de aguas (145). También se ha desarrollado otra ruta de síntesis que comienza con L-cisteína y utiliza una decarboxilasa de suelos alcalinos contaminados (155) para generar cisteamina, un importante terapéutico. Más recientemente el grupo del Prof. Jaeger ha construido un conjunto de cepas de la bacteria *E. coli* que expresan una variedad de enzimas con potencial en biotransformaciones industriales, incluidas, benzaldehído liasas, benzoilformato decarboxilasas, hidroxinitrilo liasas y alcohol dehidrogenasas, que producen benzaldehído mediante la conversión de benzoína, benzoilformato, mandelonitrilo, o benzoil alcohol, respectivamente (198). Aunque estos estudios muestran el potencial de las enzimas obtenidas por (meta-) genómica, se han de realizar esfuerzos adicionales para potenciar el uso y la aplicación de las mismas a escala industrial. Así, hasta la fecha, solo una esterasa de un microorganismo no cultivado capaz de degradar ésteres de tereftalato, un importante componente de bio-plásticos, ha sido empleado en la industria (199).

### **13. DEL DESCUBRIMIENTO A LA COMERCIALIZACIÓN: UNA VISIÓN DE FUTURO SOBRE LA (META-) GENÓMICA DEL SIGLO XXI**

En particular, El informe de las Sociedades de Microbiología señala que menos del 1% de las especies bacterianas y menos del 5% de las especies de hongos son conocidas, y que millones de especies microbianas permanecen desconocidas, por lo que es grande el potencial de los microorganismos como proveedores de biocatalizadores. Sin embargo, la complejidad de las muestras ambientales como portadoras de microorganismos ha proporcionado una visión del mundo microbiano como una “caja negra” de la cual se sabía que contenía una serie de funciones o actividades aunque no se conociese su contenido. Las aproximaciones (meta-) genómicas acopladas con nuevas herramientas de producción y empleo en biotransformaciones de los nuevos biocatalizadores identificados, permiten acceder al potencial de los organismos no cultivados. Este punto es de especial interés dado que en la actualidad el tiempo transcurrido entre la identificación de una enzima hasta su producción a escala es de aproximadamente 12 meses, casi tres veces inferior al necesario hace una década y cinco veces inferior al de principios de la década de los noventa. El descubrimiento de una enzima es seguido por un proceso de escalado, que normalmente incluye una selección y evaluación del organismo productor, un desarrollo de la fermentación y una optimización a escala de 2 L. Posteriormente, se sigue con una fermentación microbiana a escala piloto (500-2000 L), durante las cuales se producen pruebas para control de calidad y determinación del rendimiento. Finalmente la producción se realiza a escala de hasta 160.000 L para la aplicación industrial de la enzima. Pese a que este procedimiento está bien establecido, hasta la fecha no existe ningún ejemplo de producción de una enzima obtenida por métodos (meta-) genómicos a tal escala, y por lo tanto, se ha de trabajar en esta dirección.

### **14. AGRADECIMIENTOS**

Este trabajo ha contado con el apoyo económico de los siguientes proyectos: CSD2007-00005, BIO2006-11738, GEN2006-27750-C-4-E, BFU2008-04398-E/BMC, CTQ2007-60480/BQU y KBBE-226977. M-E.G., agradece al CSIC la beca JAE.

## 15. BIBLIOGRAFÍA

1. Nedwin, G. E. (1997) Using Enzymes as Benign Substitutes for Synthetic Chemicals and Harsh Conditions in Industrial Processes *Biotech* (pp. 13-32) *TiE Sustainable Environment*.
2. Kirk, O., Borchert, T. V., & Fuglsang, C. C. (2002). Industrial enzyme applications. *Current Opinion in Biotechnology*, 13(4), 345-351.
3. Beloqui, A., de Maria, P. D., Golyshin, P. N., & Ferrer, M. (2008). Recent trends in industrial microbiology. *Current Opinion Microbiology*, 11(3), 240-248.
4. Ferrer, M., Golyshina, O., Beloqui, A., & Golyshin, P. N. (2007). Mining enzymes from extreme environments. *Current Opinion Microbiology*, 10(3), 207-214.
5. Alcalde, M., Ferrer, M., Plou, F. J., & Ballesteros, A. (2006). Environmental biocatalysis: from remediation with enzymes to novel green processes. *Trends in Biotechnology*, 24(6), 281-287
6. Gernaey, K. V., Lantz, A. E., Tufvesson, P., Woodley, J. M., & Sin, G. (2010). Application of mechanistic models to fermentation and biocatalysis for next-generation processes. *Trends in Biotechnology*.
7. Prieto, M. A. (2007). From oil to bioplastics, a dream come true? *Journal of Bacteriology*, 189(2), 289-290.
8. Lee, M. Y., Park, C. B., Dordick, J. S., & Clark, D. S. (2005). Metabolizing enzyme toxicology assay chip (MetaChip) for high-throughput microscale toxicity analyses. *Proceedings of the National Academy of Sciences*, 102(4), 983-987.
9. Alvaro-Benito, M., de Abreu, M., Fernandez-Arrojo, L., Plou, F. J., Jimenez-Barbero, J., Ballesteros, A., et al. (2007). Characterization of a beta-fructofuranosidase from *Schwanniomyces occidentalis* with transfructosylating activity yielding the prebiotic 6-kestose. *Journal of Biotechnology*, 132(1), 75-81.
10. Weiner, B., Szymanski, W., Janssen, D. B., Minnaard, A. J., & Feringa, B. L. (2010). Recent advances in the catalytic asymmetric synthesis of beta-amino acids. *Chemical Society Reviews*, 39(5), 1656-1691.
11. Weiner, B., Szymanski, W., Janssen, D. B., Minnaard, A. J., & Feringa, B. L. (2010). Recent advances in the catalytic asymmetric synthesis of beta-amino acids. *Chemical Society Reviews*, 39(5), 1656-1691.
12. Fiammengo, R., & Jaschke, A. (2005). Nucleic acid enzymes. *Current Opinion in Biotechnology*, 16(6), 614-621.

13. Davis, B. G., & Boyer, V. (2001). Biocatalysis and enzymes in organic synthesis. *Natural Product Reports*, 18(6), 618-640.
14. Klibanov, A. M. (2001). Improving enzymes by using them in organic solvents. *Nature*, 409(6817), 241-246.
15. Hudson, E. P., Eppler, R. K., & Clark, D. S. (2005). Biocatalysis in semi-aqueous and nearly anhydrous conditions. *Current Opinion in Biotechnology*, 16(6), 637-643.
16. Nobeli, I., Favia, A. D., & Thornton, J. M. (2009). Protein promiscuity and its implications for Biotechnology. *Nature Biotechnology*, 27(2), 157-167.
17. Steele, H. L., Jaeger, K. E., Daniel, R., & Streit, W. R. (2009). Advances in recovery of novel biocatalysts from metagenomes. *Journal of Molecular Microbiology and Biotechnology*, 16(1-2), 25-37.
18. Ferrer, M., Beloqui, A., Golyshin, P.N. Microbial metagenomes: moving forward industrial biotechnology. *Journal of Chemical Technology and Biotechnology* (2007) 82: 421-423.
19. Bull, A. T. (2003) *Microbial Diversity and Bioprospecting American Society For Microbiology*
20. Lee, H. S., Kwon, K. K., Kang, S. G., Cha, S. S., Kim, S. J., & Lee, J. H. (2010). Approaches for novel enzyme discovery from marine environments. *Current Opinion in Biotechnology*.
21. Vorgias, C. E., and G. Antranikian. (2004) *Extremophiles: pH, Temperature, and Salinity. Microbial Diversity and Bioprospecting. ASM Press, Washington, D.C., pp. 146-153.*
22. Yarza, P., Richter, M., Peplies, J., Euzéby, J., Amann, R., Schleifer, K. H., et al. (2008). The All-Species Living Tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. *System Applied Microbiology*, 31(4), 241-250.
23. Raymond, J. (2008). Coloring in the tree of life. *Trends in Microbiology*, 16(2), 41-43.
24. Tamames, J., Abellan, J. J., Pignatelli, M., Camacho, A., & Moya, A. (2010). Environmental distribution of prokaryotic taxa. *BMC Microbiology*, 10, 85.
25. Bornscheuer, U. T. (2002). Methods to increase enantioselectivity of lipases and esterases. *Current Opinion in Biotechnology*, 13(6), 543-547.
26. Goddard, J. P., & Reymond, J. L. (2004). Enzyme assays for high-throughput screening. *Current Opinion in Biotechnology*, 15(4), 314-322.
27. Andexer, J., Guterl, J. K., Pohl, M., & Eggert, T. (2006). A high-throughput screening assay for hydroxynitrile lyase activity. *Chemical Communications (Camb)*(40), 4201-4203.

28. Pham, V. D., Palden, T., & DeLong, E. F. (2007). Large-scale screens of metagenomic libraries. *Journal of Visualized Experiments*(4), 201.
29. Ferrer, M., Beloqui, A., Vieites, J.M., Guazzaroni, M.E., Berger, I., Aarón, A. (2008) Interplay of metagenomics and in vitro compartmentalization. *Microbial Biotechnology* 2:31-39.
30. Geddie, M. L., Rowe, L. A., Alexander, O. B., & Matsumura, I. (2004). High throughput microplate screens for directed protein evolution. *Methods Enzymology*, 388, 134-145.
31. Dalbege, H., and L. Lange. (1998) Using Molecular Techniques to Identify New Microbial Biocatalysts. *Trends in Biotechnology* 16: 265-272.
32. Kotik, M. (2009). Novel genes retrieved from environmental DNA by polymerase chain reaction: current genome-walking techniques for future metagenome applications. *Journal of Biotechnology*, 144(2), 75-82.
33. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual* 2nd edn. Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press.
34. Beloqui, A., Zumárraga, M., Alcalde, M., Golyshin, P., Ferrer, M. (2008) Microbes and enzymes: recent trends and new directions to expand protein space. In: *Handbook of Protein Engineering* (John Wiley & Sons GmbH, Wiley-VCH Verlag GmbH & Co. KgaA) 1: 233–264.
35. Schiilein, M. (1996) Novel Endoglucanases. Novo Nordisk A/S. Patents WO 96/29379 and U.S. 6,001,639
36. Nielsen, P., and B. R. Jrgensen. (2001) Method for Genome Mining for secreted Protein Genes. Patent WO 01/98484, (pp. 1-43).
37. Vieites, J. M., Guazzaroni, M. E., Beloqui, A., Golyshin, P. N., & Ferrer, M. (2009). Metagenomics approaches in systems microbiology. *FEMS Microbiology Reviews*, 33(1), 236-255.
38. Gabor, E., Liebeton, K., Niehaus, F., Eck, J., & Lorenz, P. (2007). Updating the metagenomics toolbox. *Biotechnology Journal*, 2(2), 201-206.
39. Schloss, P.D. and Handelsman, J. (2003) Biotechnological prospects from metagenomics. *Current Opinion in Biotechnology* 14, 303-310
40. Gabor, E. M., de Vries, E. J., & Janssen, D. B. (2004). Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. *Environmental Microbiology*, 6(9), 948-958.
41. Hugenholtz, P., & Tyson, G. W. (2008). Microbiology: metagenomics. *Nature*, 455(7212), 481-483.



42. Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., et al. (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, 9, 386.
43. Ansorge, W. J. (2009). Next-generation DNA sequencing techniques. *N Biotechnology*, 25(4), 195-203.
44. Fox, S., Filichkin, S., & Mockler, T. C. (2009). Applications of ultra-high-throughput sequencing. *Methods in Molecular Biology*, 553, 79-108.
45. Steward, G. F., & Rappe, M. S. (2007). What's the 'meta' with metagenomics? *Isme Journal*, 1(2), 100-102.
46. Guazzaroni, M.E., Beloqui, A., Ferrer, M. (2009) Metagenomics as a new technological tool to gain scientific knowledge. *World Journal of Microbiology and Biotechnology* 25: 945–954.
47. Lorenz, P., & Eck, J. (2005). Metagenomics and industrial applications. *Nature Reviews Microbiology*, 3(6), 510-516.
48. Bertin, P.N., Medigue, C., & Normand, P. (2008). Advances in environmental genomics: towards an integrated view of micro-organisms and ecosystems. *Microbiology* 154, 347-359.
49. Ingham, C. J., Sprenkels, A., Bomer, J., Molenaar, D., van den Berg, A., van Hylckama Vlieg, J. E., et al. (2007). The micro-Petri dish, a million-well growth chip for the culture and high-throughput screening of microorganisms. *Proceedings of the National Academy of Sciences*, 104(46), 18217-18222.
50. Valdivia-Granda, W. (2008). The next meta-challenge for Bioinformatics. *Bioinformation*, 2(8), 358-362.
51. Chen, K., & Pachter, L. (2005). Bioinformatics for whole-genome shotgun sequencing of microbial communities. *PLoS Computational Biology*, 1(2), 106-112.
52. Raes, J., & Bork, P. (2008). Molecular eco-systems biology: towards an understanding of community function. *Nature Reviews Microbiology*, 6(9), 693-699.
53. Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K., & Hugenholtz, P. (2008). A bioinformatician's guide to metagenomics. *Microbiology and Molecular Biology Reviews*, 72(4), 557-578, Table of Contents.
54. Huson, D.H., Richter, D.C., Mitra, S., Auch, A.F., Schuster, S.C. (2009) Methods for comparative metagenomics. *BMC Bioinformatics* 10 (Suppl 1):S12.

55. Li W. 2009. Analysis and comparison of very large metagenomes with fast clustering and functional annotation. *BMC Bioinformatics* 10:359.
56. Sterk, P., Hirschman, L., Field, D., & Wooley, J. (2010). Genomic standards consortium workshop: metagenomics, metadata and metaanalysis (m3). *Pacific Symposium on Biocomputing*, 481-484.
57. Walker, A. (2010). A glut from the gut: metagenomics takes a giant step forward. *Nature Reviews Microbiology*, 8(5), 315.
58. Hirose, I., Sano, K., Shioda, I., Kumano, M., Nakamura, K. & Yamane K. (2000) Proteome Analysis of *Bacillus subtilis* Extracellular Proteins: a Two-Dimensional Protein Electrophoretic Study. *Microbiology. U.K.*, 146: 65-75.
59. Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., et al. (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature*, 390(6657), 249-256.
60. Dunn-Coleman, N., & Prade, R. (1998). Toward a global filamentous fungus genome sequencing effort. *Nature Biotechnology*, 16(1), 5.
61. Roberts, R. J. (2004). Identifying protein function--a call for community action. *PLoS Biology*, 2(3), E42.
62. Sleator, R. D., & Walsh, P. (2010). An overview of in silico protein function prediction. *Archives of Microbiology*, 192(3), 151-155.
63. Olsen, G. J., Lane, D. J., Giovannoni, S. J., Pace, N. R., & Stahl, D. A. (1986). Microbial ecology and evolution: a ribosomal RNA approach. *Annu Rev Microbiol*, 40, 337-365.
64. Liolios, K., Mavromatis, K., Tavernarakis, N., & Kyrpides, N. C. (2008). The Genomes On Line Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Research*, 36(Database issue), D475-479.
65. Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooseph, S., et al. (2007). The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biology*, 5(3), e77.
66. Wilmes, P., Andersson, A. F., Lefsrud, M. G., Wexler, M., Shah, M., Zhang, B., et al. (2008). Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal. *ISME Journal*, 2(8), 853-864.
67. Verberkmoes, N. C., Russell, A. L., Shah, M., Godzik, A., Rosenquist, M., Halfvarson, J., et al. (2009). Shotgun metaproteomics of the human distal gut microbiota. *ISME Journal*, 3(2), 179-189.

68. Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., et al. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, 428(6978), 37-43.
69. Belouqui, A., Guazzaroni, M.E., Ferrer, M. (2010) Procedures for protein isolation in pure culture and microbial communities. *Handbook of Hydrocarbon and Lipid Microbiology*. Springer Verlag Press.
70. Blackstock, W. P., & Weir, M. P. (1999). Proteomics: quantitative and physical mapping of cellular proteins. *Trends in Biotechnology*, 17(3), 121-127.
71. Harris, P. (2004) Proteomics as a Tool for Enzyme Discovery in Filamentous Fungi - Applications in Lignocellulose Degradation for Sugar and Fuel Ethanol Production. 1st International Fungal Proteomics Symposium, Portland, OR.
72. Benndorf, D., Vogt, C., Jehmlich, N., Schmidt, Y., Thomas, H., Woffendin, G., et al. (2009). Improving protein extraction and separation methods for investigating the metaproteome of anaerobic benzene communities within sediments. *Biodegradation*, 20(6), 737-750.
73. Parro, V., Moreno-Paz, M., & Gonzalez-Toril, E. (2007). Analysis of environmental transcriptomes by DNA microarrays. *Environmental Microbiology*, 9(2), 453-464.
74. Frias-Lopez, J., Shi, Y., Tyson, G. W., Coleman, M. L., Schuster, S. C., Chisholm, S. W., et al. (2008). Microbial community gene expression in ocean surface waters. *Proceedings of the National Academy of Sciences*, 105(10), 3805-3810.
75. Moran, A.M. (2009) Metatranscriptomics: Eavesdropping on Complex Microbial Communities. *Microbe* 4: 329-335.
76. Bailly, J., Fraissinet-Tachet, L., Verner, M. C., Debaud, J. C., Lemaire, M., Wesolowski-Louvel, M., et al. (2007). Soil eukaryotic functional diversity, a metatranscriptomic approach. *Isme Journal*, 1(7), 632-642.
77. Feist, A. M., Herrgard, M. J., Thiele, I., Reed, J. L., & Palsson, B. O. (2009). Reconstruction of biochemical networks in microorganisms. *Nature Reviews Microbiology*, 7(2), 129-143.
78. Gao, H., Yang, Z. K., Gentry, T. J., Wu, L., Schadt, C. W., & Zhou, J. (2007). Microarray-based analysis of microbial community RNAs by whole-community RNA amplification. *Applied and Environmental Microbiology*, 73(2), 563-571.
79. Daran-Lapujade, P., Rossell, S., van Gulik, W.M., Luttik, M.A., de Groot, M.J., et al. (2007). The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at posttranscriptional levels. *Proceeding of the National Academy of Science USA* 104, 15753-15758.

80. Warnecke, F., & Hess, M. (2009). A perspective: metatranscriptomics as a tool for the discovery of novel biocatalysts. *Journal of Biotechnology*, 142(1), 91-95.
81. Yaver, D. S. (2003) Global Expression Profiling of the Lignin Degrading Fungus *Ceriporiopsis subvermispom* for the Discovery of Novel Enzymes. *Applied Mycology and Biotechnology*, 4, Fungal Genomics. Elsevier Science, Amsterdam.
82. Nicholson, J. K., & Lindon, J. C. (2008). Systems biology: Metabonomics. *Nature*, 455(7216), 1054-1056.
83. Kalyuzhnaya, M. G., Lapidus, A., Ivanova, N., Copeland, A. C., McHardy, A. C., Szeto, E., et al. (2008). High-resolution metagenomics targets specific functional types in complex microbial communities. *Nature Biotechnology*, 26(9), 1029-1034.
84. Dick, G. J., Andersson, A. F., Baker, B. J., Simmons, S. L., Thomas, B. C., Yelton, A. P., et al. (2009). Community-wide analysis of microbial genome sequence signatures. *Genome Biology*, 10(8), R85.
85. Gerlach, W., Junemann, S., Tille, F., Goesmann, A., & Stoye, J. (2009). WebCARMA: a web application for the functional and taxonomic classification of unassembled metagenomic reads. *BMC Bioinformatics*, 10, 430.
86. Krause, L., Diaz, N. N., Bartels, D., Edwards, R. A., Puhler, A., Rohwer, F., et al. (2006). Finding novel genes in bacterial communities isolated from the environment. *Bioinformatics*, 22(14), e281-289.
87. Monzoorul Haque, M., Ghosh, T. S., Komanduri, D., & Mande, S. S. (2009). SOrt-ITEMS: Sequence orthology based approach for improved taxonomic estimation of metagenomic sequences. *Bioinformatics*, 25(14), 1722-1730.
88. Bork, P., Ouzounis, C., Sander, C. (1994) From Genome Sequences to Protein Function. *Current Opinion in Structural Biology* 4 (3), 393-403
89. Harrington, E. D., Singh, A. H., Doerks, T., Letunic, I., von Mering, C., Jensen, L. J., et al. (2007). Quantitative assessment of protein function prediction from metagenomics shotgun sequences. *Proceedings of the National Academy of Sciences*, 104(35), 13913-13918.
90. Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057), 376-380.
91. Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Research*, 17(3), 377-386.

92. Pop, M. (2009). Genome assembly reborn: recent computational challenges. *Briefings in Bioinformatics*, 10(4), 354-366.
93. Desai, C., Pathak, H., & Madamwar, D. (2010). Advances in molecular and "-omics" technologies to gauge microbial communities and bioremediation at xenobiotic/anthropogen contaminated sites. *Bioresource Technology*, 101(6), 1558-1569.
94. de Lorenzo, V. (2008). Systems biology approaches to bioremediation. *Current Opinion in Biotechnology*, 19(6), 579-589.
95. Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, 304(5667), 66-74.
96. Edwards, R. A., Rodriguez-Brito, B., Wegley, L., Haynes, M., Breitbart, M., Peterson, D. M., et al. (2006). Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics*, 7, 57.
97. Bidle, K. D., Lee, S., Marchant, D. R., & Falkowski, P. G. (2007). Fossil genes and microbes in the oldest ice on earth. *Proceedings of the National Academy of Sciences*, 104(33), 13455-13460.
98. Gianoulis, T. A., Raes, J., Patel, P. V., Bjornson, R., Korb, J. O., Letunic, I., et al. (2009). Quantifying environmental adaptation of metabolic pathways in metagenomics. *Proceedings of the National Academy of Sciences*, 106(5), 1374-1379.
99. Markowitz, V. M., Ivanova, N. N., Szeto, E., Palaniappan, K., Chu, K., Dalevi, D., et al. (2008). IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Research*, 36(Database issue), D534-538.
100. Seshadri, R., Kravitz, S. A., Smarr, L., Gilna, P., & Frazier, M. (2007). CAMERA: a community resource for metagenomics. *PLoS Biology*, 5(3), e75.
101. Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*, 9, 75.
102. Teeling, H., Waldmann, J., Lombardot, T., Bauer, M., & Glockner, F. O. (2004). TETRA: a web-service and a stand-alone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences. *BMC Bioinformatics*, 5, 163.
103. McHardy, A. C., & Rigoutsos, I. (2007). What's in the mix: phylogenetic classification of metagenome sequence samples. *Current Opinion Microbiology*, 10(5), 499-503.
104. Krause, L., Diaz, N. N., Goesmann, A., Kelley, S., Nattkemper, T. W., Rohwer, F., et al. (2008). Phylogenetic classification of short environmental DNA fragments. *Nucleic Acids Research*, 36(7), 2230-2239.

105. Hoff, K. J., Tech, M., Lingner, T., Daniel, R., Morgenstern, B., & Meinicke, P. (2008). Gene prediction in metagenomic fragments: a large scale machine learning approach. *BMC Bioinformatics*, 9, 217.
106. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17), 3389-3402.
107. Raes, J., Foerstner, K. U., & Bork, P. (2007). Get the most out of your metagenome: computational analysis of environmental sequence data. *Current Opinion Microbiology*, 10(5), 490-498.
108. Gomez-Alvarez, V., Teal, T. K., & Schmidt, T. M. (2009). Systematic artifacts in metagenomes from complex microbial communities. *ISME Journal*, 3(11), 1314-1317.
109. Hoff, K. J. (2009). The effect of sequencing errors on metagenomic gene prediction. *BMC Genomics*, 10, 520.
110. Garcia Martin, H., Ivanova, N., Kunin, V., Warnecke, F., Barry, K. W., McHardy, A. C., et al. (2006). Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nature Biotechnology*, 24(10), 1263-1269.
111. Sanapareddy, N., Hamp, T. J., Gonzalez, L. C., Hilger, H. A., Fodor, A. A., & Clinton, S. M. (2009). Molecular diversity of a North Carolina wastewater treatment plant as revealed by pyrosequencing. *Applied and Environmental Microbiology*, 75(6), 1688-1696.
112. Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N. N., et al. (2009). A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature*, 462(7276), 1056-1060.
113. Sharma, V. K., Kumar, N., Prakash, T., & Taylor, T. D. (2010). MetaBioME: a database to explore commercially useful enzymes in metagenomic datasets. *Nucleic Acids Research*, 38(Database issue), D468-472.
114. Cardenas, E., & Tiedje, J. M. (2008). New tools for discovering and characterizing microbial diversity. *Current Opinion in Biotechnology*, 19(6), 544-549.
115. Ferrer, M., Belouqui, A., Timmis, K. N., & Golyshin, P. N. (2009). Metagenomics for mining new genetic resources of microbial communities. *Journal of Molecular Microbiology and Biotechnology*, 16(1-2), 109-123.
116. Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews*, 68(4), 669-685.

117. Kennedy, J., Marchesi, J. R., & Dobson, A. D. (2008). Marine metagenomics: strategies for the discovery of novel enzymes with Biotechnologyogical applications from marine environments. *Microbial Cell Factories*, 7, 27.
118. Belouqui, A., Pita, M., Polaina, J., Martinez-Arias, A., Golyshina, O. V., Zumarraga, M., et al. (2006). Novel polyphenol oxidase mined from a metagenome expression library of bovine rumen: biochemical properties, structural analysis, and phylogenetic relationships. *Journal of Biological Chemistry*, 281(32), 22933-22942.
119. Bertrand, H., Poly, F., Van, V. T., Lombard, N., Nalin, R., Vogel, T. M., et al. (2005). High molecular weight DNA recovery from soils prerequisite for Biotechnologyogical metagenomic library construction. *Journal of Microbiological Methods*, 62(1), 1-11.
120. Sauer, U. (2006). Metabolic networks in motion: <sup>13</sup>C-based flux analysis. *Molecular Systems Biology*, 2, 62.
121. Courtois, S., Frostegard, A., Goransson, P., Depret, G., Jeannin, P., & Simonet, P. (2001). Quantification of bacterial subgroups in soil: comparison of DNA extracted directly from soil or from cells previously released by density gradient centrifugation. *Environmental Microbiology*, 3(7), 431-439.
122. Bodrossy, L., Stralis-Pavese, N., Konrad-Koszler, M., Weilharter, A., Reichenauer, T. G., Schofer, D., et al. (2006). mRNA-based parallel detection of active methanotroph populations by use of a diagnostic microarray. *Applied and Environmental Microbiology*, 72(2), 1672-1676.
123. Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., et al. (2004). The Pfam protein families database. *Nucleic Acids Research*, 32(Database issue), D138-141.
124. Demaneche, S., David, M. M., Navarro, E., Simonet, P., & Vogel, T. M. (2009). Evaluation of functional gene enrichment in a soil metagenomic clone library. *Journal of Microbiological Methods*, 76(1), 105-107.
125. Radajewski, S., Ineson, P., Parekh, N. R., & Murrell, J. C. (2000). Stable-isotope probing as a tool in microbial ecology. *Nature*, 403(6770), 646-649.
126. Uchiyama, T., Abe, T., Ikemura, T., & Watanabe, K. (2005). Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nature Biotechnology*, 23(1), 88-93.
127. Uchiyama, T., & Watanabe, K. (2008). Substrate-induced gene expression (SIGEX) screening of metagenome libraries. *Nature Protocols*, 3(7), 1202-1212.

128. Yamada, K., Terahara, T., Kurata, S., Yokomaku, T., Tsuneda, S., & Harayama, S. (2008). Retrieval of entire genes from environmental DNA by inverse PCR with pre-amplification of target genes using primers containing locked nucleic acids. *Environmental Microbiology*, 10(4), 978-987.
129. Boubakri, H., Beuf, M., Simonet, P., & Vogel, T. M. (2006). Development of metagenomic DNA shuffling for the construction of a xenobiotic gene. *Gene*, 375, 87-94.
130. Northen, T. R., Lee, J. C., Hoang, L., Raymond, J., Hwang, D. R., Yannone, S. M., et al. (2008). A nanostructure-initiator mass spectrometry-based enzyme activity assay. *Proceedings of the National Academy of Sciences*, 105(10), 3678-3683.
131. Knietzsch, A., Waschowitz, T., Bowien, S., Henne, A., & Daniel, R. (2003). Construction and screening of metagenomic libraries derived from enrichment cultures: generation of a gene bank for genes conferring alcohol oxidoreductase activity on *Escherichia coli*. *Applied and Environmental Microbiology*, 69(3), 1408-1416.
132. Knietzsch, A., Waschowitz, T., Bowien, S., Henne, A., & Daniel, R. (2003). Metagenomes of complex microbial consortia derived from different soils as sources for novel genes conferring formation of carbonyls from short-chain polyols on *Escherichia coli*. *Journal of Molecular Microbiology and Biotechnology*, 5(1), 46-56.
133. Lee, S. W., Won, K., Lim, H. K., Kim, J. C., Choi, G. J., & Cho, K. Y. (2004). Screening for novel lipolytic enzymes from uncultured soil microorganisms. *Applied Microbiology and Biotechnology*, 65(6), 720-726.
134. Walter, J., Mangold, M., & Tannock, G. W. (2005). Construction, analysis, and beta-glucanase screening of a bacterial artificial chromosome library from the large-bowel microbiota of mice. *Applied and Environmental Microbiology*, 71(5), 2347-2354.
135. Ranjan, R., Grover, A., Kapardar, R. K., & Sharma, R. (2005). Isolation of novel lipolytic genes from uncultured bacteria of pond water. *Biochemical and Biophysical Research Communications*, 335(1), 57-65.
136. Erwin, D. P., Erickson, I. K., Delwiche, M. E., Colwell, F. S., Strap, J. L., & Crawford, R. L. (2005). Diversity of oxygenase genes from methane- and ammonia-oxidizing bacteria in the Eastern Snake River Plain aquifer. *Applied and Environmental Microbiology*, 71(4), 2016-2025.
137. Cottrell, M. T., Yu, L., & Kirchman, D. L. (2005). Sequence and expression analyses of Cytophaga-like hydrolases in a Western arctic metagenomic library and the Sargasso Sea. *Applied and Environmental Microbiology*, 71(12), 8506-8513.



138. Ferrer, M., Golyshina, O. V., Chernikova, T. N., Khachane, A. N., Reyes-Duarte, D., Santos, V. A., et al. (2005). Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environmental Microbiology*, 7(12), 1996-2010.
139. Reyes-Duarte, D., Polaina, J., Lopez-Cortes, N., Alcalde, M., Plou, F. J., Elborough, K., et al. (2005). Conversion of a carboxylesterase into a triacylglycerol lipase by a random mutation. *Angewandte Chemie International Edition England*, 44(46), 7553-7557.
140. Ferrer, M., Beloqui, A., Golyshina, O. V., Plou, F. J., Neef, A., Chernikova, T. N., et al. (2007). Biochemical and structural features of a novel cyclodextrinase from cow rumen metagenome. *Biotechnology Journal*, 2(2), 207-213.
141. Lopez-Cortes, N., Reyes-Duarte, D., Beloqui, A., Polaina, J., Ghazi, I., Golyshina, O. V., et al. (2007). Catalytic role of conserved HQGE motif in the CE6 carbohydrate esterase family. *FEBS Letters*, 581(24), 4657-4662.
142. Ferrer, M., Golyshina, O. V., Chernikova, T. N., Khachane, A. N., Martins Dos Santos, V. A., Yakimov, M. M., et al. (2005). Microbial enzymes mined from the Urania deep-sea hypersaline anoxic basin. *Chemistry & Biology*, 12(8), 895-904.
143. Roh, C., Villatte, F., Kim, B. G., & Schmid, R. D. (2007). Screening and purification for novel cytochrome b5 from uncultured environmental micro-organisms. *Letters in Applied Microbiology*, 44(5), 475-480.
144. Lee, M. H., Lee, C. H., Oh, T. K., Song, J. K., & Yoon, J. H. (2006). Isolation and characterization of a novel lipase from a metagenomic library of tidal flat sediments: evidence for a new family of bacterial lipases. *Applied and Environmental Microbiology*, 72(11), 7406-7409.
145. Kim, Y. J., Choi, G. S., Kim, S. B., Yoon, G. S., Kim, Y. S., & Ryu, Y. W. (2006). Screening and characterization of a novel esterase from a metagenomic library. *Protein Expression and Purification*, 45(2), 315-323.
146. Suenaga, H., Ohnuki, T., & Miyazaki, K. (2007). Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. *Environmental Microbiology*, 9(9), 2289-2297.
147. Shimamura, M., Nishiyama, T., Shigetomo, H., Toyomoto, T., Kawahara, Y., Furukawa, K., et al. (2007). Isolation of a multiheme protein with features of a hydrazine-oxidizing enzyme from an anaerobic ammonium-oxidizing enrichment culture. *Applied and Environmental Microbiology*, 73(4), 1065-1072.
148. Roh, C., & Villatte, F. (2008). Isolation of a low-temperature adapted lipolytic enzyme from uncultivated micro-organism. *Journal of Applied Microbiology*, 105(1), 116-123.

149. Pearson, A., Flood Page, S. R., Jorgenson, T. L., Fischer, W. W., & Higgins, M. B. (2007). Novel hopanoid cyclases from the environment. *Environmental Microbiology*, 9(9), 2175-2188.
150. Park, S. Y., Kim, J. T., Kang, S. G., Woo, J. H., Lee, J. H., Choi, H. T., et al. (2007). A new esterase showing similarity to putative dienelactone hydrolase from a strict marine bacterium, *Vibrio* sp. GMD509. *Applied Microbiology and Biotechnology*, 77(1), 107-115.
151. Feng, Y., Duan, C. J., Pang, H., Mo, X. C., Wu, C. F., Yu, Y., et al. (2007). Cloning and identification of novel cellulase genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulases. *Applied Microbiology and Biotechnology*, 75(2), 319-328.
152. Byun, J. S., Rhee, J. K., Kim, N. D., Yoon, J., Kim, D. U., Koh, E., et al. (2007). Crystal structure of hyperthermophilic esterase EstE1 and the relationship between its dimerization and thermostability properties. *BMC Structural Biology*, 7, 47.
153. Couto, G.H., Glogauer, A., Faoro, H., Chubatsu, L.S., Souza, E.M., & Pedrosa, F.O. (2010). Isolation of a novel lipase from a metagenomic library derived from mangrove sediment from the south Brazilian coast. *Genetics and Molecular Research* 9(1), 514-523.
154. LeCleir, G. R., Buchan, A., Maurer, J., Moran, M. A., & Hollibaugh, J. T. (2007). Comparison of chitinolytic enzymes from an alkaline, hypersaline lake and an estuary. *Environmental Microbiology*, 9(1), 197-205.
155. Lammle, K., Zipper, H., Breuer, M., Hauer, B., Buta, C., Brunner, H., et al. (2007). Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. *Journal of Biotechnology*, 127(4), 575-592.
156. Jiang, C., & Wu, B. (2007). Molecular cloning and functional characterization of a novel decarboxylase from uncultured microorganisms. *Biochemical and Biophysical Research Communications*, 357(2), 421-426.
157. Hong, K. S., Lim, H. K., Chung, E. J., Park, E. J., Lee, M. H., Kim, J. C., et al. (2007). Selection and characterization of forest soil metagenome genes encoding lipolytic enzymes. *Journal of Microbiology and Biotechnology*, 17(10), 1655-1660.
158. Chen, Z. W., Liu, Y. Y., Wu, J. F., She, Q., Jiang, C. Y., & Liu, S. J. (2007). Novel bacterial sulfur oxygenase reductases from bioreactors treating gold-bearing concentrates. *Applied Microbiology and Biotechnology*, 74(3), 688-698.
159. Ono, A., Miyazaki, R., Sota, M., Ohtsubo, Y., Nagata, Y., & Tsuda, M. (2007). Isolation and characterization of naphthalene-catabolic genes and plasmids from oil-contaminated soil by using two cultivation-independent approaches. *Applied Microbiology and Biotechnology*, 74(2), 501-510.

160. van Hellemond, E. W., Janssen, D. B., & Fraaije, M. W. (2007). Discovery of a novel styrene monooxygenase originating from the metagenome. *Applied and Environmental Microbiology*, 73(18), 5832-5839.
161. Hardeman, F., & Sjoling, S. (2007). Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment. *FEMS Microbiology Ecology*, 59(2), 524-534.
162. Xu, M., Xiao, X., & Wang, F. (2008). Isolation and characterization of alkane hydroxylases from a metagenomic library of Pacific deep-sea sediment. *Extremophiles*, 12(2), 255-262.
163. Tirawongsaroj, P., Sriprang, R., Harnpicharnchai, P., Thongaram, T., Champreda, V., Tanapongpipat, S., et al. (2008). Novel thermophilic and thermostable lipolytic enzymes from a Thailand hot spring metagenomic library. *Journal of Biotechnology*, 133(1), 42-49.
164. Chu, X., He, H., Guo, C., & Sun, B. (2008). Identification of two novel esterases from a marine metagenomic library derived from South China Sea. *Applied Microbiology and Biotechnology*, 80(4), 615-625.
165. Chae, J. C., Song, B., & Zylstra, G. J. (2008). Identification of genes coding for hydrolytic dehalogenation in the metagenome derived from a denitrifying 4-chlorobenzoate degrading consortium. *FEMS Microbiology Letters*, 281(2), 203-209.
166. Mayumi, D., Akutsu-Shigeno, Y., Uchiyama, H., Nomura, N., & Nakajima-Kambe, T. (2008). Identification and characterization of novel poly(DL-lactic acid) depolymerases from metagenome. *Applied Microbiology and Biotechnology*, 79(5), 743-750.
167. Li, G., Wang, K., Liu, Y.H. (2008) Molecular cloning and characterization of a novel pyrethroid-hydrolyzing esterase originating from the Metagenome. *Microbial Cell Factories* 7 (1), 38.
168. Lee, C. M., Yeo, Y. S., Lee, J. H., Kim, S. J., Kim, J. B., Han, N. S., et al. (2008). Identification of a novel 4-hydroxyphenylpyruvate dioxygenase from the soil metagenome. *Biochemical and Biophysical Research Communications*, 370(2), 322-326.
169. Kim, S. J., Lee, C. M., Han, B. R., Kim, M. Y., Yeo, Y. S., Yoon, S. H., et al. (2008). Characterization of a gene encoding cellulase from uncultured soil bacteria. *FEMS Microbiology Letters*, 282(1), 44-51.
170. Hu, Y., Zhang, G., Li, A., Chen, J., & Ma, L. (2008). Cloning and enzymatic characterization of a xylanase gene from a soil-derived metagenomic library with an efficient approach. *Applied Microbiology and Biotechnology*, 80(5), 823-830.

171. Song, J.S., Jeon, J.H., Lee, J.H., Jeong, S.H., Jeong, B.C., Kim, S.J., Lee, J.H., & Lee, S.H. (2005). Molecular characterization of TEM-type beta-lactamases identified in cold-seep sediments of Edison Seamount (south of Lihir Island, Papua New Guinea). *Journal of Microbiology* 43(2), 172-178.
172. Riaz, K., Elmerich, C., Moreira, D., Raffoux, A., Dessaux, Y., & Faure, D. (2008). A metagenomic analysis of soil bacteria extends the diversity of quorum-quenching lactonases. *Environmental Microbiology*, 10(3), 560-570.
173. Wu, C., & Sun, B. (2009). Identification of novel esterase from metagenomic library of Yangtze river. *Journal of Microbiology and Biotechnology*, 19(2), 187-193.
174. Wei, P., Bai, L., Song, W., & Hao, G. (2009). Characterization of two soil metagenome-derived lipases with high specificity for p-nitrophenyl palmitate. *Archives of Microbiology*, 191(3), 233-240.
175. Waschkowitz, T., Rockstroh, S., & Daniel, R. (2009). Isolation and characterization of metalloproteases with a novel domain structure by construction and screening of metagenomic libraries. *Applied and Environmental Microbiology*, 75(8), 2506-2516.
176. Suenaga, H., Mizuta, S., & Miyazaki, K. (2009). The molecular basis for adaptive evolution in novel extradiol dioxygenases retrieved from the metagenome. *FEMS Microbiology Ecology*, 69(3), 472-480.
177. Pang, H., Zhang, P., Duan, C. J., Mo, X. C., Tang, J. L., & Feng, J. X. (2009). Identification of cellulase genes from the metagenomes of compost soils and functional characterization of one novel endoglucanase. *Current Microbiology*, 58(4), 404-408.
178. Morimoto, S., & Fujii, T. (2009). A new approach to retrieve full lengths of functional genes from soil by PCR-DGGE and metagenome walking. *Applied Microbiology and Biotechnology*, 83(2), 389-396.
179. Heath, C., Hu, X. P., Cary, S. C., & Cowan, D. (2009). Identification of a novel alkaliphilic esterase active at low temperatures by screening a metagenomic library from antarctic desert soil. *Applied and Environmental Microbiology*, 75(13), 4657-4659.
180. Duan, C. J., Xian, L., Zhao, G. C., Feng, Y., Pang, H., Bai, X. L., et al. (2009). Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens. *Journal of Applied Microbiology*, 107(1), 245-256.
181. Chen, I. C., Lin, W. D., Hsu, S. K., Thiruvengadam, V., & Hsu, W. H. (2009). Isolation and characterization of a novel lysine racemase from a soil metagenomic library. *Applied and Environmental Microbiology*, 75(15), 5161-5166.

182. Brennerova, M. V., Josefiova, J., Brenner, V., Pieper, D. H., & Junca, H. (2009). Metagenomics reveals diversity and abundance of meta-cleavage pathways in microbial communities from soil highly contaminated with jet fuel under air-sparging bioremediation. *Environmental Microbiology*, 11(9), 2216-2227.
183. Math, R. K., Asraful Islam, S. M., Cho, K. M., Hong, S. J., Kim, J. M., Yun, M. G., et al. (2009). Isolation of a novel gene encoding a 3,5,6-trichloro-2-pyridinol degrading enzyme from a cow rumen metagenomic library. *Biodegradation*.
184. Zhang, T., Han, W.J., Liu, Z.P. (2009) Gene cloning and characterization of a novel esterase from activated sludge metagenome. *Microbial Cell Factories*. 8(1), 67.
185. Ye, M., Li, G., Liang, W. Q., & Liu, Y. H. (2010). Molecular cloning and characterization of a novel metagenome-derived multicopper oxidase with alkaline laccase activity and highly soluble expression. *Applied Microbiology and Biotechnology*.
186. Cieslinski, H., Bialkowska, A., Tkaczuk, K., Dlugolecka, A., Kur, J., & Turkiewicz, M. (2009). Identification and molecular modeling of a novel lipase from an Antarctic soil metagenomic library. *Polish Journal of Microbiology*, 58(3), 199-204.
187. Hu, Y., Fu, C., Huang, Y., Yin, Y., Cheng, G., Lei, F., et al. (2010). Novel lipolytic genes from the microbial metagenomic library of the South China Sea marine sediment. *FEMS Microbiology Ecology*.
188. Maróti, G., Tong, Y., Yooseph, S., Baden-Tillson, H., Smith, H.O., Kovács, K.L., Frazier, M., Venter, J.C., Xu, Q. (2009) Discovery of [NiFe] hydrogenase genes in metagenomic DNA: cloning and heterologous expression in *Thiocapsa roseopersicina*. *Applied Environmental Microbiology*. 75(18), 5821-5830.
189. Marcos, M. S., Lozada, M., & Dionisi, H. M. (2009). Aromatic hydrocarbon degradation genes from chronically polluted Subantarctic marine sediments. *Letters in Applied Microbiology*, 49(5), 602-608.
190. Meilleur, C., Hupe, J. F., Juteau, P., & Shareck, F. (2009). Isolation and characterization of a new alkali-thermostable lipase cloned from a metagenomic library. *Journal of Industrial Microbiology and Biotechnology*, 36(6), 853-861.
191. Liu, K., Wang, J., Bu, D., Zhao, S., McSweeney, C., Yu, P., Li, D. (2009) Isolation and biochemical characterization of two lipases from a metagenomic library of China Holstein cow rumen. *Biochemical and Biophysical Research Communications*. 385(4), 605-611.
192. Li, L. L., McCorkle, S. R., Monchy, S., Taghavi, S., & van der Lelie, D. (2009). Bioprospecting metagenomes: glycosyl hydrolases for converting biomass. *Biotechnology Biofuels*, 2, 10.

193. Huang, Y., Lai, X., He, X., Cao, L., Zeng, Z., Zhang, J., et al. (2009). Characterization of a deep-sea sediment metagenomic clone that produces water-soluble melanin in *Escherichia coli*. *Mar Biotechnology (NY)*, 11(1), 124-131.
194. Jeon, J. H., Kim, J. T., Kim, Y. J., Kim, H. K., Lee, H. S., Kang, S. G., et al. (2009). Cloning and characterization of a new cold-active lipase from a deep-sea sediment metagenome. *Applied Microbiology and Biotechnology*, 81(5), 865-874.
195. Hjort, K., Bergstrom, M., Adesina, M. F., Jansson, J. K., Smalla, K., & Sjoling, S. (2009). Chitinase genes revealed and compared in bacterial isolates, DNA extracts and a metagenomic library from a phytopathogen-suppressive soil. *FEMS Microbiology Ecology*.
196. Sharma, S., Khan, F. G., & Qazi, G. N. (2010). Molecular cloning and characterization of amylase from soil metagenomic library derived from Northwestern Himalayas. *Applied Microbiology and Biotechnology*, 86(6), 1821-1828.
197. Kim, Y.H., Kwon, E.J., Kim, S.K., Jeong, Y.S., Kim, J., Yun, H.D., Kim, H. (2010) Molecular cloning and characterization of a novel family VIII alkaline esterase from a compost metagenomic library. *Biochemical and Biophysical Research Communications*. 393(1), 45-49.
198. Vieites, J.M., Ghazi, A., Beloqui, A., Polaina, J., Andreu, J.M., Golyshina, O.V., Nechitaylo, T.Y., Waliczek, A., Yakimov, M.M., Golyshin, P.N., Ferrer, M. (2009) Inter-conversion of catalytic abilities in a bifunctional carboxyl/feruloyl-esterase from earthworm gut metagenome. *Microbial Biotechnology*. In press.
199. Henning, H., Leggewie, C., Pohl, M., Muller, M., Eggert, T., & Jaeger, K. E. (2006). Identification of novel benzoylformate decarboxylases by growth selection. *Applied and Environmental Microbiology*, 72(12), 7510-7517.
200. Michels, A., Puetz, A., Maurer, K.H., Eggert, T., Jaeger, K.E. (2007) Use of esterases for separating plastics. WO 2007017181.