



UNIVERSIDAD NACIONAL DE COLOMBIA

Búsqueda de genes de resistencia a arsénico en el metagenoma microbiano de la Sabana de Bogotá

Sergio Mauricio Latorre Ochoa

Universidad Nacional de Colombia

Facultad de Ciencias, Instituto de Biotecnología de la Universidad Nacional (IBUN)

Bogotá, Colombia

2014

Búsqueda de genes de resistencia a arsénico en el metagenoma microbiano de la Sabana de Bogotá

Sergio Mauricio Latorre Ochoa

Tesis o trabajo de investigación presentada como requisito parcial para optar al título de:
Magister en Ciencias - Microbiología

Director (a):

Pedro Filipe de Brito Brandão, Ph.D.

Grupo de Investigación:

Grupo de Estudios para la Remediación y Mitigación de Impactos Negativos al Ambiente
(GERMINA)

Universidad Nacional de Colombia

Facultad de Ciencias, Instituto de Biotecnología de la Universidad Nacional (IBUN)

Bogotá, Colombia

2014

*Evidentemente, dedico este trabajo a mis
padres y hermanos; mis soportes incondicionales.*

Agradecimientos

Agradezco a mi maestro Pedro Filipe de Brito Brandão, quien con su guía, conocimiento, experiencia y excelente pedagogía, hizo posible la realización de este trabajo.

A las profesoras María Teresa Reguero, Martha Raquel Fontanilla y Esperanza Torres Rojas, así como al profesor Camilo Ernesto López Carrascal, quienes con sus excelentes cátedras aportaron gran parte de los conceptos y métodos centrales de este trabajo.

A Howard Junca, Cesar Osorio y Katherine Carillo quienes desde su ejercicio investigativo en *CorpoGen*, fueron soporte y ayuda fundamental para el desarrollo experimental. Así mismo al equipo de trabajo de *Manihot Biotec*, en especial a Andrea Vásquez y Elizabeth Contreras, quienes apoyaron con su experiencia investigativa parte fundamental de este trabajo.

A mis compañeros del laboratorio de Microbiología Ambiental y Aplicada y al grupo *GERMINA*, en especial a Yina Cifuentes, Sandra Montaña, Diana Tamayo, Ángela García, Iván López, Tatiana Morales y Aida Ramírez, quienes con su compañía, aportes y críticas ayudaron a mejorar este trabajo.

A la Universidad Nacional y la DIB, que mediante el proyecto de apoyo a posgrado 11973 financió parte de este trabajo.

A Colciencias, que mediante el contrato RC No. 454 - 2011 financió la mayor parte de este trabajo de investigación y por soporte con una beca de Jóvenes Investigadores.

Resumen

El arsénico, como metaloide carcinogénico, de alto riesgo para la salud humana y el medio ambiente, ha incrementado su concentración en varios ecosistemas de la Sabana de Bogotá, especialmente en la cuenca media del Río Bogotá. La alta utilización de pesticidas en los terrenos hortícolas, así como la descarga de aguas de desecho industrial y doméstico han generado condiciones *in situ* de resistencia y adaptación microbiana a la presión de selección natural de varios contaminantes, metales pesados y metaloides.

En este trabajo se realizó una búsqueda de genes involucrados en la resistencia a arsénico, presentes en las comunidades bacterianas endémicas de distintos ambientes de la Sabana de Bogotá, mediante un enfoque metagenómico. Se construyó una librería metagenómica de aproximadamente 50.000 clones a partir de muestras de suelos y sedimentos de la Sabana de Bogotá. Mediante un tamizaje que empleó arsenato como agente de selección, se logró la identificación de 11 clones de resistencia a altas concentraciones de arsénico y se identificaron genes reportados del operón *ars*, que están involucrados en la reducción de este metaloide.

Palabras clave: Arsénico, Genes de resistencia, Metagenómica, Sabana de Bogotá, Biotecnología ambiental.

Abstract

Arsenic, as a human and environmental hazardous agent, has raised its concentration in many ecosystems of the Bogota Savannah, especially in the middle basin of the Bogota River. The constant utilization of pesticides in farmlands, together with the discharge of domestic and industrial sewage, has contributed to the generation of *in situ* processes of microbial resistance and adaptation to those contaminants, including heavy metals and metalloids.

A search of arsenic resistance genes present in bacterial endemic communities of distinct ecosystems from the Bogota Savannah was conducted using a metagenomic approach. A soil and sediment based library of about 50.000 clones was performed. An arsenate based functional screening allowed the identification of 11 clones resistant to high arsenic concentrations. Its genomic sequences revealed known genes related to arsenic operon system *ars* which are involved in arsenic reduction processes.

Keywords: Arsenic, Resistance genes, Metagenomics, Bogota Savannah, Environmental biotechnology

Contenido

| | |
|--|-------------|
| Resumen..... | IX |
| Abstract..... | X |
| Lista de Símbolos..... | XIII |
| Consideraciones éticas..... | XIV |
| Introducción..... | 1 |
| | |
| 1. Estado del Arte | 5 |
| 1.1 Generalidades del arsénico..... | 5 |
| 1.2 Arsénico en el mundo y América Latina | 7 |
| 1.3 Arsénico en Colombia | 8 |
| 1.4 Metabolismo microbiano del arsénico..... | 12 |
| 1.5 Bioremediación del arsénico | 16 |
| 1.6 Metagenómica..... | 18 |
| 1.7 Metagenómica y bioremediación | 22 |
| 1.8 Metagenómica y ensayos funcionales | 24 |
| 1.9 Metagenómica y secuenciación..... | 24 |
| 1.10 Herramientas bioinformáticas para la búsqueda de genes | 26 |
| 1.11 Conclusiones del estado del arte..... | 28 |
| | |
| 2. Materiales y Métodos | 30 |
| 2.1 Muestreo | 30 |
| 2.1.1 Muestras de suelo | 32 |
| 2.1.2 Muestras de sedimentos..... | 33 |
| 2.1.3 Muestras de agua..... | 34 |
| 2.1.4 Manejo de las muestras..... | 35 |
| 2.1.5 Análisis de muestras..... | 36 |
| 2.2 Extracción de ADN..... | 36 |
| 2.3 Selección de tamaños y purificación del ADN | 37 |
| 2.4 Arreglo de extremos de los fragmentos de ADN..... | 37 |
| 2.5 Ligación..... | 38 |
| 2.6 Empaquetamiento | 38 |
| 2.7 Infección..... | 39 |
| 2.8 Librería metagenómica..... | 39 |
| 2.9 Validación de umbrales | 39 |
| 2.10 Ensayo de tamizaje funcional | 40 |
| 2.11 Búsqueda molecular de elementos conservados en los clones positivos | 41 |
| 2.12 Preparación de ADN fosmídico | 42 |
| 2.13 Estrategia de secuenciación..... | 42 |
| 2.14 Ensamblaje de las secuencias | 43 |
| 2.15 Preparación de ORF's..... | 44 |
| 2.16 Anotación funcional..... | 44 |

| | | |
|-----------|--|------------|
| 2.17 | Búsqueda e identificación de genes de resistencia a arsénico..... | 45 |
| 3. | Resultados y Discusión..... | 46 |
| 3.1 | Creación de una librería metagenómica..... | 46 |
| 3.1.1 | Concentraciones de arsénico de las muestras ambientales | 46 |
| 3.1.2 | Extracción y preparación de ADN de alto peso molecular | 47 |
| 3.1.3 | Proceso de creación de la librería metagenómica | 50 |
| 3.2 | Ensayo de tamizaje funcional..... | 52 |
| 3.2.1 | Validación de umbrales de arsénico y de medio de cultivo | 52 |
| 3.2.2 | Tamizaje funcional | 54 |
| 3.2.3 | Ensayo de validación molecular | 56 |
| 3.3 | Secuenciación de clones resistentes a arsénico y ensamblaje de secuencias.. | 57 |
| 3.3.1 | Preparación del ADN fosmídico..... | 57 |
| 3.3.2 | Obtención de secuencias | 59 |
| 3.4 | Análisis de secuencias y anotación funcional..... | 61 |
| 3.4.1 | Ensamblaje de las secuencias | 61 |
| 3.4.2 | Preparación e identificación de marcos abiertos de lectura (<i>ORF</i>) | 63 |
| 3.4.3 | Anotación jerarquizada (<i>MG-RAST</i>) | 63 |
| 3.4.4 | Anotación mediante grupos ortólogos | 66 |
| 3.4.5 | Anotación mediante dominios conservados..... | 67 |
| 3.5 | Búsqueda e identificación de genes de resistencia a arsénico..... | 70 |
| 4. | Conclusiones y Recomendaciones | 72 |
| 4.1 | Conclusiones | 72 |
| 4.2 | Recomendaciones | 73 |
| A. | Anexo: Protocolo modificado para extracción de ácidos nucleicos de alto peso molecular con PowerMax™ Soil DNA Isolation Kit (MOBIO®)..... | 75 |
| B. | Anexo: Electroferograma ejemplo de secuencias Sanger..... | 76 |
| C. | Anexo: Alineamientos <i>BLAST</i> con elementos relacionados a la resistencia de arsénico..... | 79 |
| D. | Anexo: Publicaciones y participaciones en eventos..... | 96 |
| | Bibliografía..... | 111 |

Lista de Símbolos

Abreviatura Término

| | |
|-------------------------|-----------------------------------|
| <i>ADN</i> | Ácido desoxirribonucleico |
| <i>As(III)</i> | Arsenito |
| <i>As(V)</i> | Arsenato |
| <i>ATP</i> | Adenosin trifosfato |
| <i>BAC</i> | Bacterial Artificial Chromosome |
| <i>BLAST</i> | Basic Local Alignment Search Tool |
| <i>dNTP</i> | Desoxirribonucleotido |
| <i>LB</i> | Luria-Bertani |
| <i>MgCl₂</i> | Cloruro de magnesio |
| <i>MgSO₄</i> | Sulfato de magnesio |
| <i>ORF</i> | Open Reading Frame |
| <i>PCR</i> | Polimerase Chain Reaction |
| <i>SDS</i> | Dodecilsulfato de sodio |
| <i>UV</i> | Ultra Violeta |

Consideraciones éticas

En este proyecto se consideraron los aspectos éticos relacionados en las normas científicas, técnicas y administrativas para la investigación en cuestiones ambientales. En este trabajo se incluyó el uso de muestras ambientales (suelos, sedimentos y aguas), concretamente de la Sabana de Bogotá, Colombia. Por esta razón se cumplió con los requerimientos del Decreto 309 de 2000, expedido por el Ministerio del Medio Ambiente, por el cual se reglamenta la investigación científica sobre diversidad biológica. Para la ejecución del presente proyecto se utilizaron la recolección de muestras ambientales para lo cual se requiere de permisos por parte de las autoridades gubernamentales y ambientales competentes. En este sentido, se solicitó a través de la Vicerrectoría de Investigación de la Universidad Nacional de Colombia, los respectivos permisos los cuales fueron aprobados e incluyeron: 1) certificado de la no presencia de comunidades indígenas o negras en el área de influencia del proyecto (resolución No. 117 del 19 de octubre de 2011 expedida por el Ministerio del Interior); 2) trámite de contrato marco de acceso a recursos genéticos para inclusión del proyecto y certificado de inclusión del proyecto "Búsqueda de genes resistentes a arsénico en el metagenoma microbiano de la Sabana de Bogotá" en el permiso marco de estudio con fines de investigación científica en diversidad biológica No. 8 de 2010 anexo 26 expedido por la Autoridad Nacional de Licencias Ambientales (ANLA) del Ministerio de Ambiente, Vivienda y de Desarrollo Territorial.

El proyecto no generó impacto negativo apreciable en el ambiente, ya que fue realizado a nivel de laboratorio. La mayoría de pruebas y manipulación de las muestras se adelantaron en el Laboratorio de Microbiología Ambiental y Aplicada del Departamento de Química, de la Universidad Nacional de Colombia, y por esta razón se acataron todas las normas emitidas para el trabajo en este lugar y para la disposición de residuos que se generaron en el proyecto. Además, los reactivos tóxicos conocidos o material biológico contaminante fueron desechados ajustándose a las normas nacionales y a estándares internacionales. Los investigadores que hacen parte de este proyecto no tienen conflictos de interés derivados de la investigación propuesta.

Introducción

El arsénico es un metaloide ampliamente distribuido en la corteza terrestre. Su disposición se da por fenómenos naturales como la meteorización, actividad volcánica y actividad microbiana, así como por acción antrópica, principalmente por la minería y la realización de varios procesos industriales. Su especiación y dinámica es variada dependiendo de las variaciones de los potenciales *redox* del medio exterior, así como por la presencia de otros elementos como azufre y algunos metales.

Como agente tóxico y de prioridad mundial ha sido ampliamente monitoreado y estudiado en la región Indochina y en menor medida en la región Ibérica así como en el cono sur de América Latina. No obstante, aún es escasa la información de la presencia, concentraciones y dinámica del metaloide en varios países del mundo, incluyendo Colombia.

Recientes trabajos de revisión sugieren un panorama de riesgo potencial en gran parte del territorio colombiano, magnificado por la agudización de la explotación minera en varias regiones montañosas y por los procesos crecientes de densificación demográfica en las principales ciudades.

La Sabana de Bogotá se convierte en una región en la que convergen riesgos potenciales de contaminación, dada la existencia de contaminantes en la cuenca del Río Bogotá y los desechos domésticos e industriales que generan condiciones de riesgo de toxicidad por varios metales pesados y metaloides como el arsénico. Además por la característica de sus suelos ácidos, los que generan condiciones de adsorción y retención para las especies solubles de arsénico.

Este panorama no ha estado acompañado de adelantos en mecanismos de vigilancia y políticas públicas conducentes a la eliminación de estos compuestos tóxicos, toda vez que sus inversiones son altas y sus riesgos aún no han sido cuantificados.

La bioremediación de ambientes contaminados aparece como una herramienta que ha demostrado reducir significativamente los costos de inversión y de operación en comparación con otras técnicas tradicionales. Por ello, su desarrollo e investigación son prioritarias para el panorama de riesgo potencial del arsénico en Colombia.

La búsqueda e identificación de los mecanismos genéticos de los procesos de bioremediación es el primer paso necesario para posibles diseños biotecnológicos encaminados a la solución de este tipo de problemáticas.

Un importante antecedente del presente trabajo, es el trabajo desarrollado por Carrillo (2012), quien a partir de 11 aislamientos de la Sabana de Bogotá y 17 del departamento de Cesar en Colombia, con resistencia fenotípica a arsénico, identificó la presencia de una región conservada (gen *arsB*), la cual está presente en los mecanismos descritos de reducción microbiana del arsenato en arsenito. Este trabajo abre una perspectiva para la utilización de herramientas y técnicas que permitan la búsqueda masiva de mayor diversidad genética de mecanismos de resistencia al metaloide.

La literatura sugiere una amplia diversidad de mecanismos genéticos involucrados en el metabolismo microbiano del arsénico, razón por la que es promisoría la identificación de nuevas bases genéticas que expliquen de manera suficiente la relación de los microorganismos con el metaloide y puedan generar caminos para el diseño de soluciones.

Tecnologías como la metagenómica, la secuenciación masiva ácidos nucleicos y el desarrollo de paquetes y algoritmos bioinformáticos ayudan a resolver gran parte de la búsqueda necesaria de estos mecanismos genéticos, debido a la recuperación masiva de información genética y sus relaciones, a la cual no es posible acceder a través de aislamientos de organismos individuales; enfoque que privilegian las técnicas clásicas de bioprospección.

El objetivo central del presente trabajo fue la identificación de genes de resistencia a arsénico en el metagenoma microbiano de muestras ambientales de la Sabana de Bogotá.

Para ello fue necesario: 1) generar librerías metagenómicas a partir de distintas muestras ambientales de la Sabana de Bogotá, 2) identificar y seleccionar *in vitro* clones con resistencia a arsénico en las librerías metagenómicas, 3) obtener las secuencias de los clones que confieren resistencia a arsénico en los clones seleccionados y, 4) analizar las secuencias obtenidas mediante herramientas bioinformáticas. Esta Tesis se presenta en cuatro secciones principales: la primera de ella realiza una revisión del marco teórico de los aspectos centrales en el desarrollo de esta tesis; posteriormente se presentan los materiales y métodos, resultados y discusión, sucesivamente. Finalmente se presentan las principales conclusiones y recomendaciones del trabajo de investigación.

1.Estado del Arte

1.1 Generalidades del arsénico

El arsénico es un elemento con distribución amplia en la biósfera, atmósfera e hidrósfera. Se estima que comprende alrededor de 5×10^{-4} % de la corteza terrestre. El arsénico proviene tanto de fuentes naturales como de tipo antrópico. La actividad volcánica, la meteorización de las rocas y la acción microbiana sobre las mismas, constituyen la principal fuente de disponibilidad natural del arsénico. Además, actividades como la minería, la fundición de metales, diversos tipos de procesos industriales y la utilización de plaguicidas, constituyen en gran medida, el aporte del ser humano a la dinámica de este metaloide. El arsénico se encuentra naturalmente en asociaciones con el Azufre, así como a otros minerales, entre los que se destacan Manganeso, Hierro, Cobalto, Níquel, Plata y Estaño. Estas asociaciones se encuentran en forma de minerales como la arsenopirita, el rejalgar, oropimente, arsenolita, principalmente.

El arsénico es uno de los primeros elementos caracterizados y reconocidos como agente carcinogénico. Debido al riesgo mundial a causa de este metaloide, son varios los organismos internacionales como la Agencia de Protección Ambiental de los Estados Unidos y la Organización Mundial de la Salud, que han reglamentado en varios países la reducción en el nivel máximo permitido en aguas para consumo humano pasando en la mayoría de casos de 50 a 10 $\mu\text{g/L}$ (Mirza *et al.*, 2010; Páez-Espino *et al.*, 2009). Actualmente, el arsénico se encuentra clasificado como un elemento carcinogénico del grupo I o carcinogénico humano por la Agencia Internacional de Investigaciones en Cáncer y por la Organización Mundial de la Salud (WHO & IARC, 1998; 2004).

La problemática mundial del riesgo del arsénico móvil tiende a aumentar, pues son varias las actividades humanas e industriales que producen como desecho tóxico, cargas considerables del metaloide en formas biodisponibles. De las actividades más importantes al respecto son el desarrollo de la minería, la quema de combustibles fósiles,

el uso de pesticidas y herbicidas en cultivos y la utilización de preservantes de maderas, entre otros. Dichas actividades tienen un efecto directo sobre la contaminación en cuerpos hídricos debido a la liberación del arsénico en formas solubles. Así mismo, el fenómeno de contaminación se extiende en suelos aledaños y con vocación agrícola (Mirza *et al.*, 2010).

Se ha demostrado que la forma más común de intoxicación por arsénico es por medio de la dieta diaria, así como por la ingestión de aguas contaminadas con el metaloide. Los estudios toxicológicos, demuestran que tanto el arsénico en su forma pentavalente como trivalente, son rápidamente absorbidos por el tracto gastrointestinal en animales (Phan *et al.*, 2010). La ingestión permanente de arsénico está relacionada a procesos de intoxicación crónica, los cuales pueden desencadenar en daños al sistema nervioso central, hígado, piel, y apariciones de tipos de cáncer como hiperqueratosis, arsenicosis y cáncer de piel e hígado (Li *et al.*, 2010; Litter *et al.*, 2010). Los reportes indican que son las regiones y países más pobres del planeta, quienes sufren con mayor severidad el impacto de problemas por intoxicación por arsénico debido a las condiciones naturales y antrópicas que hacen que se encuentre preferencialmente en especies solubles y por lo tanto disponibles (Litter *et al.*, 2010).

Dada la múltiple valencia del arsénico, este se puede encontrar en varios estados. Los estados redox más estables son: -3 (Gas arsino, AsH_3), -1 (Alquil arsénico), 0 (Arsénico elemental), +3 (Arsenitos) y +5 (Arsenatos), siendo los dos últimos estados los más comunes en soluciones acuosas. La forma reducida trivalente (arsenito) es común en aguas con condiciones anóxicas, mientras que la forma pentavalente (arsenato) es más común en aguas oxigenadas. El paso entre estas dos formas es termodinámicamente favorecido en condiciones óxicas, pero dependiendo su tiempo de transición de las condiciones específicas (Sullivan *et al.*, 2010). De manera similar, las formas trivalentes y pentavalentes son las más comunes en los suelos, en los cuales tienden a asociarse preferentemente a aquellos con contenidos elevados en óxidos de Fe, Mn y Al, principalmente por el As(V). La afinidad de adsorción tanto de arsenitos como arsenatos, tiende a aumentar en suelos con valores de pH ácidos (Corsini *et al.*, 2010). De esta manera, el arsénico puede estar biodisponible en cuerpos hídricos, así como en suelos agrícolas, aumentando la posibilidad de intoxicación para los seres humanos y organismos vivos en general.

El riesgo asociado a este metaloide para organismos vivos se debe a dos fenómenos bioquímicos principales: 1) a la reactividad del ion arsenito con los grupos tioles presentes en varias de las proteínas y enzimas de importancia metabólica, generándose por medio de esta reacción, especies reactivas de oxígeno que tienen un efecto deletéreo sobre componentes celulares cercanos (Shah *et al.*, 2010); y 2) al efecto mimético que tiene el ion arsenato con el ion fosfato, lo que le permite el ingreso por canales específicos para fósforo a los diferentes tipos de células (Páez-Espino *et al.*, 2009).

Las tecnologías tradicionales y más utilizadas hasta hoy en la remoción y detoxificación de cuerpos de agua contaminados se basan en aproximaciones físico-químicas, muchas de las cuales resultan tener costos muy elevados para la implementación y generalmente tasas de eficiencia muy bajas, dejando de esta manera sin salidas reales y prácticas a varios países de regiones como Latino América en donde el problema parece ser más significativo de lo que se creía. La combinación de varios métodos, así como la implementación de sistemas biológicos para programas de descontaminación, parecen tener un futuro promisorio, sobretodo en lugares que no cuentan con altos recursos de inversión (Litter *et al.*, 2010).

1.2 Arsénico en el mundo y América Latina

Dadas las características tóxicas que hacen del arsénico un agente prioritario para su manejo y monitoreo a nivel mundial, se estima que existen alrededor de 10 millones de personas que viven con una alta exposición a este metaloide a partir de las fuentes hídricas (Smith & Smith, 2004), pues están expuestas al consumo prolongado de aguas con niveles mayores a 10 µg/L, límite establecido por la Organización Mundial de la Salud (WHO & IARC, 2004).

Los principales países que albergan esta población en riesgo directo son Bangladesh, India, Vietnam, Nepal, Taiwan, China, Estados Unidos, España Argentina y Chile. No obstante la problemática es más compleja en aquellos países en vía de desarrollo en los

que no existe inversión significativa para el desarrollo de medidas efectivas de monitoreo y prevención (Mandal & Suzuki, 2002; Bundschuh *et al.*, 2009).

También, estudios en la región de América Latina, revelan que 14 de los 20 países tienen riesgos crecientes de exposición y contaminación con este metaloide, debido principalmente al aumento de las actividades mineras, las que generan movilidad del arsénico hacia formas solubles en los cuerpos hídricos superficiales y subterráneos (Bundschuh *et al.*, 2012).

Algunos de los países en donde se han reportado niveles de concentración altos y riesgos potenciales de contaminación por arsénico en América Latina son Argentina, Chile, México, Guatemala, Nicaragua, Perú, Bolivia, Brasil y Uruguay (Bundschuh *et al.*, 2008). La revisión de literatura para el caso de América Latina, es aún muy escasa para algunos países, entre ellos Colombia, país para el cual no existen reportes significativos de presencia de arsénico a partir de fuentes geológicas.

1.3 Arsénico en Colombia

Aunque existen estudios y reportes significativos sobre la presencia de arsénico en Colombia, la mayoría de estos no se encuentran referenciados en la literatura internacional, toda vez que los estudios hechos son realizados a escalas muy locales y la mayoría de ellos con fines de prospección y diagnóstico mineralógico en vastas regiones del país. Ejemplo de ello es la ausencia de la mayoría de esta información en los recientes trabajos de revisión sobre el escenario del arsénico en América Latina.

La mayoría de reportes y de estudios realizados en el país, no se han centrado en el arsénico como problema específico y significativo, sino que han evaluado la presencia de este metaloide en conjunto con otros metales pesados y minerales con diferentes enfoques, entre los que se destacan: presencia de diferentes metales pesados y metaloides en suelos y aguas, influencia de diferentes metales pesados y metaloides en la fitotoxicidad de algunos cultivos vegetales, presencia de metales y metaloides en algunos productos de la cadena alimenticia y asociaciones de diferentes metales y metaloides con minerales de importancia económica como en el caso del oro.

No obstante en el reciente artículo de revisión titulado “*Environmental occurrence of arsenic in Colombia: A review*” (Alonso *et al.*, 2014), se muestra una perspectiva más amplia de la distribución del arsénico en el país y discute sus implicaciones y potenciales riesgos al medio ambiente y la salud humana. A continuación se expone la síntesis del trabajo de búsqueda y recolección realizado.

En 2007, Ravenscroft realizó un estudio de modelación de regiones con riesgo potencial de altos contenidos de arsénico en aguas subterráneas del país. Mediante predicciones basadas en sistemas de información geográfica, se señala que alrededor de 1'664.902 personas, lo que equivale a un 5% de la población nacional, estarían en riesgo potencial de intoxicación en las zonas montañosas y valles de la región norte del país.

Nicolli (2006), señala la existencia de fenómenos de liberación de arsénico asociado a minerales de arsenopirita en cuerpos hídricos de los departamentos de Tolima, Caldas y Nariño, principalmente por la minería llevada a cabo en esta región del país, principalmente de tipo aurífera.

La Figura 1 muestra las zonas del país en donde se han realizado estudios en diferentes matrices ambientales como aguas (W), suelos - sedimentos(S) y arsénico asociado a minerales (M).

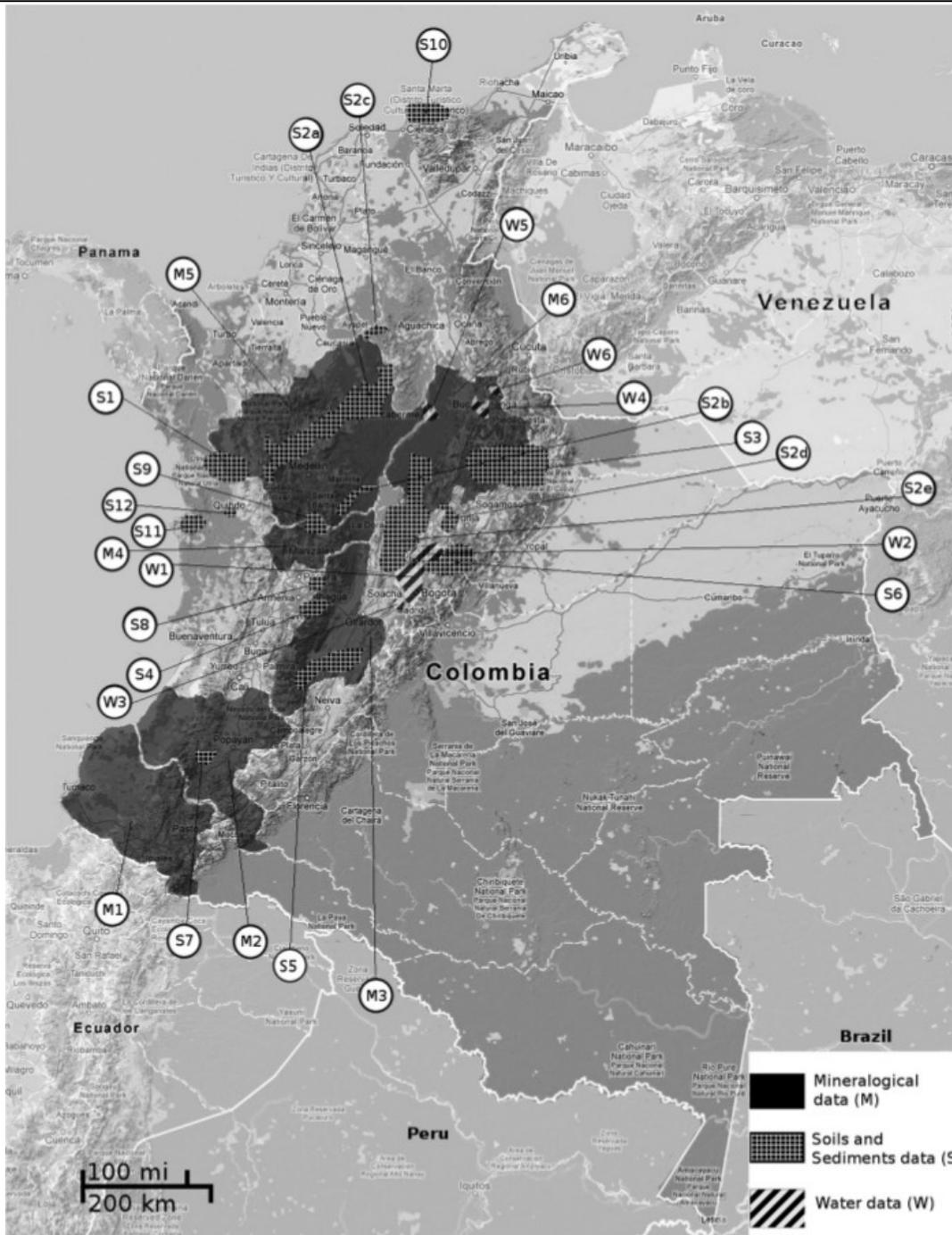


Figura 1. Reportes de estudios de arsénico en Colombia (Alonso *et al.*, 2014).

De las regiones señaladas, los departamentos de Chocó, Santander, Caldas y Tolima, tienen reportes en los que se señalan niveles de arsénico mayores a 40 mg/kg en suelos y sedimentos. Así mismo, la presencia del metaloide en estos departamentos es atribuida al desarrollo de las crecientes actividades mineras.

Por otro lado, en la Sabana de Bogotá ha sido significativo el número de trabajos realizados, en los cuales se relaciona la presencia o el efecto del arsénico con diferentes matrices biológicas, entre las que se destacan suelos, sedimentos y cultivos vegetales. La figura 2 muestra la ubicación geográfica de los trabajos reportados en las bases de datos y literatura nacional.

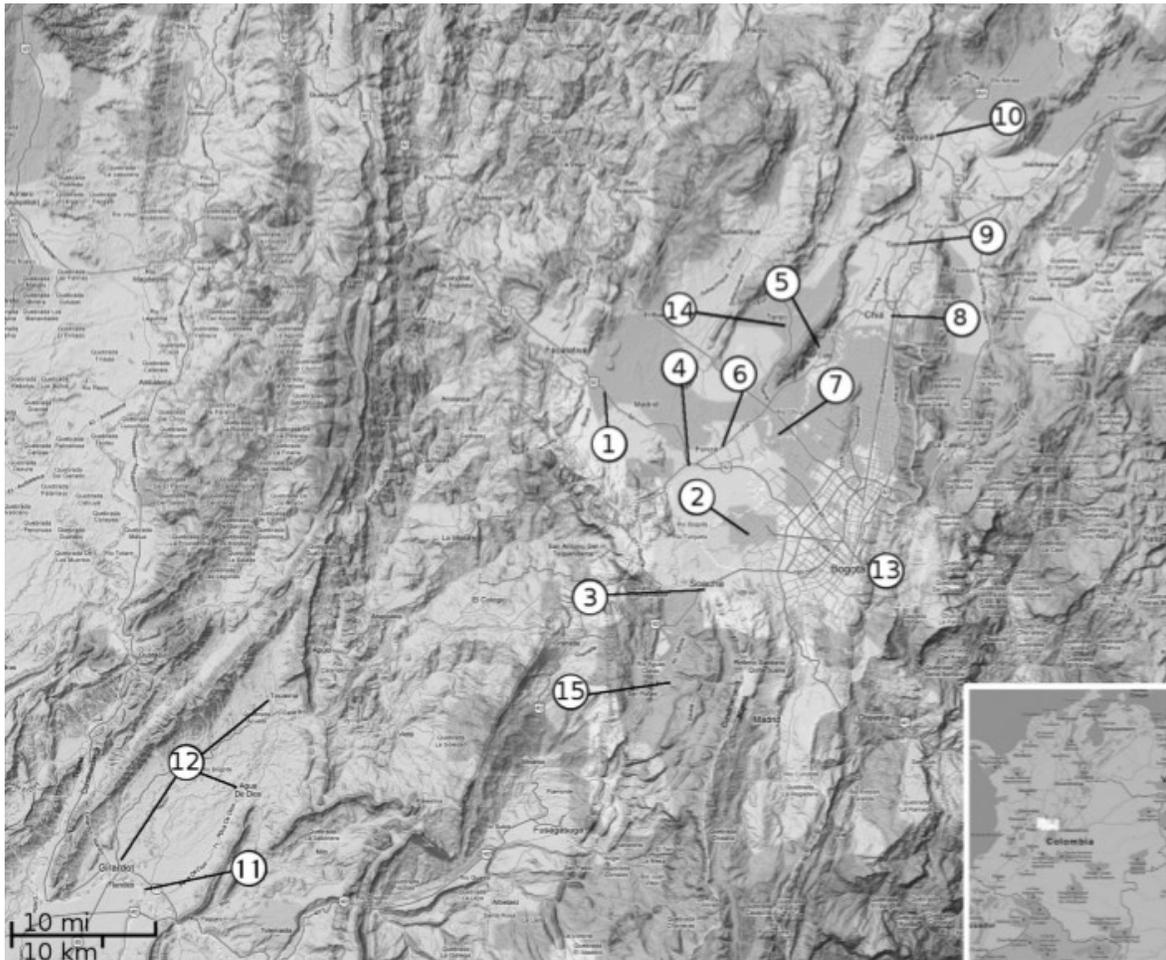


Figura 2. Estudios de arsénico en la Sabana de Bogotá (Alonso et al., 2014).

La mayoría de estudios realizados en la Sabana de Bogotá, relacionan el contenido de arsénico en algunos cultivos de interés. Es de resaltar que se encontraron altos contenidos de arsénico en cultivos de lechuga y repollo, así como valores de 0,58 mg/kg en cultivos de arroz que exceden los máximos recomendados por la OMS (0,3 mg/kg) (Figura 2, punto 12) (Alfaro et al., 2002).

Otro ejemplo es el estudio realizado por la Secretaría Distrital de Salud y el Hospital Pablo VI de Bosa (2000) en el que en un muestreo de 57 parcelas de cultivo agrícola aledañas a los ríos Tunjuelito y Bogotá, se encontró en todas las muestras, tanto de agua de riego como en hortalizas, la presencia del arsénico en un rango entre 0,5 – 1,3 mg/kg, entre otros compuestos contaminantes. En el caso de Bosa, Cota y Ricaurte, las concentraciones de arsénico en hortalizas resulta ser en la mayoría de los casos de nivel fitotóxico (Figura 2, punto 2) (Barragán, 2008).

Finalmente, los estudios y diagnósticos realizados en el país no son conducentes en el sentido de presentar un panorama de riesgo para el medio ambiente y la salud humana, toda vez que la mayoría de ellos han estado orientados hacia la descripción mineralógica de arsénico y solo algunos de ellos relacionan este metaloide con otras matrices biológicas como alimentos y suelos agrícolas.

La falta de monitoreo sobre las especies de arsénico es una limitante para la realización de diagnósticos de biodisponibilidad y riesgos toxicológicos. No obstante, dos elementos son necesarios a tener en cuenta: i) la alta concentración demográfica en la zona de la Sabana de Bogotá, genera un factor potencial de riesgo dado que existen acumulación de fuentes antrópicas de arsénico, como por ejemplo, la utilización intensiva de agrotóxicos sobre los cultivos de hortalizas en los suelos de la cuenca media del Río Bogotá, ii) el creciente *boom* minero que ha priorizado la política colombiana, es un factor de riesgo potencial debido a la liberación antrópica de arsénico asociado a fuentes minerales como la arsenopirita, mineral que es abundante en las minas auríferas del país.

1.4 Metabolismo microbiano del arsénico

Debido a la alta toxicidad de algunas especies del arsénico, este elemento actúa en el medio ambiente como un agente de selección natural, el cual es determinante para procesos de evolución en varios organismos, los cuales han generado mecanismos de interacción metabólica con el metaloide. Dichos mecanismos de metabolismo se han encontrado en plantas, levaduras y en organismos procariontes. A continuación se exponen los mecanismos más relevantes de interacción microbiana con el arsénico.

Las principales transformaciones que realizan las células bacterianas sobre el arsénico se resumen en cuatro tipos: metilación, demetilación, oxidación y reducción. Al respecto, algunas especies realizan estos procesos con el objetivo de detoxificación, más sin embargo el metabolismo del arsénico puede ser utilizado por los microorganismos con fines de ganancia energética, bien sea como donador o aceptor de electrones. Es así que varias especies procariontas han logrado sobrevivir bajo condiciones de presión de arsénico, mediante la utilización de diferentes mecanismos fisiológicos y bioquímicos (Zhou *et al.*, 2010).

Se ha reportado que el ingreso de arsénico se da al interior de la célula por medio de las especies solubles del metaloide As(III) y As(V). La forma de arsenito utiliza los canales de las gliceroporinas como forma privilegiada de ingreso y los iones arsenato, debido a su efecto mimético con los iones fosfato, utilizan los mismos mecanismos de ingreso al interior de la célula, es decir los canales transportadores de fosfatos (Páez-Espino *et al.*, 2009) (Figura 3, mecanismo #1).

Para el caso de la reducción del arsenato, algunas bacterias pueden utilizarlo como mecanismo de detoxificación o como forma de obtención de energía. En el primer caso, se ha estudiado el operón *ars* como el responsable de sintetizar proteínas capaces de convertir el As(V) a As(III), el cual es expulsado al exterior de la célula permitiendo su supervivencia en un medio externo con presión del metaloide (Corsini *et al.*, 2010) (Drewniak *et al.*, 2010) (Figura 3, mecanismo #2). El segundo caso, la presencia de una ruta disimilatoria de reducción le permite a la célula (generalmente bajo condiciones anóxicas) ganar energía mediante el acople del proceso reductivo del arsénico a la oxidación de materia orgánica (Corsini *et al.*, 2010) (Figura 3, mecanismo #5).

Por otro lado, el proceso de oxidación del As(III) es considerado un proceso de descontaminación del medio, pues el arsenito es considerado más tóxico que el arsenato. Este proceso es llevado a cabo principalmente por bacterias heterotróficas y quimioautotróficas. El mismo requiere de la utilización de una enzima arsenito oxidasa que cataliza el paso de As(III) a As(V) (Corsini *et al.*, 2010; Drewniak *et al.*, 2010) (Figura 3, mecanismo #4).

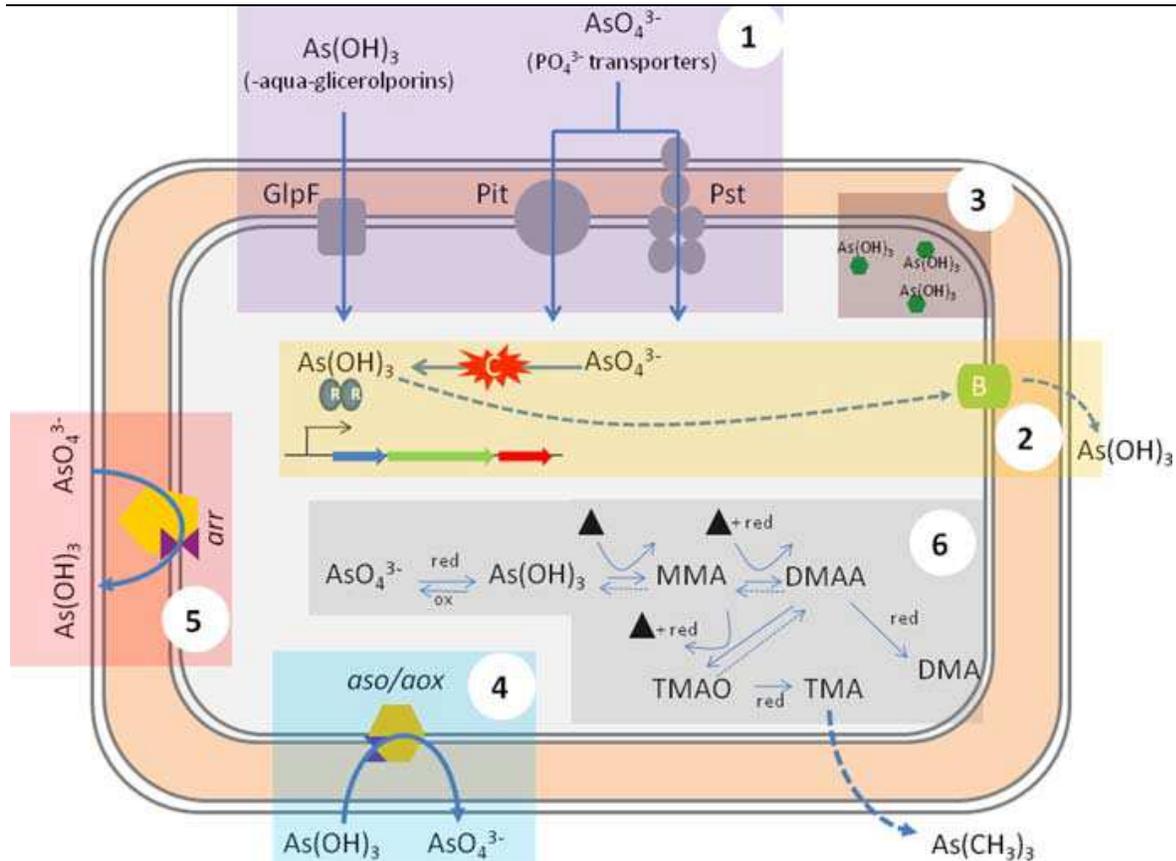


Figura 3. Metabolismo microbiano del arsénico. 1) Ingreso de arsenato al interior de la célula mediante la utilización de canales de fosfato. 2) Mecanismo de reducción del arsenato mediado por el operón *ars* y expulsión de arsenito. 3) Captura de arsénico en forma de arsenito al interior de la célula. 4) Oxidación del de arsenito mediada por genes tipo *aox*. 5) Reducción de arsenato mediada por genes tipo *arr*. 6) Mecanismos de metilación del arsénico y expulsión de arsénico en forma orgánica al exterior de la célula. Tomado de Páez-Espino *et al.* (2009).

El mecanismo de resistencia a arsénico más estudiado en microorganismos es el correspondiente al operón *ars*. Este operón fue descrito inicialmente como una región relacionada exclusivamente a plásmidos bacterianos (Hedges & Baumberg, 1973; Novick & Roth, 1968; Silver *et al.*, 1981). Sin embargo, su presencia se ha observado en más de 50 organismos entre arqueas, bacterias, levaduras y algunos otros eucariotas sugiriendo una evolución convergente de este carácter genético (Jackson & Dugas, 2003). Dentro de las Proteobacterias, la resistencia a arsénico mediada por el sistema *Ars* ha sido descrita en *Escherichia coli* (Carlin *et al.* 1995; Diorio *et al.*, 1995), *Pseudomonas aeruginosa* (Cai *et al.*, 1998), *Yersinia* spp (Bansal *et al.*, 2000), *Acidiphilum multivorum* (Susuki *et al.*, 1997; 1998), *Thiobacillus ferrooxidans* (Butcher *et al.*, 2000), *Acidithiobacillus caldus* (Dopson *et al.*, 2001), y *Corynebacterium glutamicum* (Ordóñez *et al.*, 2005), entre otras (Silver y Phung 2005).

El operón *ars* cuenta con dos componentes fundamentales: 1) el gen que codifica la arsenato reductasa (ArsC), y 2) el gen que codifica para una bomba de eflujo específica para arsenito (ArsB). ArsC es una molécula pequeña (entre 13 a 15 kDa), que cataliza la reducción de As(V) a As(III) al interior del citoplasma. Posteriormente a la reducción del As(V), la bomba de eflujo ArsB específica para As(III) tiene la posibilidad de evacuar hacia el exterior celular el producto reducido de arsenito que se encuentre al interior del citoplasma (Stolz *et al.*, 2006). Además de estos dos componentes principales, el sistema Ars puede contener ATPasas (ArsA), elementos regulatorios (ArsR, ArsD), entre otros (Mateos *et al.*, 2006). La proteína reguladora ArsR y la bomba de expulsión transmembranal ArsB confieren resistencia a arsenito, y el operón *ars* completo confiere resistencia tanto al arsenito como al arsenato. El gen *arsA* codifica una proteína ArsA intracelular, ATPasa, que se enlaza como un dímero a la proteína ArsB de la membrana permitiendo que esta bomba de expulsión de arsenito funcione quimiosmóticamente (ArsB sola) o como una ATPasa (complejo ArsAB). El gen *arsD* es sugerido como codificador de un corepresor ArsD del operón *arsRDABC* adicional a ArsR (Silver & Phung 1996). La organización del sistema de resistencia Ars es muy diverso en los microorganismos, en el cual varía tanto el tipo de gen presente dentro del operón, como el orden en que se organizan dichos genes (Butcher *et al.*, 2000; Chauhan *et al.*, 2009; Páez-Espino *et al.*, 2009; Silver & Phung, 2005; Stolz *et al.*, 2006).

Otro mecanismo de interacción con el arsénico, se ha encontrado en bacterias que realizan la reducción del arsenato utilizándolo en la cadena de transporte de electrones para el proceso de respiración celular. Genes del tipo *arr* han sido descritos en organismos como *Chrysiogenes arsenatis* y *Bacillus selenitireducens* en los cuales estas unidades genéticas codifican heterodímeros con una subunidad larga con actividad catalítica (ArrA) y una subunidad pequeña que realiza el proceso de transferencia de electrones (ArrB) (Stolz *et al.*, 2006) (Figura 3, mecanismo #5).

El proceso inverso, también es llevado a cabo en algunas bacterias, las cuales tienen la capacidad de oxidar las especies de arsenito para la obtención de arsenato, mediante la utilización de arsenito oxidasas. Estas enzimas son codificadas por genes de tipo *aox*, también conocidos en algunos reportes como genes *aro* y *aso*, los cuales están organizados en operones (Figura 3, mecanismo #4). La proteína, es un heterodímero con una subunidad catalítica (AoxB) y una subunidad asociada (AoxA), que la caracteriza la

presencia de un clúster de tipo *Rieske 2Fe-2S*, así como una secuencia líder de tipo TAT (Stolz et al, 2006).

1.5 Bioremediación del arsénico

Las tecnologías convencionales para remoción de arsénico del ambiente son principalmente de tipo físico químico, destacándose: la coagulación-filtración, el intercambio iónico, la adsorción en alúmina activada y procesos de ósmosis inversa. Si bien cualquiera de ellas tiene alta eficiencia de remoción, también tienen limitaciones en términos de sus altos costos de inversión y mantenimiento, así como por el requerimiento de personal calificado para la operación de las plantas (Gadd, 2000).

Frente a esta limitación, los procesos biológicos han aparecido como una alternativa importante, principalmente por sus bajos costos de operación requeridos. De esa forma, la inmovilización, degradación o transformación de las sustancias contaminantes mediante la actividad de microorganismos, para producir productos no tóxicos o al menos con menor toxicidad, hace de la bioremediación una herramienta que cobra mayor importancia actualmente (Gadd, 2000).

La biosorción es el fenómeno más estudiado para biorremediación de varios metales pesados. La biomasa capaz de participar en estos procesos es fácilmente extraíble y es por ello que la búsqueda de este tipo de microorganismos es un campo de investigación en constante crecimiento, así como el estudio de sistemas biosorbentes. Un ejemplo de esto es la utilización de consorcios microbianos o sistemas mixtos formados por microorganismos y macromoléculas (polímeros) sorbentes, que podrían incrementar los rendimientos en la captación de mezclas de metales pesados y metaloides (Vullo, 2003).

La utilización de microorganismos como biosorbentes de metales pesados y metaloides ofrece una alternativa potencial a los métodos ya existentes para la detoxificación y recuperación de metales tóxicos o valiosos, presentes en aguas. Muchas levaduras, hongos, algas, bacterias y cierta flora acuática tienen la capacidad de concentrar metales a partir de soluciones acuosas diluidas y de acumularlas al interior del citoplasma o al exterior de la membrana celular (Kotrba et al, 1999, Schmogger et al, 2000). Varias estrategias para remover metales pesados de ambientes contaminados han sido descritas (Valls y Lorenzo, 2002).

Con respecto a procesos de bioremediación de arsénico, han sido reportados el uso de plantas debido a la presencia de fitoquelatinas que son inducidas en presencia del agente tóxico (Schmoger *et al.* 2000), y en el uso de microorganismos para sobreexpresar genes responsables por la acumulación intracelular del metaloide (Kostal *et al.* 2004, Sauge-Merle *et al.* 2003) o su acumulación extracelular para uso como biosorbente (Kotrba *et al.* 1999, Valls *et al.* 2000). Mateos y colaboradores (2006) propusieron un modelo microbiano en *Corynebacterium glutamicum*, que combinaba la sobreexpresión de proteínas retenedoras de arsénico como la ArsR, la cual es la responsable del reconocimiento de arsénico en el medio exterior y regulación del operón *ars*, y la mutación de los canales de expulsión de arsénico hacia el exterior de la célula (ArsB), haciendo de esta manera células bacterianas con una capacidad de almacenamiento de hasta 50 veces su capacidad natural (Figura 4).

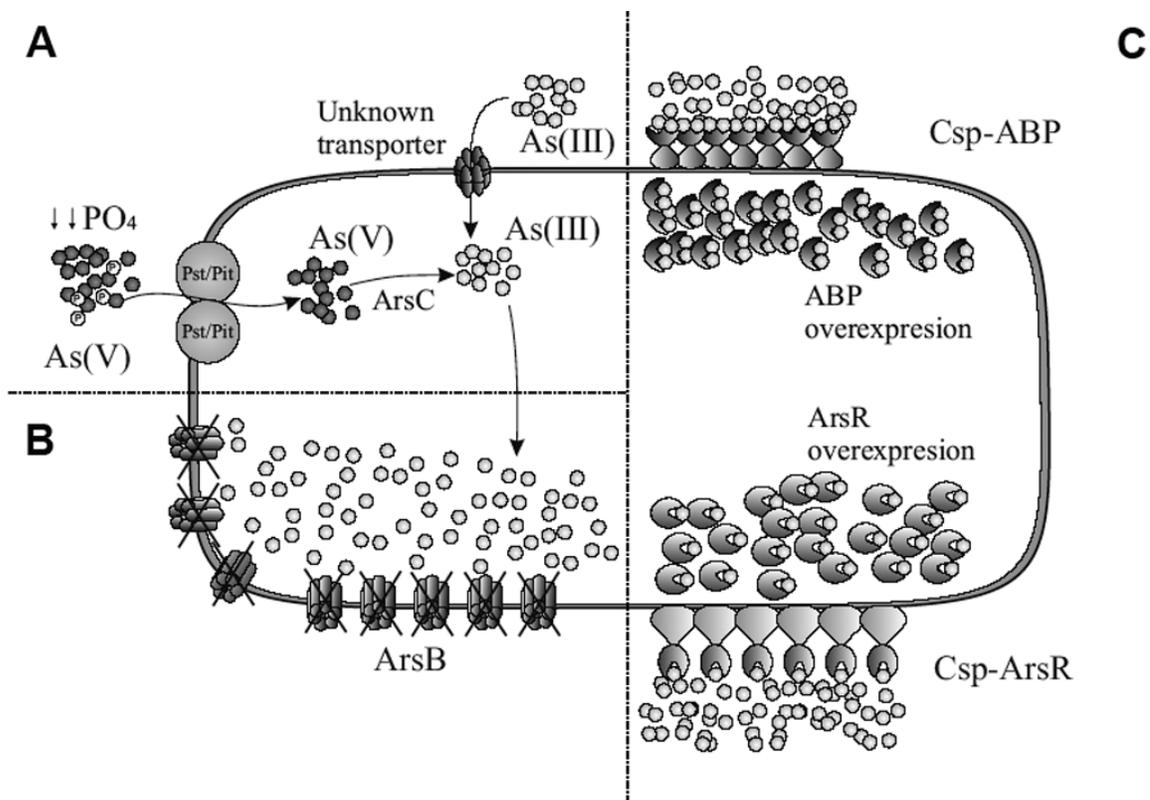


Figura 4. Modelo de bioremediación de arsénico propuesto para *C. glutamicum*. A) Ingreso de arsénico al interior de la célula; B) Mutación de canales responsables para la expulsión de arsenito; C) Sobreexpresión de proteínas retenedoras de arsénico (Mateos *et al.*, 2006).

La utilización de bacterias para la biosorción de otras formas de arsénico como el arsenito y el monometil arsenato, han sido evaluadas con éxito en especies como *Acidithiobacillus ferrooxidans*, abriendo las posibilidades del uso de bacterias para procesos de biorremediación del arsénico en ambientes contaminados (Yan *et al*, 2010). Trabajos como el de Mateos y colaboradores (2006) sugieren que la combinación de los procesos de transformación de arsénico utilizados por las bacterias por medio de técnicas de transformación genética, pueden dar lugar a organismos modelos de biorremediación con efectividad para la remoción y acumulación de arsénico de aguas contaminadas. Un ejemplo de ello es el trabajo realizado por Villadangos y colaboradores (2010), en el cual se realizó un ensayo de biorremediación, con la utilización de mutantes de bacterias corineformes. Dichas bacterias modificadas genéticamente tienen la capacidad de retención de arsenato, debido a la inactivación de los genes *arsC1* y *arsC2*, los cuales codifican las enzimas necesarias para la reducción del arsenato en arsenito, para su posterior expulsión celular.

1.6 Metagenómica

La mayor parte de la riqueza natural es desconocida. Aún existen descubrimientos de nuevas especies de animales y plantas en vastas regiones del planeta y ello se hace más significativo en los microorganismos dada la complejidad investigativa y experimental que implican sus aislamientos y procedimientos a nivel de técnicas de laboratorio.

Se estima que tan solo alrededor del 5% de los microorganismos han sido conocidos y caracterizados por algún tipo de técnica bioquímica o molecular, principalmente por el poco conocimiento actual sobre requerimientos físicos, químicos, ambientales y nutricionales de los microorganismos, lo que hace que la gran mayoría no sean cultivables bajo las técnicas convencionales de laboratorio (Amman *et al.*, 1995).

La gran dimensión de un 95% desconocido aún en el mundo microbiano, implica a su vez un desconocimiento generalizado sobre las rutas metabólicas, relaciones e interacciones especie – especie y en general la base genética que se encuentra detrás de este tipo de relaciones. Así, de la misma manera como son desconocidas las especies, son inmensas las posibilidades para la búsqueda e identificación de nuevos genes y rutas metabólicas de ambientes específicos (Turnbaugh & Gordon, 2008).

La metagenómica es principalmente un enfoque procedimental, que haciendo uso de los avances en las técnicas de extracción y purificación de ácidos nucleicos, realiza un aislamiento generalizado de un ambiente específico, dando la posibilidad de capturar la diversidad genómica que se encuentran determinadas por la presión de selección que ofrece las condiciones específicas de un entorno particular (Handelsman, 2004).

Es por ello que la metagenómica se convierte en una herramienta que posibilita el acceso a la gran mayoría de la riqueza de información microbiana de un medio ambiente específico, en el que existen relaciones metabólicas determinadas por la selección natural. El aislamiento de dicha información se logra evitando la condicionante de la realización de técnicas de cultivo específico, por lo que hace de ella una técnica de carácter masivo, que se basa en la extracción de los ácidos nucleicos totales de una muestra ambiental específica (Langer *et al.*, 2006).

En el desarrollo de la metagenómica como herramienta general se podrían discriminar dos grandes aproximaciones sobre las que se han desarrollado los desarrollos investigativos; ellas son a un nivel genotípico y a un nivel fenotípico. La primera de ellas explora las bases genómicas del funcionamiento de una comunidad específica en un medio ambiente. La segunda aborda una búsqueda o prospección en función de la expresión proteica de la comunidad microbiana (Figura 5).

Los desarrollos investigativos que se basan en el genoma suelen tener tres tipos de aproximaciones: de diversidad ecológica, a nivel de existencia de regiones genéticas específicas con funcionalidades determinadas y a nivel de exploración total bien sea del ADN o ARN. Para el caso de la exploración de ecología microbiana en un determinado ambiente, la amplificación de los genes ribosomales 16S y 18S para el caso de procariontes y eucariontes, respectivamente, suele acompañarse mediante técnicas electroforéticas que señalan una idea cualitativa de la diversidad, así como mediante secuenciación específica de este tipo de regiones conservadas. También, la identificación de regiones genéticas específicas mediante el diseño de *primers* hace de la metagenómica una herramienta para la búsqueda de elementos genéticos relacionados a funciones conservadas específicas. Finalmente, la secuenciación masiva logra generar una imagen general no solo de la diversidad microbiana del microcosmos, sino de las posibles interacciones que allí se suceden. Discriminar el ADN y ARN como molécula de estudio significa tener por un lado, la totalidad del set de intrones y exones a partir del

ADN, o las regiones codificantes y que están en expresión en un momento determinado, bajo unas condiciones específicas, cuando se parte del ARN (Handelsman, 2004).

Los desarrollos investigativos basados en la funcionalidad son dirigidos principalmente a nivel postranscripcional o a nivel de la expresión protéica del ARN que está siendo transcrito. Este tipo de aproximaciones resulta muy adecuado en los enfoques de bioprospección, toda vez que es posible encontrar funcionalidades específicas, pero con bases genéticas diferentes, las cuales sería imposible identificar con el conocimiento previo de regiones conservadas para funciones conocidas. El desarrollo de tamizajes o barridos funcionales cobra especial relevancia pues a partir de estos se realiza el proceso de selección de la funcionalidad esperada (Desai *et al.*, 2012).

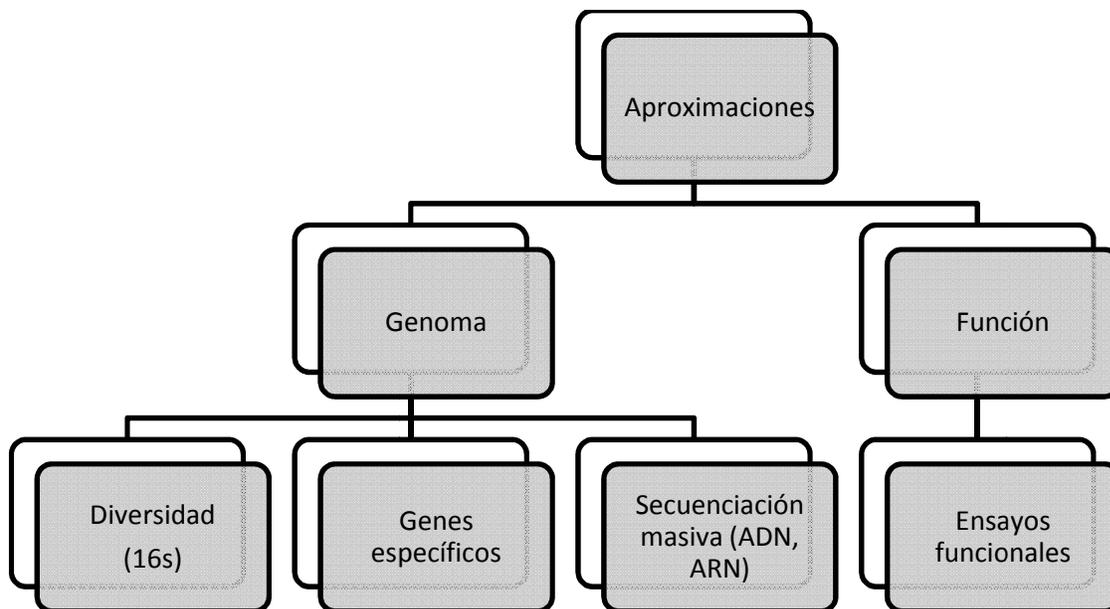


Figura 5. Aproximación general de la metagenómica.

No obstante, las aproximaciones metagenómicas de bases genéticas y de expresión funcional no son contradictorias entre sí. Por el contrario estas aproximaciones suelen ser complementarias debido a que la identificación de regiones genéticas tiene que ser verificada a nivel de expresión fenotípica y así mismo, la presencia de una función específica debe comprenderse y abordarse desde sus bases genéticas con el fin de explicar la totalidad del fenómeno de estudio.

La construcción de librerías metagenómicas logra la articulación de los enfoques genómicos y fenotípicos, pues logran organizar una muestra específica de ADN o ARN

en vectores de clonación, con el fin de ser utilizados posteriormente, bien sea para ensayos de secuenciación, amplificación de genes específicos o ensayos funcionales. Para ello, los procesos de extracción de los ácidos nucleicos, la selección del vector de clonación y el diseño del barrido funcional son pasos fundamentales para el logro satisfactorio de un proceso de bioprospección metagenómica. La construcción de librerías metagenómicas es una herramienta que permite una búsqueda masiva, relativamente rápida y de bajos costos, en comparación con otros enfoques de búsqueda de genes funcionales como por ejemplo la búsqueda mediante amplificación molecular de regiones conservadas (Stenuit *et al.*, 2008).

La extracción de ADN como primer paso de la construcción de las librerías metagenómicas, es quizá el de mayor importancia y determinante en el desarrollo de la librería. Varios autores han señalado la importancia y necesidad de obtención de ADN de alta calidad y con la remoción de agentes inhibidores y contaminantes como ácidos húmicos (Di Bella *et al.*, 2013).

Los ácidos húmicos y fúlvicos se caracterizan por su composición heterogénea y por la formación de complejos con diferentes metales que comúnmente se encuentran en suelos y sedimentos. Es por ello que su presencia en diferentes procedimientos moleculares puede considerarse como inhibitoria, debido a la formación de complejos con metales que en la mayoría de casos son cofactores necesarios para el funcionamiento de enzimas como polimerasas, fosfatasas y ligasas, entre otras (Zipper *et al.*, 2003).

Varios protocolos han sido desarrollados en función de lograr altos contenidos de ADN de alto peso molecular con remoción de la mayoría de agentes inhibitorios que aseguren una alta calidad de los ácidos nucleicos. En general, los protocolos utilizan una combinación de columnas de separación, con la utilización de técnicas electroforéticas para lograr una separación de los ácidos húmicos (Courtois *et al.*, 2001; Hurt *et al.*, 2001; Lloyd-Jones & Hunter, 2001; Santosa, 2001).

Posteriormente, la selección de los vectores de clonación, constituye un determinante para la caracterización de la librería metagenómica, en tanto el vector restringe el tamaño promedio de los insertos a ser utilizados. Las librerías construidas con la utilización de plásmidos permiten la inserción de fragmentos metagenómicos pequeños (entre 2kb y

5kb), los cuales están asociados tanto a elementos cromosómicos como plasmídicos del metagenoma estudiado. Otros enfoques han privilegiado la utilización de vectores de grandes tamaños como cósmidos, fósidos y BAC's, los cuales permiten que los elementos regulatorios y de expresión propios de los genes tengan una mayor probabilidad de resultar completos e íntegros luego del proceso de clonación. No obstante, la utilización de este tipo de vectores es restringida a la integración de ADN exclusivamente cromosómico (Kakirde *et al.*, 2010).

Posteriormente la selección del organismo hospedero, es un elemento determinante en la construcción de la librería metagenómica, dado que existe una estrecha relación entre la maquinaria genómica de un organismo con sus mecanismos de regulación, expresión y traducción proteica (Taupp *et al.*, 2011). Por ejemplo, es conocido que la mayoría de genes codificados por actinomicetos, no se expresan fácilmente en *E. coli*, debido a su diferencia significativa de contenido de G+C (Strohl, 1992). Ello genera un límite objetivo en el tipo de genes que se pueden expresar en un hospedero específico y por lo tanto cualquier hospedero utilizado en la construcción de una librería metagenómica tendrá un rango de elementos genéticos que no son compatibles con su maquinaria de expresión.

1.7 Metagenómica y bioremediación

La búsqueda de microorganismos, genes, enzimas y redes metabólicas que catalizan procesos de degradación de compuestos contaminantes o xenobióticos ha sido impulsada por el desarrollo de la metagenómica y en particular por los diseños que esta ha permitido mediante los enfoques de barridos funcionales, pues permiten encontrar a nivel fenotípico elementos genéticos expresados en condiciones particulares de contaminación simuladas *in vitro*.

Como ejemplo de ello, Daniel (2004) presenta una revisión sobre la utilización de la metagenómica en la identificación de nuevas enzimas y genes para aplicaciones industriales y de remediación, sobre las que las estrategias basadas en aproximaciones funcionales han sido mayoritarias para la identificación de nuevos mecanismos.

También, la revisión hecha por Calcagno (2005), enfatiza en las ventajas y versatilidad de la metagenómica en el campo de la remediación de ambientes contaminados y la prospección de genes relacionados con la degradación de compuestos tóxicos.

El trabajo de Eyers *et al.* (2004), presenta una revisión de dos estrategias ampliamente utilizadas en la búsqueda de genes y rutas metabólicas relacionadas con procesos de remoción bacteriana de agentes xenobióticos y contaminantes. La primera de ellas se centra en la búsqueda de regiones genómicas que potencialmente codifican genes involucrados en rutas metabólicas de detoxificación mediante la amplificación de regiones y dominios conservados; el segundo enfoque, conocido como tamizaje o barrido funcional, parte de la exploración e identificación de la capacidad descontaminante de un individuo, para desde allí conducir un proceso de identificación y caracterización a nivel genético.

En general, la metagenómica aporta la base genética sobre la cual los microorganismos metabolizan e interactúan con los compuestos xenobióticos, razón por la cual, la búsqueda de genes de resistencia a diversos tipos de contaminantes es una de las aproximaciones comúnmente utilizadas como base para futuros desarrollos biotecnológicos en el campo de la bioremediación.

Ejemplo de ello es el trabajo desarrollado por Chauhan *et al.* (2009), en el que a partir de una muestra ambiental de los desechos de una planta de tratamiento, se identificó un nuevo gen de resistencia a arsénico. La aproximación utilizada en este trabajo consistió en la creación de una librería metagenómica de 413.000 clones con la utilización de plásmidos como vectores de clonación y un tamaño promedio de insertos de ADN metagenómico de 3kb. De dos clones que presentaron resistencia a arsénico se logró el aislamiento e identificación de un nuevo gen (*arsN*) que participa junto a los genes que componen el operón *ars*, en la mediación de la reducción del arsenato en arsenito.

Finalmente, Achour *et al.* (2007), diseñaron un set de *primers* degenerados con el fin de evaluar diversidad del componente *arsB* del operón bacteriano *ars* en varios aislamientos de bacterias resistentes a arsénico. La implementación de estos *primers* pueden ayudar a dilucidar si los fenotipos positivos para un tipo de resistencia a arsénico, están ligados a algún tipo de operones descritos previamente.

1.8 Metagenómica y ensayos funcionales

La identificación de mecanismos involucrados en las rutas metabólicas a nivel de expresión fenotípica, depende en gran medida de la resolución del ensayo funcional, es decir, de la posibilidad de identificación, certeza y reproducibilidad del ensayo mismo. Por ello, el diseño del tamizaje funcional debe contar con esas características para que permitan identificar fácilmente los fenotipos deseados, máxime cuando las librerías metagenómicas construidas tienen tamaños considerables, que hacen inviables caracterizaciones bioquímicas y moleculares complejas de cada uno de los clones.

El trabajo descrito por Simeonova *et al.* (2004) establece un método que, basado en la reacción de los iones arsenito y arseniato con nitrato de plata, permite generar precipitados de color amarillo para el arsenito de plata y rojo para el arseniato de plata. De esta forma, es posible implementar dicho principio para la visualización de los cambios del estado de oxidación generados a partir de un organismo, teniendo en cuenta que uno de los principales retos es masificar y estandarizar esta técnica en función de la evaluación de miles de clones característicos de las librerías metagenómicas.

También, Chauhan *et al.* (2009), implementa un tamizaje funcional que tiene como principio la selección positiva de clones a una alta concentración de arsenato de sodio, con el fin de identificar microorganismos con un fenotipo de resistencia elevada al metaloide. En dicho trabajo se establece un valor de 50 mM de arsenato de sodio como umbral de detección y tamizaje funcional sobre el que se logran encontrar dos clones resistentes a este metaloide, sobre una librería de 413.000 clones.

1.9 Metagenómica y secuenciación

El desarrollo de la metagenómica está directamente relacionado con el desarrollo de las nuevas técnicas de secuenciación masiva (*High Throughput Sequencing*). La gran cantidad de material genético que se encuentra en un estudio metagenómico, debe acompañarse con tecnologías que posibiliten su lectura e identificación (Qin *et al.*, 2010).

La innovación hacia la secuenciación de *segunda generación*, generó un rápido despliegue de datos genómicos, haciendo en horas lo que en la secuenciación tradicional por el método de Sanger tardaba meses (Di Bella *et al.*, 2013).

Entre las tecnologías emergentes, el sistema *454 FLX+* (Roche®) se destaca por ser una tecnología basada en el principio de la *pirosecuenciación*, el cual mediante una reacción de *PCR*, identifica cada adición de nucleótidos con una reacción mediada por la luciferasa, la cual genera un patrón específico para cada nucleótido que posteriormente es captada por un sensor (Nowrousian, 2010).

Una de las desventajas de este sistema es la aparición de *extensiones homopoliméricas*, debido a la forma en que el sistema detecta la incorporación de los nucleótidos. El resultante son regiones con presuntas mutaciones de inserciones / deleciones (Kunin *et al.*, 2010). No obstante, debido a la extensión de lectura del sistema *454 FLX+* (Roche®), alrededor de 700 bases, es recomendable para el ensamblaje de fragmentos metagenómicos, pues está por encima de la extensión generada por el resto de las tecnologías de segunda y tercera generación desarrolladas hasta el momento (Di Bella *et al.*, 2013).

Otras tecnologías desarrolladas de segunda generación son: el sistema *HiSeq 2000/2005* (Illumina®) el cual, basado en el principio de un terminador reversible, genera fragmentos de alrededor 150 pares de bases y una lectura total por corrida de hasta 600 Gb; el sistema *5500xl W SOLiD* (Life Technologies®) el que, a partir de una reacción de ligación, genera fragmentos de alrededor 75 pares de bases y una lectura total por corrida de hasta 320 Gb (Liu *et al.*, 2012). Finalmente el sistema *Ion PGM* (Life Technologies®) el que, con base al principio de detección protónica, genera lecturas entre 100 – 200 pares de bases con una lectura total por corrida de hasta 2 Gb (Rusk, 2011).

Sin embargo, la secuenciación por terminación de didesoxirribonucleótidos desarrollada por Sanger *et al.* (1997) aún es ampliamente utilizada en el campo de la metagenómica. Por un lado para la identificación a nivel de especie utilizando la región conservada del gen 16S ARNr (Cole *et al.*, 2009), pero también para la realización de un marcaje específico con el fin de utilizar de manera complementaria la secuenciación por el método de Sanger y otras alternativas de secuenciación masiva.

1.10 Herramientas bioinformáticas para la búsqueda de genes

Analizar la gran cantidad de información proveniente de la secuenciación masiva se convierte, naturalmente en el principal reto que la metagenómica enfrenta. El desarrollo de la capacidad de análisis computacional mediante algoritmos ágiles y certeros es una necesidad que debe acompañar de manera paralela el desarrollo del acceso a la información genética masiva (Handelsman, 2004).

La ventaja que ofrece la metagenómica de dar un amplio panorama de la diversidad genómica de un ambiente particular, es a su vez su principal reto y cuello de botella en el análisis y esclarecimiento de tal cantidad de datos. En primer lugar la metagenómica debe enfrentar el reto del ensamblaje de las múltiples lecturas (Scholz *et al.*, 2012).

Las tecnologías de segunda y tercera generación de secuenciación masiva que generan lecturas por debajo de las 200 pares de bases tiene la mayor dificultad, pues debido a la redundancia de funciones y rutas metabólicas existentes en un ambiente específico, se eleva la probabilidad de formación de quimeras genómicas.

Las lecturas más grandes, generadas por los sistemas *454* (Roche®) que se encuentran entre 400 y 1000 pares de bases, reducen significativamente la probabilidad de generación de quimeras, pero a su vez incorporan la posibilidad de errores en el proceso de ensamblaje debido a la presencia de regiones homopoliméricas.

Por ello, varios algoritmos han sido adaptados y otros nuevos desarrollados con el fin de optimizar el ensamblaje de fragmentos de lecturas cortos, provenientes de muestras metagenómicas. Algunos de ellos son *Meta-IDBA* (Peng *et al.*, 2011), *MetaVelvet* (Namiki *et al.*, 2012), *Metagenomic Assembly Program (MAP)* (Lai *et al.*, 2012), *AbySS* (Simpson *et al.*, 2009), *Bambus 2* (Koren *et al.*, 2011), *Genovo* (Laserson *et al.*, 2011) y *MIRA* (Chevreux *et al.*, 1999; 2004).

Particularmente, el algoritmo *MIRA* (Chevreux *et al.*, 1999; 2004), es importante para el trabajo de lecturas provenientes de pirosecuenciación. Esta herramienta bioinformática ha desarrollado una optimización para el ensamblaje específico de lecturas provenientes de sistemas de Roche®, incluyendo paquetes para el trabajo específico de los formatos de este tipo de secuencias. Algunos trabajos recientes han hecho uso de este algoritmo

para el ensamblaje de este tipo de secuencias (Ekström *et al.*, 2013; Hook *et al.*, 2014; Xavier *et al.*, 2012).

Existen sin embargo, algunos enfoques de análisis metagenómico que obvian el proceso de ensamblaje y realizan un proceso de asignación funcional directamente sobre las lecturas generadas a partir de los sistemas de secuenciación masiva. Por esencia, este tipo de enfoques suelen perder genes de interés y funciones asociadas, pues el tamaño corto de las lecturas no permite la identificación de tales regiones con la mayoría de algoritmos existentes. Por ello, entre más grandes sean los fragmentos de lecturas, mayor es la probabilidad de identificar funcionalidad, así que se hace deseable un proceso previo de ensamblaje con el fin de generar *contigs* y a su vez regiones más grandes o *scaffolds* (Desai *et al.*, 2012).

El segundo reto bioinformático, es la asignación funcional sobre los *contigs* o *scaffolds* generados, bien sea para la identificación de genes o la asignación de funciones y rutas metabólicas. En general existen dos posibilidades: la primera de ellas, consiste en la búsqueda de genes a partir de las propiedades intrínsecas de las secuencias, principalmente en la identificación de marcos abiertos de lecturas o *ORF's* (del inglés *Open Reading Frames*); la segunda estrategia es mediante la utilización de algoritmos de alineamiento como *BLAST* y *BLASTX*, con el fin de realizar búsquedas de similaridad en las bases de datos genómicas.

En el primer caso, algunos de los algoritmos desarrollados son: *FragGeneScan* (Rho *et al.*, 2010), *Orphelia* (Hoff *et al.*, 2009), *MetaGeneAnnotator* (Noguchi *et al.*, 2008), *Glimmer-MG* (Kelley *et al.*, 2012), *MetaGeneMark* (Zhu *et al.*, 2010) y *Metagenomics Gene Caller (MGC)* (El Allali & Rose, 2013). Estos algoritmos realizan un proceso de identificación de genes potenciales sin ningún tipo de información adicional más que las características internas de la secuencia; son algoritmos *de novo*.

También, es posible realizar un proceso de anotación en función de la búsqueda de patrones similares de las secuencias con las bases de datos existentes. En general, el proceso consiste en la utilización de un algoritmo de alineamiento múltiple con el fin de ubicar regiones similares en bases de datos específicas y herramientas específicas para la anotación funcional como: *Clusters of Orthologous Groups system (COG)* (Tatusov *et al.*, 2003), *Pfam* (Bateman *et al.*, 2004), *the Kyoto Encyclopedia of Genes and Genomes*

(KEGG) (Kanehisa & Goto, 2000), *the Conserved Domains Database* (CDD) (Marchler-Bauer *et al.*, 2005), *SEED* (Overbeek *et al.*, 2005), *TIGRFAM* (Selengut *et al.*, 2007).

El último desarrollo de procesos de anotación, ha estado centrado en la automatización y combinación de los dos enfoques anteriormente descritos, muchos de ellos para el fácil acceso en línea. Algunos de los más destacados son: *MG-RAST* (Meyer *et al.*, 2009), *Metagenome Analyzer (MEGAN)* (Huson *et al.*, 2007), *the Integrated Microbial Genomes system for metagenomes (IMG-M)* (Markowitz *et al.*, 2008), *Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis (CAMERA)* (Sun *et al.*, 2011), *KEEG Automatic Annotation Server (KAAS)* (Moriya *et al.*, 2007).

1.11 Conclusiones del estado del arte

La presencia y riesgo potencial de arsénico a nivel mundial, ha sido monitoreada principalmente en la región Indochina y en menor medida en la región Ibérica así como en el cono sur de América Latina. Aún son escasos los reportes a nivel de la literatura universal que señalen la magnitud de la presencia y problemática del arsénico en Colombia.

No obstante, el reciente trabajo de revisión desarrollado por Alonso *et al.* (2014), es ilustrativo frente a los estudios realizados de presencia del metaloide en el país, resaltando particularmente las zonas de departamentos como Chocó, Tolima, Santander y Caldas, dada la importante liberación mineralógica desde las fuentes de arsenopirita hacia formas solubles de arsenato y arsenito. Dinámica que tiende a agudizarse dadas las dinámicas antrópicas en aumento relacionadas con la explotación minera a gran escala.

Así mismo, se señala la zona de la Sabana de Bogotá como un punto en el que convergen riesgos potenciales de contaminación dada la existencia de contaminantes en la cuenca del Río Bogotá y los desechos domésticos e industriales que generan condiciones de riesgo de toxicidad por varios metales pesados y metaloides como el arsénico. Gran parte de la despensa alimentaria de esta región del país se encuentra ubicada en zonas de la cuenca media de la Sabana de Bogotá, la que se caracteriza por tener suelos son pH's ácidos, sobre los que se aumenta la afinidad de adhesión y retención por parte de los minerales y coloides de los suelos y sedimentos.

No obstante el riesgo potencial creciente en el país, no ha sido acompañado por un monitoreo permanente que conduzca al diseño de suficientes mecanismos de descontaminación, en gran medida por los altos costos de inversión de infraestructura, monitoreo y costos operacionales que supone la presencia de arsénico en cuerpos hídricos, suelos y sedimentos de importancia. Así mismo, la mayoría de tecnologías convencionales desarrolladas hasta el momento, aunque logran resultados satisfactorios de remoción de arsénico, aún mantienen costos muy elevados. Es por ello que las alternativas biotecnológicas son adecuadas pues suponen menos costos operacionales que las tecnologías tradicionales.

Para el desarrollo de procesos de bioremediación se hace necesaria la búsqueda de la base genética y metabólica que soporta los procesos de descontaminación microbiana. Al respecto, se amplía la diversidad de mecanismos metabólicos microbianos descritos en relación al arsénico, razón por la que hay mayores probabilidades de encontrar genes y rutas novedosas que sienten las bases para futuros desarrollos.

De esta manera, herramientas como la metagenómica, la secuenciación masiva y la utilización de herramientas de análisis bioinformático, son centrales para el aprovechamiento de la gran diversidad biológica y metabólica existente en un ambiente con una presión de selección natural de contaminantes.

Por ello, el presente trabajo tiene por objetivo central la identificación de genes de resistencia a arsénico en el metagenoma microbiano de muestras ambientales de la Sabana de Bogotá, por lo que se plantearon como objetivos específicos: 1) generar librerías metagenómicas a partir de distintas muestras ambientales de la Sabana de Bogotá, 2) identificar y seleccionar *in vitro* clones con resistencia a arsénico en las librerías metagenómicas, 3) obtener las secuencias de los genes que confieren resistencia a arsénico en los clones seleccionados y, 4) analizar las secuencias obtenidas mediante herramientas bioinformáticas.

2. Materiales y Métodos

2.1 Muestreo

Se seleccionaron dos puntos de muestreo en la cuenca media del Río Bogotá, en un área comprendida entre el municipio de Mosquera (Cundinamarca) y la localidad de Bosa en Bogotá. A continuación se presentan las coordenadas de los predios seleccionados para la realización del muestreo:

- Finca Santa Ana: 4°38'N, 74°14'W
- Finca Santa Anita: 4°38'N, 74°37'W

La Figura 6 muestra una ubicación general de los sitios de muestreo con respecto a la Sabana de Bogotá. Allí se puede apreciar la influencia de Río Bogotá y del Río Tunjuelo en el área seleccionada. La Figura 7 permite ver con mayor detalle, la ubicación de los sitios de muestreo con respecto al cauce del Río Bogotá. Los predios seleccionados corresponden a fincas con uso exclusivo de explotación agrícola con utilización intensiva de plaguicidas.

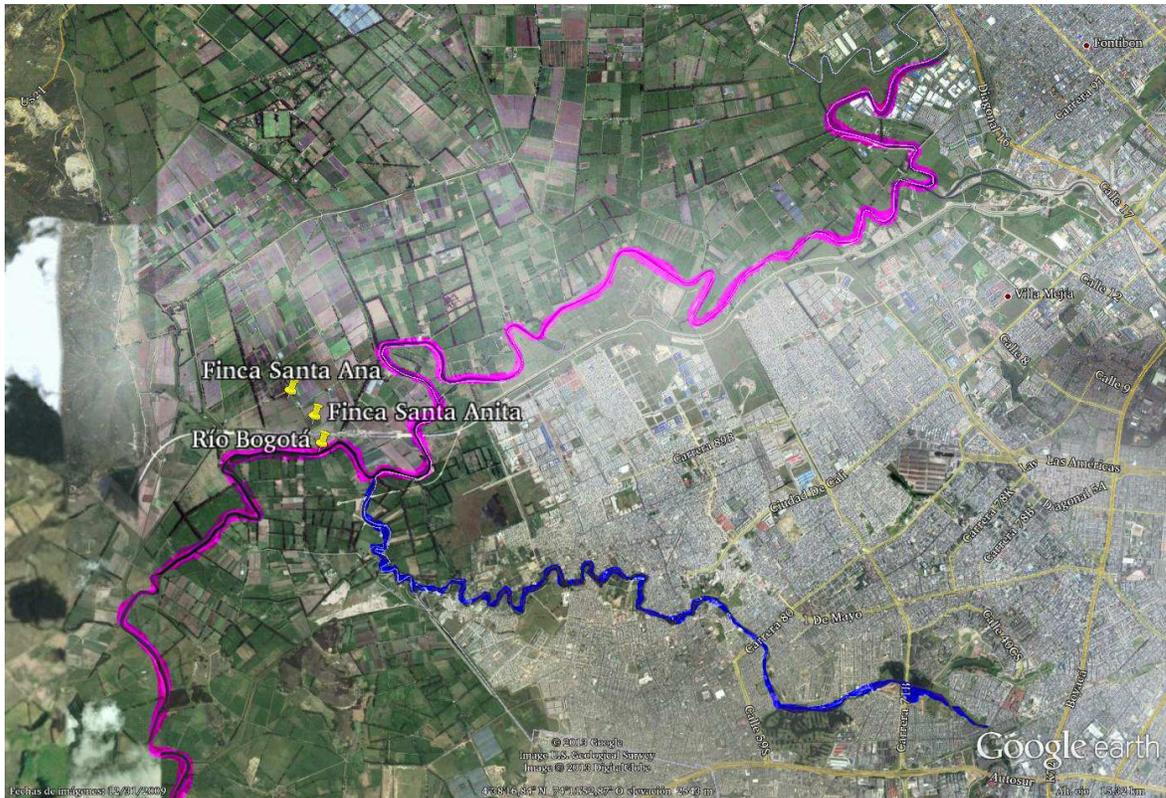


Figura 6. Sitios de muestreo de este trabajo. Río Bogotá (color violeta) y Río Tunjuelo (color azul)



Figura 7. Detalle de los sitios exactos de muestreo de este trabajo.

2.1.1 Muestras de suelo

Las muestras de suelo fueron tomadas de los predios Santa Ana y Santa Anita. Para el momento de la toma de las muestras, sistemas de cultivo de Acelga y Cebolla estaban establecidos en las fincas, respectivamente.

En cada predio se tomaron 4 muestras compuestas de suelos. Así mismo, cada muestra estuvo conformada por 4 submuestras. Estas submuestras corresponden a suelo rizosférico y fueron tomadas con la ayuda de una pala, a profundidades entre los 0 – 30 cm, dada la existencia de cultivos de hortalizas los cuales tienen un sistema radicular que profundiza hasta los 30 cm. Los sitios de submuestreo fueron seleccionados de manera aleatoria.

Cada muestra compuesta fue separada en dos unidades: la primera de ellas, de alrededor 50 g con el fin de ser utilizada en los procedimientos metagenómicos (Figura 8) y la otra, de alrededor 1 kg para los procedimientos analíticos de contenido total de arsénico que se describen posteriormente.



Figura 8. Muestra de suelo recolectada para uso en los protocolos metagenómicos.

2.1.2 Muestras de sedimentos

Las muestras de sedimentos fueron tomadas del cauce del Río Bogotá (Figura 7). Allí, se tomaron dos tipos de muestra: la primera de ellas del borde del cauce y la otra del centro del mismo. Las muestras del borde del cauce fueron tomadas con la ayuda de una pala de mano. Las muestras del centro del cauce fueron tomadas con la ayuda de una máquina dragadora, la cual estaba siendo utilizada en labores de remoción de sedimentos del Río (Figura 9). De igual manera, las muestras fueron separadas en dos unidades para ser utilizadas en los procedimientos metagenómicos y analíticos.



Figura 9. Extracción de sedimentos del Río Bogotá.

2.1.3 Muestras de agua

Con la ayuda de un equipo de recolección de muestras de agua, se tomaron dos muestras en recipientes de polietileno de 1 L de capacidad cada uno (Figura 10).



Figura 10. Recolección de muestras de agua del Río Bogotá.

2.1.4 Manejo de las muestras

Las muestras de suelos, sedimentos y aguas, fueron inmediatamente guardadas en una nevera de enfriamiento con bloques de frío a su interior. Luego del proceso de muestreo, las muestras fueron llevadas hacia las instalaciones del Departamento de Química de la Universidad Nacional de Colombia – sede Bogotá.

Las muestras destinadas a los procedimientos analíticos fueron llevadas al Laboratorio de Química Analítica en donde se realizó un pre-secado de las muestras de suelos y sedimentos a lo largo de un día a temperatura ambiente. Luego de ello, las muestras de suelos fueron trituradas y posteriormente cernidas, utilizando una malla *mesh* 100. Tanto las muestras de suelo cernidas, como de sedimentos fueron llevadas al laboratorio de suelos de la Facultad de Agronomía con el fin de realizar el proceso de deshidratación durante 24 horas a una temperatura de 65 °C.

Para la realización de los procedimientos metagenómicos, las muestras de suelos y sedimentos fueron almacenadas a una temperatura de -70 °C y las muestras de agua a una temperatura de 4 °C.

2.1.5 Análisis de muestras

Las muestras tomadas de las diferentes matrices ambientales, fueron analizadas en el Laboratorio de Análisis Químico del Departamento de Química de la Universidad Nacional de Colombia. Mediante la técnica de Absorción Atómica con Generación de Hidruros, se realizó una cuantificación de arsénico total para cada una de las muestras (Alonso, 2014).

2.2 Extracción de ADN

Dos kits comerciales fueron utilizados para optimizar el proceso de extracción y purificación de ADN sobre los tres tipos de matrices ambientales (suelos, sedimento y agua): *Meta-G-Nome™ DNA Isolation Kit (Epicentre®)* y *PowerMax™ Soil DNA Isolation Kit (MO-BIO®)*.

Para los ensayos realizados con el kit *Meta-G-Nome™ DNA Isolation Kit (Epicentre®)*, se utilizaron muestras de 1 g para suelos y sedimentos y se siguieron las indicaciones del protocolo del kit. Así mismo, para las muestras de agua, se utilizaron volúmenes iniciales de 50 mL.

Para los procedimientos realizados con el kit *PowerMax™ Soil DNA Isolation Kit (MO-BIO®)*, se utilizaron 10g de muestra inicial de suelo y sedimento. Con el fin de mantener la mayor integridad posible del ADN así como su alto peso molecular, se realizaron algunos cambios en el protocolo sugerido (Anexo A). Brevemente, los pasos de agitación (*vortex*) fueron suprimidos por completo y a su vez fueron cambiados por un tratamiento que incluye intervalos de calor a temperatura de 70 °C y agitación suave. El volumen final de 5 mL de ADN fue reconcentrado adicionando 0,2 mL de NaCl 5 M y 10,5 mL de etanol absoluto seguido de una centrifugación durante 30 min a 2500x g y una resuspensión final en 200 µl.

Para las muestras de agua del Río Bogotá se utilizó el Kit *Meta-G-Nome™ DNA Isolation Kit (Epicentre®)*, siguiendo las recomendaciones del protocolo.

2.3 Selección de tamaños y purificación del ADN

Se realizó una electroforesis a 30 V constantes y con una concentración de agarosa de 1,5 % sobre las muestras de ADN provenientes de suelos y sedimentos. Una alícuota de ADN de referencia (40 kb) fue cargada junto a una alícuota del ADN metagenómico, seguida del resto de esta misma muestra. Ni las muestras cargadas, ni la agarosa fueron teñidas con agentes intercalantes, con el fin de preservar la integridad del ADN metagenómico. Este procedimiento también aseguró la total remoción de impurezas y contaminantes como ácidos húmicos provenientes de la materia orgánica de las muestras.

Posteriormente se realizó una disección del gel de los dos primeros carriles, que contenían las alícuotas de la muestra de referencia, así como de una pequeña alícuota de la muestra metagenómica. Esta sección fue bañada en una concentración de bromuro de etidio y expuesta a luz UV, con el fin de visualizar tanto el sitio promedio de fragmentos de 40 kb, como la presencia de ADN metagenómico de la alícuota. Las secciones de gel fueron unidas y se realizó el corte de la región en la que migraron los fragmentos con este peso molecular.

Finalmente, se utilizó el Kit de recuperación y purificación *UltraClean™ GelSpin DNA Purification Kit (MO-BIO®)* con el fin de recuperar el ADN metagenómico seleccionado a partir de la electroforesis. El proceso realizado siguió las recomendaciones del protocolo establecido por el producto.

2.4 Arreglo de extremos de los fragmentos de ADN

La fosforilación y arreglo de los extremos de los fragmentos de ADN metagenómico fue realizada bajo las indicaciones sugeridas por el protocolo y con las siguientes condiciones y concentraciones:

- 8 μ L Buffer de reparación 10X
- 8 μ L dNTP's 2,5mM
- 8 μ L ATP 10mM
- 52 μ L ADN de alto peso molecular (80 ng / μ L)
- 4 μ L Mix enzimático de reparación

El volumen total de la reacción fue 80 μL . La temperatura de reacción fue de 16 $^{\circ}\text{C}$ y su tiempo de 1 hora. Finalmente, la reacción fue sometida a 70 $^{\circ}\text{C}$ durante 10 minutos con el objetivo de desactivar las enzimas.

Con el fin de eliminar los residuos de reactivos del proceso de fosforilación y con el objetivo de reconcentrar el ADN, se utilizó nuevamente el Kit de recuperación y purificación *UltraClean™ GelSpin DNA Purification Kit (MO-BIO®)* y luego de ello se resuspendió el ADN purificado en 12 μL agua estéril y desionizada.

2.5 Ligación

La clonación de los fragmentos de ADN metagenómico de peso seleccionado, purificados y con extremos fosforilados, fue realizada con el vector de clonación pCC2FOS (*Epicentre®*). Para ello se ligaron los fragmentos con las siguientes condiciones y concentraciones:

| | |
|--------------------|--|
| 1 μL | Buffer de ligación (10X <i>Fast-Link Ligation Buffer</i>) |
| 1 μL | ATP 10mM |
| 3,35 μL | Vector pCC2FOS (<i>Epicentre®</i>) |
| 6 μL | ADN metagenómico de alto peso molecular |
| 1 μL | Ligasa |

La reacción de ligación se llevó a cabo a una temperatura constante de 16 $^{\circ}\text{C}$ durante 16 horas y posteriormente a una temperatura de 70 $^{\circ}\text{C}$ durante 10 minutos con el fin de desactivar la ligasa.

2.6 Empaquetamiento

El producto de ligación fue empaquetado utilizando 50 μL de *MaxPlax Packaging Extracts (Epicentre®)* durante 4 horas a una temperatura constante de 30 $^{\circ}\text{C}$. De esta manera, los fagos resultantes del empaquetamiento fueron utilizados para el posterior proceso de infección.

2.7 Infección

El contenido total del producto de empaquetamiento se utilizó para la infección de células *E. coli* EPI300-T1®, las cuales fueron pre-tratadas con suplementos de MgSO₄ y Maltosa durante 16 horas, con el fin de activar los receptores para la infección. Las células fueron re-inoculadas en 50 mL de medio LB (10 g de Triptosa, 5 g NaCl, 10 g de extracto de levadura; en 1 litro de agua) con los suplementos y tras un periodo aproximado de 3 horas, hasta alcanzar una densidad óptica (600 nm) entre 0,8 y 1,0, fueron infectadas con los fagos durante un periodo de 80 minutos a 37 °C.

2.8 Librería metagenómica

Alícuotas del producto final de 1 mL fueron sembradas en cajas de polietileno de 22 cm x 22 cm con medio LB + Cloranfenicol (12,5 µg/mL) e incubadas durante 24 horas. Las colonias resultantes fueron seleccionadas y organizadas en microplacas de 384 pozos [con medio LB + Cloranfenicol (12,5 µg/mL) + Glicerol (20 %)] con la implementación del sistema robótico *Q-Pix* (*Genetix*®), el cual a partir de un sistema de toma de imágenes acoplado a un brazo mecánico, realiza un proceso de organización automatizada de colonias en cajas de polietileno. Luego del proceso de organización de las librerías, las cajas fueron incubadas a 37 °C durante 20 horas para los análisis de tamizaje funcional y su posterior almacenamiento.

2.9 Validación de umbrales

Con el fin de validar la metodología empleada por Chauhan *et al.* (2009), se tomó una concentración de 50 mM como umbral de referencia, añadiendo a esta una concentración de 20 mM de arsenato de sodio, con el fin de determinar la sensibilidad de cepas control a estos niveles del metaloide.

Tres cepas tipo de *E. coli* fueron solicitadas al CGSC (*Coli Genetic Stock Center, USA*) y utilizadas como controles negativos y positivos de procedimiento. Como control positivo, fueron utilizadas las cepas *MG1655* y *W3110* (Carlin *et al.*, 1995), las que tienen resistencia cromosómica a arsenato debido a la presencia del operón *ars*; como control

negativo se utilizó el genotipo *AW3110*, la cual tiene una mutación en el operón *ars* que le confiere un fenotipo de susceptibilidad a arsénico.

En este ensayo fue incluido el genotipo *E. coli EPI300-T1s*, el cual fue el organismo utilizado como hospedero para la creación de la librería metagenómica. Además, se utilizó un control externo de crecimiento, empleando arsenito de sodio como agente inhibidor, pues no existen reportes para ninguno de los genotipos, de resistencia a esta especie del arsénico.

2.10 Ensayo de tamizaje funcional

A partir de los umbrales seleccionados, fueron implementados ensayos de tamizajes funcionales con altas concentraciones de arsenato de sodio como agente seleccionador de los clones resistentes.

Como control de procedimiento, se utilizó medio LB, con suplemento de Cloranfenicol (12,5 µg/mL) y sin la adición de arsénico, seleccionando cuatro de las cajas de la librería metagenómica, con el fin de determinar que la inhibición de crecimiento a los clones es debido al efecto del arsénico en el medio de cultivo. Las cajas fueron dispuestas a 37°C durante 16 horas.

Posteriormente, el medio base (LB + Cloranfenicol), fue suplementado con concentraciones de 20 mM y 50 mM de arsenato de sodio. Con la ayuda de un repicador manual, se logró sembrar por caja de polietileno el correspondiente a seis cajas de la librería metagenómica, para un total de 2304 clones por caja.

Todas las cajas fueron incubadas durante 24 horas a 37 °C. Al cabo de ese tiempo, aquellos clones que demostraron crecimiento, fueron seleccionados y aislados nuevamente de manera individual en medios de cultivo con las mismas características. Así mismo, los clones seleccionados, o clones positivos a resistencia a arsenato de sodio, fueron guardados a una temperatura de -70 °C en un medio líquido LB, con 20 % de glicerol.

2.11 Búsqueda molecular de elementos conservados en los clones positivos

A partir de la información reportada en el trabajo de Archour *et al.* (2007), se utilizaron los *primers* degenerados denominados como *darsB1F* y *darsB1R*, los que amplifican regiones conservadas del gen *arsB* el cual codifica para un transportador de arsenito acoplado al mecanismo de la reducción bacteriana de arsénico. Las secuencias de los *primers* son:

darsB1F: 5' – GGTGTGGAACATCGTCTGGAAYGCNAC – 3'

darsB1R: 5' – CAGGCCGTACACCACCAGRTACATNCC – 3'

Se realizó una amplificación por *PCR* de los clones que presentaron resistencia a arsénico en el tamizaje funcional, siguiendo las condiciones optimizadas por Carrillo (2012), variando la adición de ADN bacteriano por un pequeño fragmento de colonia bacteriana (*Colony PCR*). Las tablas 3 y 4 resumen las condiciones generales del proceso de amplificación del fragmento del gen *arsB*.

Tabla 3. Concentraciones finales de reactivos para la amplificación por *PCR* de un fragmento del gen *arsB* (Carrillo, 2012).

| Reactivo | Concentración en reacción de <i>PCR</i> final |
|---|---|
| Buffer Enzima <i>Taq</i> (<i>Bioline</i>) | 1X |
| MgCl ₂ | 1,5 mM |
| dNTP's | 0,2 mM |
| <i>Primer F</i> | 0,2 μM |
| <i>Primer R</i> | 0,2 μM |
| <i>Taq</i> Polimerasa | 0,1 U/μL |

Tabla 4. Condiciones generales de la *PCR* para la amplificación de un fragmento del gen *arsB* (Carrillo, 2012).

| Número de ciclos | Temperatura | Tiempo |
|------------------|-------------------------------|--------|
| 1 | 94 °C | 5 min |
| 10 | 94 °C | 45 s |
| | 57-52 °C (TD* - 0,5 °C/ciclo) | 45 s |
| | 72 °C | 30 s |
| 20 | 94 °C | 45 s |
| | 52 °C | 45 s |
| | 72 °C | 30 s |
| 1 | 72 °C | 7 min |

* TD: *Touchdown*

Así mismo, se utilizaron los *primers* P52F y P323R (Bachate *et al.*, 2009) los cuales amplifican una región del sistema de la arsenato reductasa (ArsC). Se realizó el procedimiento de amplificación siguiendo el protocolo optimizado de Carrillo (2012). Las secuencias de los *primers* son:

P52F: 5' – AGCCAAATGGCAGAAGC – 3'

P323R: 5' – GCTGGRTCRTCAAATCCCCA – 3'

2.12 Preparación de ADN fosmídico

Se realizó un crecimiento para cada uno de los clones positivos en 50 ml de medio líquido LB + Cloranfenicol (12,5 µg/mL) + solución de autoinducción para la generación de múltiples copias del fósido durante 20 horas a 37 °C. Posteriormente se realizaron extracciones por medio del método de lisis alcalina con SDS *Midiprep* (Sambrook y Russell, 2001). Con el fin de mantener la mayor integridad posible del ADN fosmídico, los tiempos de espera con la solución de acetato de potasio y ácido acético fueron reducidos a 3 minutos.

Así mismo, se realizó un crecimiento de los clones positivos, escalando el volumen de medio de cultivo a 200 ml. Mediante centrifugaciones consecutivas, se reconcentraron a un volumen final de 50 ml y posteriormente se realizó la extracción mediante lisis alcalina.

Finalmente, se realizó la contratación del servicio de *Macrogen Inc.* (Corea del Sur), con el fin de realizar un nuevo proceso de extracción y purificación de ADN fosmídico de los clones positivos por separado, así como un proceso de extracción de una mezcla del crecimiento de los clones positivos, con el fin de obtener una *pool* de ADN para ser utilizado en el proceso de secuenciación masiva.

2.13 Estrategia de secuenciación

Se realizó una estrategia combinada utilizando los métodos de *Sanger* y pirosecuenciación, con los sistemas de capilaridad y 454 *FLX+* (*Roche*®), respectivamente.

La mezcla de ADN de los clones positivos fue utilizada para la creación de otra librería para la realización del proceso de pirosecuenciación. No obstante, la secuenciación sobre dicha mezcla genera lecturas de la totalidad de los fragmentos fosmídicos, sin identificar al clon específico del cuál proviene.

Por ello, se planeó una complementación de la pirosecuenciación con fragmentos de secuencias por el método capilar. Teniendo en cuenta que los fragmentos de ADN metagenómico de cada uno de los fósidos tiene un tamaño aproximado de 40.000 pares de bases, los fragmentos de la secuenciación por *Sanger* actúan como identificadores para poder organizar las pirosecuencias en cada uno de los clones provenientes. A continuación se señalan los *primers* utilizados para el proceso de secuenciación por *Sanger*. Estos *primers* corresponden a los extremos de clonación del fósido pCC2FOS (*Epicentre*®):

pCC2™ Forward Sequencing Primer 5'-GTACAACGACACCTAGAC-3'
pCC2™ Reverse Sequencing Primer 5'-CAGGAAACAGCCTAGGAA-3'

2.14 Ensamblaje de las secuencias

Los datos genómicos generados por el sistema *454 FLX+* (*Roche*®) son comprimidos en un archivo binario con extensión *.sff*. Con el fin de descomprimir los datos y realizar la conversión del archivo binario a un formato *.fastq*, se utilizó el *script sff_extract* (*GPLv2*, 1991) desarrollado con el paquete *MIRA* por Chevreux *et al.* (1994; 2004). El formato *.fastq* contiene en un mismo archivo las lecturas de cada uno de los nucleótidos, así como sus correspondientes valores de calidad *Phred*. Así mismo, la mayoría de algoritmos y paquetes bioinformáticos destinados a la realización de procesos de ensamblaje reconocen este formato como fuente de datos primaria.

Se utilizó el algoritmo *MIRA v. 3.9* (Chevreux *et al.*, 1994; 2004) para realizar el proceso de ensamblaje. Se realizó un test de memoria con la utilización del *script miramem*. Luego de ello el archivo fue fraccionado en 7 subarchivos, con el fin de facilitar su procesamiento en la memoria *RAM*. De esta manera, se generaron 6 archivos con 25.000 lecturas cada uno y un archivo adicional con 7.583 lecturas.

Cada uno de los subarchivos fue ensamblado mediante el algoritmo *MIRA* con las siguientes especificaciones para el archivo de configuración:

```
job = genome,denovo,accurate
technology = 454
```

Varias iteraciones fueron realizadas con el fin de reagrupar las secuencias en *contigs* cada vez más grandes hasta la conformación de los *scaffolds*.

2.15 Preparación de ORF's

Se realizó la traducción y la posterior identificación de *ORF's* a partir del set de *scaffolds*, mediante el programa *getorf* de la suite bioinformática *EMBOSS* (Rice *et al.*, 2000) con los siguientes parámetros para la selección de uso codónico bacteriano y restricción de tamaño mínimo del marco de lectura:

```
$ getorf -table 11 -minsize 60 scaffolds.fasta
```

2.16 Anotación funcional

A partir de la preparación de los *ORF's*, fueron implementados tres sistemas de anotación funcional: i) mediante grupos ortólogos, ii) mediante dominios conservados, iii) mediante anotación jerarquizada.

La anotación mediante grupos ortólogos fue realizada utilizando el sistema automatizado *KAAS* (Moriya *et al.*, 2007), el cual utiliza la base de datos de *KEGG* con el fin de realizar una anotación jerarquizada de funciones y rutas metabólicas.

La anotación mediante dominios conservados se realizó mediante un alineamiento múltiple automatizado entre los *ORF's* y la base de datos de los dominios conservados o *CDD* (*Conserved Domain Database*) (Marchier-Bauer & Bryant, 2004; Marchier-Bauer *et al.*, 2005)

Finalmente, la anotación jerarquizada fue realizada con la utilización de la plataforma virtual automatizada de *MG-RAST* (Meyer *et al.*, 2009).

2.17 Búsqueda e identificación de genes de resistencia a arsénico

Con el fin de realizar la búsqueda de genes de resistencia a arsénico al interior de los *scaffolds* construidos, se creó una base de datos local con estas secuencias, utilizando el paquete bioinformático *BLAST+* (v. 2.2.28) (Camacho *et al.*, 2009).

```
$ makeblastdb -in scaffolds.fasta -dbtype nucl -out scaffoldsdb -name scaffoldsdb
```

Se generaron tres archivos *multifasta* con elementos relacionados con el metabolismo y la resistencia bacteriana a arsénico. El primero de ellos, con las secuencias de los genes de la arsenato reductasa *arsC* (ID. BAE77791.1), la bomba de expulsión *arsB* (ID. BAE77792.1) y el gen regulador del operón *arsR* (ID. BAE77793.1) del genoma de *E. coli* W3110. Se realizó un alineamiento con el algoritmo *blastn* (Zhang *et al.*, 2000) con el fin de buscar el operón *ars* al interior de las secuencias.

```
$ blastn -db scaffolds -query 3genes.fasta -out 3genes.out
```

También se tomaron las secuencias de cada una de las proteínas que componen el operón *arsRDABC* en *Klebsiella pneumoniae* (emb|CAG29009.1; emb|CAG29010.1; emb|CAG29011.1; emb|CAG29012.1; emb|CAG29013.1). A partir de ellas se realizó un alineamiento utilizando el algoritmo *tblastn* (Altschul *et al.*, 1997) con el fin de buscar alineamientos en la traducción proteica de la base de datos, partiendo de una secuencia de proteína para la búsqueda.

```
$ tblastn -db scaffolds -query arsrabc.fasta -out arsrabc.out
```

Finalmente se utilizaron secuencias proteicas de regiones conservadas de la subunidad del sitio *rieske* para la arsenito oxidasa (cd03476) y la subunidad pequeña de la arsenito oxidasa.

3. Resultados y Discusión

3.1 Creación de una librería metagenómica

3.1.1 Concentraciones de arsénico de las muestras ambientales

La Tabla 1 presenta el resultado de las mediciones analíticas de arsénico total para las muestras tomadas en suelos, sedimentos y agua.

Tabla 1. Concentraciones de arsénico total en muestras de los distintos sitios de muestreo.

| Tipo de muestra | Contenido total de arsénico |
|--------------------------|-----------------------------|
| Suelo (Finca Sta. Ana) | 2,70 ± 0,11 mg / kg |
| Suelo (Finca Sta. Anita) | 2,38 ± 0,10 mg / kg |
| Sedimento | 1,01 ± 0,15 mg / kg |
| Agua | 1,34 ± 0,17 µg / L |

Los resultados analíticos muestran mayores concentraciones de arsénico en la matriz de los suelos. La intensa actividad agrícola en estas fincas que cuentan con un historial de explotación de más de cincuenta años, ha requerido de una alta utilización de plaguicidas de diverso origen. Muchos de estos plaguicidas contienen arsénico como componente, lo que ha ocasionado que este metaloide se acumule preferencialmente en este tipo de matriz, como por ejemplo el metanoarsonato monosódico el cual es utilizado como controlador de arvenses en Colombia. Corsini *et al.* (2010), han señalado que en suelos con valores ácidos de pH, la afinidad de las especies solubles de arsénico tiende a aumentar, debido a la asociación principalmente de los iones arsenato a minerales de Fe, Mn y Al.

El proceso de lixiviación y drenaje natural de los suelos genera un flujo que se dirige al cauce del Río Bogotá. No obstante, teniendo en cuenta la corriente hídrica, es explicable la baja concentración de arsénico total en la muestra recuperada del agua y un mayor contenido que ha sido depositado en los sedimentos del lecho del Río, debido a su afinidad por los elementos coloidales y demás minerales como hierro que forman nuevos compuestos asociados con el arsénico.

3.1.2 Extracción y preparación de ADN de alto peso molecular

La muestra inicial para la creación de la librería metagenómica, se efectuó en el sector de la cuenca media del Río Bogotá, caracterizado por los altos niveles de contaminantes provenientes de las descargas de los residuos urbanos, así como la lixiviación de los pesticidas utilizados en la producción hortícola que allí se realiza. El lecho del río ha aportado históricamente sedimentos a los suelos ribereños mediante un constante aporte de materia orgánica.

De esta manera, altos contenidos de materia orgánica y de contaminantes como metales pesados y metaloides, hacen parte constitutiva de este particular ambiente. Estos dos factores son los principales problemas relacionados con la manipulación de los ácidos nucleicos, desde su extracción y purificación, hasta las reacciones centrales para el proceso de generación de una librería genómica (Di Bella *et al.*, 2013).

La Figura 11 muestra el resultado de la comparación de las extracciones utilizando dos kits comerciales. Las lecturas realizadas por espectrofotometría, resultaron en concentraciones finales de ADN que oscilaron entre 5 ng / μ l – 15 ng / μ l (Figura 8, líneas 5 y 6) para el kit *Meta-G-Nome™ DNA Isolation Kit (Epicentre®)* y concentraciones entre 80 ng/ μ l – 190 ng/ μ l (Figura 8, líneas 3 y 4) para el kit *PowerMax™ Soil DNA Isolation Kit (MO-BIO®)*.



Figura 11. Comparación de métodos de extracción. Línea 1, control positivo con adición de 100 ng de ADN lambda (tamaño molecular de 48,5kb); Línea 2, control positivo con adición de 20 ng de ADN Control (*Epicentre*®) (tamaño molecular de 40kb); Líneas 3 y 4, productos de extracción de ADN con *PowerMax*TM *Soil DNA Isolation Kit* (*MO-BIO*®), cada pozo con 5 µl; Líneas 5 y 6, productos de extracción de ADN con *Meta-G-Nome*TM *DNA Isolation Kit* (*Epicentre*®), cada pozo con 5µl. Gel de agarosa (1%); 80V constantes; 90 minutos.

Para la extracción de ADN proveniente de las muestras de agua del Río Bogotá se utilizó el Kit *Meta-G-Nome*TM *DNA Isolation Kit* (*Epicentre*®), siguiendo las recomendaciones del protocolo. Sin embargo, no se lograron concentraciones de ADN visibles en la electroforesis (no mostrado), razón por la cual, las muestras de agua fueron descartadas para los procedimientos posteriores.

A pesar de que se lograron altas concentraciones de ADN favorables para la creación de la librería, el barrido observado en los pozos 3 y 4 de la Figura 11, es consistente con degradación de los ácidos nucleicos, así como impurezas presentes en el ADN, provenientes de la materia orgánica, ácidos húmicos y fúlvicos presentes en las muestras de suelos y sedimentos. Por esta razón, se hizo necesario un procedimiento que

purificase las muestras de ADN al mismo tiempo separara los fragmentos de alto peso molecular.

Dado que el proceso de creación de la librería metagenómica está condicionado por el tamaño del vector de clonación, así como el tamaño óptimo para los insertos, fue necesaria la separación de ADN de alto peso molecular. El fósmdido de clonación pCC2FOS (*Epicentre*®) (Figura 12), requiere insertos de alrededor 40 kb de tamaño para la realización de una clonación exitosa.

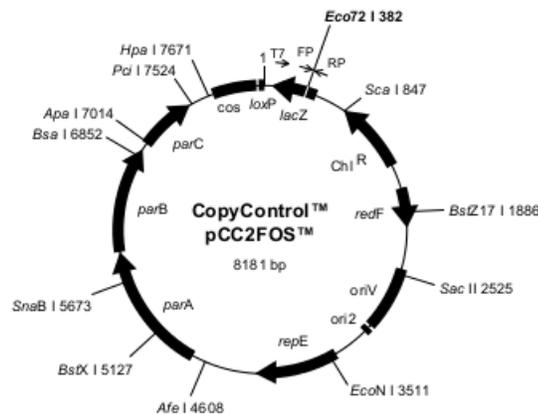


Figura 12. Mapa del vector de clonación pCC2FOS (*Epicentre*®).

El proceso de electroforesis resultó adecuado para la remoción de los agentes inhibidores así como para la selección del tamaño adecuado del ADN requerido (Figura 13). Las muestras fueron posteriormente purificadas, proceso luego del cual se obtuvo una concentración aproximada de 80 ng/μl, cantidad apropiada para la construcción de la librería metagenómica. Estos fragmentos fueron posteriormente fosforilados, requisito necesario para su proceso de ligación al vector de clonación dado que la enzima ligasa reconocerá solo fragmentos con extremos terminados y fosforilados.

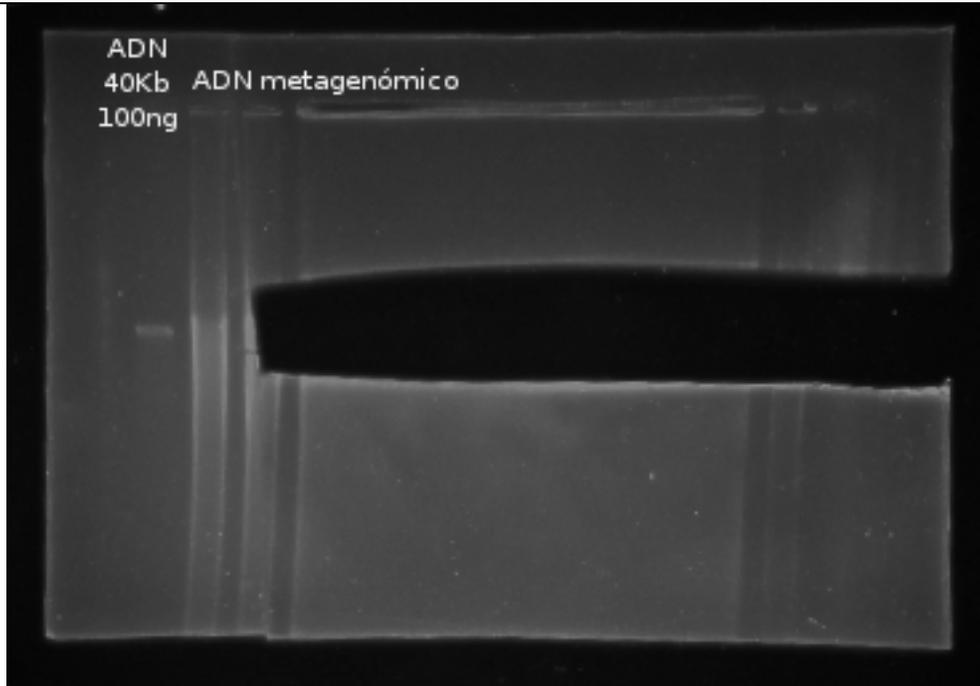


Figura 13. Gel de electroforesis mostrando la región donde se realizó la selección y corte de ADN metagenómico de alto peso molecular (40 kb aprox.) para construcción de librería metagenómica. Gel de agarosa (1,5%); 30V constantes; 16 horas.

El método implementado para extracción de ADN está basado en la separación por membranas y columnas. De esta manera el Kit *PowerMax™ Soil DNA Isolation Kit (MO-BIO®)* demostró ser efectivo en para este propósito. Al respecto, el trabajo realizado por Sagar *et al.* (2014) compara varios métodos de extracción de ADN de muestras de suelos, concluyendo que la tecnología ofrecida por *MO-BIO®* resulta ser la más eficiente en la remoción de ácidos húmicos. Sin embargo, dado que la remoción no es total, la realización de la electroforesis y la posterior elución de las bandas, mejoró la separación de los agentes inhibidores. El método por electroforesis es ampliamente señalado como una de las mejores estrategias de remoción (Hurt *et al.*, 2001; Santosa, 2001).

3.1.3 Proceso de creación de la librería metagenómica

Luego de la realización de los procesos de ligación, empaquetamiento e infección en la célula hospedera *E. coli* EPI300-T1®, fue creada la librería metagenómica de suelos y sedimentos de un ambiente de la Sabana de Bogotá, específicamente de la cuenca media del Río Bogotá.

La organización de la librería metagenómica se realizó en 130 cajas de polietileno de 384 pozos cada una para un número de hasta 49.920 clones. Teniendo en cuenta un tamaño teórico esperado de 40 kb por inserto metagenómico de cada clon, la metagenoteca representa una cobertura teórica aproximada de 2.000 Mb de genoma microbiano de este ambiente. La caracterización de la cobertura es un valor estimado, toda vez que no se realizó una medición real a partir de digestiones aleatorias de una muestra representativa de clones.



Figura 11. Ejemplo de caja de cultivo con clones transformantes de muestra del río de Bogotá luego de proceso de infección.

Finalmente, la metagenoteca construida tuvo un tamaño aproximado de 2.000 Mb. Si se tiene en cuenta un tamaño aproximado de 4 Mb por genoma bacteriano (*E. coli*), la librería metagenómica logró una cobertura potencial teórica de alrededor 500 genomas de bacterias.

Estudios con enfoques metagenómicos similares han utilizado tamaños de librerías metagenómicas que varían entre 23.400 clones con fósmidos (Kim *et al.*, 2010), 200.000

clones con plásmidos (Wang *et al.*, 2010) y 92.000 clones con plásmidos (Yao *et al.*, 2010), lo que significa coberturas estimadas alrededor de 936 Mb, 0,8 Mb y 0,37 Mb, respectivamente; coberturas inferiores a la realizada en el presente trabajo.

No obstante, una caracterización con mayor resolución de esta librería podría sugerir un valor de cobertura más aproximado, pues aunque se realizó un proceso de selección de tamaños de ADN metagenómico, estos pueden encontrarse entre un rango de 35 – 45 kb. Se asume el valor de 40 kb como valor medio, sin embargo una caracterización más exhaustiva con la selección y digestión aleatoria de algunos clones, deberá ser realizada con el fin de estimar un valor empírico y no teórico.

Así mismo, la selección de fragmentos de ADN metagenómico de alto peso molecular, delimitó la librería metagenómica a fragmentos de ADN cromosómico, toda vez que los tamaños seleccionados exceden el tamaño de ADN extracromosómico como por ejemplo ADN plasmídico. Por ello, el diseño de una librería metagenómica construida a partir de fósidos como vector de clonación, deberá ser complementada con una librería de plásmidos con el fin de abarcar ADN plasmídico, en donde muchos de los mecanismos reportados de resistencia a arsénico podrían ser identificados.

3.2 Ensayo de tamizaje funcional

3.2.1 Validación de umbrales de arsénico y de medio de cultivo

Con el fin de verificar la viabilidad de la célula hospedera *E. coli* EPI300-T1®, fueron realizados ensayos con adición de arsenato y arsenito y teniendo como referencia cepas de *E. coli* controles para resistencia y susceptibilidad a arsenato. La Tabla 2 y la Figura 12, presentan los resultados de los ensayos de crecimiento con las diferentes cepas control a diferentes concentraciones de arsénico. Estos resultados, permiten concluir, que las concentraciones de 20 mM y 50 mM de arsenato de sodio, tienen un efecto de inhibición de crecimiento sobre el genotipo *E. coli* EPI300-T1®, el cual es utilizado como organismo hospedero para la creación de la librería metagenómica y por lo tanto, es posible afirmar que cualquier variación de crecimiento de los clones en el desarrollo del proceso de barrido funcional, corresponderá a las nuevas funciones fenotípicas que aporta el ADN metagenómico en relación a mecanismos de resistencia a arsénico.

Tabla 2. Ensayos de crecimiento de genotipos de *E. Coli* EPI300-T1, W3110, AW3110 y AW3110 a diferentes concentraciones de arsenato de sodio (NaH_2AsO_4).

| GENOTIPO de <i>E. coli</i> | LB + Arsenato de sodio (20 mM) | LB + Arsenato de sodio (50 mM) | LB | LB + Arsenito de sodio (15 mM) |
|----------------------------|--------------------------------|--------------------------------|-------------|--------------------------------|
| EPI300-T1 | No crecimiento | No crecimiento | Crecimiento | No crecimiento |
| W3110 | Crecimiento | Crecimiento | Crecimiento | No crecimiento |
| MG1655 | Crecimiento | Crecimiento | Crecimiento | No crecimiento |
| AW3110 | No crecimiento | No crecimiento | Crecimiento | No crecimiento |

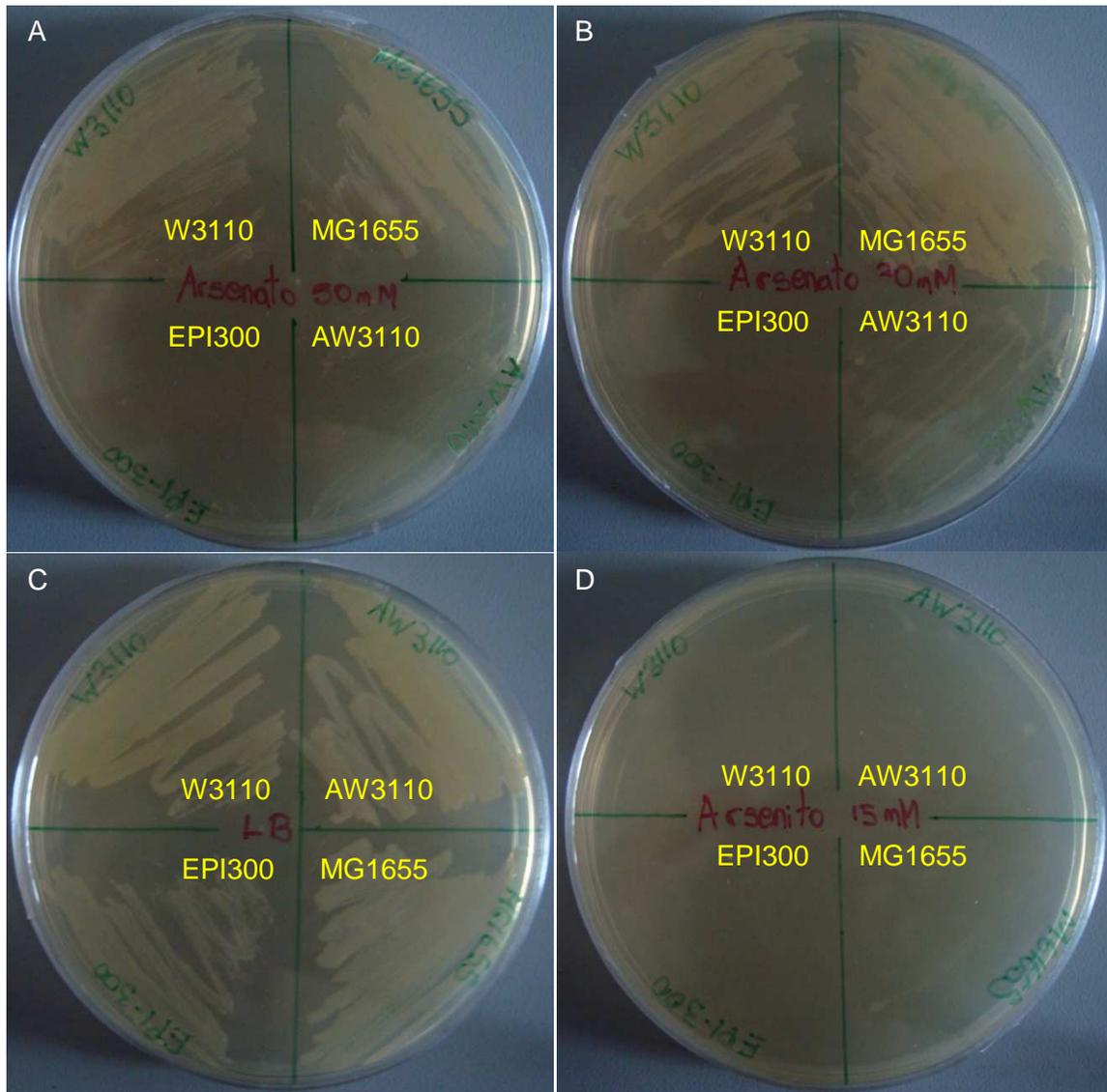


Figura 12. Crecimiento de cuatro genotipos de *E. coli* (W3110 con resistencia a arsénico; MG1655 con resistencia a arsénico; AW3110 con mutación en operón *ars* – baja o nula resistencia a arsénico; EPI300-T1® célula hospedera de la librería metagenómica): A) Ensayo con medio LB suplementado con arsenato de

sodio a 50 mM. B) Ensayo con medio LB suplementado arsenato de sodio 20 mM. C) Ensayo con medio LB sin suplemento de arsénico. D) Ensayo con medio LB suplementado con arsenito de sodio 15 mM.

Luego de la organización de la librería, fueron tomadas cuatro cajas, las que corresponden a una muestra de 1536 clones y crecidas a medio de cultivo LB suplementado con cloranfenicol, con el fin de validar el medio de cultivo a utilizar. La Figura 13 demuestra que las condiciones básicas del medio de cultivo, no presentan factores inhibitorios sobre el crecimiento de los clones de la librería metagenómica y por lo tanto es posible utilizar este medio de cultivo para desarrollar el ensayo de tamizaje funcional con altas concentraciones de arsenato.

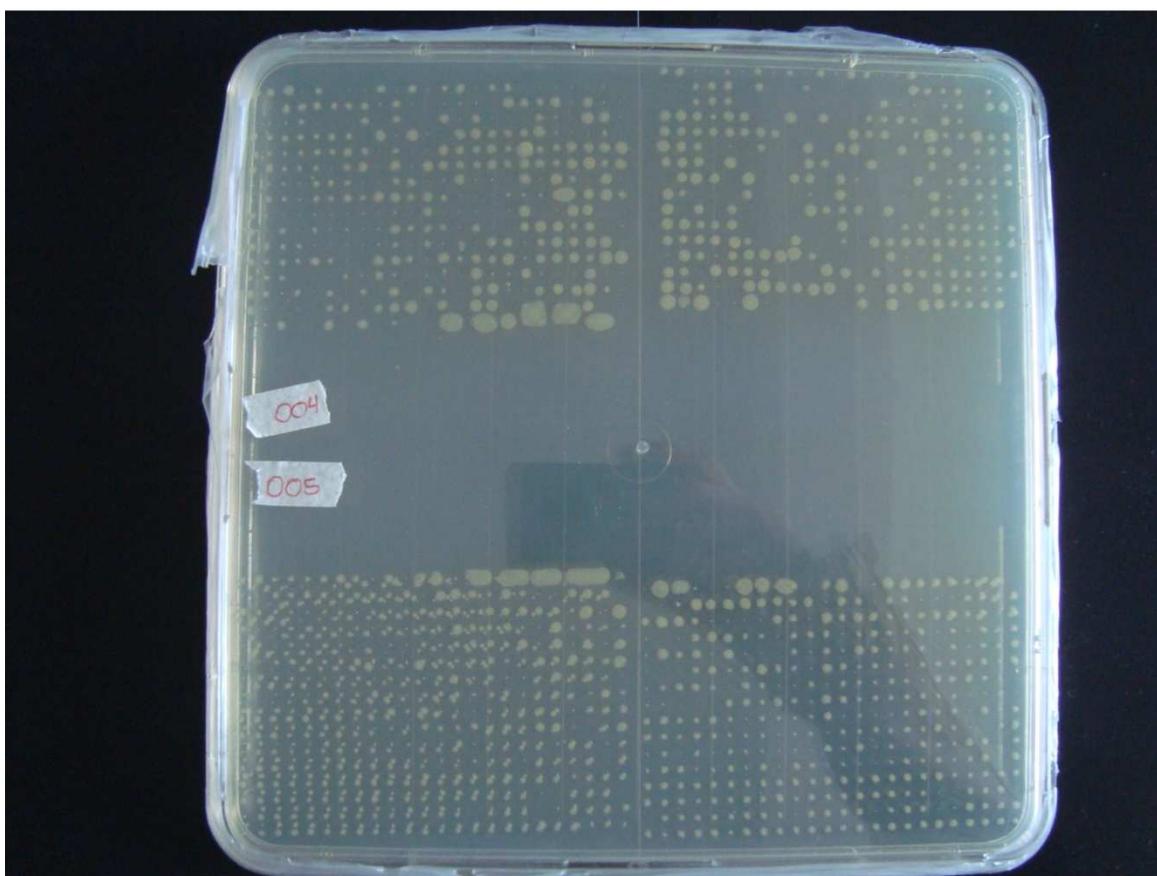


Figura 13. Control de procedimiento del tamizaje funcional. Se sembraron 1536 clones en medio LB + Cloranfenicol, sin la adición de arsénico. Incubación durante 24 horas a 37 °C.

3.2.2 Tamizaje funcional

Finalmente, de 50.000 clones evaluados, se identificaron mediante este procedimiento 9 clones con resistencia a la concentración de 20 mM de arsenato de sodio y 2 clones con

resistencia al umbral de 50 mM para un total de 11 clones con resistencia fenotípica a arsenato de sodio (Figura 14).



Figura 14. Ensayo de tamizaje funcional. Se señalan dos clones con resistencia a una concentración de 50 mM de arsenato de sodio.

De una librería metagenómica de alrededor 50.000 clones, 11 de ellos exhibieron actividad de resistencia a una concentración alta de arsénico; por ello estos clones fueron denominados como clones positivos. La baja tasa de clones positivos, es uno de los principales retos de los tamizajes funcionales, pues allí influyen varios factores: i) la capacidad del vector en relación a su tamaño y sitios de expresión, ii) el tipo de organismo hospedero utilizado, debido a la especificidad de su metabolismo natural, iii) el diseño del barrido funcional, teniendo en cuenta temperatura, condiciones de aerobiosis, luminosidad, etc.

Es por ello que la mayoría de trabajos con enfoques de barridos funcionales logran bajas tasas de recuperación de clones positivos. Ejemplo de ello es el trabajo de Wang *et al.* (2013), en el que a partir de una librería de 48.000 clones, identificaron 19 con actividad positiva para el ensayo; relación similar a la de este trabajo.

3.2.3 Ensayo de validación molecular

A partir de la identificación de once clones resistentes a arsénico, se realizó una búsqueda molecular de elementos reportados, principalmente del operón *ars*. La Figura 15 muestra la amplificación exitosa del gen *arsB*, el cual codifica para la bomba transmembranal de expulsión de arsenito (Figura 3, mecanismo #2).

El fragmento amplificado es consistente con el peso molecular reportado por Achour *et al.* (2007), en el que se obtienen fragmentos de alrededor 750 kb. Este resultado sugiere que los clones resistentes a arsénico tienen un metabolismo relacionado con la reducción de arsenato a arsenito por medio del operón bacteriano *ars*.

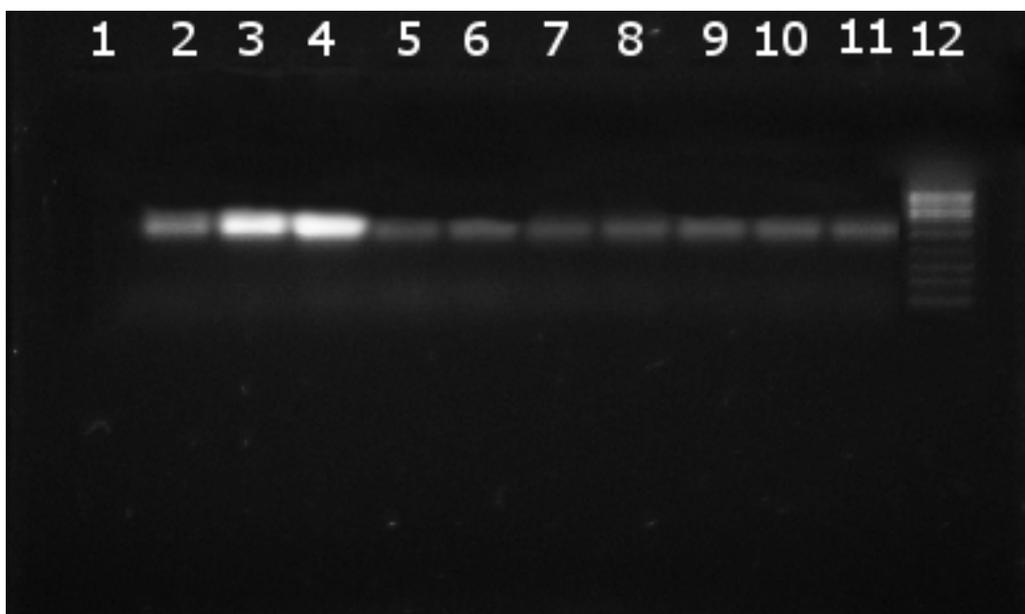


Figura 15. Amplificación de fragmento del gen *arsB* en los clones resistentes a arsénico. La línea 1 es un control negativo de procedimiento con agua desionizada y *buffer* de carga. Las líneas 2 – 11 representan diez clones resistentes a arsénico. La línea 12 representa un marcador de peso molecular de 1Kb (*Axygen*[®]).

No se encontró amplificación para el gen *arsC* el cual es el responsable de la conversión de arsenato a arsenito mediante la acción enzimática de la arsenato reductasa. Estos resultados sugieren, bien sea, la existencia de polimorfismos en el gen que hacen el proceso de amplificación por PCR poco específico con los *primers* utilizados, o la posible existencia de una nueva ruta metabólica de resistencia a arsénico alternativa a la reducción de arsenato a arsenito.

3.3 Secuenciación de clones resistentes a arsénico y ensamblaje de secuencias

3.3.1 Preparación del ADN fosmídico

Las Figuras 16 y 17 muestran los resultados de las extracciones de ADN fosmídico realizadas consecutivamente siguiendo los protocolos *Miniprep* y *Midiprep*, respectivamente. Las extracciones por el método *Miniprep*, resultaron en concentraciones de ADN con valores por debajo de 15 ng/μl, lo cual hizo estas muestras inviables para el proceso de secuenciación. Con las extracciones por el método *Midiprep*, los resultados no variaron significativamente con respecto al primer ensayo, pues se obtuvieron valores de concentraciones por debajo de 20 ng/μl.

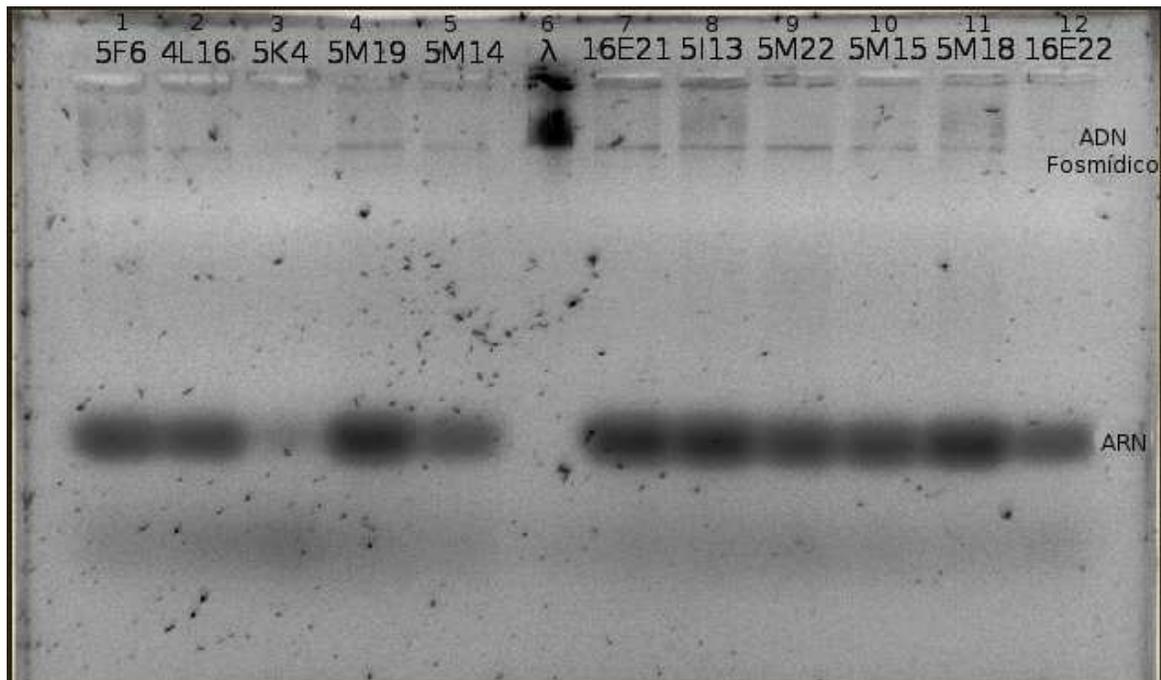


Figura 16. Extracción de fósmidos mediante el método *Midiprep* (Sambrook y Russell, 2001). El carril 6 corresponde a una muestra de ADN de fago lambda (100ng). Los restantes carriles corresponden a 5 µl de ADN fosmídico de cada uno de los 11 clones positivos.

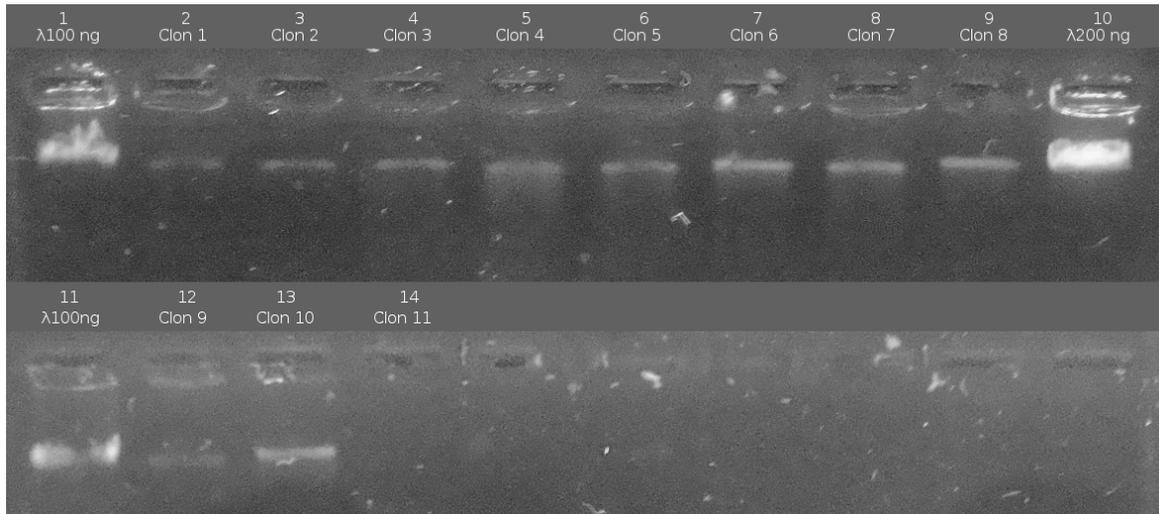


Figura 17. Extracción de fósmidos mediante el método *Midiprep* (Sambrook y Russell, 2001). Los carriles 1 y 10 (arriba) y el carril 1 (abajo), corresponden a muestras de ADN de fago lambda. Los restantes carriles corresponden a 5 µl de ADN fosmídico de cada uno de los 11 clones positivos.

La Figura 18 muestra el resultado de la extracción y purificación de ADN fosmídico de los 11 clones positivos realizado mediante la contratación del servicio de extracción y purificación por *Macrogen Inc.* La concentración obtenida resultó suficiente para el posterior proceso de secuenciación.

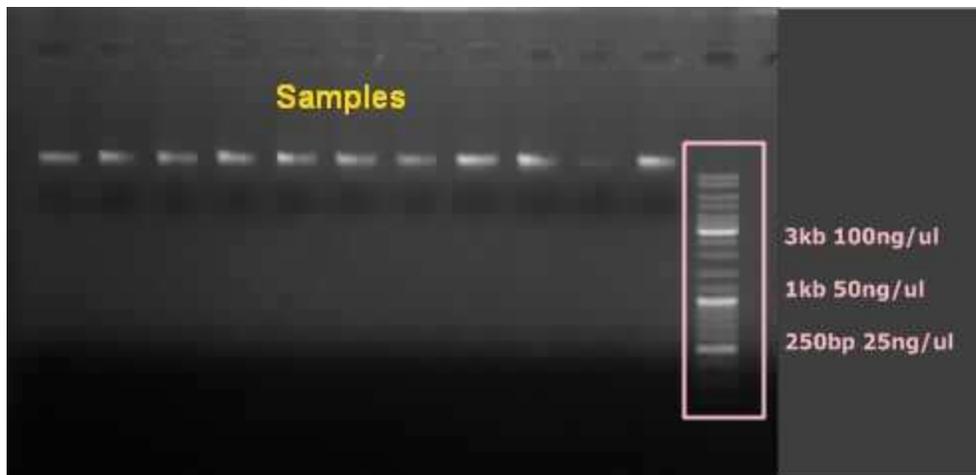


Figura 18. Extracción de ADN fosmídico de los 11 clones con ADN metagenómico positivos para el crecimiento en 20 y 50 mM de arsenato de sodio. Imagen tomada por *Macrogen Inc.*

3.3.2 Obtención de secuencias

Fueron recibidas 22 secuencias *Sanger*, correspondientes a las secuencias de los extremos de los fósidos de cada clon. Los resultados no arrojaron secuencias confiables, pues sus valores de calidad resultaron muy bajos y su electroferograma es típico de una lectura errada (Anexo B). Más del 70 % de las bases tienen valores de calidad *Phred* por debajo de 10, lo cual quiere decir que más del 70 % de las lecturas tiene una probabilidad de error de lectura de por lo menos 1 en 10 bases (Ewing & Green, 1998).

Se recibió también el resultado de la secuenciación masiva por pirosecuenciación, arrojando como resultado 157.583 secuencias, que representan 7'113.435 bases en total. El tamaño promedio por lectura fue de 451 bases (Figura 19). Una submuestra aleatoria de los valores de calidad del total de las bases, señala un promedio de calidad de un valor *Phred* de 32, lo cual quiere decir que las lecturas tienen una probabilidad de error de lectura inferior a 0,1 %.

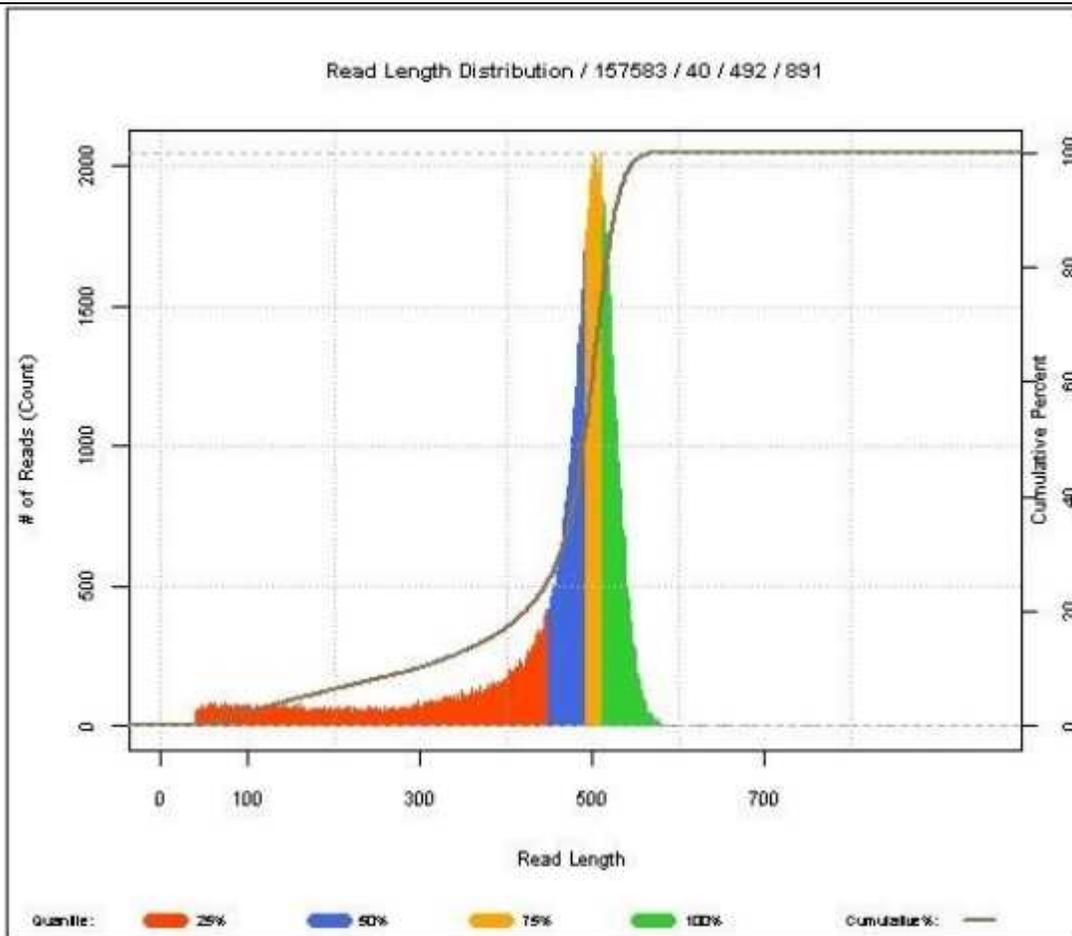


Figura 19. Diagrama de distribución de tamaño de las lecturas de secuencias por el sistema 454 FLX+ (Roche®). La mayoría de las lecturas se ubican en un rango (eje x) de 450 – 550 bases por lectura. Imagen de MacroGen Inc.

La estrategia de secuenciación utilizada en el proyecto consistió en complementar las posibilidades de secuenciación masiva del sistema 454 FLX+ (Roche®) para la obtención del contenido genómico total de los 11 clones positivos a la resistencia a arsenato de sodio, con la posibilidad de utilización del sistema de secuenciación por el método de *Sanger* con el fin de ubicar y organizar las lecturas de pirosecuenciación con su clon de origen, utilizando los extremos de los fósmidos a manera de etiquetas.

No obstante, las secuencias *Sanger* registraron valores muy bajos de calidad, que sugieren que no existe un proceso específico de secuenciación con los *primers* utilizados. Sin embargo el proceso de secuenciación por el método masivo 454 FLX+ (Roche®), registró valores de alta calidad de las secuencias.

Los resultados positivos para la secuenciación por el sistema *454 FLX+* (Roche®) y negativos para el sistema *Sanger* sugieren un daño en el mecanismo de acople de los *primers* a los extremos de clonación del vector. Añadiendo a ello la imposibilidad de conseguir cantidades suficientes de ADN fosmídico, es probable que exista un daño en el mecanismo mediado por la arabinosa, el cual es el inductor de la replicación de múltiples copias del fósido. Tanto el mecanismo de replicación múltiple como el sitio de unión de los *primers* son adyacentes, lo que sugiere una región del vector de clonación defectuosa.

A pesar que no fue posible la obtención de las etiquetas de los extremos del vector, las pirosecuencias aportan gran cantidad de información genética sobre la totalidad de los clones resistentes a arsénico encontrado, razón por la que se procedió a la realización del ensamblaje de dichas secuencias y la búsqueda e identificación de genes de resistencia a arsénico

3.4 Análisis de secuencias y anotación funcional

3.4.1 Ensamblaje de las secuencias

De las 157.583 lecturas recibidas, se lograron ensamblar en la primera iteración del algoritmo utilizado, un resultado de 8.835 *contigs* con un tamaño promedio de 1.014 pares de bases (Tabla 5).

Tabla 5. Resumen de primer proceso de ensamblaje de secuencias

| Subarchivo | Número de lecturas | Contigs generados | Tamaño medio de contigs |
|-------------------|---------------------------|--------------------------|--------------------------------|
| 1 | 25.000 | 1.894 | 1.175 |
| 2 | 25.000 | 1.847 | 1.238 |
| 3 | 25.000 | 622 | 912 |
| 4 | 25.000 | 1.867 | 1.154 |
| 5 | 25.000 | 1.854 | 1.079 |
| 6 | 25.000 | 630 | 924 |
| 7 | 7.583 | 121 | 616 |
| Total | 157.583 | 8.835 | 1.014 |

La segunda iteración realizada a partir de los *contigs* generados, los cuales fueron reagrupados en dos archivos (*a* y *b*), generaron un resultado de 3.216 nuevos *contigs* con un tamaño promedio de 1.116 pares de bases (Tabla 6).

Tabla 6. Resumen del segundo proceso de ensamblaje de las secuencias

| Subarchivo | Número de lecturas | Contigs generados | Tamaño medio de contigs |
|-------------------|---------------------------|--------------------------|--------------------------------|
| a | 4.480 | 1.488 | 1.130 |
| b | 5.811 | 1.728 | 1.101 |
| Total | 10.691 | 3.216 | 1.116 |

Finalmente, los *contigs* generados a partir de la segunda iteración del algoritmo, fueron procesados bajo las mismas condiciones. El proceso de ensamblaje concluyó con 951 *scaffolds*, con un tamaño promedio de 1.075 pares de bases, los cuales fueron utilizados para la realización de los análisis y procedimientos bioinformáticos (Tabla 7) (Figura 20).

Tabla 7. Resumen del proceso final de ensamblaje

| Subarchivo | Número de lecturas | Scaffolds generados | Tamaño medio de scaffolds |
|-------------------|---------------------------|----------------------------|----------------------------------|
| | 3.216 | 951 | 1.075 |

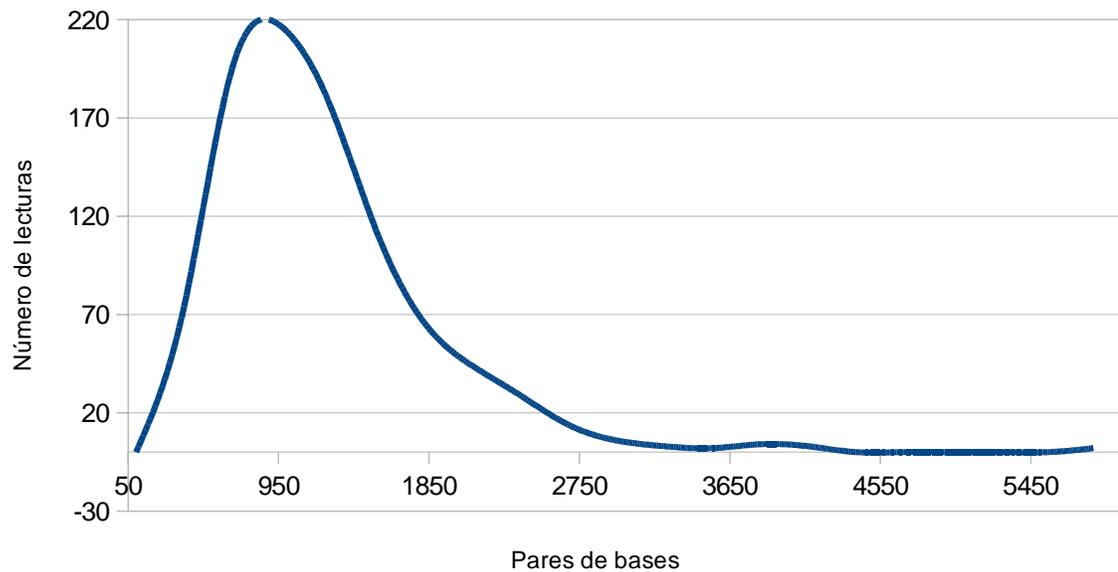


Figura 20. Histograma de tamaño de *scaffolds* construidos.

Debido a que los parámetros para el algoritmo *MIRA* fueron ajustados para el ensamblaje *de novo* de secuencias 454 las cuales son propensas a errores y a generación de quimeras, ello resultó en una mayor selectividad al momento de la creación de los *contigs* y posteriormente de los *scaffolds*, razón por la cual no se lograron ensambles de mayor longitud. No obstante, el tamaño medio de los *scaffolds* generados fue adecuado para los análisis bioinformáticos posteriores.

3.4.2 Preparación e identificación de marcos abiertos de lectura (*ORF*)

Se logró la identificación de 46.082 marcos abiertos de lectura (*ORF*), de los cuales 30.761 corresponden al sentido 5'-3' y los restantes 15.321 en antisentido, en los seis marcos de lectura probables existentes.

3.4.3 Anotación jerarquizada (*MG-RAST*)

Se utilizó la plataforma virtual *MG-RAST* (Meyer *et al.*, 2009) con el fin de realizar un proceso automatizado de anotación funcional jerarquizada de las secuencias organizadas en *scaffolds*.

Las anotaciones realizadas están relacionadas con especies bacterianas que en su mayoría pertenecen al *phylum* Proteobacteria, destacándose el orden de las Enterobacteriales, especies que sobreviven generalmente en sitios contaminados con altos contenidos de materia orgánica y poca disponibilidad de oxígeno como es el caso de la cuenca media del Río Bogotá, sus sedimentos y suelos aledaños (Figura 21).

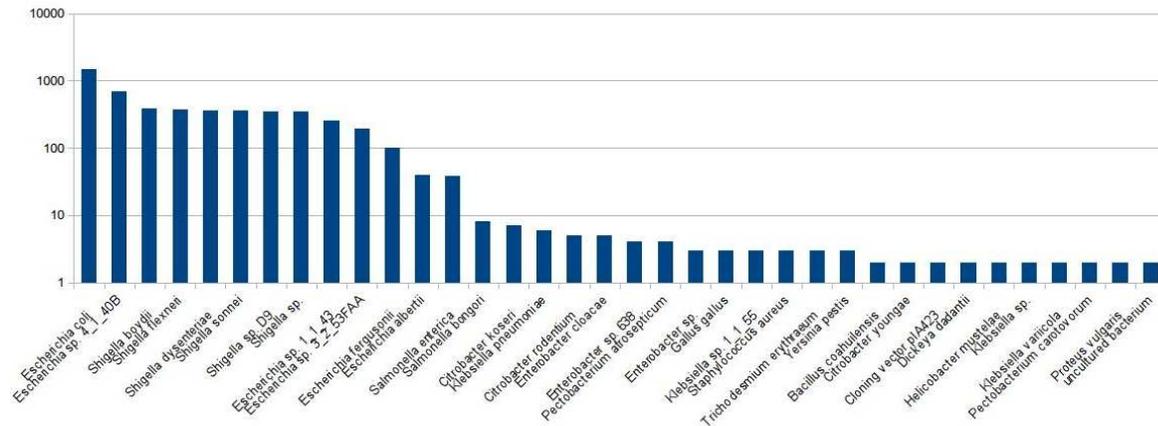


Figura 21. Principales especies relacionadas con la anotación funcional.

Las principales categorías funcionales analizadas teniendo en cuenta las bases de datos COG y KO, están relacionadas al metabolismo celular y procesamiento de información ambiental (Figura 22).

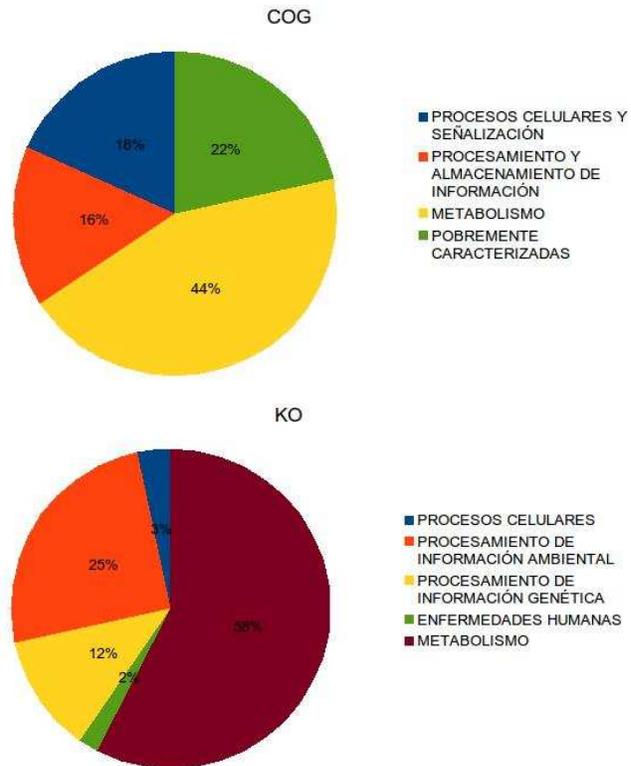


Figura 22. Principales grupos funcionales de COG y KO anotados mediante MG-RAST.

Finalmente, las anotaciones funcionales relacionadas con el metabolismo del arsénico fueron encontradas en la búsqueda con las bases de datos de *IMG*, *KEGG*, *PATRIC*, *SEED*, *SwissProt*, *TrEMBL*, *eggNOG*. Los elementos encontrados, están relacionados con el mecanismo del operón *ars*, el cual es responsable de la reducción del ion arsenato a arsenito (Tabla 8) (Figura 3, mecanismo #2).

Tabla 8. Anotaciones relacionadas con el metabolismo del arsénico, realizadas por medio de *MG-RAST*.

| FUENTE | FUNCIÓN |
|---------------|--|
| IMG | glutaredoxin family-like arsenate reductase |
| KEGG | Anion permease ArsB/NhaD |
| PATRIC | Arsenate reductase |
| PATRIC | Arsenic efflux pump protein |
| SEED | Arsenate reductase (EC 1.20.4.1) |
| SEED | Arsenic efflux pump protein |
| SwissProt | Putative arsenate reductase |
| SwissProt | Putative arsenical pump membrane protein |
| TrEMBL | Glutaredoxin family-like arsenate reductase |
| eggNOG | Arsenate reductase and related proteins, glutaredoxin family |
| eggNOG | Na ⁺ /H ⁺ antiporter NhaD and related arsenite permeases |

La significancia de anotaciones generadas, sugieren que las secuencias provienen de organismos de tipo enterobacteriales, resultados que también son consistentes con las anotaciones funcionales generadas, las cuales identifican la mayoría de secuencias con bacterias de este tipo. De manera consistente, en el trabajo realizado por Chauhan *et al.* (2009), el tipo de genes identificados estuvo estrechamente relacionado con secuencias reportadas para Proteobacterias, las cuales son mayoritarias para este tipo de ambientes.

3.4.4 Anotación mediante grupos ortólogos

Se realizó una selección de grupos ortólogos de *KEGG* relacionados de manera directa o indirecta con el metabolismo del arsénico (Figura 23). Los resultados señalan que existe anotación de secuencias relacionadas con sistemas de transporte, de secreción y de metabolismo microbiano en diversos ambientes. Estos resultados son congruentes con los ensayos funcionales, pues al existir una presión de selección mediada por altas concentraciones de arsénico en el entorno exterior, los microorganismos que utilicen preferencialmente mecanismos de interacción con el medio ambiente, detoxificación, transporte y secreción de compuestos tóxicos, tendrán las posibilidades de asegurar su crecimiento en los ensayos funcionales.

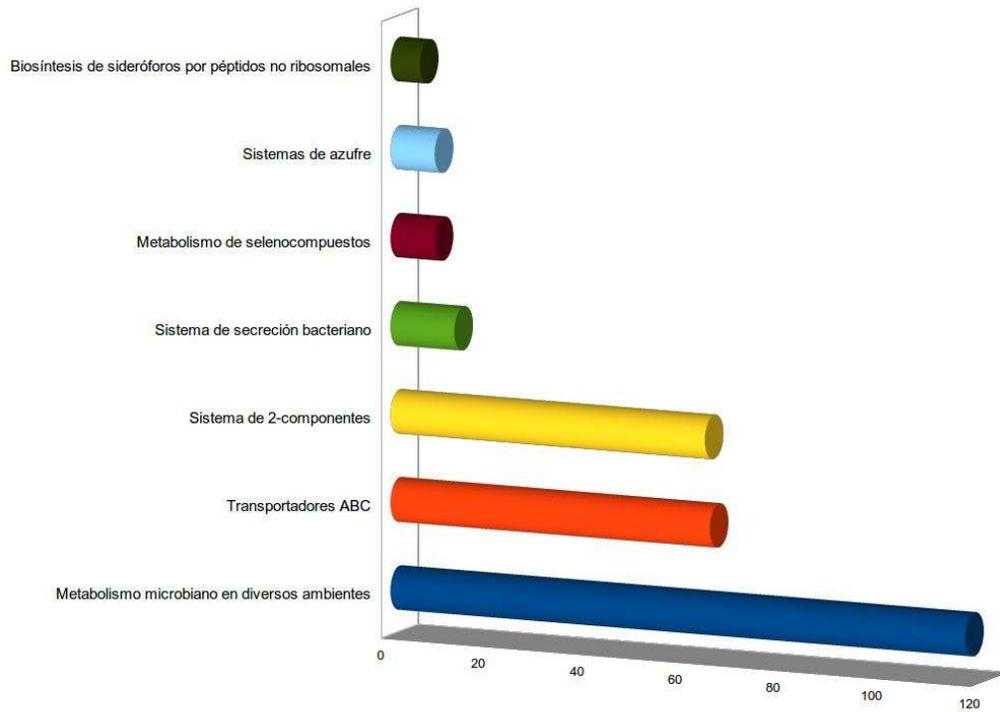


Figura 23. Principales grupos ortólogos relacionados con el metabolismo del arsénico. Los valores numéricos representan el número de anotaciones realizadas para cada categoría.

Inicialmente, las anotaciones funcionales por categorías funcionales mostraron una marcada tendencia de la prevalencia de genes microbianos relacionados con el medio ambiente. La categoría “*Microbial metabolism in diverse environments*” implementada por *KEGG*, incluye una gran variedad de funciones microbiológicas entre las que se encuentran funciones principales de los ciclos biogeoquímicos, pero también la expresión de la mayoría de metabolitos secundarios relacionados con rutas metabólicas de degradación de los compuestos tóxicos mayormente reportados y estudiados. Otras funciones con un número significativo de anotaciones, fueron aquellas relacionadas con mecanismos de secreción y transporte.

3.4.5 Anotación mediante dominios conservados

Se realizó un complemento a la búsqueda de funciones, realizando un alineamiento múltiple automatizado de los *ORF*'s con la base de datos de los dominios conservados o *CDD* (*Conserved Domain Database*) (Marchier-Bauer & Bryant, 2004; Marchier-Bauer *et al.*, 2005), los cuales brindan una información más detallada que distingue dominios

específicos, multidominios y superfamilias de proteínas encontradas. Fueron seleccionados 49 registros que están relacionados con el metabolismo directo del arsénico en bacterias (Tabla 9).

Tabla 9. Registro de dominios conservados encontrados en los *scaffolds* construidos.

| Short name | Hit type | Accession | E-Value | Query |
|-----------------------------------|-------------|-----------|--------------|--|
| ArsB_NhaD_permease superfamily | superfamily | cl17221 | 0.00155183 | Q#18555 ->Scafflod_c31 |
| ArsB_NhaD_permease superfamily | superfamily | cl17221 | 0.00155183 | Q#18555 ->Scafflod_c31[Scafflod_c310_9 [304 - 456]] |
| ArsB_NhaD_permease superfamily | superfamily | cl17221 | 4.09943e-76 | Q#34997 ->Scafflod_c64 |
| ArsB_NhaD_permease superfamily | superfamily | cl17221 | 4.09943e-76 | Q#34997 ->Scafflod_c64[Scafflod_c641_12 [2 - 727]] |
| ArsB_NhaD_permease superfamily | superfamily | cl17221 | 2.31487e-59 | Q#36291 ->Scafflod_rep[Scafflod_rep_c668_8 [1 - 471]] |
| ArsB_NhaD_permease superfamily | superfamily | cl17221 | 1.20026e-147 | Q#3830 ->Scafflod_rep[Scafflod_rep_c42_104 [1926 - 595] (REVERSE SENSE)] |
| ArsB_permease | specific | cd01118 | 2.31487e-59 | Q#36291 ->Scafflod_rep |
| ArsB_permease | specific | cd01118 | 1.20026e-147 | Q#3830 ->Scafflod_rep |
| ArsC | multi-dom | pfam03960 | 2.34794e-42 | Q#3841 ->Scafflod_rep |
| ArsC_ArsC | specific | cd03034 | 5.15748e-61 | Q#3841 ->Scafflod_rep |
| HTH_ARSR | specific | cd00090 | 3.61048e-06 | Q#19446 ->Scafflod_c32 |
| HTH_ARSR | specific | cd00090 | 5.03901e-08 | Q#32175 ->Scafflod_c59 |
| HTH_ARSR | specific | cd00090 | 6.87584e-19 | Q#3802 ->Scafflod_rep |
| HTH_ARSR superfamily | superfamily | cl17220 | 0.00620022 | Q#15357 ->Scafflod_c24[Scafflod_c247_48 [305 - 3] (REVERSE SENSE)] |
| HTH_ARSR superfamily | superfamily | cl17220 | 3.61048e-06 | Q#19446 ->Scafflod_c32 |
| HTH_ARSR superfamily | superfamily | cl17220 | 3.61048e-06 | Q#19446 ->Scafflod_c32[Scafflod_c326_30 [595 - 80] (REVERSE SENSE)] |
| HTH_ARSR superfamily | superfamily | cl17220 | 5.03901e-08 | Q#32175 ->Scafflod_c59 |
| HTH_ARSR superfamily | superfamily | cl17220 | 5.03901e-08 | Q#32175 ->Scafflod_c59[Scafflod_c599_72 [261 - 1] (REVERSE SENSE)] |
| HTH_ARSR superfamily | superfamily | cl17220 | 1.49022e-07 | Q#34059 ->Scafflod_c62 |
| HTH_ARSR superfamily | superfamily | cl17220 | 1.49022e-07 | Q#34059 ->Scafflod_c62[Scafflod_c624_27 [989 - 1270]] |
| HTH_ARSR superfamily | superfamily | cl17220 | 0.0059968 | Q#34774 ->Scafflod_c63[Scafflod_c635_36 [989 - 189] (REVERSE SENSE)] |
| HTH_ARSR superfamily | superfamily | cl17220 | 6.87584e-19 | Q#3802 ->Scafflod_rep[Scafflod_rep_c42_76 [2354 - 1938] (REVERSE SENSE)] |
| HTH_ARSR superfamily | superfamily | cl17220 | 0.00426708 | Q#6860 ->Scafflod_c92[Scafflod_c92_38 [450 - 1] (REVERSE SENSE)] |
| HTH_ARSR superfamily | superfamily | cl17220 | 0.00367027 | Q#9565 ->Scafflod_c13[Scafflod_c139_18 [466 - 744]] |
| Molybdopterín-Binding superfamily | superfamily | cl09928 | 3.50344e-13 | Q#12433 ->Scafflod_c19[Scafflod_c196_56 [1372 - 38] (REVERSE SENSE)] |
| Molybdopterín-Binding superfamily | superfamily | cl09928 | 9.21819e-11 | Q#12433 ->Scafflod_c19[Scafflod_c196_56 [1372 - 38] (REVERSE SENSE)] |
| Molybdopterín-Binding superfamily | superfamily | cl09928 | 5.96046e-67 | Q#19161 ->Scafflod_c32[Scafflod_c321_20 [575 - 1069]] |
| Molybdopterín-Binding superfamily | superfamily | cl09928 | 1.0475e-69 | Q#20913 ->Scafflod_c35[Scafflod_c356_18 [3 - 656]] |
| Molybdopterín-Binding superfamily | superfamily | cl09928 | 1.23977e-165 | Q#29962 ->Scafflod_c55[Scafflod_c551_32 [780 - 1] (REVERSE SENSE)] |
| Molybdopterín-Binding superfamily | superfamily | cl09928 | 5.35054e-147 | Q#33975 ->Scafflod_rep[Scafflod_rep_c623_42 [3 - 1949]] |
| Molybdopterín-Binding superfamily | superfamily | cl09928 | 9.9271e-35 | Q#33975 ->Scafflod_rep[Scafflod_rep_c623_42 [3 - 1949]] |
| Molybdopterín-Binding superfamily | superfamily | cl09928 | 1.53611e-85 | Q#36667 ->Scafflod_c67[Scafflod_c677_20 [293 - 895]] |
| Molybdopterín-Binding superfamily | superfamily | cl09928 | 6.40461e-81 | Q#9273 ->Scafflod_c13[Scafflod_c133_18 [2 - 946]] |
| MopB_CT superfamily | superfamily | cl09929 | 2.45302e-85 | Q#12433 ->Scafflod_c19[Scafflod_c196_56 [1372 - 38] (REVERSE SENSE)] |
| MopB_CT superfamily | superfamily | cl09929 | 5.92881e-67 | Q#21853 ->Scafflod_c37[Scafflod_c377_10 [2 - 337]] |
| MopB_CT superfamily | superfamily | cl09929 | 2.28152e-56 | Q#33975 ->Scafflod_rep[Scafflod_rep_c623_42 [3 - 1949]] |

| | | | | |
|------------------------------|-------------|---------|-------------|---|
| MopB_CT superfamily | superfamily | cl09929 | 8.67205e-59 | Q#9273 ->Scafflod_c13[Scafflod_c133_18 [2 - 946]] |
| Thioredoxin_like superfamily | superfamily | cl00388 | 5.11239e-19 | Q#30300 ->Scafflod_c55[Scafflod_c558_41 [108 - 1] (REVERSE SENSE)] |
| Thioredoxin_like superfamily | superfamily | cl00388 | 0.000239179 | Q#31266 ->Scafflod_c58[Scafflod_c582_30 [1358 - 1435]] |
| Thioredoxin_like superfamily | superfamily | cl00388 | 5.07641e-51 | Q#3211 ->Scafflod_c34[Scafflod_c34_56 [2094 - 1711] (REVERSE SENSE)] |
| Thioredoxin_like superfamily | superfamily | cl00388 | 7.2123e-31 | Q#3553 ->Scafflod_c39[Scafflod_c39_85 [581 - 3] (REVERSE SENSE)] |
| Thioredoxin_like superfamily | superfamily | cl00388 | 1.12946e-49 | Q#35646 ->Scafflod_c65[Scafflod_c651_39 [897 - 1] (REVERSE SENSE)] |
| Thioredoxin_like superfamily | superfamily | cl00388 | 8.56793e-39 | Q#35646 ->Scafflod_c65[Scafflod_c651_39 [897 - 1] (REVERSE SENSE)] |
| Thioredoxin_like superfamily | superfamily | cl00388 | 4.73956e-24 | Q#36296 ->Scafflod_rep |
| Thioredoxin_like superfamily | superfamily | cl00388 | 4.73956e-24 | Q#36296 ->Scafflod_rep[Scafflod_rep_c668_13 [475 - 618]] |
| Thioredoxin_like superfamily | superfamily | cl00388 | 5.15748e-61 | Q#3841 ->Scafflod_rep[Scafflod_rep_c42_115 [588 - 157] (REVERSE SENSE)] |

A nivel de dominio específico y superfamilia, existe la identificación del componente *arsB*, el cual codifica para la expulsión del ion arsenito mediante la membrana celular hacia el exterior, por lo que está catalogado como una permeasa. Este gen está acoplado al mecanismo de reducción del arsenato, por lo que generalmente está acompañado por el gen *arsC*.

En efecto, la búsqueda de dominios encontró registros del dominio conservado para el gen *arsC*, el cual es responsable de codificar la arsenato reductasa, mecanismo que se encarga de hacer la reducción del ión arsenato a arsenito. Así mismo, las arsenato reductasas hacen parte de la superfamilia de las thioredoxinas.

También, una parte estructural del complejo del operón *ars* fue identificada. Esta corresponde al gen *arsR* el cual funciona como regulador genético, que en presencia de compuestos de arsénico activa la expresión de los genes *downstream* como el *arsB* y *arsC*, entre otros (Mateos *et al.*, 2006) (Figura 3, mecanismo #2).

Finalmente, se identificaron molibdopterinas, las cuales están asociadas a los cofactores de molibdeno utilizados en enzimas oxidoreductasas como la arsenito oxidasa (Figura 3, mecanismo #4).

Se logró identificar mediante anotaciones, cinco funciones estrechamente relacionados con el metabolismo del arsénico: tres de ellos como parte constitutiva del operón *ars*, otra anotación que sugiere la presencia de la familia thioredoxina, de la cual el operón *ars* hace parte y por último una anotación relacionada con el metabolismo de las molibdopterinas, cofactores necesarios para el proceso de oxidación del ion arsenito a

arsenato. Las anotaciones funcionales, son solamente agrupaciones que por medio de similitud establecen categorías funcionales generales, las cuales deben ser validadas.

3.5 Búsqueda e identificación de genes de resistencia a arsénico

A partir de la búsqueda de genes de resistencia a arsénico reportados mediante el algoritmo *blastn*, se encontró la existencia de los genes *arsR*, *arsB* y *arsC*, los cuales son la base necesaria para el funcionamiento del operón *ars* fueron identificados al interior de las secuencias (Figura 24 y Anexo C1).

También, a partir de la implementación del algoritmo *tblastn*, se identificaron los genes de resistencia a arsénico *arsR*, *arsD*, *arsA*, *arsB* y *arsC* del operón *ars* tipo de *Klebsiella pneumoniae* (Figura 24 y Anexo C2).

En el caso de la subunidad del sitio *rieske*, no existió ningún alineamiento y para la subunidad pequeña de la arsenito oxidasa, se encontraron alineamientos con valores *E* desde 1,4 hasta 8, lo que sugiere que son alineamientos de carácter aleatorio y su hallazgo en la base de datos de dominios conservados se debe a la asociación aleatoria con los sitios de regulación metabólica de arsénico (Anexo C3).

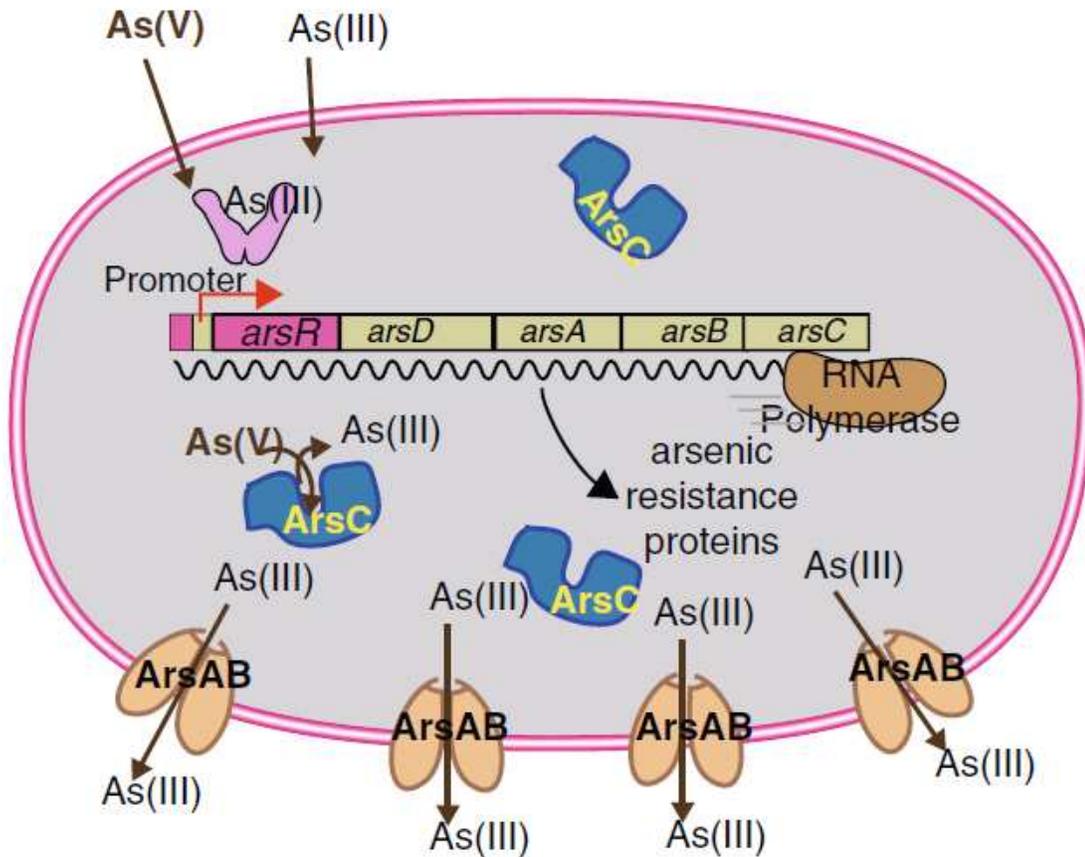


Figura 24. Operón bacteriano de resistencia a arsénico *ars*. Tomado de Diesel *et al.* (2009).

Los resultados reportados son consistentes con la identificación de genes de resistencia a arsénico en el metagenoma ambiental de muestras de la Sabana de Bogotá representado en los once clones que presentaron resistencia a arsénico.

Consistentes con las anotaciones funcionales, los alineamientos específicos validaron la presencia del operón *ars*. No solamente se encontraron los genes *arsB*, *arsC* y *arsR*, identificados inicialmente, sino que se encontraron además los genes complementarios *arsA* y *arsD* los cuales hacen más eficiente el proceso de reducción del ion arsenato a arsenito y su posterior expulsión hacia el exterior. Chauhan *et al.* (2009) identificaron la presencia de los genes *arsC*, *arsD* y *arsN*, en dos clones resistentes a arsénico en un metagenoma de una planta de tratamiento.

4. Conclusiones y Recomendaciones

4.1 Conclusiones

Los resultados presentados dan cuenta de la búsqueda e identificación de genes de resistencia a arsénico en el metagenoma microbiano de la Sabana de Bogotá y especialmente del Río Bogotá.

Los retos que suponía la realización de una librería metagenómica de un sitio contaminado como la cuenca media del Río Bogotá, fueron desarrollados principalmente en la selección y estandarización de un método de extracción y purificación de ácidos nucleicos, así como en el proceso de la selección a partir de un barrido funcional.

De esta manera, se logró comprobar la alta eficiencia del método de extracción a partir del kit comercial *PowerMax™ Soil DNA Isolation Kit (MO-BIO®)* que logró cantidades suficientes para los procesos de la creación de la librería metagenómica. Así mismo, las modificaciones realizadas sobre el protocolo establecido del kit comercial, lograron evitar la fragmentación de los ácidos nucleicos en virtud de garantizar fragmentos de alto peso molecular.

La realización de la electroforesis como paso posterior a la extracción de los ácidos nucleicos logró la separación de agentes inhibidores como ácidos húmicos y fúlvicos, así como otros contaminantes provenientes de los ambientes particulares, toda vez que la gran mayoría de estos residuos fueron eliminados por medio de la selección de los tamaños específicos de alto peso molecular.

La realización de la librería metagenómica logró la generación de más de 50.000 clones con fragmentos metagenómicos de un ambiente contaminado de la Sabana de Bogotá.

La cobertura teórica lograda fue de aproximadamente 2.000 Mb. No obstante, se requiere de una mayor caracterización molecular con el fin de determinar un valor empírico más ajustado al resultado final de la librería.

La realización del tamizaje funcional con altos contenidos de arsenato de sodio, permitió discriminar aquellos clones que presentaron resistencia fenotípica al ión arsenato. La técnica resultó adecuada para el manejo de grandes cantidades de clones de la librería metagenómica. A partir de ello, fueron recuperados once clones con resistencia fenotípica a arsénico, nueve de ellos a un umbral de 20 mM de arsenato de sodio y dos restantes a una concentración de 50 mM.

El proceso de secuenciación masiva logró generar más 157.583 secuencias. Luego de la realización de un proceso de ensamblaje se lograron 951 *scaffolds* con un tamaño promedio de 1.075 pares de bases. Las secuencias ensambladas resultaron adecuadas para los procesos de anotación funcional e identificación de genes de resistencia a arsénico.

Las condiciones de selección que ofrece el ambiente de la cuenca media del Río Bogotá, como baja concentración de oxígeno, acumulación en sedimentos y suelos de contaminantes y alta e intensiva utilización de plaguicidas, propician el desarrollo y la supervivencia de microorganismos que han generado mecanismos de resistencia a tales condiciones, destacándose varias enterobacterias, las cuales tienen amplios reportes en la literatura de mecanismos de resistencia a arsénico.

Finalmente se lograron identificar genes reportados de resistencia a arsénico en un metagenoma microbiano de la Sabana de Bogotá. Los genes encontrados corresponden al *arsB*, *arsC* y *arsR*, *arsA* y *arsD* todos ellos como parte constitutiva del operón *ars*.

4.2 Recomendaciones

La librería metagenómica generada es una fuente de datos genómicos que deberá ser aprovechada a futuro. Por ello, es necesario el desarrollo de trabajos investigativos que se centren en el análisis de datos masivos los cuales pueden provenir a partir de una secuenciación masiva de la totalidad de la librería establecida. La identificación de genes,

funciones y poblaciones microbianas en relación a ambientes altamente contaminados debe dar luces sobre un nuevo enfoque investigativo que se puede desarrollar.

El sustrato para el tamizaje funcional fue el ion arsenato, por lo cual, la actividad encontrada de reducción de este ion, estuvo en relación al tipo de barrido funcional realizado. Por otro lado, no se logró la identificación de genes que median relaciones de oxidación del ion arsenito a arsenato. La amplia diversidad genómica de esta función biológica hace que nuevos enfoques para la identificación específica de esta actividad, deban ser utilizados. El sustrato para el tamizaje funcional fue el ion arsenato, por lo cual, la actividad encontrada de reducción de este ion, estuvo en relación al tipo de barrido funcional realizado.

Así mismo, es necesaria la identificación de cada mecanismo de resistencia a arsénico existente en cada uno de los clones positivos con el fin de identificar nuevas rutas y genes potenciales de resistencia a arsénico. Para ello es necesaria la adaptación de secuencias de marcaje (*tags*) que permitan el vínculo entre las secuencias generadas y los clones de los cuales provienen.

Los genes de resistencia a arsénico identificados hasta el momento, permitirán la continuación de la investigación en miras al desarrollo de aplicaciones biotecnológicas enfocadas hacia procesos de biodetección y biorremediación de ambientes contaminados de arsénico.

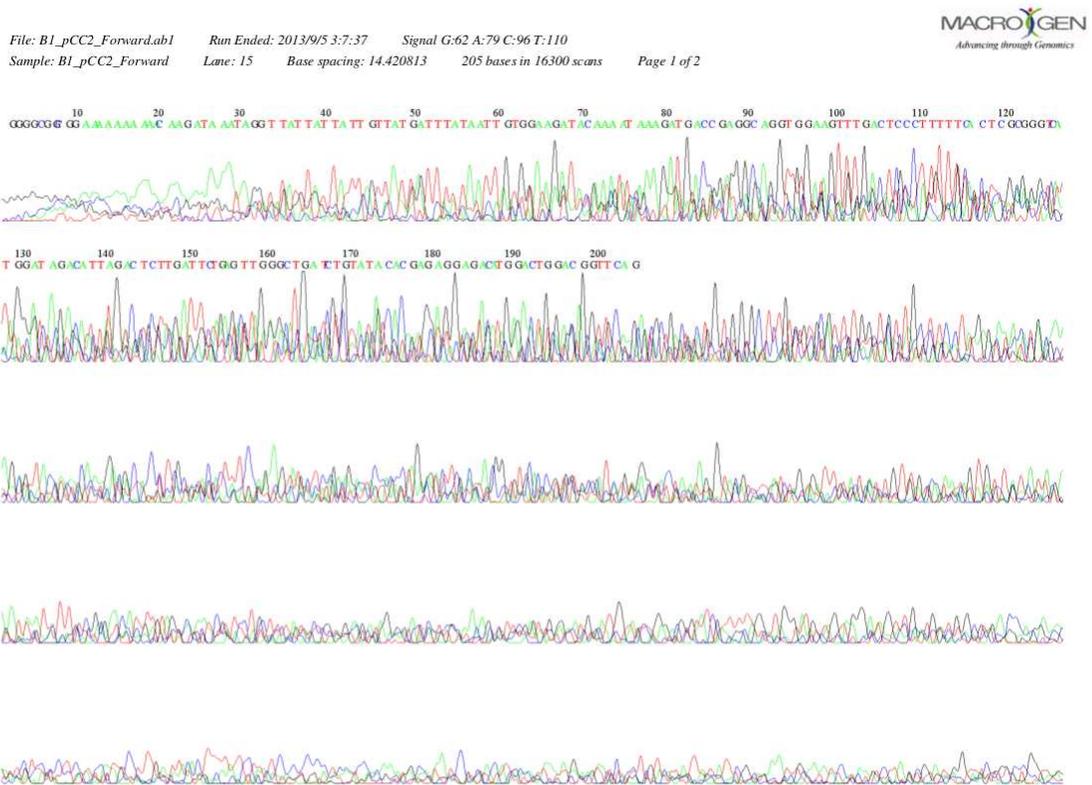
El desarrollo de biosensores para la detección de arsénico es una posibilidad de aplicación biotecnológica. En países en los que hay carencia de de monitoreo sobre contaminación potencial de arsénico como en el caso de Colombia (Alonso *et al.*, 2013), el desarrollo de este tipo de tecnologías podría ser un aporte valioso para un diseño de monitoreo del metaloide a bajo costo de operación.

La utilización del mecanismo de regulación del operón *ars* brinda las posibilidades, para que por medio de la adaptación con un gen reportero, se puedan realizar detecciones de manera rápida y a bajo costo a partir del diseño de un biosensor (French *et al.*, 2011). Este tipo de estudios están siendo desarrollados actualmente en el grupo de investigación.

A. Anexo: Protocolo modificado para extracción de ácidos nucleicos de alto peso molecular con *PowerMax*TM *Soil DNA Isolation Kit* (MOBIO®)

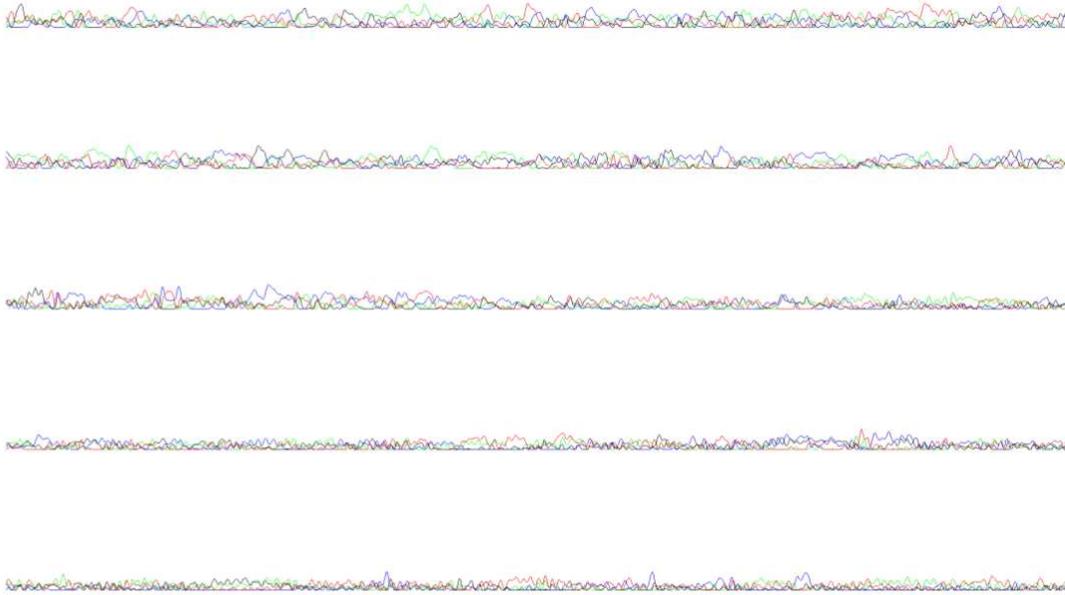
1. Precalentar un baño termoestático a 70°C.
2. Agregar 15 ml de *PowerBead Solution* al tubo rotulado como *PowerBead Tube*, el cual contiene las partículas o *beads*.
3. Agregar 10 g de muestra (suelo o sedimento) al *PowerBead Tube*.
4. Agregar 1.2 ml de la Solución C1 al *PowerBead Tube*.
5. Agitar con *vortex* durante 5 segundos.
6. Calentar el tubo en el baño termoestático a 70°C, durante 5 minutos.
7. Retirar el tubo del baño y mezclar el tubo por inversión suave, durante 5 minutos.
8. Repetir los pasos 6 y 7 tres veces.
9. Centrifugar el tubo a 2500 x *g* y 4°C durante 6 minutos.
10. Transferir el sobrenadante a un nuevo tubo de 50ml. Descartar el *pellet*.
11. Agregar 5ml de la Solución C2 y mezclar por inversión.
12. Incubar a 4°C durante 1 hora.
13. Centrifugar a 2500 x *g* y 4°C durante 6 minutos.
14. Transferir el sobrenadante a un nuevo tubo de 50ml. Descartar el *pellet*.
15. Agregar 4ml de la Solución C3 y mezclar por inversión.
16. Incubar a 4°C durante 1 hora.
17. Centrifugar a 2500 x *g* y 4°C durante 6 minutos.
18. Transferir el sobrenadante a un nuevo tubo de 50ml. Descartar el *pellet*.
19. Agregar 30ml de la solución C4 y mezclar por inversión.
20. Llenar el filtro del tubo *Spin Filter* con la solución. Centrifugar a 2500 x *g* durante 2 minutos y descartar la solución. Hacer este mismo proceso hasta pasar todo el contenido a través del filtro.
21. Agregar 10ml de la Solución C5 al filtro y centrifugar a 2500 x *g* durante 3 minutos. Descartar la solución.
22. Centrifugar nuevamente a 2500 x *g*, durante 5 minutos.
23. Transferir el filtro a un nuevo tubo de 50ml y agregar 5 ml de la Solución C6 sobre el filtro.
24. Descartar el filtro.

B. Anexo: Electroferograma ejemplo de secuencias Sanger



78 Búsqueda de genes de resistencia a arsénico en el metagenoma microbiano de la Sabana de Bogotá

File: B1_pCC2_Forward.ab1 Run Ended: 2013/05 3:7:37 Signal G:62 A:79 C:96 T:110
Sample: B1_pCC2_Forward Lane: 15 Base spacing: 14.420813 205 bases in 16300 scans Page 2 of 2



C. Anexo: Alineamientos *BLAST* con elementos relacionados a la resistencia de arsénico

1. Blastn para operón *ars* (genes *arsB*, *arsC*, *arsR*)

```
BLASTN 2.2.28+
Reference: Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb
Miller (2000), "A greedy algorithm for aligning DNA sequences", J
Comput Biol 2000; 7(1-2):203-14.

Database: scaffolds.fasta
951 sequences; 1,022,791 total letters

Query= lc1|AP009048.1_cdsid_BAE77791.1 [gene=arsC] [protein=arsenate
reductase] [protein_id=BAE77791.1]
[location=complement(3989753..3990178)]
Length=426
Score      E

Sequences producing significant alignments:                      (Bits) Value

TodosTodos_rep_c42                                           787    0.0
> TodosTodos_rep_c42
Length=2517
Score = 787 bits (426), Expect = 0.0
Identities = 426/426 (100%), Gaps = 0/426 (0%)
Strand=Plus/Minus

Query 1   ATGAGCAACATTACCATTTATCACAACCCGGCCTGCGGCACGTCGCGTAATACGCTGGAG 60
|||||
Sbjct 579 ATGAGCAACATTACCATTTATCACAACCCGGCCTGCGGCACGTCGCGTAATACGCTGGAG 520

Query 61  ATGATCCGCAACAGCGGCACAGAACCGACTATTATCCATTATCTGGAAACTCCGCCAACG 120
|||||
Sbjct 519 ATGATCCGCAACAGCGGCACAGAACCGACTATTATCCATTATCTGGAAACTCCGCCAACG 460

Query 121 CGCGATGAACTGGTCAAACCTATTGCCGATATGGGGATTTCCGTACGCGCGCTGCTGCGT 180
|||||
Sbjct 459 CGCGATGAACTGGTCAAACCTATTGCCGATATGGGGATTTCCGTACGCGCGCTGCTGCGT 400

Query 181 AAAAACGTGCAACCGTATGAGGAGCTGGGCCTTGCGGAAGATAAATTTACTGACGATCGG 240
|||||
```

80 Búsqueda de genes de resistencia a arsénico en el metagenoma microbiano de la Sabana de Bogotá

Sbjct 399 AAAAACGTCGAACCGTATGAGGAGCTGGGCCTTGCGGAAGATAAATTTACTGACGATCGG 340

Query 241 TTAATCGACTTTATGCTTCAGCACCCGATTCTGATTAATCGCCCGATTGTGGTGACGCCG 300

|||||

Sbjct 339 TTAATCGACTTTATGCTTCAGCACCCGATTCTGATTAATCGCCCGATTGTGGTGACGCCG 280

Query 301 CTGGGAACTCGCCTGTGCCGCCCTTCAGAAGTGGTGTGAAATTTCTGCCAGATGCGCAA 360

|||||

Sbjct 279 CTGGGAACTCGCCTGTGCCGCCCTTCAGAAGTGGTGTGAAATTTCTGCCAGATGCGCAA 220

Query 361 AAAGCGCATTCTCCAAGGAAGATGGCGAGAAAGTGGTTGATGAAGCGGTAAGCGCCTG 420

|||||

Sbjct 219 AAAGCGCATTCTCCAAGGAAGATGGCGAGAAAGTGGTTGATGAAGCGGTAAGCGCCTG 160

Query 421 AAATAA 426

|||||

Sbjct 159 AAATAA 154

Lambda K H

1.33 0.621 1.12

Gapped

Lambda K H

1.28 0.460 0.850

Effective search space used: 407531026

Query= lcl|AP009048.1_cdsid_BAE77792.1 [gene=arsB]
 [protein=arsenite/antimonite transporter] [protein_id=BAE77792.1]
 [location=complement(3990191..3991480)]

Length=1290
 Score E

Sequences producing significant alignments: (Bits) Value
 TodosTodos_rep_c42 2383 0.0

> TodosTodos_rep_c42

Length=2517

Score = 2383 bits (1290), Expect = 0.0

Identities = 1290/1290 (100%), Gaps = 0/1290 (0%)

Strand=Plus/Minus

Query 1 ATGTTACTGGCAGGCGCTATCTTTGTCTGACCATCGTATTGGTTATCTGGCAGCCGAAA 60

|||||

Sbjct 1881 ATGTTACTGGCAGGCGCTATCTTTGTCTGACCATCGTATTGGTTATCTGGCAGCCGAAA 1822

Query 61 GGTTTAGGCATCGGCTGGAGTGCAACGCTCGGCCGAGTACTGGCGTTAGTTACGGGCGTG 120

|||||

Sbjct 1821 GGTTTAGGCATCGGCTGGAGTGCAACGCTCGGCCGAGTACTGGCGTTAGTTACGGGCGTG 1762

Query 121 GTCCATCCGGGTGATATTCGGTGGTGGAAATATCGTCTGGAACGCGACGGCTGCGTTT 180

|||||

Sbjct 1761 GTCCATCCGGGTGATATTCGGTGGTGGAAATATCGTCTGGAACGCGACGGCTGCGTTT 1702

Query 181 ATCGCCGTCATTATCATCAGCCTGCTGCTGGATGAGTCCGGCTTTTTTGAATGGGCGCG 240

|||||

```
Sbjct 1701 ATGCCGTCATTATCATCAGCCTGCTGCTGGATGAGTCCGGCTTTTTGGAATGGGCGCG 1642

Query 241 CTGCACGCTCTCACGCTGGGGTAATGGTCGTGGTCGCTTGCTGTTTACCTGGATTGTCCTG 300
|||||
Sbjct 1641 CTGCACGCTCTCACGCTGGGGTAATGGTCGTGGTCGCTTGCTGTTTACCTGGATTGTCCTG 1582

Query 301 CTCGGTGTGCCGTTGCCGCCCTGTTTGCCAATGATGGCGCGCGCTTATTTTGACACCG 360
|||||
Sbjct 1581 CTCGGTGTGCCGTTGCCGCCCTGTTTGCCAATGATGGCGCGCGCTTATTTTGACACCG 1522

Query 361 ATTGTCATCGCCATGCTGCTGGCTTTAGGGTTCAGTAAAGGCACTACGCTGGCGTTCGTG 420
|||||
Sbjct 1521 ATTGTCATCGCCATGCTGCTGGCTTTAGGGTTCAGTAAAGGCACTACGCTGGCGTTCGTG 1462

Query 421 ATGGCGCCCGGATTTCATTGCCGATACCGCCAGCCTGCCGCTTATTGTCTCCAACCTGGTG 480
|||||
Sbjct 1461 ATGGCGCCCGGATTTCATTGCCGATACCGCCAGCCTGCCGCTTATTGTCTCCAACCTGGTG 1402

Query 481 AATATCGTTTCCGCTGATTTCTTTGGCCTCGGCTTTCGCGAATACGCCCTCGGTGATGGTG 540
|||||
Sbjct 1401 AATATCGTTTCCGCTGATTTCTTTGGCCTCGGCTTTCGCGAATACGCCCTCGGTGATGGTG 1342

Query 541 CCGTGGATATCGCCCGATTGTTGCCACGCTGGTATGTTACATCTCTATTTTCGCAA 600
|||||
Sbjct 1341 CCGTGGATATCGCCCGATTGTTGCCACGCTGGTATGTTACATCTCTATTTTCGCAA 1282

Query 601 GATATCCGCGAACTACGATATGGCGCTGCTGAAATCTCCCGCAGAAGCGATCAAAGAT 660
|||||
Sbjct 1281 GATATCCGCGAACTACGATATGGCGCTGCTGAAATCTCCCGCAGAAGCGATCAAAGAT 1222

Query 661 CCTGCTACGTTCAAACCTGGCTGGGTTGTTTACTGCTTCTGCTGGGATTTTTCGTC 720
|||||
Sbjct 1221 CCTGCTACGTTCAAACCTGGCTGGGTTGTTTACTGCTTCTGCTGGGATTTTTCGTC 1162

Query 721 CTGGAACCGCTCGGCATTCGGTGAGCGCCATTGCAGCTGTGGCGCGCTGATATTATTT 780
|||||
Sbjct 1161 CTGGAACCGCTCGGCATTCGGTGAGCGCCATTGCAGCTGTGGCGCGCTGATATTATTT 1102

Query 781 GTCGTCGCTAAACGCGGTCATGCGATTAATACGGGTAAGTCTGCGCGGTGCCCCCTGG 840
|||||
Sbjct 1101 GTCGTCGCTAAACGCGGTCATGCGATTAATACGGGTAAGTCTGCGCGGTGCCCCCTGG 1042

Query 841 CAGATTGTCATCTTCTCGCTCGGCATGTATCTGGTGGTTTATGGCCTGCGCAATGCCGGA 900
|||||
Sbjct 1041 CAGATTGTCATCTTCTCGCTCGGCATGTATCTGGTGGTTTATGGCCTGCGCAATGCCGGA 982

Query 901 TTAACGGAATATCTTTCTGGCGTACTCAACGTGCTGGCGGATAACGGCCTGTGGGCGCG 960
|||||
Sbjct 981 TTAACGGAATATCTTTCTGGCGTACTCAACGTGCTGGCGGATAACGGCCTGTGGGCGCG 922

Query 961 ACGCTCGGCACCGGATTCCTCACCGCTTCTCTTCTATTATGAACAATATGCCGACG 1020
|||||
Sbjct 921 ACGCTCGGCACCGGATTCCTCACCGCTTCTCTTCTATTATGAACAATATGCCGACG 862
```

82 Búsqueda de genes de resistencia a arsénico en el metagenoma microbiano de la Sabana de Bogotá

```

Query 1021 GTACTGGTTGGCGCGTTGTCATTGATGGCAGCACGGCATCTGGCGTTATCAAAGAAGCG 1080
|||||
Sbjct 861 GTACTGGTTGGCGCGTTGTCATTGATGGCAGCACGGCATCTGGCGTTATCAAAGAAGCG 802

Query 1081 ATGGTTTATGCCAATGTGATTGGCTGCGATTTGGGACCGAAAATTACCCCAATGGTAGC 1140
|||||
Sbjct 801 ATGGTTTATGCCAATGTGATTGGCTGCGATTTGGGACCGAAAATTACCCCAATGGTAGC 742

Query 1141 CTGGCTACGCTACTCTGGCTGCACGTACTTTCGCAGAAGAATATGACTATCAGCTGGGGA 1200
|||||
Sbjct 741 CTGGCTACGCTACTCTGGCTGCACGTACTTTCGCAGAAGAATATGACTATCAGCTGGGGA 682

Query 1201 TATTACTTCCGTACAGGGATTATCATGACCCTGCCTGTGCTGTTTGTGACGCTGGCTGCG 1260
|||||
Sbjct 681 TATTACTTCCGTACAGGGATTATCATGACCCTGCCTGTGCTGTTTGTGACGCTGGCTGCG 622

Query 1261 CTGGCGCTACGTCTCTCTTTCACCTTTGTAA 1290
|||||
Sbjct 621 CTGGCGCTACGTCTCTCTTTCACCTTTGTAA 592

```

Lambda K H
1.33 0.621 1.12

Gapped

Lambda K H
1.28 0.460 0.850

Effective search space used: 1272578580

Query= lcl|AP009048.1_cdsid_BAE77793.1 [gene=arsR] [protein=DNA-binding
transcriptional regulator] [protein_id=BAE77793.1]
[location=complement(3991534..3991887)]

Length=354
Score E

Sequences producing significant alignments: (Bits) Value

TodosTodos_rep_c42 654 0.0

> TodosTodos_rep_c42

Length=2517

Score = 654 bits (354), Expect = 0.0

Identities = 354/354 (100%), Gaps = 0/354 (0%)

Strand=Plus/Minus

```

Query 1 ATGTCATTTCTGTTACCCATCCAATTGTTCAAATCTTGCTGATGAAACCCGCTGGGC 60
|||||
Sbjct 2288 ATGTCATTTCTGTTACCCATCCAATTGTTCAAATCTTGCTGATGAAACCCGCTGGGC 2229

Query 61 ATCGTTTTACTGCTCAGCGAACTGGGAGAGTTATGCGTCTGCGATCTCTGCACTGCTCTC 120
|||||
Sbjct 2228 ATCGTTTTACTGCTCAGCGAACTGGGAGAGTTATGCGTCTGCGATCTCTGCACTGCTCTC 2169

Query 121 GACCAGTCGACGCCAAGATCTCCCGCCACCTGGCATTGCTGCGTGAAAGCGGGCTATTG 180
|||||
Sbjct 2168 GACCAGTCGACGCCAAGATCTCCCGCCACCTGGCATTGCTGCGTGAAAGCGGGCTATTG 2109

```

```

Query 181 CTGGACCGCAAGCAAGGTAAGTGGGTTTCATTACCGCTTATCACCGCATATCCAGCATGG 240
|||||
Sbjct 2108 CTGGACCGCAAGCAAGGTAAGTGGGTTTCATTACCGCTTATCACCGCATATCCAGCATGG 2049

Query 241 GCGGCGAAAATTATTGATGAGGCCTGGCGATGTGAACAGGAAAAGGTTTCAGGCGATTGTC 300
|||||
Sbjct 2048 GCGGCGAAAATTATTGATGAGGCCTGGCGATGTGAACAGGAAAAGGTTTCAGGCGATTGTC 1989

Query 301 CGCAACCTGGCTCGACAAAAGTTCGCGGGACAGTAAGAACATTTGCAGTTAA 354
|||||
Sbjct 1988 CGCAACCTGGCTCGACAAAAGTTCGCGGGACAGTAAGAACATTTGCAGTTAA 1935

Lambda      K      H
1.33      0.621  1.12

Gapped

Lambda      K      H
1.28      0.460  0.850

Effective search space used: 335259514

Database: contigs.fasta
Posted date: Sep 29, 2013 11:16 AM
Number of letters in database: 1,022,791
Number of sequences in database: 951

Matrix: blastn matrix 1 -2
Gap Penalties: Existence: 0, Extension: 2.5

```

2. *tblastn* para operón *arsRDABC* de *K. Pneumoniae*

TBLASTN 2.2.27+

Reference: Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402.

Database: contigs.fasta

951 sequences; 1,022,791 total letters

Query= gi|149850118|emb|CAG29009.1| ArsR [Klebsiella pneumoniae]

Length=117

Score E

| Sequences producing significant alignments: | (Bits) | Value |
|---|--------|-------|
| TodosTodos_rep_c42 | 182 | 5e-56 |
| TodosTodos_c185 | 26.2 | 0.23 |
| TodosTodos_c475 | 24.6 | 0.65 |
| TodosTodos_c477 | 23.9 | 1.1 |
| TodosTodos_c910 | 23.5 | 1.6 |

84 Búsqueda de genes de resistencia a arsénico en el metagenoma microbiano de la Sabana de Bogotá

TodosTodos_c541

21.6 6.1

> TodosTodos_rep_c42

Length=2517

Score = 182 bits (463), Expect = 5e-56, Method: Compositional matrix adjust.

Identities = 86/116 (74%), Positives = 99/116 (85%), Gaps = 0/116 (0%)

Frame = -2

```
Query 1  MPEIASLQLFKILSDETRLGIVLLREMGELCVCDLCTALEQSQPKTSRHLAMLRESGLL 60
          M + +QLFKIL+DETRLGIVLLL E+GELCVCDLTAL+QSQPK SRHLA+LRESGLL
Sbjct 2288 MSFLLPIQLFKILADETRLGIVLLSELGELCVCDLTALDQSQPKISRHLALLRESGLL 2109
```

```
Query 61  LDRKQKGWVHYRLSPHIPSWAALVIEQAWLSQQDDVQAIARKLASANCSGSGKAVC 116
          LDRKQKGWVHYRLSPHIP+WAA +I++AW +Q+ VQAI R LA NCSG K +C
Sbjct 2108 LDRKQKGWVHYRLSPHIPAWAAKIIDEAWRCEQEKVQAIVRNLRQNCSGDSKNIC 1941
```

> TodosTodos_c185

Length=2566

Score = 26.2 bits (56), Expect = 0.23, Method: Compositional matrix adjust.

Identities = 9/23 (39%), Positives = 15/23 (65%), Gaps = 0/23 (0%)

Frame = -1

```
Query 7  LQLFKILSDETRLGIVLLREMG 29
          LQL K+L E +G++ + +MG
Sbjct 1975 LQLIKVLQKEMSMGVIFITHDMG 1907
```

> TodosTodos_c475

Length=696

Score = 24.6 bits (52), Expect = 0.65, Method: Compositional matrix adjust.

Identities = 10/27 (37%), Positives = 16/27 (59%), Gaps = 0/27 (0%)

Frame = +1

```
Query 28  MGELCVCDLCTALEQSQPKTSRHLAML 54
          + ++ CDL +EQ P+T+ LA L
Sbjct 55  VAKISECDLTATIEQRGPQTADL 135
```

> TodosTodos_c477

Length=710

Score = 23.9 bits (50), Expect = 1.1, Method: Compositional matrix adjust.

Identities = 9/21 (43%), Positives = 13/21 (62%), Gaps = 0/21 (0%)

Frame = +3

```
Query 54  LRESGLLLDRKQKGWVHYRLS 74
          R S LL+D + WV YR++
Sbjct 120 FRGSRLMDFRNSSWVRYRIA 182
```

> TodosTodos_c910

Length=1391

Score = 23.5 bits (49), Expect = 1.6, Method: Compositional matrix adjust.

Identities = 13/40 (33%), Positives = 23/40 (58%), Gaps = 8/40 (20%)

```

Frame = -2
Query 37 CTALEQSQPKTSRHLMLRESGLLLDRKQGK-----WVH 70
          C+A+ +QP+TS + +L + + DR+Q K WV+
Sbjct 142 CSAV--AQPETSSQIRLLSAASIGDDRRQSKNCFAS*WVN 29

> TodosTodos_c541
Length=596
Score = 21.6 bits (44), Expect = 6.1, Method: Compositional matrix adjust.
Identities = 11/47 (23%), Positives = 20/47 (43%), Gaps = 1/47 (2%)
Frame = +3
Query 45 PKTSRHLMLRESGLLLDRKQGKWWHYRLSPHIPSWAALVIEQAWLS 91
          P +RH R +K G+ +H+++ P W A + W +
Sbjct 381 PALARHRKSPRPD-HFPHQKDGRRIFHQIRPAPRGWCAPGVGTWVRT 518

Lambda K H a alpha
0.322 0.133 0.407 0.792 4.96

Gapped
Lambda K H a alpha sigma
0.267 0.0410 0.140 1.90 42.6 43.6
Effective search space used: 14843498

Query= gi|149850119|emb|CAG29010.1| ArsD [Klebsiella pneumoniae]

Length=120
Score E

Sequences producing significant alignments: (Bits) Value
TodosTodos_rep_c713 24.6 0.53
TodosTodos_c665 24.3 1.00
TodosTodos_c746 23.5 1.8
TodosTodos_c251 22.3 4.3
TodosTodos_c58 22.3 4.3
TodosTodos_c403 22.3 4.6
TodosTodos_c224 21.2 9.4

> TodosTodos_rep_c713
Length=581
Score = 24.6 bits (52), Expect = 0.53, Method: Compositional matrix adjust.
Identities = 10/20 (50%), Positives = 13/20 (65%), Gaps = 0/20 (0%)
Frame = +3
Query 25 VLVNFSADVQWLKGRGVQIE 44
          ++ + A QWL GRGV IE
Sbjct 45 IIASLEAGYQWLPGRGVVIE 104

> TodosTodos_c665
Length=1223
Score = 24.3 bits (51), Expect = 1.00, Method: Compositional matrix adjust.
Identities = 11/21 (52%), Positives = 14/21 (67%), Gaps = 0/21 (0%)
Frame = -3
Query 61 AKAFLEASGAEGLPLLLLDGE 81
          A AF E S + LP LL++DG

```

86 Búsqueda de genes de resistencia a arsénico en el metagenoma microbiano de la Sabana de Bogotá

```
Sbjct 1215 AMAFAEGSANDWLP LLMDGH 1153

> TodosTodos_c746
Length=951

Score = 23.5 bits (49), Expect = 1.8, Method: Compositional matrix adjust.
Identities = 10/20 (50%), Positives = 11/20 (55%), Gaps = 0/20 (0%)
Frame = +3
Query 83 VMAGRYPKRAELARWFGIPL 102
          V+ G YP L RWF PL
Sbjct 888 VLLGAYPMVLLLRWF EKPL 947

> TodosTodos_c251
Length=1171

Score = 22.3 bits (46), Expect = 4.3, Method: Compositional matrix adjust.
Identities = 10/23 (43%), Positives = 12/23 (52%), Gaps = 0/23 (0%)
Frame = +1
Query 83 VMAGRYPKRAELARWFGI PLEKV 105
          MAG+YP R +LA F L
Sbjct 838 AMAGQYPHRTKLAI SFTRALSST 906

> TodosTodos_c58
Length=2123

Score = 22.3 bits (46), Expect = 4.3, Method: Compositional matrix adjust.
Identities = 10/23 (43%), Positives = 13/23 (57%), Gaps = 0/23 (0%)
Frame = -1
Query 74 PLLLLDGETVMAGRYPKRAELAR 96
          P DG+T MAG P+R A+
Sbjct 2060 PAFSPDGDTCMAGSAPERRRTAK 1992

> TodosTodos_c403
Length=1586

Score = 22.3 bits (46), Expect = 4.6, Method: Composition-based stats.
Identities = 10/18 (56%), Positives = 13/18 (72%), Gaps = 1/18 (6%)
Frame = +1
Query 4 LTVFDP-AMCCSTGVCGS 20
          L+ F P AM CST +CG+
Sbjct 469 LSCFRPKAMFCSTVICGN 522

> TodosTodos_c224
Length=1823

Score = 21.2 bits (43), Expect = 9.4, Method: Compositional matrix adjust.
Identities = 14/37 (38%), Positives = 19/37 (51%), Gaps = 0/37 (0%)
Frame = -3
Query 60 KAKAFLEASGAEGLP LLLLLDGETVMAGRYPKRAELAR 96
          K+ A L S A L +L + R+P+RAE AR
Sbjct 1455 KSPANLSYSEAA MLAVLPQAPSRLRPDRWPERAEAR 1345
```

Lambda K H a alpha
 0.321 0.137 0.415 0.792 4.96

Gapped

Lambda K H a alpha sigma
 0.267 0.0410 0.140 1.90 42.6 43.6

Effective search space used: 15351325

Query= gi|149850120|emb|CAG29011.1| ArsA [Klebsiella pneumoniae]

Length=583

Score E

| Sequences producing significant alignments: | (Bits) | Value |
|---|--------|-------|
| TodosTodos_c693 | 32.0 | 0.029 |
| TodosTodos_c460 | 28.1 | 0.42 |
| TodosTodos_c244 | 28.1 | 0.42 |
| TodosTodos_c221 | 27.7 | 0.65 |
| TodosTodos_c522 | 26.9 | 1.0 |
| TodosTodos_c271 | 26.6 | 1.1 |
| TodosTodos_c75 | 26.6 | 1.5 |
| TodosTodos_rep_c5 | 25.8 | 2.6 |
| TodosTodos_rep_c949 | 25.8 | 2.8 |
| TodosTodos_rep_c8 | 25.8 | 2.9 |
| TodosTodos_rep_c950 | 25.8 | 2.9 |

> TodosTodos_c693

Length=1084

Score = 32.0 bits (71), Expect = 0.029, Method: Compositional matrix adjust.

Identities = 15/41 (37%), Positives = 23/41 (56%), Gaps = 0/41 (0%)

Frame = +2

Query 71 LSALEIDPQDAAQQYRARIIVDPKGLLPDDVVNSISEQLSG 111

 ++A E+ PQ+A + YR R V P + DDV+ + Q G

Sbjct 650 MNAFELTPQEAEWYGRDVPYQAAPVADDVLVTFQHQPIG 772

> TodosTodos_c460

Length=977

Score = 28.1 bits (61), Expect = 0.42, Method: Compositional matrix adjust.

Identities = 13/36 (36%), Positives = 19/36 (53%), Gaps = 0/36 (0%)

Frame = +3

Query 329 LILMGKGGVGKTTMAAAIAVSLADKGFVHLLTSD 364

 +I+++G GVGKTT +A +G V L D

Sbjct 801 VILMVGVNGVGKTTTIGKLARQFEQQKSVMLAAGD 908

> TodosTodos_c244

Length=853

Score = 28.1 bits (61), Expect = 0.42, Method: Compositional matrix adjust.

Identities = 21/61 (34%), Positives = 31/61 (51%), Gaps = 11/61 (18%)

Frame = +3

88 Búsqueda de genes de resistencia a arsénico en el metagenoma microbiano de la Sabana de Bogotá

Query 330 IMLMGKGGVGGKTTMAAAIAVSLADKGFVHL-TTSDPAHLSTTLNGSLKNLQVSRINPH 388
+++ G G+GKTT+A +A + V+L TTS P + L L NL+ PH
Sbjct 27 LLIFGPPGLGKTTLANIVANEMG----VNLRTTSGPVLEKAGDLAAMLNLE----PH 176

Query 389 D 389
D
Sbjct 177 D 179

> TodosTodos_c221
Length=1247

Score = 27.7 bits (60), Expect = 0.65, Method: Compositional matrix adjust.
Identities = 20/69 (29%), Positives = 32/69 (46%), Gaps = 3/69 (4%)
Frame = +2

Query 328 GLIMLMGKGGVGGKTTMAAAIAVSLADKGFVHLTTSDPAHLSTTLNGSLKNLQVSRINP 387
G + L+ +GG+ + A A +G +L AAH + L +L+ +VSRI
Sbjct 929 GALWLVDGGLSQQLSQAVWAGFRFRPRG---NLVAQSLAAHGNNPLAATLRGRRVRSIAV 1099

Query 388 HDETERYRQ 396
H +R R
Sbjct 1100 HPARQRERH 1126

> TodosTodos_c522
Length=749

Score = 26.9 bits (58), Expect = 1.0, Method: Compositional matrix adjust.
Identities = 14/43 (33%), Positives = 25/43 (58%), Gaps = 3/43 (7%)
Frame = -1

Query 75 EIDPQDAAQQYRARIQVDPKGLLPDDVVNSISEQLSGACTTEI 117
+++ Q AA ++RAR+VD +K P V + S+ ++ EI
Sbjct 488 DVEYQAAASEWRARVVDALKSRAP---VAAPSQSVATGAVNEI 369

> TodosTodos_c271
Length=645

Score = 26.6 bits (57), Expect = 1.1, Method: Compositional matrix adjust.
Identities = 11/41 (27%), Positives = 23/41 (56%), Gaps = 0/41 (0%)
Frame = -2

Query 430 QAFSRVIREAGKRFFVMDTAPTGHLLLLLDATGAYHREIAR 470
+ ++ +RE+ + ++ T P + LLL++ H +IAR
Sbjct 353 ELMAKTLRESAEPVIVSTGPQTNVALLLNHPHLSKIAR 231

> TodosTodos_c75
Length=757

Score = 26.6 bits (57), Expect = 1.5, Method: Compositional matrix adjust.
Identities = 16/54 (30%), Positives = 25/54 (46%), Gaps = 0/54 (0%)
Frame = -3

Query 515 ERAGIHPWGIINNSLSIAQTQSPLLCQRALQERPQIEVVKHQHASRIALVPVM 568
+RA + W N A + LCQR + + +IE V A IA++ V+
Sbjct 401 KRAANFVYSW*CGNEFVGATLPAAFQCQREMLTQSEIEFVIGF*AITIAIILVL 240

```

> TodosTodos_rep_c5
Length=1560

Score = 25.8 bits (55), Expect = 2.6, Method: Compositional matrix adjust.
Identities = 10/18 (56%), Positives = 13/18 (72%), Gaps = 0/18 (0%)
Frame = +1

Query 151  RLLQLPGAWSSFIENPD 168
          RLL +P +WS  +SNPD
Sbjct 1306 RLLAIPTSWSRVADSNPD 1359

> TodosTodos_rep_c949
Length=2794

Score = 25.8 bits (55), Expect = 2.8, Method: Compositional matrix adjust.
Identities = 10/18 (56%), Positives = 13/18 (72%), Gaps = 0/18 (0%)
Frame = -1

Query 151  RLLQLPGAWSSFIENPD 168
          RLL +P +WS  +SNPD
Sbjct 1828 RLLAIPTSWSRVADSNPD 1775

> TodosTodos_rep_c8
Length=5676

Score = 25.8 bits (55), Expect = 2.9, Method: Compositional matrix adjust.
Identities = 10/18 (56%), Positives = 13/18 (72%), Gaps = 0/18 (0%)
Frame = +3

Query 151  RLLQLPGAWSSFIENPD 168
          RLL +P +WS  +SNPD
Sbjct 3963 RLLAIPTSWSRVADSNPD 4016

> TodosTodos_rep_c950
Length=5824

Score = 25.8 bits (55), Expect = 2.9, Method: Compositional matrix adjust.
Identities = 10/18 (56%), Positives = 13/18 (72%), Gaps = 0/18 (0%)
Frame = +3

Query 151  RLLQLPGAWSSFIENPD 168
          RLL +P +WS  +SNPD
Sbjct 3960 RLLAIPTSWSRVADSNPD 4013

Lambda      K      H      a      alpha
0.317      0.132    0.376    0.792    4.96

Gapped
Lambda      K      H      a      alpha      sigma
0.267      0.0410   0.140    1.90     42.6     43.6

Effective search space used: 133219550

Query= gi|149850121|emb|CAG29012.1| ArsB [Klebsiella pneumoniae]

```

90 Búsqueda de genes de resistencia a arsénico en el metagenoma microbiano de la Sabana de Bogotá

```

Length=429
Score      E
Sequences producing significant alignments:          (Bits) Value

TodosTodos_rep_c42                               698    0.0
TodosTodos_rep_c668                              295    2e-99
TodosTodos_c938                                  25.4    2.4
TodosTodos_c700                                  23.9    6.0
    
```

> TodosTodos_rep_c42

```

Length=2517

Score = 698 bits (1801), Expect = 0.0, Method: Compositional matrix adjust.
Identities = 403/429 (94%), Positives = 420/429 (98%), Gaps = 0/429 (0%)
Frame = -1
    
```

```

Query 1      MLLAGAI FVLTIVLVIWQPKGLGIGSATLGAVLALISGVVHFGDIPVVWNIWNATAIF 60
             MLLAGAI FVLTIVLVIWQPKGLGIGSATLGAVLAL++GVVH GDIPVVWNIWNATA F
Sbjct 1881    MLLAGAI FVLTIVLVIWQPKGLGIGSATLGAVLALVTVGVVHFGDIPVVWNIWNATAAF 1702

Query 61     IAVIIISLLLDESQFFEWAAALHVSRWGNGRGRLLFTYIvllgaavaalfandgaalILTP 120
             IAVIIISLLLDESQFFEWAAALHVSRWGNGRGRLLFT+IVLLGAVAALFANDGAALILTP
Sbjct 1701    IAVIIISLLLDESQFFEWAAALHVSRWGNGRGRLLFTWIVLVLGAVAALFANDGAALILTP 1522

Query 121    IVIAMLLALGFSKGTTLAFVMAAGFIADTASLPLIVSNLVNIVSADFFGLGFTEYASVMV 180
             IVIAMLLALGFSKGTTLAFVMAAGFIADTASLPLIVSNLVNIVSADFFGLGF EYASVMV
Sbjct 1521    IVIAMLLALGFSKGTTLAFVMAAGFIADTASLPLIVSNLVNIVSADFFGLGFREYASVMV 1342

Query 181    PVDIAAIATLVMLHLFFRKDIPPTYDlallkapakaikdlatFRTGWivlilllvvgffv 240
             PVDIAAI+ATLVMLHL+FRKDIP YD+ALLK+PA+AIKD ATF+TGW+VL+LLLVGFFV
Sbjct 1341    PVDIAAIVATLVMLHLFYFRKDIPQNYDMALLKSPAIAIKDPATFKTGWVLLLLLVGFFV 1162

Query 241    LEPLGIPVSAIAAVGAVILFAVAKRGHAINTEGKVLRGAPWQIVIFSLGMYLVVYGLRNAG 300
             LEPLGIPVSAIAAVGA+ILF VAKRGHAINTEGKVLRGAPWQIVIFSLGMYLVVYGLRNAG
Sbjct 1161    LEPLGIPVSAIAAVGALILFVAVAKRGHAINTEGKVLRGAPWQIVIFSLGMYLVVYGLRNAG 982

Query 301    LTEYLSGVLNVLADKGLWAATFGTGFLTAFLSSIMNMPTVLVGALSIDGSTATGVIKEA 360
             LTEYLSGVLNVLAD GLWAAT GTGFLTAFLSSIMNMPTVLVGALSIDGSTA+GVIKEA
Sbjct 981     LTEYLSGVLNVLADNGLWAATLGTGFLTAFLSSIMNMPTVLVGALSIDGSTATGVIKEA 802

Query 361    MIFANVIGCDLGPKITPIGSLATLLWLHVLSQKNMTITWGYFRTGIVMTLPVLFVTLAA 420
             M++ANVIGCDLGPKITPIGSLATLLWLHVLSQKNMTI+WGYYFRTGI+MTLPVLFVTLAA
Sbjct 801     MVYANVIGCDLGPKITPIGSLATLLWLHVLSQKNMTISWGYFRTGIIMTLPVLFVTLAA 622

Query 421    LALRLSFTL 429
             LALRLSFTL
Sbjct 621    LALRLSFTL 595
    
```

> TodosTodos_rep_c668

```

Length=619

Score = 295 bits (756), Expect = 2e-99, Method: Compositional matrix adjust.
Identities = 146/157 (93%), Positives = 154/157 (98%), Gaps = 0/157 (0%)
Frame = +1

Query 273    KVLRGAPWQIVIFSLGMYLVVYGLRNAGLTEYLSGVLNVLADKGLWAATFGTGFLTAFLS 332
    
```

```

KVLRGAPWQIVIFSLGMY+VVYGLRNAG TEYLSGVLN+LADKGLWAATFGTGFLTAFLS
Sbjct 1 KVLRGAPWQIVIFSLGMYIVVYGLRNAGFTEYLSGVLNLLADKGLWAATFGTGFLTAFLS 180

Query 333 SIMNMPTVLVGALSIDGSTATGVKEAMIFANVIGCDLGPKITPIGSLATLLWLHVLSQ 392
S+MNMPTVL+GALSIDGSTATGV+KEAMI+ANVIGCDLGPKITPIGSLATLLWLHVL+Q
Sbjct 181 SVMNMPTVLIGALSIDGSTATGVVKEAMIYANVIGCDLGPKITPIGSLATLLWLHVLAQ 360

Query 393 KNMTITWGYFRTGIVMTLPVLFVTLAALRLSFTL 429
KNMTITWGYFRTGIVMT+PVLVTLAALA RLS TL
Sbjct 361 KNMTITWGYFRTGIVMTVPVLFVTLAALAWRLSVTL 471

```

> TodosTodos_c938

Length=1274

Score = 25.4 bits (54), Expect = 2.4, Method: Compositional matrix adjust.

Identities = 18/54 (33%), Positives = 25/54 (46%), Gaps = 16/54 (30%)

Frame = +3

```

Query 30 LGAVLALISGVVHFGDIPVVVNIWNATATFIAVIIISLLDESGFWEAALHV 83
LG LAL SG+ H GD+ V+++S L SG A++HV

```

Sbjct 1071 LGKELALQSGLAHKGDV-----VVMVSGALVPSGTTNTASVHV 1184

> TodosTodos_c700

Length=565

Score = 23.9 bits (50), Expect = 6.0, Method: Compositional matrix adjust.

Identities = 13/21 (62%), Positives = 15/21 (71%), Gaps = 0/21 (0%)

Frame = -1

```

Query 406 GIVMTLPVLFVTLAALRLS 426
GIV T+P+L T AA LRLS

```

Sbjct 256 GIVTTVPLLCFTAATRLRLS 194

| Lambda | K | H | a | alpha |
|--------|-------|-------|-------|-------|
| 0.328 | 0.142 | 0.438 | 0.792 | 4.96 |

Gapped

| Lambda | K | H | a | alpha | sigma |
|--------|--------|-------|------|-------|-------|
| 0.267 | 0.0410 | 0.140 | 1.90 | 42.6 | 43.6 |

Effective search space used: 94231456

Query= gi|149850122|emb|CAG29013.1| ArsC [Klebsiella pneumoniae]

Length=141

Score E

| Sequences producing significant alignments: | (Bits) | Value |
|---|--------|-------|
| TodosTodos_rep_c42 | 270 | 3e-87 |
| TodosTodos_rep_c668 | 78.6 | 8e-20 |
| TodosTodos_c214 | 23.1 | 2.9 |
| TodosTodos_c709 | 21.6 | 9.2 |

> TodosTodos_rep_c42

Length=2517

Score = 270 bits (689), Expect = 3e-87, Method: Compositional matrix adjust.

Identities = 129/141 (91%), Positives = 138/141 (98%), Gaps = 0/141 (0%)

92 Búsqueda de genes de resistencia a arsénico en el metagenoma microbiano de la Sabana de Bogotá

```
Frame = -1
Query 1 MSNITIYHNPACGTSRNTLEMIRNSGIEPTVILYLETPPSRDELKLIADMGISVRALLR 60
      MSNITIYHNPACGTSRNTLEMIRNSG EPT+I YLETPP+rDEL+KLIADMGISVRALLR
Sbjct 579 MSNITIYHNPACGTSRNTLEMIRNSGTEPTIIHYLETPTTRDELVKLIADMGISVRALLR 400
Query 61 KNVEPEEELGLAEDKFTDDQLIDFMLQHPILINRPIVVTPLGTLKCRPSEVVLIDILPDAQ 120
      KNVEPEEELGLAEDKFTDD+LIDFMLQHPILINRPIVVTPLGT+LCRPSEVVL+ILPDAQ
Sbjct 399 KNVEPEEELGLAEDKFTDDRLIDFMLQHPILINRPIVVTPLGTRLCRPSEVVLLEILPDAQ 220
Query 121 KAAFTKEDGKVVDDSGKRLK 141
      K AF+KEDGKVVDD+GKRLK
Sbjct 219 KGAFSKEDGKVVDEAGKRLK 157
```

```
> TodosTodos_rep_c668
Length=619
Score = 78.6 bits (192), Expect = 8e-20, Method: Compositional matrix adjust.
Identities = 35/44 (80%), Positives = 40/44 (91%), Gaps = 0/44 (0%)
Frame = +1
Query 1 MSNITIYHNPACGTSRNTLEMIRNSGIEPTVILYLETPPSRDEL 44
      MSNITIYHNPACGTSRNTLEM+ N+G EPT+I YL+ PP+rDEL
Sbjct 487 MSNITIYHNPACGTSRNTLEMLHNNNGNEPTIINYLDMPTRDEL 618
```

```
> TodosTodos_c214
Length=1660
Score = 23.1 bits (48), Expect = 2.9, Method: Compositional matrix adjust.
Identities = 10/23 (43%), Positives = 14/23 (61%), Gaps = 0/23 (0%)
Frame = -2
Query 87 QHPILINRPIVVTPLGTLKCRPS 109
      Q+P+ RP+ +P LCRPS
Sbjct 672 QNPVRKGRPLFSSPRLRILCRPS 604
```

```
> TodosTodos_c709
Length=926
Score = 21.6 bits (44), Expect = 9.2, Method: Compositional matrix adjust.
Identities = 12/28 (43%), Positives = 16/28 (57%), Gaps = 0/28 (0%)
Frame = -3
Query 113 LDILPDAQKAAFTKEDGKVVDDSGKRL 140
      LDI + Q AA E +VD +GK+L
Sbjct 909 LDISVNWQGAAKAYSFDEVIVDSNGKKL 826
```

| Lambda | K | H | a | alpha |
|--------|-------|-------|-------|-------|
| 0.318 | 0.138 | 0.392 | 0.792 | 4.96 |

Gapped

| Lambda | K | H | a | alpha | sigma |
|--------|--------|-------|------|-------|-------|
| 0.267 | 0.0410 | 0.140 | 1.90 | 42.6 | 43.6 |

```
Effective search space used: 20513762
Database: contigs.fasta
Posted date: Sep 29, 2013 11:16 AM
Number of letters in database: 1,022,791
Number of sequences in database: 951
Matrix: BLOSUM62
```

Gap Penalties: Existence: 11, Extension: 1
 Neighboring words threshold: 13
 Window for multiple hits: 40

3. *tblastn* para dominios conservados de arsenito oxidasa

TBLASTN 2.2.27+

Reference: Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402.

Database: contigs.fasta

951 sequences; 1,022,791 total letters

Query= cd03476: Rieske_ArOX_small

Length=126

***** No hits found *****

| Lambda | K | H | a | alpha |
|--------|-------|-------|-------|-------|
| 0.318 | 0.135 | 0.418 | 0.792 | 4.96 |

Gapped

| Lambda | K | H | a | alpha | sigma |
|--------|--------|-------|------|-------|-------|
| 0.267 | 0.0410 | 0.140 | 1.90 | 42.6 | 43.6 |

Effective search space used: 17026015

Query= TIGR02694: arsenite_ox_S arsenite oxidase, small subunit

Length=129

Score E

| Sequences producing significant alignments: | (Bits) | Value |
|---|--------|-------|
| TodosTodos_c584 | 23.9 | 1.4 |
| TodosTodos_c63 | 23.5 | 1.9 |
| TodosTodos_c112 | 23.5 | 2.0 |
| TodosTodos_c195 | 23.5 | 2.1 |
| TodosTodos_rep_c698 | 23.1 | 2.6 |
| TodosTodos_c339 | 22.7 | 3.3 |
| TodosTodos_c324 | 22.3 | 4.4 |
| TodosTodos_c194 | 21.6 | 8.0 |

> TodosTodos_c584

Length=1023

Score = 23.9 bits (50), Expect = 1.4, Method: Compositional matrix adjust.

Identities = 7/18 (39%), Positives = 12/18 (67%), Gaps = 0/18 (0%)

Frame = -2

Query 92 VWGQATQNLFPQIVLRVAD 109

+W + T +PQ++LRV

Sbjct 809 IWHRRTLMPQLLRVKQ 756

94 Búsqueda de genes de resistencia a arsénico en el metagenoma microbiano de la Sabana de Bogotá

```
> TodosTodos_c63
Length=1572
Score = 23.5 bits (49), Expect = 1.9, Method: Composition-based stats.
Identities = 7/13 (54%), Positives = 10/13 (77%), Gaps = 0/13 (0%)
Frame = -3

Query 79 HFSVFDPEKGGQQ 91
          HF +F PE+G Q+
Sbjct 916 HFHIFSPERGRQR 878

> TodosTodos_c112
Length=1368
Score = 23.5 bits (49), Expect = 2.0, Method: Compositional matrix adjust.
Identities = 7/14 (50%), Positives = 10/14 (71%), Gaps = 0/14 (0%)
Frame = -1

Query 63 CPVSYSDNKTFC 76
          CP + +DN+TF C
Sbjct 96 CPYRFVSDNRTFEC 55

> TodosTodos_c195
Length=1688
Score = 23.5 bits (49), Expect = 2.1, Method: Compositional matrix adjust.
Identities = 7/15 (47%), Positives = 13/15 (87%), Gaps = 0/15 (0%)
Frame = -2

Query 16 KLNEPLDFNYPDASS 30
          +L+EP +FN+PD ++
Sbjct 1624 RLSEP*EFNWPDKTA 1580

> TodosTodos_rep_c698
Length=2677
Score = 23.1 bits (48), Expect = 2.6, Method: Compositional matrix adjust.
Identities = 17/61 (28%), Positives = 27/61 (44%), Gaps = 4/61 (7%)
Frame = +1

Query 41 VEGGVGPDGDIVAFSTLCTHMGCPSVSYSDNKTFCNCPCHFSVFDPEKGGQQVWGQATQNL 100
          ++GG P+ DIVA L V A + N P + DP+ +++ T N+
Sbjct 370 IDGGSAPENDIVAIARELRPTRLIV--DATDMGLN-PGEIRIIDPDDIA-EMFMMTTHNM 537

Query 101 P 101
P
Sbjct 538 P 540

> TodosTodos_c339
Length=1226
Score = 22.7 bits (47), Expect = 3.3, Method: Compositional matrix adjust.
Identities = 9/38 (24%), Positives = 18/38 (47%), Gaps = 0/38 (0%)
Frame = -2

Query 47 PDGDIVAFSTLCTHMGCPSVSYSDNKTFCNCPCHFSVFD 84
          G I S +C H + Y+A+ + F C ++ ++
Sbjct 811 KQGVTCLSRVCLHEIHQIDYTAEAEELFACNLYWPFYE 698
```

```
> TodosTodos_c324
Length=1708

Score = 22.3 bits (46), Expect = 4.4, Method: Composition-based stats.
Identities = 6/14 (43%), Positives = 11/14 (79%), Gaps = 0/14 (0%)

Frame = -2
Query 75  NCPCHFSVFDPEKG 88
          +C C F++ DP++G
Sbjct 990  SCLCSFAIVDPQRG 949

> TodosTodos_c194
Length=1516

Score = 21.6 bits (44), Expect = 8.0, Method: Compositional matrix adjust.
Identities = 18/70 (26%), Positives = 26/70 (37%), Gaps = 16/70 (23%)
Frame = +2

Query 52  VAFSTLCTHMG--PVSYSADNKTFNCPCHFSVFD-----PEKGGQVWGQATQNLPI 103
          V T C++ C + + D T PCHF F E+ G W Q
Sbjct 1019 VTAGTACSYNDCLDRLEHHPDGLTTGFPCHFQCFHGKPAVRRERHGPAAHWFSCQ----- 1183

Query 104  VLRVADNGDI 113
          V +NG++
Sbjct 1184 ---VENNGEV 1204

Lambda      K      H      a      alpha
0.317      0.140    0.435    0.792    4.96

Gapped

Lambda      K      H      a      alpha      sigma
0.267      0.0410   0.140    1.90     42.6     43.6

Effective search space used: 17863360

Database: contigs.fasta
Posted date: Sep 29, 2013 11:16 AM
Number of letters in database: 1,022,791
Number of sequences in database: 951

Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Neighboring words threshold: 13
Window for multiple hits: 40
```

D. Anexo: Publicaciones y participaciones en eventos

1. Realización de un artículo de revisión

Environmental Pollution 186 (2014) 272–281

Contents lists available at ScienceDirect

Environmental Pollution

journal homepage: www.elsevier.com/locate/envpol



Review

Environmental occurrence of arsenic in Colombia: A review

David L. Alonso^a, Sergio Latorre^b, Elianna Castillo^a, Pedro F.B. Brandão^{b, *}

^aLaboratorio de Química Ambiental, Grupo de Estudios para la Remedación y Mitigación de Impactos Negativos al Ambiente (GERMINA), Departamento de Química, Facultad de Ciencias, Universidad Nacional de Colombia, Ciudad Universitaria, Avenida Carrera 30 N° 45-03, Bogotá, Colombia

^bLaboratorio de Microbiología Ambiental y Aplicada, Grupo de Estudios para la Remedación y Mitigación de Impactos Negativos al Ambiente (GERMINA), Departamento de Química, Facultad de Ciencias, Universidad Nacional de Colombia, Ciudad Universitaria, Avenida Carrera 30 N° 45-03, Bogotá, Colombia

ARTICLE INFO

Article history:
Received 11 August 2013
Received in revised form 3 December 2013
Accepted 9 December 2013

Keywords:
Arsenic
Colombia
Latin America

ABSTRACT

The international literature on the presence of arsenic (As) in Latin America does not disclose the true magnitude of the presence of As in Colombia. In this paper, we summarize the literature on As occurrence in Colombia. The data reveal that As is present in matrices such as soil, sediments and water and in the food chain. Some of the As concentrations exceed the limits specified by national and international regulations. Arsenic higher concentrations are associated with mining regions (e.g., soils, up to 148 mg/kg; sediments, up to 1400 mg/kg) and agricultural areas (e.g., vegetables, up to 5.40 mg/kg; irrigation water, up to 255 µg/L), and underscore the potential human and environmental risks associated with the presence of As in the country. This review highlights the importance of focusing research on understanding the occurrence, origin and distribution of As in Colombia to better understand its environmental and public health impact.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Arsenic (As) is an element of natural origin that is widely distributed in the atmosphere, hydrosphere and biosphere. Environmental As is associated mainly with two types of sources: a) natural processes such as weathering, biological activity and volcanic emissions and b) human activities such as mining, industrial processes, smelting of metals, production of pesticides and wood preservatives, and use of fossil fuels. Although natural mineralization and microbial activities increase the mobilization of As in the environment, human activities exacerbate As contamination in soil and in water supplies (Litter et al., 2009; Lage et al., 2006).

Arsenic ingestion and/or inhalation at high concentrations may cause various health problems, including diabetes, skin lesions, gastrointestinal disorders, and genetic effects, especially due to its carcinogenic effects (IARC, 2004). In general, the populations most affected by As ingestion are those who live in countries with low incomes where access to clean drinking water is limited. People living in these countries must rely on groundwater exploitation to meet their water supply needs (Bundschuh et al., 2008). In Latin America, the estimated population at risk of As exposure exceeds 14

million people, with hundreds of recorded cases of exposure in countries such as Mexico, Chile and Argentina and potential risks of contamination to populations in Guatemala, Nicaragua, Peru, Bolivia, Brazil, Honduras, Cuba, El Salvador, Ecuador and Uruguay (Litter et al., 2009; Bundschuh et al., 2012a).

Many articles have been published on the problem of As in Latin America. However, there is little information in the international literature on the presence of this toxic metalloid in Colombia. For example, in a special edition of the journal *Science of the Total Environment* (vol. 429, July 2012), several review articles compile the most up-to-date information on As in Latin America. Nevertheless, only Bundschuh et al. (2012a) presents specific values for As in Colombia. The absence of data on the occurrence of As in Colombia is understandable because many of the results of studies carried out in Colombia are only available in a limited number of databases of Colombian public and private institutions. In addition, most of these study results are published in Spanish, which limits their distribution and international exposure. We conducted an extensive and in-depth literature review of books, national and international journals, undergraduate and graduate theses produced at Colombian universities, online databases and libraries of academic institutions and Colombian government agencies such as the Colombian Geological Survey (formerly the Colombian Institute of Geology and Mining, INGEOMINAS) and the Institute of Hydrology, Meteorology and Environmental Studies of Colombia (IDEAM).

* Corresponding author.

E-mail addresses: dalonso@unal.edu.co (D.L. Alonso), slatorre@unal.edu.co (S. Latorre), ecastillo@unal.edu.co (E. Castillo), pfb@unatedu.co (P.F.B. Brandão).

2. Overview of arsenic worldwide and in Latin America

It is estimated that tens of millions of people worldwide are at risk of exposure to excessive As levels in contaminated water, which constitutes a major public health problem (Smith and Smith, 2004). Arsenic contamination of groundwater has been reported extensively in many countries, including Bangladesh, India, Vietnam, Nepal, Taiwan, China, United States, Spain, Argentina and Chile (Mandal and Suzuki, 2002; Bundschuh et al., 2009). One of the main diseases caused by the consumption of water and/or food contaminated with As is chronic endemic regional hydroarsenicism (CERHA), which occurs after a variable period of exposure to As contents greater than 10 µg/L in drinking water (WHO, 1996). Latin American countries in which As groundwater pollution studies and risk assessments for CERHA in humans populations have been conducted include Argentina, Chile, Mexico Guatemala, Nicaragua, Peru, Bolivia, Brazil and Uruguay (Lillo, 2003; Nicolli, 2006; Bundschuh et al., 2008).

Recently published reviews of the occurrence of As in Latin America reveal a dramatic environmental problem of As water contamination in 14 of the 20 Latin America countries, where human activities in mining areas are the principal cause of As pollution because they increase the mobilization of As in surface water and groundwater (Bundschuh et al., 2012a). Arsenic in water can potentially enter the human food chain through bioaccumulation in fish and seafood, cattle meat, milk and cheese, and edible plants. Contaminated irrigation water can also contribute to accumulation of As in agricultural soils, representing a health risk to local residents (Bundschuh et al., 2012b). Conventional treatments for As removal, as well as newer and cheaper methods, suitable for use in Latin America, have been recently compiled, and some of these represent viable alternatives for reducing As exposure risk for poor and isolated populations of scattered regions of Latin America (Litter et al., 2010).

3. Overview of arsenic in Colombia

Colombia is a country located in the northwestern part of South America, with a land area of 1,141,748 km² and a marine extension of 928,660 km². The population of Colombia as of 2011 was 46,581,823 (National Administrative Department of Statistics – Departamento Administrativo Nacional de Estadística – DANE, 2011), making it the fourth most populous country in the American continent. Most of the population is located in the central (Andean) and northern (Caribbean) regions of the country. Colombia is composed of 32 geographic regions (Departments) and a Capital District (Bogotá D.C.). In large areas of the country, groundwater resources, which are prone to As contamination, are used for human consumption due to deficiencies in the aqueduct system and/or lack of surface water (Callejas, 2007). Although there have been a number of studies on the geology and hydrogeology of

Colombia, mainly carried out by the Colombian Geological Service (INGEOMINAS, 2004), scant research has been carried out on As pollution of water in Colombia. It is known that As contamination of groundwater in Colombian aquifers is of mixed origin (natural and anthropogenic) and mainly affects the departments of Tolima and Caldas, which are located in the central mountain ranges, and Nariño in the western Cordillera, which are areas of Neo-Tertiary volcanism. In these areas, the As comes from deposits containing arsenopyrite or other minerals with traces of As (Nicolli, 2006).

Since 2001 and 2003, the United States Environmental Protection Agency (US-EPA) and the World Health Organization (WHO), respectively, have recommend a maximum allowable As level of 10 µg/L in drinking water (US-EPA, 2001; WHO, 2003), a threshold level that Colombia adopted in 2007 as the maximum permissible concentration of As in drinking water (Ministry of Social Welfare and Ministry of Environment, Housing and Territorial Development – Ministerio de la Protección Social y Ministerio de Ambiente, Vivienda y Desarrollo Territorial, 2007). Other limits have been established in Colombian legislation for the maximum allowable As level in irrigation water (0.1 mg/L) and livestock drinking water (0.2 mg/L). Table 1 lists several maximum values currently established in Colombian legislation with respect to the presence, levels and toxicity of As in various matrices.

Ravenscroft (2007) reported a model based on a geographic information system (GIS) database developed by the Environmental Systems Research Institute (ESRI) to predict areas in the world with As risk in groundwater. For Colombia, the model predicts As risk primarily for the northern intermountain basins, where As may be mobilized by reductive dissolution, affecting a population of 1,480,000, with lesser components of risk deriving from and by alkali desorption (180,000 people) in drier areas and from sulfide oxidation (250,000 people) in the intensely mineralized volcanic terrain of the central mountains. The model predicts that the population that is at risk from As pollution in alluvial groundwater is approximately 1,664,092, which is equivalent to 5% of the total population of Colombia.

3.1. Arsenic in minerals, soils and sediments

Most of the mineralogical, geochemical and hydrogeological studies of soils and sediments in Colombia date to the second half of the twentieth century. The existing information has been the result of investigations by INGEOMINAS, which has also produced localization maps for those areas where studies have been conducted. That information is available online in the SICAT database of the Colombian Geological Survey (<http://www.sgc.gov.co>) and in the "Catalog of mineral deposits, prospects and manifestations in Colombia" (Mutis, 1983). In addition, a report by INGEOMINAS (1987) showed that As is usually present in the form of arsenopyrite associated with veins of ore minerals in several Departments, mostly along the central and western Cordillera of Colombia (sites:

Table 1
Maximum arsenic levels in different matrices established by the Colombian legislation.

| Matrix type | Limit value | Reference |
|---|-------------|--|
| Water used in Agric culture | 0.1 mg/L | Decree 1594 from 1984 (Ministerio de Agricultura, 1984) |
| Flora and fauna preservation in cold or warm freshwaters, and in estuarine or marine waters | 0.1 mg/L | |
| Water used for livestock | 0.2 mg/L | Resolution 2115 from 2007 (Ministerio de la Protección Social y Ministerio de Ambiente, Vivienda y Desarrollo Territorial, 2007) |
| Drinking water, for human and domestic use | 0.01 mg/L | |
| Waste water disposal | 0.01 mg/L | Resolution 3956 from 2009 (Secretaría Distrital de Ambiente, 2009) |

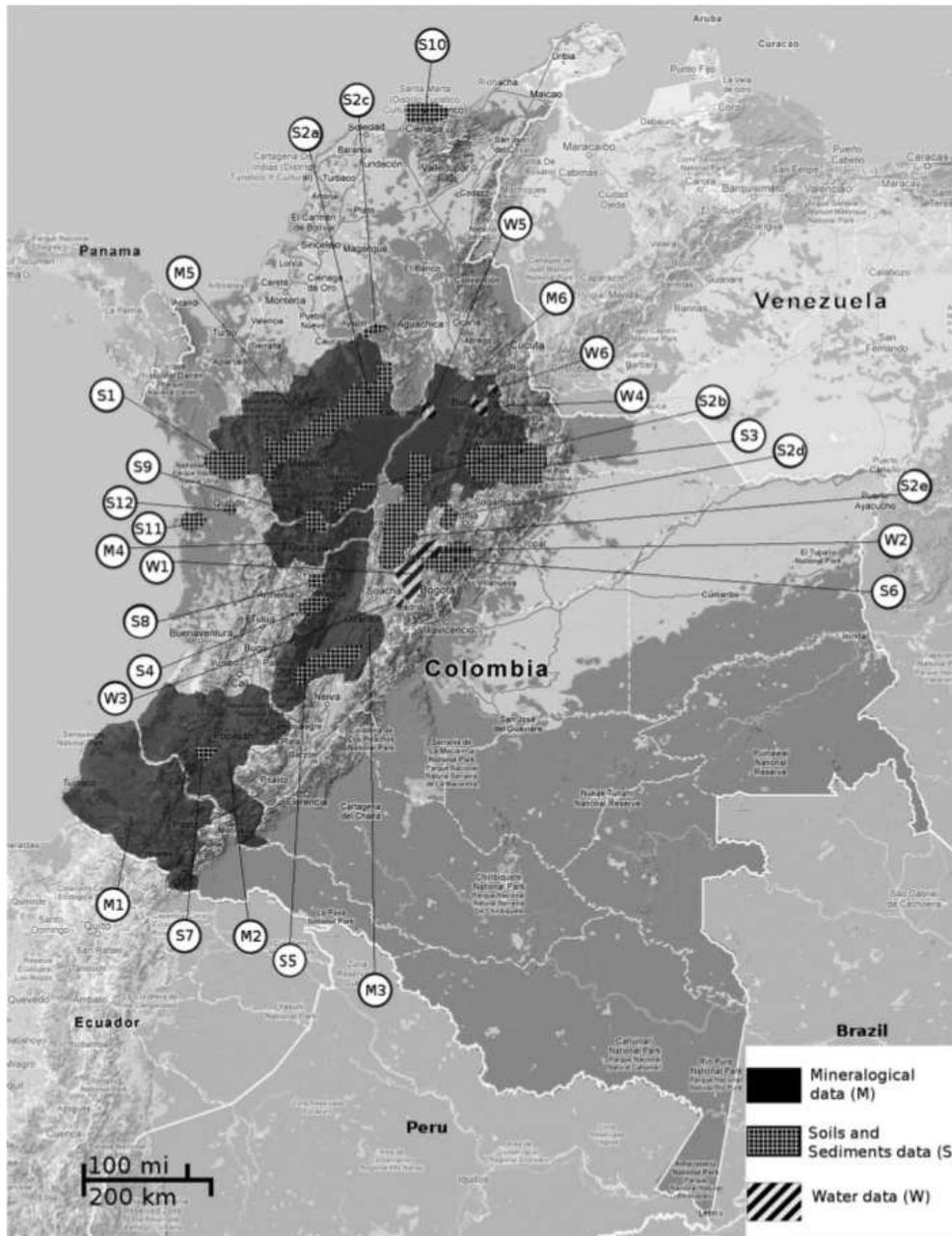


Fig. 1. Sites in Colombia where studies of arsenic have been reported. Numbers M1 –M6 (dark shaded areas): regions where the presence of arsenopyrite has been reported; numbers S1–S12 (square white lines on dark shaded areas): sites where studies of As have been reported in soil and sediment samples (see Table 2 for details); numbers W1–W6 (dash white lines shaded areas): sites where studies of As have been reported in surface water samples (see Table 3 for details). Adapted from a map of Colombia (Google Maps®).

M1, Nariño; M2, Cauca; M3, Tolima; M4, Caldas; and M5, Antioquia; Fig. 1). Recently, Ángel and Fierro (2012) studied the presence of As in the Vetas-California mining district (Santander Department), known for gold and silver vein-type mineralizations, in the eastern Cordillera of Colombia (site M6, Fig. 1). They demonstrated that sulfur minerals found in this area (pyrite, rosickyite, arsenopyrite, erignilite and semseyite) contained a higher proportion of As than those minerals containing gold in their structure.

The following information about the presence of As in soils and sediments in Colombia has been organized by geographic region (Departments). The information is summarized in Table 2 and Fig. 1 (sites S1–S13). The references referred to by the term SICAT (1–15), corresponding to the studies cited below, are listed in the Supplementary Material.

In 2005, González et al. (SICAT-1) reported As levels of 1.0–4.0 mg/kg in fine active sediments (site S1, Fig. 1) in the southwest zone of the department of Antioquia, mainly in the municipalities of Urrao and Vigía del Fuerte. To produce the Geochemical Atlas of Colombia, Prieto et al. (SICAT-2) sampled large areas of the central, northern and eastern regions of Colombia in 2005. The highest As levels, in the range of 13.7–35.3 mg/kg, were found in the Departments of Antioquia, Bolívar, Boyacá, Cundinamarca and Santander (sites S2a–S2e and S3, Fig. 1). As part of the same study, Vargas et al. (SICAT-3) studied a large area of the Department of Santander (site S3, Fig. 1) to obtain information about the local As levels. They reported maximum As levels of 40.7 mg/kg in soils and 13.6 mg/kg in stream sediments.

Several studies have been conducted in the Department of Tolima. In 1975 and 1985, Lozano et al. (SICAT-4) and INGEOMINAS (SICAT-5), respectively, reported on As levels in soils and sediments

in this Department. Their results for the municipality of Cajamarca are particularly noteworthy (site S4, Fig. 1). The first study reported As levels in the range of 200–3000 mg/kg in the heavy fraction of stream sediments located near mining zones. The second study reported As levels of 0.93–148 mg/kg in soils located near a gold mining project in the municipality of San Antonio. In 2003, Camacho et al. (SICAT-6) conducted an extensive study of various sectors of the Ataco zone (site S5, Fig. 1) and found As levels of 0.7–45.2 mg/kg.

Gúiza et al. (SICAT-7) conducted geochemical mapping in the upper basin of the Bogotá River (site S6, Fig. 1) in the Cundinamarca Department and found As levels of 1.2–8.6 mg/kg in soil and 1.6–10.5 mg/kg in sediments.

Studies conducted in the southern and western parts of Colombia in the 1970's found As levels in pan-concentrate sediments, and most of these As levels were high. León et al. (SICAT-8) reported As levels of 70–5000 mg/kg at a site known as El Bordo in the Cauca Department (site S7, Fig. 1). Lozano et al. (SICAT-4) found As levels of 200–5000 mg/kg around mining areas at Salento in the Quindío Department (site S8, Fig. 1). Kassem et al. (SICAT-9) reported As levels of 1000–10,000 mg/kg for the upper basin of the Quindío River (site S8, Fig. 1). Pérez and Lozano (SICAT-10) reported As levels of 200–7000 mg/kg detected in the municipalities of Samaná and Pensilvania in the Caldas Department (site S9, Fig. 1). Although these levels are considered high, it should be noted they were determined using qualitative pre-concentration procedures that have low reproducibility. Moreover, no recent similar studies are available for comparison with the results of the cited studies.

Perdomo et al. (1998) studied the sediments of the Ciénaga Grande de Santa Marta swamp (site S10, Fig. 1) on the Caribbean

Table 2
Arsenic studies in soil and sediment samples of Colombia.

| Department | Site | Sample type | Analytical method | As range (mg/kg) | As mean (mg/kg) | SD | Median (mg/kg) | n | Reference ^a | Site on map ^b |
|--------------|--|---------------------------|---------------------|------------------|-----------------|-------|----------------|-----|------------------------|--------------------------|
| Antioquia | Urrao and Vigía del Fuerte | Stream sediments | AAS ^c | 1.0–4.0 | 2.3 | 0.9 | 2.0 | 196 | SICAT-1 | S1 |
| | | Center and northwest | Floodplain sediment | AFS ^d | 13.7–35.3 | 5.0 | nd | 3.7 | 389 | SICAT-2 |
| | Southwest | Floodplain sediment | AFS ^d | 13.7–35.3 | 5.0 | nd | 3.7 | 389 | SICAT-2 | S2b |
| Bolívar | South | Floodplain sediment | AFS | 13.7–35.3 | 5.0 | nd | 3.7 | 389 | SICAT-2 | S2c |
| Boyacá | Midwest and northeast | Floodplain sediment | AFS | 13.7–35.3 | 5.0 | nd | 3.7 | 389 | SICAT-2 | S2d |
| Cundinamarca | Northwest | Floodplain sediment | AFS | 13.7–35.3 | 5.0 | nd | 3.7 | 389 | SICAT-2 | S2e |
| | Zipaquirá, Chocontá and Valle de Ubaté | Soils | INAA ^e | 1.2–8.6 | 5.9 | nd | 5.1 | 95 | SICAT-7 | S6 |
| Santander | Southwest | Stream sediments | INAA | 1.6–10.5 | 8.3 | nd | 6.2 | 95 | | |
| | | Floodplain sediment | AFS | 13.7–35.3 | 5.0 | nd | 3.7 | 389 | SICAT-2 | S3 |
| | Barrancabermeja–California–Málaga | Horizon A soils | INAA | <0.5–34.4 | 5.62 | nd | 4.75 | 56 | SICAT-3 | S3 |
| | | Horizon C soils | INAA | <0.5–40.7 | 7.64 | nd | 6.25 | 56 | | |
| | | Stream sediments | INAA | <0.5–13.6 | 13.21 | nd | 5.15 | 56 | | |
| Tolima | Cajamarca | Pan-concentrate sediments | AES ^f | 200–5000 | nd | nd | nd | 49 | SICAT-4 | S4 |
| | San Antonio | Soils | CM ^g | 0.93–148 | 7.2 | nd | 5.89 | 60 | SICAT-5 | S4 |
| | Ataco Zone, Tapias | Soils | nd | 1.0–32.3 | nd | nd | nd | 65 | SICAT-6 | S5 |
| | Ataco Zone, Río Meche | Soils | nd | 0.7–18.8 | nd | nd | nd | 47 | | |
| | Ataco Zone, Cerro Copete | Soils | nd | 1.0–45.2 | nd | nd | nd | 37 | | |
| Cauca | El Bordo | Pan-concentrate sediments | AES | 70–5000 | nd | nd | nd | 20 | SICAT-8 | S7 |
| Quindío | Salento | Pan-concentrate sediments | AES | 200–5000 | nd | nd | nd | 49 | SICAT-4 | S8 |
| | | Pan-concentrate sediments | nd | 1000–10,000 | nd | nd | nd | 80 | SICAT-9 | S8 |
| Caldas | Samaná and Pensilvania | Pan-concentrate sediments | AES | <200–7000 | nd | nd | 200 | 35 | SICAT-10 | S9 |
| | | Marmato Mining District | Stream sediments | nd | 10–1400 | nd | nd | nd | nd | Bundschuh et al. (2012a) |
| Magdalena | Santa Marta | Sediments | AAS | 0.015–0.087 | 0.044 | 0.027 | 0.029 | 12 | Perdomo et al. (1998) | S10 |
| Chocó | Coquí | Fine sediments | AAS | 7–83 | nd | nd | nd | 58 | SICAT-11 | S11 |
| | | Ichó and Comitó rivers | Sediments | AES | 1–40 | nd | nd | nd | nd | SICAT-12 |
| | Quibdó | Stream sediments | CM | 1–24 | 6 | nd | 4 | 739 | SICAT-13 | |
| | Quibdó | Stream sediments | AAS | <1.0 | nd | nd | nd | 196 | SICAT-1 | S12 |

^a See reference SICAT database.

^b Refer to Fig. 1 nd: not described.

^c AAS: Atomic Absorption Spectroscopy.

^d AFS: Atomic Fluorescence Spectroscopy.

^e INAA: Instrumental Neutron Activation Analysis.

^f AES: Atomic Emission Spectroscopy.

^g CM: Colorimetric Method.

coast of Colombia and found maximum As concentrations of 0.087 mg/kg in the dry season and 0.059 mg/kg in the rainy season. Other studies have reported the following As levels in sediments near the Pacific coast: 7–83 mg/kg at the *Coquí River* (site 11, Fig. 1; SICAT-11), 1–40 mg/kg at the *Comitó River* and *Ichó River* (site 12, Fig. 1; SICAT-12; SICAT-13), and less than 1.0 mg/kg in the municipality of *Quibdó* in the *Chocó* Department (site 12, Fig. 1; SICAT-1).

An examination of the above-mentioned findings indicates that sites in the Departments of *Tolima* and *Santander*, which are located around mining areas, contained As levels up to 148 and 41 mg/kg, respectively (Table 2). These values are high in comparison to values obtained in studies of similar soils of Brazil (15–100 mg/kg), Canada (19 mg/kg) and the USA (10–100 mg/kg) (Deschamps et al., 2002). Given the current mining boom in Colombia, more studies are necessary to evaluate As levels near mining areas.

3.2. Arsenic in surface water and groundwater

Although groundwater is one of the major sources of drinking water on the planet, Colombia uses preferentially surface water, which satisfies most of the national demand, as the main source of drinking water. However, in some areas, small amounts of water are drawn from aquifers of recent deposits and Tertiary and Cretaceous sedimentary units through wells ranging in depth from 50 to 300 m. This practice is customary in areas where the flow of surface water is not constant, when surface water availability is limited or where deficiencies in the water distribution system exist (Cacua, 2004; INGEOMINAS, 2004; Callejas, 2007). The following information (summarized in Table 3) complements the few studies reported by Bundschuh et al. (2012a) on the presence of As in Colombian water resources.

In 1997, González and Vargas (SICAT-14) reported high levels of As and heavy metals in phreatic water samples from sites of intense irrigation of horticultural crops in several municipalities of the *Sabana de Bogotá* in the *Cundinamarca* Department (site W1, Fig. 1). The arsenic content in the groundwater ranged from 14 to 255 µg/L, suggesting a need to take control measures to prevent further quality degradation of agricultural soils in the region. In 2005, Espinosa and Gúiza (SICAT-15) reported As contents of 0.03–2.73 µg/L in surface water supplies in the municipalities of *Zipaquirá*, *Chocontá* and *Valle de Ubaté* (site W2, Fig. 1).

Sarmiento et al. (1999a) carried out studies in the municipality of *Sibaté*, where the *Muña* reservoir is located (site W3, Fig. 1). This area, south of Bogotá, has suffered increasing amounts of urban and industrial waste discharge, which have been associated with respiratory and dermatological diseases and skin cancer in the *Sibaté*

population. These authors reported As contents of 32–52 µg/L (mean ± SD: 41.8 ± 5.3 µg/L; n = 16) in the water used by people living near the *Muña* reservoir. This concentration range exceeds the maximum allowable level specified in Colombian legislation for drinking water (10 µg/L; Table 1).

Cacua (2004) evaluated the quality of two groundwater wells located in *Bucaramanga* (site W4, Fig. 1) in the *Santander* Department and found As levels of 3.0–4.0 µg/L. Based on these levels, Cacua concluded that these two wells could be used as sources of drinking water. In 2005, Vargas et al. (SICAT-3) analyzed surface water samples from a large area located between the municipalities of *Barrancabermeja*, *Suratá* and *Málaga* (site W5, Fig. 1) and reported As contents ranging from less than 0.03 to 3.54 µg/L. In an analysis of the environmental effects of gold mining, Ángel and Fierro (2012) used various parameters to evaluate the water quality in the upper basin of the *Suratá River* near the municipality of *California* (site W6, Fig. 1) and reported As concentrations up to 13 µg/L, which again is above the As limit specified in Colombian legislation for drinking water.

Bundschuh et al. (2012a) reported high levels of As in the water of the *Marmato River* in the *Marmato* mining district (site S9, Fig. 1) to the northwest of the *Caldas* Department. The amount of As in the river's sediments was found to vary seasonally (dry season, 10–850 mg/kg; wet season, 440–1400 mg/kg). The levels suggest that acid mine drainage can be identified as one of the main sources of contamination of surface water in the area, influencing large cities, such as *Cali*, where a high percentage of the population depends on water from the river.

In general, the few available studies mentioned above indicate that at the locations studied, As is present in water at levels that, despite being low, exceed in several cases the maximum national recommended level of As allowed in water used for drinking, irrigation, livestock and aquatic life (Table 1). Therefore, taking into account the regions evaluated, it can be inferred that the occurrence of As in these water resources can be attributed to human activities. Nevertheless, research is required to obtain further evidence to confirm potential contamination predicted by the Ravenscroft (2007) model for many of these areas.

3.3. Arsenic in the agricultural sector

Information about As in the Colombian agricultural sector is very limited. The studies available on this subject are mainly related to the *Cundinamarca* Department and particularly to the *Sabana de Bogotá*, which includes the city of *Bogotá* (Fig. 2). Horticulture is one of the main agricultural activities in this area. Its soils are continually exposed to water from the highly contaminated *Bogotá River*

Table 3
Arsenic studies in surface water and groundwater samples of Colombia.

| Department | Site | Analytical method | As range (µg/L) | As mean (µg/L) | SD | Median (µg/L) | n | Reference ^a | Site on map ^b |
|--------------|---|----------------------|-----------------|----------------|------|---------------|----|--------------------------|--------------------------|
| Cundinamarca | Madrid, Cota, Chia, Cajicá, Zipaquirá, Mosquera, Funza and Soacha | ICP-AES ^c | 14–255 | 93.8 | 54.2 | 90 | 16 | SICAT-14 | W1 |
| | Zipaquirá, Chocontá and Valle de Ubaté | ICP-MS ^d | 0.03–2.73 | 0.32 | nd | 0.21 | 95 | SICAT-15 | W2 |
| | Sibaté | AAS-HG | 32–52 | 41.8 | 5.3 | 41 | 16 | Sarmiento et al. (1999a) | W3 |
| Santander | Bucaramanga | AAS-HG ^e | 3.0–4.0 | 3.5 | 0.5 | nd | 2 | Cacua (2004) | W4 |
| | Barrancabermeja, Suratá and Málaga | ICP-MS | <0.03–3.54 | 0.05 | nd | 0.4 | 56 | SICAT-3 | W5 |
| | California | nd | <13.0 | nd | nd | nd | 1 | Ángel and Fierro (2012) | W6 |

^a See reference SICAT database.

^b Refer to Fig. 1 nd: not described.

^c ICP-AES: Inductively Coupled Plasma-Atomic Emission Spectroscopy.

^d ICP-MS: Inductively Coupled Plasma-Mass Spectrometry.

^e AAS-HG: Atomic Absorption Spectroscopy-Hydride Generation.

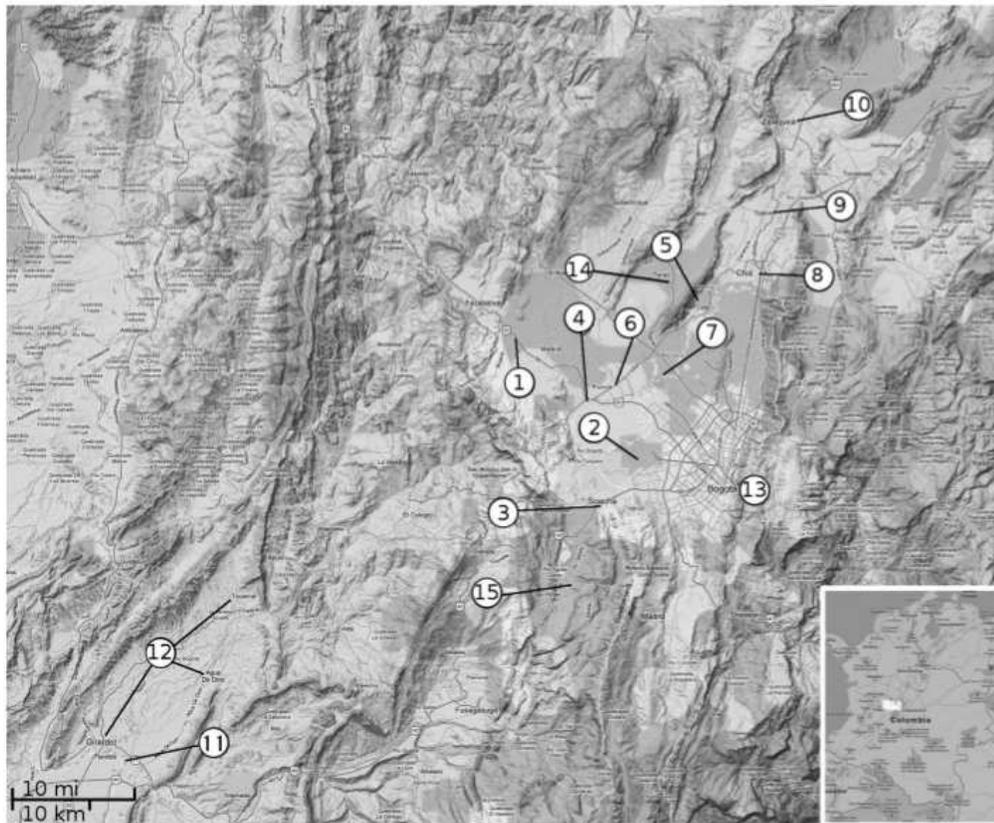


Fig. 2. Sites in the Department of Cundinamarca (Colombia) where agricultural studies reported the determination of total arsenic in examined samples. Numbers 1 to 15 correspond to the locations where the studies were made (see Table 4 for details). Cleared area in Colombian map (bottom right) represents the enlarged region shown in figure. Adapted from a map of Colombia (Google Maps[®]).

and its effluents, which is used for irrigation purposes. A summary of the information available on the total As content in agricultural samples is presented in Table 4, and the sites from which the samples were collected are shown in Fig. 2.

In 1997, González and Vargas (SICAT-14) conducted a risk assessment study in which the concentrations of As, Cd, Co, Cr, Cu, Hg, Mn, Mo, Ni, Pb and Zn in soils from the *Sabana de Bogotá* were determined by *aqua regia* extraction. The various municipalities from which soils were sampled are shown in Fig. 2 and listed in Table 4: *Bojacá* and *Madrid* (site 1), *Bosa* (site 2), *Soacha* (site 3), *Mosquera* (site 4), *Cota* (site 5), *Funza* (site 6), *Engativá* (site 7), *Chía* (site 8), *Cajicá* (site 9) and *Zipaquirá* (site 10). The maximum As concentrations found were 3.2 mg/kg in the northern area (*Zipaquirá*) and 12.0 mg/kg in the southern area (*Soacha*) (SICAT-14). The results showed that the areas of highest As concentration were located in the southern part of the *Sabana de Bogotá*, where the *Bogotá River* receives contamination loads from the industrial and urban areas of the city, as well as waste from the upstream leather industry (SICAT-14).

In 2000, the Bogotá District Health Secretariat (Secretaría Distrital de Salud de Bogotá, 2000) assessed the contamination levels of As, Cd, Cr, Hg and Pb at vegetable plantations in *Bosa* (site 2, Fig. 2) irrigated with domestic waste water from the surrounding neighborhoods, rainwater, and *Tunjuelito River* water. The authors reported that among the samples analyzed, only 2.6% and 5.9% of lettuce and chard, respectively, had As levels above 0.14 mg/kg. The As levels in the soils sampled varied from 0.5 to 1.3 mg/kg—much lower than the levels reported for other soils from the *Sabana de Bogotá*.

Miranda et al. (2008) studied a large zone of horticultural crops and analyzed the presence of some pollutants such as As in irrigation water, soil and plant tissue (lettuce, celery, cabbage and broccoli) in the municipalities of *Soacha* and *Mosquera* (sites 3 and 4, respectively; Fig. 2). The As levels found were less than 3.0 µg/L for irrigation water, up to 1.7 mg/kg for soil, and 0.02–0.19 mg/kg (wet weight) in plant tissue. The authors concluded that, in general, the As contents in these samples were not excessive, although the As levels found in cabbage were higher than the world average.

Table 4
Agricultural studies of arsenic in the *Sabana de Bogotá* (Department of Cundinamarca).

| Site | Sample type | Analytical method | As range (mg/kg) | As mean (mg/kg) | SD | Median (mg/kg) | n | Reference ^a | Site on map ^b |
|------------------------------------|----------------------------|--------------------------------|------------------|-----------------|-------|----------------|----|--------------------------------------|--------------------------|
| Bojacá – Madrid | Agricultural soil | ICP-AES ^c | 1.1–6.2 | 2.97 | 1.34 | 2.95 | 20 | SICAT-14 | 1 |
| | Agricultural soil | ICP-AES | 0.2–12.0 | 2.15 | 2.49 | 1.1 | 21 | SICAT-14 | 2 |
| | Agricultural soil | AAS-HG ^d | 0.5–1.3 | nd | nd | nd | 10 | Secretaría Distrital de Salud (2000) | |
| Soacha | Vegetables | AAS-HG | 0.07–0.13 | nd | nd | nd | 72 | Secretaría Distrital de Salud (2000) | |
| | Agricultural soil | AAS-HG | 0.98–1.69 | 1.34 | 0.5 | nd | 2 | Barragán (2008) | |
| | Vegetables | AAS-HG | 0.116–0.199 | 0.158 | 0.03 | 0.156 | 7 | Barragán (2008) | |
| | Agricultural soil | ICP-AES | 0.2–12.0 | 2.2 | 2.5 | 1.1 | 21 | SICAT-14 | 3 |
| | Agricultural soil | AAS-HG | 1.35–1.75 | 1.51 | 0.21 | 1.42 | 3 | Miranda et al. (2008) | |
| | Agricultural soil | ICP-AES | 3.65–4.90 | 3.96 | 0.48 | 3.75 | 12 | Bello and Lesmes (2011) | |
| Mosquera | Vegetables | AAS-HG | 0.020–0.190 | 0.055 | 0.047 | 0.05 | 15 | Miranda et al. (2008) | |
| | Vegetables | ICP-AES | 0.002–0.011 | 0.004 | 0.004 | 0.003 | 9 | Bello and Lesmes (2011) | |
| | Irrigation water | AAS-HG | <3 µg/L | nd | nd | nd | 3 | Miranda et al. (2008) | |
| | Agricultural soil | ICP-AES | 1.4–5.1 | 3.2 | 1.1 | 3.3 | 19 | SICAT-14 | 4 |
| | Irrigation water | AAS-HG | <3 µg/L | nd | nd | nd | 3 | Miranda et al. (2008) | |
| | Agricultural soil | AAS-HG | 1.35–1.75 | 1.51 | 0.21 | 1.42 | 3 | Miranda et al. (2008) | |
| Cota | Vegetables | AAS-HG | 0.020–0.190 | 0.055 | 0.047 | 0.05 | 15 | Miranda et al. (2008) | |
| | Agricultural soil | AAS-HG | 0.93–3.70 | 2.05 | 0.91 | 2.07 | 16 | Muñoz (2008) | |
| | Irrigation water | AAS-HG and AAS-GF ^e | <5.0 µg/L | nd | nd | nd | 8 | Muñoz (2008) | |
| | Agricultural soil | ICP-AES | 0.2–8.0 | 2.7 | 1.8 | 2.2 | 21 | SICAT-14 | 5 |
| | Agricultural soil | AAS-HG | 1.10–2.54 | 1.66 | 0.66 | 1.49 | 4 | Barragán (2008) | |
| | Vegetables | AAS-HG | 0.116–0.199 | 0.158 | 0.03 | 0.156 | 7 | Barragán (2008) | |
| Funza | Agricultural soil | ICP-AES | 1.4–5.1 | 3.2 | 1.1 | 3.3 | 19 | SICAT-14 | 6 |
| | Agricultural soil | ICP-AES | 0.2–4.3 | 2.4 | 1.3 | 2.8 | 25 | SICAT-14 | 7 |
| Engativá – La Florida | Agricultural soil | ICP-AES | 0.2–7.2 | 2.1 | 1.9 | 2.0 | 17 | SICAT-14 | 8 |
| | Agricultural soil | ICP-AES | 1.0–4.2 | 2.0 | 1.0 | 2.0 | 21 | SICAT-14 | 9 |
| Chía | Agricultural soil | ICP-AES | 0.2–3.2 | 1.5 | 1.1 | 1.2 | 17 | SICAT-14 | 10 |
| | Agricultural soil | AAS-HG | 2.54 | nd | nd | nd | 1 | Barragán (2008) | 11 |
| Cajicá | Agricultural soil | nd | 2.5–14.5 | 7.1 | 3.1 | 7.2 | 23 | Alfaro et al. (2002) | |
| | Irrigation water | nd | 20–53 µg/L | 37.7 | 7.6 | 37 | 25 | Alfaro et al. (2002) | |
| | Vegetables | AAS-HG | 0.116–0.199 | 0.158 | 0.03 | 0.156 | 7 | Barragán (2008) | |
| | Rice, plantain and grasses | nd | 0.23–1.95 | 0.72 | 0.37 | 0.65 | 27 | Alfaro et al. (2002) | |
| | Agricultural soil | nd | 2.5–14.5 | 7.1 | 3.1 | 7.2 | 23 | Alfaro et al. (2002) | 12 |
| | Irrigation water | nd | 20–53 µg/L | 37.7 | 7.6 | 37 | 25 | Alfaro et al. (2002) | |
| Zipaquirá | Rice, plantain and grasses | nd | 0.23–1.95 | 0.72 | 0.37 | 0.65 | 27 | Alfaro et al. (2002) | |
| | Agricultural soil | nd | 5.8–12.5 | 8.1 | 1.8 | 8.1 | 12 | González and Mejía (1995) | 13 |
| Ricaurte | Agricultural soil | ICP-AES | 2.55–3.86 | 3.2 | 0.7 | 3.18 | 6 | Bello and Lesmes (2011) | 14 |
| | Vegetables | ICP-AES | 0.003–0.024 | 0.006 | 0.009 | 0.003 | 9 | Bello and Lesmes (2011) | |
| Girardot, Agua de Dios and Tocaima | Vegetables | AAS ^f | 0.42–5.40 | 2.16 | 1.80 | 1.27 | 12 | Sarmiento et al. (1999a) | 15 |

^a See reference SICAT database.

^b Refer to Fig. 2 nd: not described.

^c ICP-AES: Inductively Coupled Plasma-Atomic Emission Spectroscopy.

^d AAS-HG: Atomic Absorption Spectroscopy-Hydride Generation.

^e AAS-GF: Atomic Absorption Spectroscopy-Graphite Furnace.

^f AAS: Atomic Absorption Spectroscopy.

However, the As levels detected in cabbage are below the Chilean National Standard for As in plants for human consumption (0.50 mg/kg) and below the levels reported in other Latin American countries, such as Bolivia (Bundschuh et al., 2012b).

The irrigation district called *La Ramada*, which covers 6400 ha in the vicinity of the municipalities of *Mosquera* and *Soacha* (site 4, Fig. 2), was studied by Muñoz (2008). Agricultural soils and irrigation water used in the cultivation of vegetables (broccoli, cauliflower and lettuce), were evaluated. As levels of 0.93–3.70 mg/kg were detected for the soils and levels up to 5.0 µg/L were detected in the irrigation water.

Barragán (2008) used Tessier's extraction procedure to analyze the As mobile fractions in agricultural soils in the *Cundinamarca* Department in municipalities near the *Bogotá River*, such as *Bosa*, *Cota* and *Ricaurte* (sites 2, 5 and 11, respectively, in Fig. 2). According to the author, a direct relationship was found between the As absorption by the plants and the As contents in the iron and manganese oxide fractions. The total As contents in the soils were 0.98–2.54 mg/kg, and those in the plants were 0.116–0.199 mg/

kg. Although the levels detected are below the limits for As in agricultural soils and plant tissues, this study is important because it is one of the few conducted in Colombia that evaluated As bioavailability in soils using various extraction procedures and examined the relation of As bioavailability to the As concentration in plants.

Between 1996 and 1998, Alfaro et al. (2002) monitored the levels of As, Cd, Hg and Pb in the lower basin of the *Bogotá River* (*Ricaurte*, site 11; *Girardot, Agua de Dios* and *Tocaima*, site 12; Fig. 2). The region evaluated comprises approximately 2000 ha of rice, 300 ha of banana and 4200 ha of pasture that are irrigated with water from the *Bogotá River*. The authors reported that consumers of products from this region, such as rice, are at risk of chronic intoxication by As and heavy metals. The As contents in rice grains from this region ranged from 0.46 to 0.58 mg/kg, exceeding the maximum recommended level (0.30 mg/kg; WHO, 2012) and the levels reported for other countries (Bangladesh, Cambodia and India; Wang et al., 2013). However, the metalloid contents in the soils sampled (2.5–14.5 mg/kg) are considered to be within acceptable

levels and are consistent with levels reported for agricultural soils in other Latin American countries (Bundschuh et al., 2012a).

González and Mejía (1995) analyzed the presence of As, Cd, Cr, Hg and Pb in soils from six horticultural areas of the *Sabana de Bogotá*. The As levels in the soils ranged from 5.8 to 12.5 mg/kg (site 13, Fig. 2). They also studied the accumulation of these elements in the edible parts of vegetable (lettuce, cucumber and carrot) grown in these areas. Analysis of As accumulated in the vegetables using three different metalloid concentrations in the irrigation water (0.1, 1.0 and 10.0 mg/L) showed that lettuce accumulated As at greater concentrations (0.97, 3.82 and 13.50 mg/kg, respectively) than cucumber (0.04, 0.14 and 1.26 mg/kg, respectively) and carrot (0.07, 0.36 and 0.94 mg/kg, respectively). These results highlight the need to continue studying As contents in vegetables, especially lettuce, because it is consumed in great quantities in various regions of Colombia.

Belo and Lesmes (2011) tested samples of agricultural soil and vegetable tissues (celery, lettuce and chard) from the municipality of Tenjo (site 14, Fig. 2). Arsenic was detected in both agricultural soil (2.55–3.86 mg/kg) and plant tissues (0.003–0.024 mg/kg). The As contents in vegetable tissues were lower than those determined in the above-mentioned studies, possibly because Tenjo is located north of the city of Bogotá, where the Bogotá River receives lower contamination loads.

3.4. Arsenic in the human food chain

Several studies have been conducted on the use of biomarkers to assess human As exposure in Latin America (e.g., As levels in blood, hair, nail and urine samples), but few such studies have been conducted in Colombia. The only comprehensive work on this subject in Colombia is that reported by Sarmiento et al. (1999a, 1999b), who used hair as a biomarker for individuals living in areas close to the Muña reservoir (Sibaté, site 15; Fig. 2). The As contents detected in hair samples ranged from 10.5 to 2078 µg/kg ($n = 50$). Nevertheless, these results are considered important because these As levels are comparable to those reported by McClintock et al. (2012) in Bolivia (37–2110 µg/kg) and considerably lower than those detected in a highly As-exposed area in Argentina (400–20,000 µg/kg). Furthermore, Sarmiento et al. reported alarmingly high As levels in water used for human consumption and irrigation (up to 52 µg/L) and high values in strawberries (0.36 mg/kg), vegetables (0.42–5.40 mg/kg), milk (38–76 µg/L) and bovine meat (1.58 mg/kg). The As content in milk depends on nutritional and environmental factors, and its presence in livestock water raises serious concerns about As in cow's milk. The As content detected in milk in the study by Sarmiento et al. exceeded the limit recommended by the WHO for As in drinking water (10 µg/L) and exceeded the values reported for As in cow's milk in other Latin American regions, such as Comarca Lagunera (Mexico, 1.0–8.6 µg/L) and the Chaco–Pampean plain (Argentina, 0.5–7.8 µg/L) (Bundschuh et al., 2012b). The As concentration detected in bovine meat is also higher than that reported by Alfaro et al. (2002) (mean value: 0.25 mg/kg) for a site near the area studied by Sarmiento et al. (1999a, 1999b) (site 15, Fig. 2). In addition, the As content exceeds the range of values reported for llama muscle meat in an area of high As exposure in the Andean highlands of Argentina (0.06–0.28 mg/kg), which does not exceed the maximum permissible limit set by the US Food and Drug Administration (0.5 mg/kg) (Bundschuh et al., 2012b). In Colombia, as in other Latin American countries, cattle are a major type of livestock raised and supply milk and meat for the people living in the region. Thus, it is important to perform additional studies in other regions of the country to establish relationships between the drinking water used for cattle, the forage irrigation water quality,

and the As contents in livestock products, given the role of cattle in As exposure via the human food chain.

4. Conclusions and recommendations

The studies mentioned in this review describe what is presently known about the risk of As exposure in Colombia and help to complete the picture of risk of exposure to this problematic toxic metalloid in Latin America. The published literature shows that the studies of As conducted in Colombia have been mostly geological and agricultural studies.

Most geological studies on the presence of As in Colombia have been carried out by the Colombian Geological Survey (*Servicio Geológico Colombiano*, SGC) and have shown that arsenic typically occurs in Colombia in the form of arsenopyrite associated with gold ores in various parts of the country where there is volcanic activity. The arsenic contents of soils reported by the SGC are within natural soil As levels. Of the country's 32 departments, high levels of As (>40 mg/kg) in soils and sediments have been reported in only four (Caldas, Chocó, Santander and Tolima). These concentrations can be attributed to the release of As by mining activities in those areas. However, based on the available data, it is not possible to draw firm conclusions concerning the impact of mining activities in these areas on the risk of As exposure because studies on the effects of As on the human food chain and in the surrounding towns have not been conducted. Therefore, considering that the country is going through a mining boom, we believe more research should be conducted on ecosystem-integrated studies in these areas.

The studies that have been conducted in relation to the agricultural sector have been mostly restricted to a single area of the country, the *Sabana de Bogotá*, a region that is home to approximately a sixth of the country's population. Studies have reported high levels of As in vegetables such as lettuce and cabbage. This information is very important because these are among the most consumed vegetables in Colombia, apart from onions and tomatoes. Only a single study (Alfaro et al., 2002) conducted in the same area was found that was related to the content of As in rice, the most widely consumed product in the country, and the As content found (0.58 mg/kg) exceeded the maximum recommended by the WHO (0.30 mg/kg). In that study, other substances considered to be pollutants, such as Cd and Pb, also were found in high concentrations in food samples, soils and irrigation water. Considering that As can be accumulated in organisms and that synergistic effects might occur when other pollutants are taken up simultaneously, the As concentrations in food samples collected in the *Sabana de Bogotá* constitute an alarming public health concern.

Studies of As in soils in the *Sabana de Bogotá* have typically shown low concentrations of the metalloid. Nevertheless, the considerable As accumulation in plants is associated with its presence in irrigation water taken from the Bogotá River. It has been suggested that the metalloid accumulates in the first soil horizon, where it can become bioavailable and thus enter plant tissue. The presence of As in the studied foodstuffs can therefore be attributed to irrigation water from the middle and lower basins of the Bogotá River and/or its tributaries. This area is highly contaminated by effluents from the industrial and urban areas of the city of Bogotá, as well as waste from leather processing facilities upstream. It is clear that the high levels of As detected represent a high risk of chronic As intoxication for more than 8 million people in the *Sabana de Bogotá*. However, there may be other sources of As contamination in vegetables that should be studied further. In another study (SICAT-14), conducted in the upper basin of the Bogotá River (Chía, Cota and Zipaquirá) where the influence of pollutants is much lower, an As concentration of 255 µg/L in irrigation water, which exceeds the level allowed by Colombian

legislation (100 µg/L), was reported. This water is from an underground source, which highlights the need to investigate more thoroughly the presence of As in groundwater resources in Colombia.

As for the presence of arsenic in the food chain and its effect on human health, although there have been only a few studies on this subject conducted in Colombia, these studies have reported alarmingly high levels of As in the studied areas (*Sibaté* and *Ricarte*): in human hair (up to 2078 µg/kg), strawberries (0.36 mg/kg), vegetables (up to 5.40 mg/kg), milk (up to 76 µg/L) and bovine meat (1.58 mg/kg). We therefore conclude that more studies on the presence of As in rural and urban areas should be performed in Colombia. An assessment of As occurrence in different areas of the country should be conducted to correlate As contents in foods with those in environmental matrices (e.g., water and soil), which can reveal the pathways of As exposure via the human food chain. The widespread presence of As in Colombia underscores the importance of these studies, as there might be people who are exposed to the metalloid but are unaware of their exposure because of the lack of consistent evaluation. In addition, the current governmental push to develop the country's mining activities might increase the population's exposure to toxic compounds such as As.

Another important aspect of the presence of As in Colombia and its environmental impact is the relationship between the total As concentration and the concentrations of As(III) and As(V) species, both organic and inorganic. No studies conducted in Colombia have reported on the chemical speciation of As, mainly because research in this area is very limited, as is the availability of the necessary specialized instrumentation, except for the SGC. Therefore, studies on the speciation of As are necessary to assess its danger and its environmental impact on human health.

Finally, this review highlights the need for more research focused on understanding the occurrence, origin, distribution, speciation and remediation of As in Colombia given the country's ongoing significant developments in industrial, mining and agricultural activities, all of which might lead to increased As pollution in Colombia.

Acknowledgments

We are grateful to the Department of Science, Technology and Innovation of Colombia, COLCIENCIAS, and the *Dirección de Investigación sede Bogotá, DIB*, of the *Universidad Nacional de Colombia* for their financial support of various research projects related to the study of toxic metals in the environment by the GERMINA research group (COLCIENCIAS Contrato RC No. 454-2011 and DIB projects No. 9140, 11973 and 12789). SL thanks Colciencias for a "Jóvenes Investigadores" scholarship.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/doi:10.1016/j.envpol.2013.12.009>.

References

- Allaro, R., García, E., Montenegro, O., 2002. Niveles de contaminación de mercurio, cadmio, arsénico y plomo en subsistemas de producción de la cuenca baja del Río Bogotá. *Rev. U.D.C.A. Actual. Div. Cient.* 4, 66–71. Spanish.
- Ángel, A., Fierro, J., 2012. Análisis y modelamiento del comportamiento de fluidos líquidos de pilas de escorbos en minería de oro. Universidad Nacional de Colombia, Departamento de Geociencias, Bogotá, Spanish.
- Barraquán, O., 2008. Estudio de diferentes metodologías para determinar la bio-disponibilidad de cadmio y arsénico en suelos y su relación con la concentración en plantas. *NOVA* 6 (9), 35–39. Spanish.
- Belk, P., Lesmes, L., 2011. Determinación de metales pesados en apio (*Apium graveolens* L.), lechuga (*Lactuca sativa* L. var. Batavia) y acelga (*Beta vulgaris* L.)

- mediante ICP-AES (Inductively Coupled Plasma Emission Spectrometry) en dos zonas de producción de hortalizas de la Sabana de Bogotá. Universidad Nacional de Colombia, Facultad de Agronomía, Bogotá, Spanish.
- Bundschuh, J., Pérez, A., Litter, M. (Eds.), 2008. Distribución del arsénico en la región Ibérica e Iberoamericana. CYTED, Buenos Aires, Spanish.
- Bundschuh, J., Armentia, M.A., Bhattacharya, P., Matschullat, J., Birke, P., Mukherjee, A.B. (Eds.), 2009. Natural Arsenic in Groundwater of Latin America – Occurrence, Health Impact and Remediation. CRC Press.
- Bundschuh, J., Litter, M., Parvez, F., Román-Ross, G., Nicolli, H., Jean, J., et al., 2012a. One century of arsenic exposure in Latin America: a review of history and occurrence from 14 countries. *Sci. Total Environ.* 429, 2–35.
- Bundschuh, J., Nath, B., Bhattacharya, P., Liu, C., Armentia, M., Moreno, M., et al., 2012b. Arsenic in the human food chain: the Latin American perspective. *Sci. Total Environ.* 429, 92–106.
- Cacua, I., 2004. Caracterización físico-química y microbiológica para aguas subterráneas en zona de influencia del Acueducto Metropolitano de Bucaramanga e implementación de los métodos de análisis para arsénico, selenio y flúor. Universidad Industrial de Santander, Facultad de Ciencias, Bucaramanga, Spanish.
- Callejas, D., 2007. Detección de arsénico de origen natural del agua subterránea en Colombia. Universidad de los Andes, Departamento de Ingeniería Ambiental, Bogotá, Spanish.
- DANE – Departamento Administrativo Nacional de Estadística, 2011. Estimación y proyección de población nacional, departamental y municipal total por área 1985–2020. DANE, Spanish. Available at: http://www.dane.gov.co/index.php?option=com_content&view=article&id=75&Itemid=72 (last accessed 02.12.13).
- Deschamps, E., Grinelli, V., Lange, F., Matschullat, J., Rauer, B., Schmidt, H., 2002. Soil and sediment geochemistry of the Iron Quadrangle, Brazil: the case of arsenic. *J. Soils Sediments* 2, 216–222.
- González, S., Mejía, L., 1995. Contaminación por cadmio y arsénico en suelos y hortalizas en un sector de la cuenca del Río Bogotá. *Rev. Suelos Ecuator.* 25, 51–56. Spanish.
- IARC – International Agency for Research on Cancer, 2004. Some drinking-water disinfectants and contaminants, including arsenic. In: IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, vol. 84. Available at: <http://monographs.iarc.fr/ENG/Monographs/vol84/mono84.pdf> (last accessed 02.12.13).
- INGEOMINAS – Instituto Colombiano de Geología y Minería, 1987. Recursos Minerales de Colombia, second ed. Tomo I. Publicaciones Geológicas Especiales de INGEOMINAS, Bogotá, Spanish.
- INGEOMINAS – Instituto Colombiano de Geología y Minería, 2004. Programa de Exploración de Aguas Subterráneas. INGEOMINAS, Spanish. Available at: <http://www.ingeominas.gov.co/getattachment/92fd1dc3-e4a6-4450-96b5-b19ab276144/Programa-exploracion-aguas-subterranas.aspx> (last accessed 02.12.13).
- Lage, C.R., Nayak, A., Kim, C., 2006. Arsenic ecotoxicology and innate immunity. *Integr. Comp. Biol.* 46 (6), 1040–1054.
- Lillo, J., 2003. Peligros Geoquímicos: arsénico de origen natural en las aguas. Universidad Complutense de Madrid, Facultad de Ciencias Geológicas, Grupo de Estudios de Minería y Medioambiente, Madrid, Spanish. Available at: http://www.ucm.es/info/crismine/Ambiente_Serena/Peligros_As_2.pdf (last accessed 02.12.13).
- Litter, M., Armentia, A., Fariás, S., 2009. Metodologías analíticas para la determinación y especiación de Arsénico en aguas y suelos. IBEROARSEN. CYTED, Buenos Aires, Spanish.
- Litter, M., Morgada, M., Bundschuh, J., 2010. Possible treatments for arsenic removal in Latin American waters for human consumption. *Environ. Pollut.* 158, 1105–1118.
- Mandal, B., Suzuki, K., 2002. Arsenic round the world: a review. *Talanta* 58, 201–235.
- McClintock, T., Chen, Y., Bundschuh, J., Oliver, J., Navoni, J., Olmos, V., et al., 2012. Arsenic exposure in Latin America: biomarkers, risk assessments and related health effects. *Sci. Total Environ.* 429, 76–91.
- Ministerio de la Protección Social y Ministerio de Ambiente, Vivienda y Desarrollo Territorial, 2007. Resolución 2115 de 2007. República de Colombia, Bogotá, Spanish.
- Ministerio de Agricultura, 1984. Decreto No. 1594 de 1984. República de Colombia, Bogotá, Spanish.
- Miranda, D., Carranza, C., Rojas, C., Jerez, C., Fischer, G., Zurita, J., 2008. Acumulación de metales pesados en suelo y plantas de cuatro cultivos hortícolas, regados con agua del río Bogotá. *Rev. Col. Cienc. Hort.* 2 (2), 180–191. Spanish.
- Muñoz, E., 2008. Determinación de arsénico por técnicas de Absorción Atómica en vegetales, suelos y aguas de riego. Universidad Nacional de Colombia, Departamento de Química, Bogotá, Spanish.
- Mutis, J., 1983. Catálogo de los yacimientos, prospectos y manifestaciones minerales de Colombia. INGEOMINAS, Bogotá, Spanish.
- Nicolli, H., 2006. Arsénico en aguas subterráneas de Latinoamérica: panorama y perspectivas. Instituto de Geoquímica, Centro de Investigaciones San Miguel, CONICET, Buenos Aires, Spanish. Available at: <http://www.cnea.gov.ar/xxi/ambiental/iberoarsen/docs/mesaredondamadelplata/PresentacionNicolli.pdf> (last accessed 02.12.13).
- Perdomo, L., Ensminger, L., Espinosa, L., Elster, C., Wallner, M., Schmetter, M., 1998. The mangrove ecosystem of the Genaga Grande de Santa Marta (Colombia): observations on regeneration and trace metals in sediment. *Mar. Pollut. Bull.* 37, 393–403.

- Ravenscroft, P., 2007. Predicting the Global Extent of Arsenic Pollution of Ground-water and its Potential Impact on Human Health. Final Report. UNICEF, New York.
- Samiento, M., Idrovo, A., Restrepo, M., 1999a. Determinación de arsénico total en cabello de individuos expuestos al embalse contaminado del Muña (Sibaté, Cundinamarca). *Rev. Fac. Med. UN* 47, 205–209. Spanish.
- Samiento, M., Idrovo, A., Restrepo, M., Díaz, M., González, A., 1999b. Evaluación del impacto del embalse del Muña sobre la salud humana. *Rev. Salud Pùb.* 1 (2), 157–171. Spanish.
- Secretaría Distrital de Ambiente, 2009. Resolución 3956 de 2009. Alcaldía Mayor de Bogotá. Bogotá. Spanish.
- Secretaría Distrital de Salud, 2000. Diagnostico de los niveles de contaminación por plaguicidas y metales pesados en los cultivos de hortalizas y trabajadores de los mismos localidad 7 Bosa. In: Convenio 1096/99 Hospital Pablo VI Bosa ESE. Secretaría Distrital de Salud, Bogotá.
- SICAT Database (The original references for the SICAT numeration [1 to 15] mentioned in this paper can be found in the online supplementary material, and were retrieved at the Servicio Geológico Colombiano (<http://www.sgc.gov.co>) from the database *Sistema de Información para el inventario, catalogación, valoración y administración de la información técnico-científica (SICAT)* – Information System for inventory, cataloging, assessment and management of technical and Scientific information, which can be accessed through the following web address: <http://aplicaciones1.ingominas.gov.co/sicat/html/consultaBasica.aspx>.
- Smith, A., Smith, M., 2004. Arsenic drinking water regulations in developing countries with extensive exposure. *Toxicology* 198, 39–44.
- US-EPA – United States Environmental Protection Agency, 2001. National Primary Drinking Water Regulations: Arsenic and Clarifications to Compliance and New Source Contaminants Monitoring. In: Final Rule 40 CFR Parts 9, 141, and 142. US-EPA, Washington D.C. Available at: <http://www.epa.gov/fedrgstr/EPA-WATER/2001/January/Day-22/w1668.htm> (last accessed 02.12.13.).
- Wang, H., Sthianopkao, S., Chen, Z., Man, Y., Du, J., Xing, G., et al., 2013. Arsenic concentration in rice, fish, meat and vegetables in Cambodia: a preliminary risk assessment. *Environ. Geochem. Health* 35, 745–755.
- WHO – World Health Organization, 2003. Arsenic in Drinking-water. Background Document for Preparation of WHO Guidelines for Drinking-water Quality (WHO/SDE/WSH/03.04/75). WHO, Geneva.
- WHO – World Health Organization, 2012. Proposed Draft Maximum Levels for Arsenic in Rice. Joint FAO/WHO Food Standards Programme Codex Committee on Contamination in Foods: Sixth Session Maastricht, The Netherlands, 26–30 March 2012, p. 16.
- WHO – World Health Organization, 1996. International Programs on Chemical Safety. Guidelines for Drinking Water Quality. Health Criteria and Other Supporting Information, second ed. Geneva. Available at: http://www.who.int/water_sanitation_health/dwq/gdwq2v1/en/index1.html (last accessed 02.12.13.).

2. Realización de un artículo sobre el desarrollo de este trabajo de investigación

A partir del presente trabajo de investigación se está realizando la elaboración de un artículo científico titulado: “*Presence of arsenic resistance genes in a colombian metagenome*”

3. Presentación oral en congreso internacional

Exposición de la ponencia titulada “**Arsenic resistance genes from a Colombian microbial metagenome**” en “V International Conference on Environmental, Industrial and Applied Microbiology – BioMicroWorld 2013” desarrollada en Madrid (España), los días 2, 3 y 4 de Octubre de 2013.

Arsenic resistance genes from a Colombian microbial metagenome

S. Latorre¹ and P. F. B. Brandão²

¹GERMINA Group, Institute of Biotechnology, Faculty of Sciences, Universidad Nacional de Colombia, Bogotá, Colombia

²GERMINA Group (Grupo de Estudios para la Remediación y Mitigación de Impactos Negativos al Ambiente), Department of Chemistry, Faculty of Sciences, Avenida Carrera 30 N° 45-03, Universidad Nacional de Colombia, Bogotá, Colombia

Arsenic is a metalloid associated with severe adverse effects on the environment and human health. Its presence in different countries such as Bangladesh and India has stimulated a global prevention interest from international organizations like the World Health Organization [1]. Recent studies alert about the presence of arsenic in Latin America, although the knowledge about its occurrence in Colombia is very scarce [2]. However, geological and environmental studies published at national level in this country reveal that arsenic distribution can be potentially large, particularly in mining areas and in regions surrounding contaminated river basins, which in some cases report arsenic concentrations exceeding limits established by law [3].

This work aimed to search genes involved in arsenic resistance in the metagenome of the Bogotá River, Colombia, in order to provide knowledge on the existence of possible different metabolic interactions with this metalloid in this environment. DNA from soil, sediment and water samples of the middle basin of the Bogotá River were purified in order to make a fosmid based library of more than 50,000 clones. DNA fragments of about 40 Kb were cloned in pCC2FOS vector in order to obtain coverage of 2,000 Mb of the metagenome samples. Functional screenings were performed to evaluate the phenotypic response of the clones to sodium arsenate concentrations of 20 and 50 mM. Nine clones showed a positive response by growing in 20 mM arsenate and two clones could grow up to 50 mM concentration. Fosmid DNA of the clones was purified and

a mixed Sanger and Pyrosequencing strategy was performed in order to obtain the sequences of the metagenomic fragments. Further, Open Reading Frames were searched to identify candidate genes for the arsenic resistance activity and those regions were compared with genomic public databases. Potentially new mechanisms and genetic elements of metabolic interaction with arsenic have been found in the microbial metagenome of the Bogotá River. The newly described genes could be useful for future biotechnological applications such as arsenic bioremediation strategies or the development of biosensor devices to detect this contaminant in the environment.

Keywords: Arsenic; resistance genes; metagenomics; Colombia.

References

- [1] Litter MI, Morgada ME, Bundschuh J. (2010). Possible treatments for arsenic removal in Latin American waters for human consumption. *Environmental Pollution* 158: 1105-1118. doi: 10.1016/j.envpol.2010.01.028
- [2] Bundschuh J, Litter MI, Parvez F, Román-Ross G, Nicolli HB, Jean JS, Liu CW, López D, Armienta MA, Guilherme LR, Cuevas AG, Cornejo L, Cumbal L, Toujaguez R. (2012). One century of arsenic exposure in Latin America: a review of history and occurrence from 14 countries. *Science of the Total Environment* 429: 2-35. doi: 10.1016/j.scitotenv.2011.06.024
- [3] Alonso DL, Latorre S, Castillo E, Brandão PFBB (2013) Arsenic in Latin America: Colombia (submitted for publication - July 2013)



4. Presentación de poster en congreso internacional

Presentación de poster titulado “Búsqueda de genes de resistencia a arsénico en el metagenoma microbiano del Río Bogotá” en “V Congreso Internacional de Microbiología Industrial”, desarrollado en Bogotá los días 8, 9 y 10 de Mayo de 2013. A continuación se presenta el resumen consignado en el libro de memorias.

MNA P13. Búsqueda de genes de resistencia a Arsénico en el metagenoma microbiana del río Bogotá

Latorre Orchoa Sergio Mauricio¹, De Brito Brandao Pedro Filipe¹

¹Universidad Nacional de Colombia

Correo electrónico: smiatorreo@gmail.com

El arsénico es un metaloide asociado a fuertes efectos nocivos sobre el medio ambiente y al ser humano. Su presencia en distintos países incentivó un interés mundial de prevención desde organismos internacionales como la Organización Mundial para la Salud.

Recientes trabajos alertan de su presencia en América Latina, aunque en Colombia poco se conoce sobre su problemática. Estudios geológicos y ambientales muestran que las dimensiones pueden ser grandes, pues en zonas de explotación minera y en zonas aledañas a cuencas de ríos contaminados, existen concentraciones que superan las permitidas por la legislación.

En este trabajo se realizó una búsqueda de genes involucrados en la resistencia a arsénico en muestras metagenómicas del río Bogotá.

Se realizaron extracciones de ADN de muestras de suelos, sedimentos y aguas recolectadas de la cuenca media del río y a partir de ellas se construyeron librerías metagenómicas (>50.000 clones).

Se realizaron tamizajes funcionales evaluando el crecimiento de los clones en concentraciones hasta 50mM de arsenato de sodio. Nueve clones mostraron crecimiento a 20mM arsénico y dos a 50mM. Utilizando iniciadores específicos se detectó la presencia en estos clones de genes que codifican para la bomba de expulsión de arsénico, ArsB.

La secuenciación de los fragmentos insertados en los clones con resistencia a arsénico permitirá determinar si son nuevos elementos genéticos. Se pretende con el trabajo aportar conocimiento sobre la existencia de diferentes interacciones metabólicas con este metaloide con el ánimo de sentar bases para futuros desarrollos biotecnológicos en miras a la remediación de ambientes contaminados.

5. Aceptación de presentación oral en congreso nacional

Aceptación a la realización de presentación oral de la ponencia titulada “Detección por metagenómica de genes de resistencia a arsénico en el Río Bogotá” en el “Primer Congreso Colombiano de Bioquímica y Biología Molecular – C2B2”, que se llevará a cabo en Bogotá.



Bogotá, D.C., 03 de mayo de 2014.

Autor (a):
Sergio Mauricio Latorre Ochoa
Universidad Nacional de Colombia

Su trabajo *Detección por metagenómica de genes de resistencia a arsénico en el río bogotá* ha sido **ACEPTADO** para participar en el **Primer Congreso Colombiano de Bioquímica y Biología Molecular – C2B2**, el cual se llevará a cabo en el Auditorio Alfonso López Pumarejo de la Universidad Nacional de Colombia, sede Bogotá, entre el 4 y el 7 de junio de 2014.

La modalidad de presentación de su trabajo será **ORAL** en el área **Bioquímica y Biología Molecular Ambiental**. Por esta razón y pretendiendo el desarrollo exitoso del Congreso, le solicitamos confirmar su asistencia al evento a más tardar el próximo martes 6 de mayo. Adicionalmente, requerimos su contacto telefónico para ultimar los detalles de la presentación.

Finalmente, le sugerimos tramitar su inscripción al C2B2 a la mayor brevedad posible para ingresar su resumen a las Memorias del Congreso.

Cordialmente,

Moisés Wasserman
Presidente Comité Científico
Primer Congreso Colombiano de Bioquímica y Biología Molecular
Http: www.c2b2.com.co

DETECCIÓN POR METAGENÓMICA DE GENES DE RESISTENCIA A ARSÉNICO EN EL RÍO BOGOTÁ

Sergio M. Latorre Ochoa¹. Pedro F. B. Brandão².

¹ Grupo de investigación GERMINA, Facultad de Ciencias, Universidad Nacional de Colombia. smlatorreo@gmail.com

² Grupo de investigación GERMINA, Facultad de Ciencias, Universidad Nacional de Colombia. pfdeb@unal.edu.co

El arsénico es un metaloide asociado a impactos negativos al ambiente y a la salud humana. Es considerado por instituciones como la Agencia Internacional para la Investigación del Cáncer (IARC), como carcinogénico para los seres humanos.

En países como Bangladesh e India, los niveles de contaminación por este metaloide han hecho de este problema una de las principales preocupaciones de salud pública. En América Latina, se tienen registros de niveles tóxicos en la región del cono sur. Sin embargo, en Colombia, es aún muy reciente la publicación de trabajos de revisión que alertan sobre la potencialidad de riesgo que existe sobre el país, en el que no existen medidas preventivas ni regulatorias para enfrentar los riesgos potenciales de contaminación.

En el presente trabajo se realizó una búsqueda de genes de resistencia a arsénico en el metagenoma microbiano de la Sabana de Bogotá, con especial énfasis en la cuenca media del Río Bogotá, por ser esta zona un significativo depósito de contaminantes provenientes de la actividad industrial, agrícola y doméstica.

Partiendo de muestras ambientales de suelos, sedimentos y aguas, se realizaron extracciones de ADN metagenómico. Con la utilización del vector de clonación *pCC2FOS* y de la célula hospedera *E. coli* *EPI-300*, se realizó la generación de una librería de aproximadamente 50.000 clones con una cobertura aproximada de 2000 Mb.

Con el fin de identificar expresión fenotípica de resistencia a arsénico, se implementó un barrido funcional con concentraciones de 20 mM y 50 mM de arsénico, que arrojaron como resultado once clones con resistencia al metaloide.

Una estrategia de pirosecuenciación permitió identificar las secuencias de los clones resistentes. Con la ayuda de algoritmos y paquetes bioinformáticos se logró la identificación de los genes *arsA*, *arsB*, *arsC*, *arsD* y *arsR* del operón encargado de la reducción del arsenato.

Con la identificación de estos elementos genéticos, actualmente se realizan experimentos conducentes a la implementación biotecnológica mediante la creación de biosensores que puedan ser implementados a futuro en diferentes ecosistemas con el fin de generar mecanismos de detección de arsénico sencillos y de bajo costo.

Palabras clave: Arsénico, metagenómica, Río Bogotá, bioprospección.

Área temática: Ambiental.

Bibliografía

Achour, A., Bauda, P., Billard, P. 2007. Diversity of arsenite transporter genes from arsenic-resistant soil bacteria. *Res. Microbiol.* 158:128-137

Alfaro, R., García, E., Montenegro, O. 2002. Niveles de contaminación de mercurio, cadmio, arsénico y plomo en subsistemas de producción de la cuenca baja del Río Bogotá. *Rev. U.D.C.A. Actual Div Cient.* 4:66-71

Alonso, David. 2014. "Determinación de arsénico (As) total en muestras ambientales mediante análisis por activación neutrónica (AAN) y absorción atómica con generador de hidruros (AAS-GH)". Tesis de Maestría (No publicada). Facultad de Ciencias. Universidad Nacional de Colombia.

Alonso, D., Latorre, S., Castillo, E., Brandão, P. 2013. Environmental occurrence of arsenic in Colombia: a review. *Environ Pollut.* 186:272-281

Altschul, S., Madden, T., Schäffer, A., Zhang, Z., Miller, W., Lipman, D. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389-3402

Amann, R., Ludwig, W., Schleifer, K. 1995. Phylogenetic identification *in situ* detection of individual microbial cells without cultivation. *Microbial Rev.* 59:143-169

Bachate, S., Cavalca, L., Andreoni, V. 2009. Arsenic-resistant bacteria isolated from agricultural soils of Bangladesh and characterization of arsenate-reducing strains. *J Appl Microbiol.* 107:145-156

Bansal N., Sinha I., Viridi, J. 2000. Arsenic and cadmium resistance in environmental isolates of *Yersinia enterocolitica* and *Yersinia intermedia*. *Can J Microbiol.* 46:481-484

Barragán, O. 2008. Estudio de diferentes metodologías para determinar la biodisponibilidad de cadmio y arsénico en suelos y su relación con la concentración en plantas. NOVA. 6:35-39

Bateman, A., Coin, L., Durbin, R., Finn, R., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. 2004. The Pfam protein families database. Nucleic Acids Res. 32:D138-D141

Bundschuh, J., Armienta, M., Bhattacharya, P., Matschullat, J., Birkle, P., Mukherjee, A. (Eds). 2009. Natural arsenic in groundwater of Latin America occurrence, health impact and remediation. CRC Press.

Bundschuh, J., Litter, M., Parvez, F., Román-Ross, G., Nicolli, H., Jean, J., *et al.* 2012. One century of arsenic exposure in Latin America: a review of history and occurrence from 14 countries. Sci Total Environ. 429:2-35

Bundschuh, J., Pérez, A., Litter, M. (Eds.). 2008. Distribución del arsenic en la región Ibérica e Iberoamericana. CYTED, Buenos Aires.

Butcher, B., Deane, S., Rawlings, D. 2000. The chromosomal arsenic resistance genes of *Thiobacillus ferrooxidans* have an unusual arrangement and confer increased arsenic and antimony resistance to *Escherichia coli*. Appl Environ Microbiol. 66:1826-1833

Cai, J., Salmon, K., Dubow, M. 1998. A chromosomal *ars* operon homologue of *Pseudomonas aeruginosa* confers increased resistance to arsenic and antimony in *Escherichia coli*. Microbiology 144:2705-2729

Calcagno, T., Mohn, W., de Lorenzo, V. 2005. Exploring the microbial biodegradation and biotransformation gene pool. Trends in Biotechnol. 23:497-506

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T. 2009. BLAST+: architecture and applications. BMC Bioinformatics. 10:421

Carlin, A., Shi, W., Dey, S., Rosen, B. 1995. The *ars* operon of *Escherichia coli* confers arsenical and antimonial resistance. *J Bacteriol.* 177:981-986

Carrillo, K. Identificación de genes involucrados en la transformación y resistencia a arsenic en microorganismos recuperados de zonas de Colombia con la presencia del metal. 2012. Tesis de Maestría. Facultad de Ciencias. Universidad Nacional de Colombia.

Chauhan, N., Ranjan, R., Purohit, H., Kalia, V., Sharma, R. 2009. Identification of genes conferring arsenic resistance to *Escherichia coli* from an effluent treatment plant sludge metagenomic library. *FEMS Microbiol Ecol.* 67:130-139

Chevreux, B., Pfisterer, T., Drescher, B., Driesel, A., Muller, W., Wetter, T., Suhai, S. 2004. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequences ESTs. *Genome Res.* 14:1147-1159

Chevreux, B., Wetter, T., Guhai, S. 1999. Genome sequence assembly using signals and additional sequence information. En: *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB).* 99:45-56

Cole, J., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R., Kulam-Syed-Mohideen, A., McGarrel, D., Marsh, T., Garrity, G. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* 37:D141-D145

Corsini, A., Cavalca, L., Crippa, L., Zaccheo, P., Andreoni, V. 2010. Impact of glucose on microbial community of a soil containing pyrite cinders: Role of bacteria in arsenic mobilization under submerged condition. *Soil Biol Biochem.* 42:699-707

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. *Environ Microbiol.* 3:431-439

Daniel, R. 2004. The soil metagenome – a rich resource for the discovery of novel natural products. *Current Opin Biotech.* 15:199-204

Desai, N., Antonopoulos, D., Gilbert, J., Glass, E., Meyer, F. 2012. From genomics to metagenomics. *Current Opin Biotech.* 23:72-76

Di Bella, J., Bao, Y., Gloor, G., Burton, J., Reid, G. 2013. High throughput sequencing methods and analysis for microbiome research. *J Microbiol Meth.* 95:401-414

Diesel, E., Schreiber, M., van der Meer, J. 2009. Development of bacteria-based bioassays for arsenic detection in natural waters. *Anal Bioanal Chem.* 394:687-693

Diorio, C., Cai, J., Marmor, J., Shinder, R., Dubow, M. S. 1995. An *Escherichia coli* chromosomal *ars* operon homolog is functional in arsenic detoxification and is conserved in gram-negative bacteria. *J Bacteriol.* 177:2050-2056

Dopson, M., Lindstrom, E., Hallberg, K. 2001. Chromosomally encoded arsenical resistance of the moderately thermophilic acidophile *Acidithiobacillus caldus*. *Extremophiles.* 5:247-255

Drewniak, L., Matlakowska, R., Rewerski, B., Sklodowska, A. 2010. Arsenic release from gold mine rocks mediated by the activity of indigenous bacteria. *Hydrometallurgy* 104:437-442

Ekström, J., Arroyo, L., Bzhalava, D., Söderlund-Strand, A., Hultin, E., Nordin, P., Stenquist, B., Paoli, J., Forslund, O., Dillner, J. 2013. Diversity of human papillomaviruses in skin lesions. *Virology.* 447:300-311

El Allali, A., Rose, J. 2013. MGC: a metagenomic gene caller. *BMC Bioinforma.* 14:S6-S6

Ewing, B., Green, P. 1998. Base-calling of automated sequencer traces using *Phred. II.* error probabilities. *Genome Res.* 8:186-194

Eyers, L., George, I., Schuler, L., Stenuit, B., Agathos, S., El Fantroussi, S. 2004. Environmental genomics: exploring the unmined richness of microbes to degrade xenobiotics. *Appl Microbiol Biot.* 66:123-130

Frechh, C., de Mora, K., Joshi, N., Elfick, A., Haseloff, J., Ajioka, J. 2011. A5 Synthetic biology and the art of biosensor design. En: Choffnes, E. (coord). The science and applications of synthetic and systems biology: workshop summary. The National Academies Press. Washington, D.C.

Gadd, G. 2000. Bioremedial potential of microbial mechanisms of metal mobilization and immobilization. *Curr Opin Biotechnol.* 11:271-279

GPLv2. 1991. GNU General Public License, version 2. <http://www.gnu.org/licenses/gpl-2.0.html>

Handelsman, J. 2004. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol. R.* 68:669-685

Hedges, R., Baumberg, S. 1973. Resistance to arsenic compounds conferred by a plasmid transmissible between strains of *Escherichia coli*. *J Bacteriol.* 115:459-460

Hoff, K., Lingner, T., Meinicke, P., Tech, M. 2009. Orphelia: predicting genes in metagenomic sequencing reads. *Nucleic Acids Res.* 37:W101-W105

Hook, S., Twine, N., Simpson, S., Spadaro, D., Moncuquet, P., Wilkins, M. 2014. 454 pyrosequencing-based analysis of gene expression profiles in the amphipod *Melita plumulosa*: Transcriptome assembly and toxicant induced changes. *Aquat.Toxicol.* <http://dx.doi.org/10.1016/j.aquatox.2013.11.022>

Huson, D., Auch, A., Qi, J., Schuster, S. 2007. MEGAN analysis of metagenomic data. *Genome Res.* 17:377-386

Hurt, R., Qiu, X., Wu, L., Roh, Y., Palumbo, A., Tiedje, J., Zhou, J. 2001. Simultaneous recovery of RNA and DNA from soils and sediments. *Appl Environ Microb.* 67:4495-4503

Jackson, C., Dugas, S. 2003. Phylogenetic analysis of bacterial and archeal *arsC* gene sequences suggest an ancient, common origin for arsenate reductase. *BMC Evol Biol.* 3:18-28

Kanehisa, M., Goto, S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28:27-30

Kelley, D., Liu, B., Delcher, A., Pop, M., Salzberg, S. 2012. Gene prediction with Glimmer for metagenomic sequences augmented by classification and clustering. *Nucleic Acids Res.* 40:e9-e9

Kim, Y., Kwon, E., Kim, S., Jeong, Y., Kim, J., Yun, H., Kim, H. 2010. Molecular cloning and characterization of a novel family VIII alkaline esterase from a compost metagenomic library. *Biochem Bioph Res Co.* 393:45-49

Koren, S., Treangen, T., Pop, M. 2011. Bambus 2: scaffolding metagenomics. *Bioinformatics.* 27:2964-2971

Kostal, J., Yang, R., Wu, C.H., Mulchandani, A., Chen, W. 2004. Enhanced arsenic accumulation in engineered bacterial cells expressing *ArsR*. *Appl Environ Microbiol.* 70:4582-4587

Kotrba, P., Dolecková, L., de Lorenzo, V., Ruml, T. 1999. Enhanced bioaccumulation of heavy metal ions by bacterial cells due to surface display of short metal binding peptides. *Appl Environ Microbiol.* 65:1092-1098

Kunin, V., Engelbrektsen, A., Ochman, H., Hugenholtz, P. 2010. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ Microbiol.* 12:118-123

Lai, B., Ding, R., Li, Y., Duan, L., Zhu, H. 2012. A *de novo* metagenomic assembly program for shotgun DNA reads. *Bioinformatics.* 28:1455-1462

Langer, M., Gabor, E., Liebeton, K., Meurer, G., Niehaus, F., Schulze, R., Eck, J., Lorenz, P. 2006. Metagenomics: An inexhaustible access to nature's diversity. *Biotechnol J.* 1:815- 821

Laserson, J., Jojic, V., Koller, D. 2011. *Genovo: de novo* assembly for metagenomes. *J Comput Biol.* 18:429-443

Li, Y., Wang, J., Luan, Z., Liang, Z. 2010. Arsenic removal from aqueous solution using ferrous based red mud sludge. *J Hazard Mater.* 177:131-137

Litter, M., Morgada, M., Bundschuh, J. 2010. Possible treatments for arsenic removal in Latin American waters for human consumption. *Environ Pollut.* 158:1105-1118

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., Law, M. 2012. Comparison of next-generation sequencing systems. *J Biomed Biotechnol.* 2012:251364

Lloyd-Jones, G., Hunter, D. 2001. Comparison of rapid DNA extraction methods applied to contrasting New Zealand soils. *Soil Biol Biochem.* 33:2053–2059

Mandal, B., Morgada, M., Bundschuh, J. 2010. Possible treatments for arsenic removal in Latin American waters for human consumption. *Environ Pollut.* 158:1105-1118

Marchler-Bauer, A., Anderson, J., Cherukuri, P., DeWeese-Scott, C., Geer, L., Gwadz, M., He, S., Hurwitz, D., Jackson, J., Ke, Z. 2005. CDD: a conserved domain database for protein classification. *Nucleic Acids Res.* 33:D192-D196

Marchler-Bauer, A., Bryant, S. 2004. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.* 32:W327-331

Markowitz, V., Ivanova, N., Szeto, E., Palaniappan, K., Chu, K., Dalevi, D., Chen, I., Grechkin, Y., Dubchak, I., Anderson, I. 2008. IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res.* 36:D534-D538

Mateos, L., Ordóñez, E., Letek, M., Gil, J. 2006. *Corynebacterium glutamicum* as a model bacterium for the bioremediation of arsenic. *Int Microbiol.* 9:207-215

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., Edwards, R. 2008. The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics.* 9:386-386

Miranda, D., Carranza, C., Rojas, C., Jerez, C., Fischer, G., Zurita, J. 2008. Acumulación de metales pesados en suelo y plantas de cuatro cultivos hortícolas, regados con agua del río Bogotá. *Revista Colombiana de Ciencias Hortícolas.* 2:180-191

Mirza, N., Mahmood, Q., Pervez, A., Ahmad, R., Farooq, R., Shah, M., Azim, M. 2010. Phytoremediation potential of *Arundo donax* in arsenic-contaminated synthetic wastewater. *Bioresour Technol.* 101:5815-5819

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A., Kanehisa, M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 30:W182-W185

Namiki, T., Hachiya, T., Tanaka, H., Sakakibara, Y. 2012. MetaVelvet: an extension of Velvet assembler to *de novo* metagenome assembly from short sequence reads. *Nucleic Acids Res.* 40:e155-e155

Nicolli, H. 2006. Arsénico en aguas subterráneas de Latinoamérica: panorama y perspectivas. Instituto de Geoquímica, Centro de Investigaciones San Miguel. CONICET, Buenos Aires.

Noguchi, H., Taniguchi, T., Itoh, T. 2008. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res.* 15:387-396

Novick, R., Roth, C. 1968. Plasmid-like resistance to inorganic salts in *Staphylococcus aureus*. *J Bacteriol.* 95:1135-1342

Nowrousian, M. 2010. Next-generation sequencing techniques for eukaryotic microorganisms: sequencing-based solutions to biological problems. *Eukaryot Cell*. 9: 1300-13010

Ordóñez, E., Letek, M., Valbuena, N., Gil, J., Mateos, L. 2005. Analysis of genes involved in arsenic resistance in *Corynebacterium glutamicum* ATCC 13032. *Appl Environ Microbiol*. 71:6206-6215

Overbeek, R., Begley, T., Butler, R., Choudhuri, J., Chuang, H., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R. 2005. The subsystem approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res*. 33:5691-5702

Páez-Espino, D., Tamames, J., de Lorenzo, V., Cánovas, D. 2009. Microbial responses to environmental arsenic. *Biometals* 22:117-130

Peng, Y., Leung, H., Yiu, S., Chin, F. 2011. Meta-IDBA: a *de novo* assembler for metagenomic data. *Bioinformatics*. 27:i94-i101

Phan, K., Sthiannopkao, S., Kim, K., Wong, M., Sao, V., Hashim, J., Yasin, M., Aljunid, S. 2010. Health risk assessment of inorganic arsenic intake of Cambodia residents through groundwater drinking pathway. *Water Res*. 44:5777-5788

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zhen, H, Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Jian, M., Zhou, Y., Li, Y., Zhang, X., Qin, N., Yang, H., Wang, J., Brunak, S., Dore, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Bork, P., Ehrlich, S. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 464: 59-65

Ravenscroft, P. 2007. Predicting the global extent of arsenic pollution of groundwater and its potential impact on human health. Final report. UNICEF, New York.

Rho, M., Tang, H., Ye, Y. 2010. FragGeneScan: predicting genes in short error-prone reads. *Nucleic Acids Res.* 38:e191-e191

Rice, P., Longden, I., Bleasby, A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* 16:276-277

Rusk, N. 2011. Torrents of sequence. *Nat Methods.* 8:44

Sagar, K., Singh, S., Goutam, K., Konwar, B. 2014. Assessment of five soil DNA extraction methods and a rapid laboratory-developed method for quality soil DNA extraction for 16s rDNA-based amplification and library construction. *J Microbiol Methods.* 97:68-73

Sambrook, J., Russell, D. 2001. *Molecular Cloning. A laboratory manual.* Third Edition. Cold Spring Harbor, New York

Sanger, F., Air, G., Barrel, B., Brown, N., Coulson, A., Fiddes, C., Hutchison, C., Slocombe, P., Smith, M. 1997. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature.* 256: 687-695

Santosa, D. 2001. Rapid extraction and purification of environmental DNA for molecular cloning applications and molecular diversity studies. *Mol Biotechnol.* 17:59–64

Sauge-Merle, S., Cui n , S., Carrier, P., Lecomte-Pradines, C., Luu, D.T., Peltier, G. 2003. Enhanced toxic metal accumulation in engineered bacterial cells expressing *Arabidopsis thaliana* phytochelatin synthase. *Appl Environ Microbiol.* 69:490-494

Scholz, M., Lo, C., Chain, P. 2012. Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Current Opin Biotech.* 23:9-15

Schmoger, M.E., Oven, M., Grill, E. 2000. Detoxification of arsenic by phytochelatin in plants. *Plant Physiol.* 122:793-801

Secretaria Distrital de Salud, Hospital Pablo VI Bosa ESE. 2000. Diagnóstico de los niveles de contaminación por plaguicidas y metales pesados en los cultivos de hortalizas y trabajadores de los mismos, Localidad 7 Bosa, Bogotá, D.C.. [en línea] <http://190.25.230.149:8080/dspace/bitstream/123456789/438/13/NIVELES%20DE%20PLAGUICIDAS%20Y%20METALES%20PESADOS%20BOSA.pdf> (Consulta: 27 Septiembre 2010)

Selengut, J., Haft, D., Davidsen, T., Ganapathy, A., Gwinn-Giglio, M., Nelson, W., Richter, A., White, O. 2007. TIGRFAMs and genome properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. *Nucleic Acids Res.* 35:D260-D264

Shah, D., Shen, M., Chen, W., Da Silva, N. 2010. Enhanced arsenic accumulation in *Saccharomyces cerevisiae* overexpressing transporters Fps1p or Hxt7p. *J. Biotechnol.* 150: 101-107

Silver, S., Budd, K., Leahy, K., Shaw, W., Hammond, D., Novick, R., Willsky, M., Malamy., Rosenberg, H. 1981. Inducible plasmid-determined resistance to arsenate, arsenite, and antimony (III) in *Escherichia coli* and *Staphylococcus aureus*. *J Bacteriol.* 146:983-996

Silver, S., Phung, L. 1996. Bacterial heavy metal resistance: new surprises. *Ann Rev Microbiol.* 50:753-789

Silver, S., Phung, L. 2005. Genes and enzymes involved in bacterial oxidation and reduction of inorganic arsenic. *Appl Env Microbiol.* 71:599-608

Simeonova, D., Lièvreumont, D., Lagarde, F., Muller, D., Groudeva, V., Lett, M. 2004. Microplate screening assay for the detection of arsenite-oxidizing and arsenate-reducing bacteria. *FEMS Microbiol Lett.* 237:249-253

Simpson, J., Wong, K., Jackman, S., Schein, J., Jones, S., Birol, I. 2009. AbySS: a parallel assembler for short read sequence data. *Genome Res.* 19:1117-1123

Smith, A., Smith, M. 2004. Arsenic drinking water regulations in developing countries with extensive exposure. *Toxicology.* 198:39-44

Stenuit, B., Eysers, L., Schuler, L., Agathos, S., George, I. 2008. Emerging high-throughput approaches to analyze bioremediation of sites contaminated with hazardous and/or recalcitrant wastes. *Biotechnol Adv.* 26:561-575

Stolz, J., Basu, P., Santini, J., Oremland, R. 2006. Arsenic and selenium in microbial metabolism. *Annu Rev Microbiol.* 60:107-130

Strohl, W. 1992. Compilation and analysis of DNA sequences associated with apparent streptomycete promoters. *Nucleic Acids Res.* 20:961-974

Sullivan, C., Tyrer, M., Cheeseman, C., Graham, N.J.D. 2010. Disposal of water treatments wastes containing arsenic - A review. *Sci Total Environ.* 408:1770-1778

Sun, S., Chen, J., Li, W., Altintas, I., Lin, A., Peltier, S., Stocks, K., Allen, E., Ellisman, M., Grethe, J. 2011. Community cyberinfrastructure for advanced microbial ecology research and analysis: the CAMERA resource. *Nucleic Acids Res.* 39:D546-D551

Susuki, K., Wakao, N., Kimura, T., Sakka, K., Ohmiya, K. 1998. Expression and regulation of the arsenic resistance operon of *Acidiphilium multivorum* AIU 301 Plasmid pKW301 in *Escherichia coli*. *Appl Environ Microbiol.* 64:411

Tatusov, R., Fedorova, N., Jackson, J., Jacobs, A., Kiryutin, B., Koonin, E., Krylov, D., Mazumder, R., Mekhedov, S., Nikolskaya, A. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics.* 4:41-41

Taupp, M., Mewis, K., Hallam, S. 2011. The art and desing of functional metagenomic screens. *Current Opin Biotech.* 22:465-472

Turnbaugh, P., Gordon, J. 2008. An invitation to the marriage of metagenomics and metabolomics. *Cell*. 134:708-713

Valls, M., de Lorenzo, V. 2002. Exploiting the genetic and biochemical capacities of bacteria for the remediation of heavy metal pollution. *FEMS Microbiol Rev*. 26:327-338

Villandagos, A.F., Ordóñez, E., Muñoz, M.I., Pastrana, I.M., Fiuza, M., Gil, J.A., Mateos, L.M., Aller, A.J. 2010. Retention of arsenate using genetically modified coryneform bacteria and determination of arsenic in solid samples by ICP-MS. *Talanta*. 80:1421-1427

Vullo, D.L. 2003. Microorganismos y metales pesados: una interacción en beneficio del medio ambiente. *Revista Química Viva*, volumen 2, número 3 [en línea] <http://www.quimicaviva.qb.fcen.uba.ar/Actualizaciones/metales/metales.html> (Consulta: Octubre 2010)

Wang, S., Wang, K., Li, L., Liu, Y. 2013. Isolation and characterization of a novel organic solvent-tolerant and halotolerant esterase from soil metagenomic library. *J Mol Catal B-Enzym*. 95:1-8

WHO & IARC. WORLD HEALTH ORGANIZATION and INTERNATIONAL AGENCY FOR RESEARCH ON CANCER. 1998. IARC Monographs on the evaluation of carcinogenic risks to humans. Volume 23. Some Metals and Metallic Compounds

WHO & IARC. WORLD HEALTH ORGANIZATION and INTERNATIONAL AGENCY FOR RESEARCH ON CANCER. 2004. IARC Monographs on the evaluation of carcinogenic risks to humans. Volume 84. Some Drinking-water Disinfectants and Contaminants, including Arsenic

Xing, M., Zhang, X., Huang, H. 2012. Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis. *Biotechnol Adv*. 30:920-929

Yan, L., Yin, H., Zhang, S., Leng, F., Nan, W., Li, H. 2010. Biosorption of inorganic and organic arsenic from aqueous solution by *Acidithiobacillus ferrooxidans* BY-3. *J Hazard Mater*. 178:209-217

Yao, J., Chen, Q., Shen, A., Cao, W., Liu, Y. A novel feruloyl esterase from a soil metagenomic library with tannase activity. *J Mol Catal B-Enzym.* 95:55-61

Zhang, Z., Schwartz, S., Wagner, L., Miller, W. 2000. A greedy algorithm for aligning DNA sequences. *J Comput Biol.* 7:203-214

Zhou, Y., Yao, J., He, M., Choi, M., Feng, L., Chen, H., Wang, F., Chen, K., Zhuang, R., Maskow, T., Wang, G., Zaray, G. 2010. Reduction in toxicity of arsenic(III) to *Hallobacillus* sp. Y35 by kaolin and their related adsorption studies. *J Hazard Mater.* 176:487-494

Zhu, W., Lomsadze, A., Borodovsky, M. 2010. *Ab initio* gene identification in metagenomic sequences. *Nucleic Acids Res.* 38:e132-e132

Zipper, H., Buta, C., Lämmle, K., Brunner, H., Bernhagen, J., Vitzthum, F. 2003. Mechanisms underlying the impact of humic acids on DNA quantification by SYBR Green I and consequences for the analysis of soils and aquatic sediments. *Nucl Acids Res.* 31:e39