



Ministerio
de Relaciones
Laborales

Registro Oficial N°. 11 del 10 de junio del 2013

INFORME DE VIÁTICOS, MOVILIZACIONES, SUBSISTENCIAS Y ALIMENTACIÓN AL EXTERIOR

FECHA DE INFORME (dd-mmm-aaaa)

13-ene-16

DATOS GENERALES

APELLIDOS-NOMBRES DE LA O EL SERVIDOR:

Leda Carolina Restrepo Cardona

PUESTO:

Especialista en Biología Molecular

CIUDAD - PROVINCIA DE LA COMISION

Bogota Colombia

NOMBRE DE LA UNIDAD DE LA O EL SERVIDOR

Cenalm-Espol

INFORME DE ACTIVIDADES Y PRODUCTOS ALCANZADOS (Enumerar, según Art. 23)

curso de actualización en técnicas moleculares (preparación de librerías para secuenciación masiva) para el proyecto "Desarrollo e implementación de métodos de control y prevención de enfermedades en especies acuáticas de uso comercial y uso potencial en maricultura o repoblación", del Centro de costo No. 2135-50-03.

ITINERARIO	SALIDA / CASA O TRABAJO	LLEGADA / CASA O	NOTA			
FECHA dd-mmm-aaaa	16-nov-15	27-dic-15	Estos datos se refieren al tiempo efectivamente utilizado en el comisión, desde la salida del lugar de residencia o trabajo habituales o del cumplimiento de la licencia según sea el caso, hasta su llegada de estos sitios.			
HORA hh:mm	8:00	17:00				
HORA Inicio de Labores el día de retorno						
TRANSPORTE UTILIZADO			SALIDA	LLEGADA		
TIPO DE TRANSPORTE (Aéreo, terrestre, otros)	NOMBRE DEL TRANSPORTE	RUTA	FECHA dd-mmm-aaaa	HORA hh:mm	FECHA dd-mmm-aaaa	HORA hh:mm
Aereo	Avianca	Guayaquil-Bogota	16-nov-15	8:00	27-dic-15	17:00

NOTA: En caso de haber utilizado: 1) transporte público, aéreo, fluvial o terrestre, se deberá adjuntar obligatoriamente los pases a bordo o boletos, de acuerdo a lo que establece en artículo N°. 23 y 2) vehículos institucionales, se adjuntará la hoja de ruta con tipo de vehículo, número de placa, kilometraje recorrido y los nombres apellidos del conductor.

OBSERVACIONES

FIRMA SERVIDOR/A COMISIONADO

Leda Restrepo Cardona

NOTA

El presente informe deberá presentarse dentro del término máximo de 4 días de cumplida la licencia, caso contrario la liquidación se demorará e incluso de no presentarlo tendría que restituir los valores pagados. Cuando la licencia sea superior al número de horas o días autorizados, se deberá adjuntar la autorización por escrito de la Máxima Autoridad o su Delegado/a.

FIRMAS DE APROBACIÓN

JEFE/E INMEDIATO DE LA O EL RESPONSABLE DE LA UNIDAD

NOMBRE: Bonny Bayot PhD.

MÁXIMA AUTORIDAD O SU DELEGADO/A

NOMBRE

Informe de viaje
Asistencia al Curso-Entrenamiento “Ensamblaje y Anotación de Secuencias Genómicas”
Bogotá, Colombia
23 de noviembre a 23 de diciembre de 2015

Realizado por M.SC. Leda Restrepo
Centro Nacional de Acuicultura e Investigaciones Marinas (CENAIM)

1. Antecedentes

Entre el 16 de noviembre y el 27 de diciembre del 2015 la suscrita realizó una estancia en la ciudad de Bogotá (Colombia) cuyo objetivo fue asistir al Curso-Entrenamiento “Ensamblaje y Anotación de Secuencias Genómicas”, impartido por el Dr. Andrés Mauricio Pinzón Velasco en la Universidad Nacional de Colombia. El curso tuvo lugar entre el 23 de noviembre y el 23 de diciembre del 2015. El viaje de ida a Colombia y regreso a Ecuador se realizó los días 16 de noviembre y 27 de diciembre, respectivamente. El Curso-Entrenamiento fue dictado por el Dr. Andrés Mauricio Pinzón Velasco, director del Grupo de Bioinformática y Biología de Sistemas del Instituto de Genética de la Universidad Nacional de Colombia donde se realizó el análisis bioinformático de un conjunto de genomas provenientes de aislados bacterianos que se encuentran asociados a problemas de mortalidad en sistemas de producción. Se realizó el ensamblaje de las secuencias genómicas de las diferentes cepas bacterianas que se tenían (ensamblajes de novo y por referencia), se disociaron las librerías genómicas y se ejecutó el ensamblaje de los plásmidos presentes en dichos aislados bacterianos (ensamblajes de novo y por referencia), se evaluó la calidad de los genomas bacterianos y sus plásmidos asociados, se determinó el número de genes presentes en el genoma bacteriano y en los plásmidos, se evaluaron genes específicos o de interés, se determinó la posible función asociada a dichos genes y se realizó una predicción preliminar de caracteres asociados a los transcritos de diversos genes (anotación funcional). Todo esto se realizó mediante un análisis in silico de cepas secuenciadas mediante técnicas nueva generación provenientes de Ecuador. Esta actividad se encuentra enmarcada en uno de los objetivos del proyecto de Investigación financiado por SENESCYT: PIC-14-CENAIM003 - “Desarrollo e implementación de métodos de control y prevención de enfermedades en especies acuáticas de uso comercial y uso potencial en maricultura o repoblación”. A continuación, se resume brevemente las principales actividades realizadas en el Curso-Entrenamiento “Ensamblaje y Anotación de Secuencias Genómicas” realizado con el Dr. Pinzón. Cabe destacar que los gastos de pasaje, hospedaje y alimentación fueron cubiertos por el Proyecto SENESCYT PIC-14CENAIM-003.

2. Curso-Entrenamiento “Ensamblaje y Anotación de Secuencias Genómicas”

El curso tuvo una duración de 30 días (entre el 23 de noviembre y 23 de diciembre) para un total de 176 horas. Fue impartido por el Dr. Andrés Mauricio Pinzón Velasco líder del grupo en Bioinformática y Biología de Sistemas Computacional (GIBBS) del Instituto de Genética de la Universidad Nacional de Colombia (<http://www.genetica.unal.edu.co/gibbs/>).

El objetivo de la pasantía fue trabajar bajo la supervisión y asesoramiento del Dr. Pinzón, en el ensamblaje y la anotación de cuatro genomas bacterianos que habíamos aislado en el CENAIM mediante la plataforma de secuenciación Illumina MiSeq. Los aislados fueron

- **Evaluación del control de calidad**

Se realizó un control de calidad de los datos crudos para identificar y excluir datos con problemas serios de calidad, lo cual me permite ahorrar gran cantidad de tiempo en los análisis posteriores. Las herramientas que emplee para chequear la calidad de las secuencias, fue para evaluar la probabilidad de que la base asignada en cada read sea la base correcta, que la distribución de los nucleótidos sea la adecuada, la distribución del contenido de GC se real, evaluación de las secuencias repetidas, entre otros parámetros. Luego de la evaluación de la calidad inicial de los datos, se realizaron los procesos de corrección pertinentes para cada una de las cepas, filtrando los reads que presentaban baja calidad y fragmentando a partir de la posición en la cual la calidad comienza a decaer.

En la sección de anexos se muestra las imágenes del control de calidad (Anexo 3) de una de las cepas ecuatorianas con las que se trabajó en el Curso-Entrenamiento "Ensamblaje y Anotación de Secuencias Genómicas", donde se muestra el control de calidad de los datos crudos recibidos para dicha cepa y posteriormente los mismos datos de calidad luego de la corrección de los datos realizada, obteniendo y dejando los datos limpios para el ensamblaje.

- **Ensamblaje de novo**

Se evaluaron diferentes métodos para el ensamblaje de novo. Todas las metodologías utilizadas se basan en la suposición de que fragmentos de ADN altamente similares se originan de la misma posición dentro del genoma, pero los algoritmos utilizados por cada programa para conectar los fragmentos individuales en secuencias contiguas más largas, son diferentes para cada uno de éstos. Luego de ensamblar los reads se obtuvieron unas secuencias consenso denominadas contigs. Se evaluaron diferentes metodologías para el ensamblaje de novo, debido a que el reto en este tipo de análisis está dado desde el punto de vista computacional, puesto que la corta longitud de los reads, dificulta los ensamblajes en genomas bacterianos que tienen un gran número de repeticiones, ya que no se sabe con certeza de que repetición se obtuvo determinado read.

Con las secuencias de las bacterias que se estaban analizando se tenían dos tipos de librerías: Single-End y Paired-Ends. Dos de las muestras fueron secuenciadas con la técnica de extremos pareados (Paired-Ends), las cuales permiten alargar los reads usando herramientas computacionales, conociendo la secuencia de ambos extremos de un fragmento, la distancia aproximada entre ellos se logra un solapamiento de los reads, lo que permite unirlos para formar una secuencia más larga. Los ensambladores que utilizamos basan sus análisis en los grafos De Bruijn. Los cuales modelan la relación entre subcadenas exactas de longitud k dentro de los reads, los nodos en el gráfico representan como k -mers, y los conectores indican que k -mers adyacentes se solapan por $k-1$ letras, por lo que la longitud del k -mer correlaciona con la longitud del solapamiento que el ensamblador es capaz de detectar (Zerbino, D.R & Birney, E., 2008). En esta metodología no se modelan directamente los reads, sino que están implícitamente representados por los conectores en el gráfico de Bruijn.

Cuando se tuvieron los contigs se procedió a realizar el proceso de Scaffolding, mediante el cual se conectaron los contigs obtenidos. A pesar de no conocer la secuencia entre ellos, se pueden conectar, infiriendo cuándo dos contigs son adyacentes, dejando una distancia aproximada, determinada por la longitud del inserto (en PE).

- **Ensamblaje por referencia**

La metodología utilizada para en ensamblaje por referencia es similar a las usadas en los ensamblajes de novo, debido a que se basan en la suposición de los fragmentos de ADN (reads) pero lo hacen en base a la comparación contra un genoma de referencia determinado. Luego del ensamblaje por referencia obtuvimos contigs de mayor tamaño que los que obtuvimos con los ensamblajes de novo, gracias a que el número de repeticiones y su estructuración se tenía previamente cargado.

- **Control de calidad luego del ensamblaje**

Luego de los ensamblajes se evaluó el número de fragmentos obtenidos con el fin de observar que se haya cubierto el cromosoma completo de cada bacteria y también la secuencia completa de los plásmidos accesorios. Se evaluaron parámetros métricos como: la talla mínima, máxima y media de los contigs, así como la talla total del ensamblaje, el cual coincidió con la talla reportada del genoma. De igual manera, se tuvo en cuenta el valor N50, el cual muestra el menor número de los contigs que cubren el genoma.

Finalmente, se realizó remapeo de los contigs obtenidos en el ensamblaje de novo contra los resultados obtenidos con el ensamblaje contra el genoma de referencia. Al visualizar los resultados de estas comparaciones pudimos observar la consistencia del genoma y si la estructuración de los contigs es confiable.

- **Cierre de genomas bacterianos**

La terminación de los genomas es el proceso en el cual se hace un cierre de los gaps presentes en cada genoma ensamblado. Este paso es laborioso computacionalmente, pero permitió corregir varios de los errores de las secuencias presentes por la secuenciación y algunos causados luego de los ensamblajes. Aunque en nuestro caso el cierre de algunos genomas no pudo completarse en su totalidad y quedaron algunos gaps entre los contigs.

- **Anotación estructural y funcional de los genomas**

Luego de tener cada genoma ensamblado se procedió a realizar la anotación estructural y la anotación funcional, los cual nos permitió identificar las principales características del genoma. Para la anotación estructural se utilizaron dos metodologías: el método *ab initio* y el método por comparación. El método *ab initio* utiliza algoritmos estadísticos para el reconocimiento de patrones que determinan si la secuencia de interés es codificante o no, mostrando la detección de motivos específicos en la secuencia. El método por comparación identifica zonas de alta similitud en organismos relacionados para reconocer las regiones codificantes (Brosch R., et al., 2001)

Para la anotación funcional, también utilizamos dos métodos diferentes: Por homología y dominios funcionales. La comparación por homología busca inferir la función de un gen comparándolo con secuencias homólogas en bases de datos. La búsqueda de motivos y dominios funcionales busca asignar los genes encontrados en familias génicas o indicar el grupo de procesos en los que pueda estar involucrado (Darling A, Miklos I & Ragan M., 2008).