**REPORTE ENTREGA DEL**

Project\_NCalderon\_2020\_08\_05\_07\_29\_45

**AGOSTO 2023**

Unidad Universitaria de Secuenciación Masiva y Bioinformática

[**IDENTIFICACIÓN DEL REPORTE O PROYECTO**](#_heading=h.gjdgxs)[**3**](#_heading=h.gjdgxs)

[**IDENTIFICACIÓN DEL USUARIO Y COLABORADORES:**](#_heading=h.1fob9te)[**3**](#_heading=h.1fob9te)

[**DESCRIPCIÓN HISTÓRICA DE LAS MUESTRAS**](#_heading=h.3znysh7)[**4**](#_heading=h.3znysh7)

[**DESCRIPCIÓN METODOLÓGICA**](#_heading=h.2et92p0)[**5**](#_heading=h.2et92p0)

[CUANTIFICACIÓN DE LAS MUESTRAS](#_heading=h.tyjcwt) [5](#_heading=h.tyjcwt)

[PREPARACIÓN DE LAS BIBLIOTECAS](#_heading=h.3dy6vkm) [6](#_heading=h.3dy6vkm)

[SECUENCIACIÓN DE LAS MUESTRAS](#_heading=h.1t3h5sf) [**7**](#_heading=h.1t3h5sf)

[**REPORTE DE CALIDAD**](#_heading=h.4d34og8)[**7**](#_heading=h.4d34og8)

[**ANÁLISIS BIOINFORMÁTICO**](#_heading=h.2s8eyo1)[**7**](#_heading=h.2s8eyo1)

[**RESULTADOS**](#_heading=h.17dp8vu)[**7**](#_heading=h.17dp8vu)

[**EVALUACIÓN DEL SERVICIO**](#_heading=h.3rdcrjn) **7**

|  |
| --- |
| El presente documento es un reporte de entrega del proyecto descrito en el siguiente segmento y asegura que la UUSMB está comprometida con la imparcialidad y la confidencialidad de los datos aquí descritos: |

## IDENTIFICACIÓN DEL REPORTE O PROYECTO

|  |  |
| --- | --- |
| ID del proyecto: | Project\_NCalderon\_2020\_08\_05\_07\_29\_45 |
| Nombre del proyecto: | Análisis de la expresión diferencial de genes en el tracto digestivo de escarabajos barrenadores |
| Descripción: | El proyecto contempla analizar los transcriptomas del tracto digestivo de 3 especies de escarabajos barrenadores que usan el mismo tejido hospedero para identificar los genes de enzimas digestivas que participan en la degradación de madera. Este proyecto forma parte las líneas de investigación de mi laboratorio. |
| Fecha de solicitud: | 2020/08/05 |
| Fecha de informe | 2023/08/24 |
| Fecha de Análisis | 2023/08/24 |
| Número de muestras: | 12 |

## IDENTIFICACIÓN DEL USUARIO Y COLABORADORES:

En esta sección se describen los responsables y colaboradores ligados a este proyecto:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **User** | **Nombre** | **Dependencia** | **teléfono** | **email** | **role** |
| NCalderon | Nancy Calderon Cortes | Escuela Nacional de Estudios Superiores Unidad Morelia | 443-689-3500 ext-80559 | ncalderon@enesmorelia.unam.mx | Responsable |

DESCRIPCIÓN HISTÓRICA DE LAS MUESTRAS

Para las muestras donde la aceptación es “CONDICIONADA” o “RECHAZADA”, el servicio aún puede llevarse a cabo, previa autorización del usuario, quedando la UUSMB libre de toda responsabilidad sobre cualquier falla en la construcción de la biblioteca, generación de datos incompletos o nulos que se obtengan en el proyecto.

En esta sección se definen los identificadores únicos de cada muestra, su nombre, identificación del tubo, fecha de recepción, fecha de análisis de calidad y evaluación de Aceptación:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Numero consecutivo** | **Nombre de la Muestra** | **Identificador del tubo** | **Fecha de recepción** | **Fecha análisis calidad** | **Estatus de**  **aceptación** |
| 5930 | NCalderon\_200909\_BOST1 | BA | 2020/09/11 |  | Aceptada |
| 5932 | NCalderon\_200909\_BOST2 | BB | 2020/09/11 |  | Aceptada |
| 5935 | NCalderon\_200909\_BOST3 | BI | 2020/09/11 |  | Aceptada |
| 5939 | NCalderon\_200909\_BOST4 | BII | 2020/09/11 |  | Aceptada |
| 5940 | NCalderon\_200909\_ONC1 | O3 | 2020/09/11 |  | Aceptada |
| 5941 | NCalderon\_200909\_ONC2 | 06 | 2020/09/11 |  | Aceptada |
| 5942 | NCalderon\_200909\_ONC3 | O7 | 2020/09/11 |  | Aceptada |
| 5943 | NCalderon\_200909\_ONC4 | OA | 2020/09/11 |  | Aceptada |
| 5944 | NCalderon\_200909\_CER1 | CA | 2020/09/11 |  | Aceptada |
| 5945 | NCalderon\_200909\_CER2 | CB | 2020/09/11 |  | Aceptada |
| 5946 | NCalderon\_200909\_CER3 | CC | 2020/09/11 |  | Aceptada |
| 5947 | NCalderon\_200909\_CER4 | CD | 2020/09/11 |  | Aceptada |

En esta sección se define el volumen reportado por el usuario y el volumen y concentración encontrados por los miembros de la unidad:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Número consecutivo** | **Nombre de la**  **Muestra** | **Identificador del tubo** | **vol(µL) usuario** | **vol(µL)**  **uusmb** | **concentracion total (ng)** |
| 5930 | NCalderon\_200909\_BOST1 | BA | 50 | 40.0 | 220.0 |
| 5932 | NCalderon\_200909\_BOST2 | BB | 50 | 40.0 | 244.7 |
| 5935 | NCalderon\_200909\_BOST3 | BI | 50 | 40.0 | 170.0 |
| 5939 | NCalderon\_200909\_BOST4 | BII | 50 | 40.0 | 408.0 |
| 5940 | NCalderon\_200909\_ONC1 | O3 | 50 | 40.0 | 891.0 |
| 5941 | NCalderon\_200909\_ONC2 | 06 | 50 | 40.0 | 417.0 |
| 5942 | NCalderon\_200909\_ONC3 | O7 | 50 | 40.0 | 366.0 |
| 5943 | NCalderon\_200909\_ONC4 | OA | 50 | 40.0 | 430.0 |
| 5944 | NCalderon\_200909\_CER1 | CA | 50 | 40.0 | 641.0 |
| 5945 | NCalderon\_200909\_CER2 | CB | 50 | 40.0 | 576.0 |
| 5946 | NCalderon\_200909\_CER3 | CC | 50 | 40.0 | 474.0 |
| 5947 | NCalderon\_200909\_CER4 | CD | 50 | 40.0 | 380.0 |

En esta sección se reportan los valores de absorbancia reportados por el usuario y los encontrados por la Unidad. El tipo de muestra, rendimiento esperado y observaciones o comentarios

**abs usuario**

**abs uusmb**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Numero** | **Nombre de la muestra** | **Identificación del tubo** | **260\_280 1.8 a 2** | **260\_230 1.8 a 2** | **260\_280**  **2 a 2.2** | **260\_230**  **2 a 2.2** | **tipo** | **rendimiento** |
| 5930 | NCalderon\_200909\_BOST1 | BA | 2.11 | 2.09 |  |  | RNA | 10 millones |
| 5932 | NCalderon\_200909\_BOST2 | BB | 2.07 | 1.91 |  |  | RNA | 10 millones millones |
| 5935 | NCalderon\_200909\_BOST3 | BI | 2.03 | 1.86 |  |  | RNA | 10 millones millones millones millones |
| 5939 | NCalderon\_200909\_BOST4 | BII | 2.05 | 1.92 |  |  | RNA | 10 millones millones millones millones millones millones millones |
| 5940 | NCalderon\_200909\_ONC1 | O3 | 2.13 | 1.9 |  |  | RNA | 10 millones millones millones millones millones millones millones |
| 5941 | NCalderon\_200909\_ONC2 | 06 | 2.09 | 2.17 |  |  | RNA | 10 millones |
| 5942 | NCalderon\_200909\_ONC3 | O7 | 2.11 | 2.1 |  |  | RNA | 10 millones |
| 5943 | NCalderon\_200909\_ONC4 | OA | 2.11 | 2.16 |  |  | RNA | 10 millones |
| 5944 | NCalderon\_200909\_CER1 | CA | 2.11 | 2.28 |  |  | RNA | 10 millones |
| 5945 | NCalderon\_200909\_CER2 | CB | 2.11 | 2.19 |  |  | RNA | 10 millones |
| 5946 | NCalderon\_200909\_CER3 | CC | 2.14 | 2.29 |  |  | RNA | 10 millones |
| 5947 | NCalderon\_200909\_CER4 | CD | 2.12 | 2.23 |  |  | RNA | 10 millones |

En esta sección se muestran los comentarios para cada una de las muestras

|  |  |  |  |
| --- | --- | --- | --- |
| **Número consecutivo** | **Nombre de la muestra** | **Identificación del tubo** | **Comentariox** |
| 5930 | NCalderon\_200909\_BOST1 | BA | El rendimiento recomendado |
| 5932 | NCalderon\_200909\_BOST2 | BB |  |
| 5935 | NCalderon\_200909\_BOST3 | BI |  |
| 5939 | NCalderon\_200909\_BOST4 | BII |  |
| 5940 | NCalderon\_200909\_ONC1 | O3 |  |
| 5941 | NCalderon\_200909\_ONC2 | 06 |  |
| 5942 | NCalderon\_200909\_ONC3 | O7 |  |
| 5943 | NCalderon\_200909\_ONC4 | OA |  |
| 5944 | NCalderon\_200909\_CER1 | CA |  |
| 5945 | NCalderon\_200909\_CER2 | CB |  |
| 5946 | NCalderon\_200909\_CER3 | CC |  |
| 5947 | NCalderon\_200909\_CER4 | CD |  |

DESCRIPCIÓN METODOLÓGICA.

• CUANTIFICACIÓN DE LAS MUESTRAS

Las muestras originales fueron cuantificadas utilizando el RNA Nano 6000 Kit de la compañía Agilent en el Bioanalyzer 2100 también de la compañía Agilent. La construcción de las bibliotecas se realizó siguiendo el protocolo de TruSeq Stranded mRNA Sample Preparation Kit de la compañía Illumina.

• PREPARACIÓN DE LAS BIBLIOTECAS

La construcción de las bibliotecas se realizó utilizando el TruSeq Stranded mRNA Library Prep Kit, siguiendo a detalle el protocolo sin ninguna modificación. Se adicionó un paso final de purificación de las bibliotecas a través de un gel de Low Range Agarose al 1.5%. Posteriormente, las bibliotecas fueron analizadas para determinar el tamaño promedio de cada biblioteca, a través de un ensayo de electroforesis en capilar utilizando un equipo Bioanalyzer 2100 de la compañía Agilent, utilizando un chip de DNA 1000. Las bibliotecas analizadas, fueron cuantificadas utilizando el Qubit dsDNA HS Assay Kit en un fluorómetro DeNovix DS-11 o Qubit. Ya cuantificadas por fluorescencia, las muestras se volvieron a cuantificar en esta ocasión a través de una reacción de qPCR, utilizando el 2x SYBR Green qPCR Master Mix (High ROX) de la compañía Bimake.com en un equipo Step One de la compañía Applied Biosystems.

• SECUENCIACIÓN DE LAS MUESTRAS

Las muestras cuantificadas por qPCR, se normalizaron a una concentración de 4 nM y se hizo una mezcla de todas las bibliotecas a secuenciarse en una sola dilución. Se prepararon 1.5 ml de la solución con las bibliotecas a una concentración de 1.8 pM para cargar en el equipo de secuenciación. La secuencia se realizó en un equipo de secuenciación NextSeq 500 de la compañía Illumina, utilizando un NextSeq 500/550 High Output Kit v2.5 de 150 cycles para una secuencia de 2x75 ciclos.

## REPORTE DE CALIDAD

El siguiente reporte de calidad reporta el % de GC, rendimiento y % de nucleótidos por base:

<http://www.sisbi-sin-liga.com>

## En este reporte se aprecia que el rendimiento quedó 1 de secuencias solicitadas. Como resultado de la revisión de estos reportes de calidad, y dado que el promedio de calidad se encuentra para todas las muestras a lo largo de todas las secuencias por arriba de Q30, el porcentaje de GC se apega al teórico y los porcentajes de nucleótidos a lo largo de todas las bases se comparten dentro de lo esperado, no se realizó ninguna limpieza por calidad o de adaptadores o para eliminar secuencias repetidas.

## ANÁLISIS BIOINFORMÁTICO DE EXPRESIÓN DIFERENCIAL

El análisis bioinformático inicio con el alineamiento pareado de las 12 muestras vs el 2, 3. 4 de 5, mismo que fue indexado con un k-mero de tamaño 6 y un step de 7 para el programa de alineamiento de secuencias cortas 8. Luego llevamos a cabo el alineamiento, con el programa 9 solicitando un porcentaje de semejanza de 10. Se marcaron los duplicados ópticos usando MarkDuplicates VN:2.18.14-SNAPSHOT. Luego se realizó el cálculo de coberturas usando el programa 11. 12 de coberturas 13.

Estas coberturas se subieron al sitio http://www.uusmb.unam.mx/ideamex/ para realizar el análisis de expresión diferencial y se solicitaron las comparaciones:

14.

**Definir comparaciones**

|  |  |  |
| --- | --- | --- |
|  | Control | mettl3 |
| control |  |  |
| mettl3 |  |  |

Para reportar los resultados se utilizaron como líneas de corte un FDR o pvalue menor a 15, una magnitud de cambio superior a 16 (LogFC >= 17) y una cuenta por Millon igual a 18 (CPM= 19).

Se aplicaron 4 métodos de análisis de expresión diferencial:

* [edgeR Results](http://www.uusmb.unam.mx/ideamex/avila.dela195537/EdgeR.php)
* [DESeq2 Results](https://bioconductor.org/packages/release/bioc/html/DESeq2.html)
* [limma Results](https://www.bioconductor.org/packages/release/bioc/html/limma.html)
* [NOISeq Results](https://bioconductor.org/packages/release/bioc/html/NOISeq.html)

## RESULTADOS

Los resultados se encuentran disponibles en la liga de IDEAmex: Para la tabla solo con los transcritos principales se encuentra en . En la sección de análisis de datos, en la gráfica 20, podemos observar una muy buena distancia entre los grupos a comparar. En la gráfica de 21, con las muestras ya normalizadas, se observan todas las réplicas con distribuciones muy semejantes, cosa que se corrobora con la 22.

Dentro de cada una de las carpetas:

• 23

• 24

• 25

• 26

Encontrará los archivos que contienen en la primera columna los identificadores de todos transcritos con los valores de expresión diferencial diferente de cero, seguido de los parámetros de cada método seleccionado, y luego las cuentas crudas y normalizadas para cada réplica de cada muestra.

27

Los archivos siguientes, contienen sólo aquellos genes con una magnitud de cambio mayor a 28 (LogFC > 29 ). y un pvalue < 30, para aquellos genes con cuentas por millón > 31 (CPM> 32):

33

Cada método puede generar una lista de genes que pueden coincidir o no, con los otros métodos. en los archivos siguientes están aquellos genes que fueron reportados como diferencialmente expresados por EdgeR y algún otro método:

FIELDMANUALZ

En la sección 35, se hace un análisis de integrativo de los resultados dados por cada método, para cada par de muestras comparadas.

Si quisieran revisar todos los resultados pueden consultarlos en:

36

37

## EVALUACIÓN DEL SERVICIO

Como parte de nuestro proceso de mejora, sería muy útil para nosotros saber su opinión sobre nuestro servicio, por lo que agradeceremos nos retroalimenta vía el cuestionario:

[Encuesta de satisfacción](https://forms.gle/TGVhbikTi8az5sQw5)

|  |  |  |  |
| --- | --- | --- | --- |
| **CONTROL DE EMISIÓN** | | | |
|  | **Elaboró:** | **Revisó:** | **Autorizó:** |
| **Nombre** | Lucio Ricardo Montero | REVISEUSER | AUTHORIZEUSER |
| **Firma** | /var/www/html/reportProjects/imgSign/4Lucio Ricardo11.jpg | REVISESIGN | AUTHORIZESIGN |
| **Fecha** | 2023/08/24 | REVISEDATE | AUTHORIZEDATE |

Nota: La representatividad de la muestra no es responsabilidad del laboratorio ya que no realizamos muestreo. Queda prohibida la reproducción parcial de este reporte.