

ÍNDICE DE CONTENIDOS

1. INTRODUCCIÓN	1
3. 1 Marco Conceptual	1
3.1.1 Secuenciamiento de ADN	1
3.1.2 Sequencing by Synthesis (SBS) desarrollado por SOLEXA	2
3.1.3 Illumina	4
3.1.4 RNA-Seq.....	5
3.1.5 Reporte de errores en Illumina	6
3.1.6 Silenciamiento génico mediado por ARN.....	7
3.1.7 <i>Nannochloropsis salina</i>	7
3.1.8 Secuenciación de <i>N. salina</i>	8
3.1.9 Análisis inicial de <i>N. salina</i>	9
3.2 Planteamiento del problema.....	11
4. HIPÓTESIS DE TRABAJO	11
5. OBJETIVOS	11
5.1 Objetivo general	11
5.2 Objetivos específicos	12
6. MATERIALES	13
6.1 Secuencias	13
6.2 Software y Herramientas Computacionales	13
6.3 Computadores	14
7. METODOLOGÍA	16
7.1 Cálculo de cobertura y profundidad con BEDTools.....	16
7.2 Definición de zonas anómalas	18
7.3 Clasificación de Pícs a partir de archivo GFF3	19
7.4 BLAST para búsqueda de patrones de expresión	21
7.5 Evaluación de errores de secuenciación por contenido de GC	24
7.6 Determinar presencia de duplicaciones con CD-HIT.....	27

7.7 TopHat para mapeo de lecturas sobre Transcriptoma de Referencia	29
7.8 Cuffdiff para cálculo de Expresión Diferencial	32
7.9 Efectos de Pics sobre Expresión Génica de un organismo.....	33
8. RESULTADOS Y DISCUSIÓN	35
8.1 Universo de estudio	35
8.2 Transcritos donde existen Pics	37
8.3 Patrones de Expresión	39
8.4 Sesgos por contenido de GC	40
8.5 Lecturas duplicadas.....	42
8.6 Pics causados por duplicaciones	44
8.7 Efectos de Pics sobre expresión génica de <i>Nannochloropsis salina</i>	46
8.7.1 Diferencias en Expresión Diferencial.....	46
8.7.2 Test Estadístico	47
9. CONCLUSIÓN	48
10. REFERENCIAS.....	49

I. ÍNDICE DE TABLAS

Tabla 1	Comparación entre tres plataformas de Secuenciación de Nueva Generación	5
Tabla 2	Secuencias Illumina para <i>N. salina</i>	13
Tabla 3	Especificaciones de softwares utilizados	14
Tabla 4	Especificaciones de Computador Mac Mini	15
Tabla 5	Especificaciones Servidor Bio-5	15
Tabla 6	Detalles de archivo de salida entregado por el programa BEDTools y su herramienta <i>coverageBed</i>	17
Tabla 7	Información almacenada en un archivo GFF3.	19
Tabla 8	Genes que pierden su Expresión Diferencial luego de clusterizar lecturas duplicadas	46
Tabla 9	Resultado del T-test aplicado sobre los Transcritos de <i>N. salina</i> previo y posterior al clustering de lecturas duplicadas.	47

II. ÍNDICE DE FIGURAS

Figura 1	Representación de la amplificación de una muestra de ADN	2
Figura 2	Representación del etiquetado de nucleótidos con fluorescencia	3
Figura 3	Fluctuación de intensidad de fluorescencia de cada nucleótido a medida que avanzan los ciclos de secuenciación	4
Figura 4	Disposición de las lecturas de Illumina bajo el genoma de referencia en un estado de Expresión Diferencial	10
Figura 5	Disposición de las lecturas de Illumina bajo el genoma de referencia	13
Figura 6	Representación de Función <i>coverageBed</i>	16
Figura 7	Proceso para calcular cobertura y profundidad en todo el transcriptoma de <i>N. salina</i>	17
Figura 8	Diagrama de flujo del pipeline destinado a la detección de zonas anómalas o Pics	18
Figura 9	Diagrama de flujo, para la obtención de las posiciones de los Pics con respecto al Transcriptoma de referencia.	21
Figura 10	Parámetros para ejecutar el programa BLAST	22
Figura 11	Método de procesamiento de resultados de BLAST	23
Figura 12	Flujo de trabajo para asociar Contenido de GC y Cobertura	25
Figura 13	Correlación existente entre la cobertura de lecturas Illumina y el contenido	26
Figura 14	Diagrama de flujo para la búsqueda de lecturas duplicadas	28
Figura 15	Parámetro para la ejecución de los software CD-HIT y Tophat	31
Figura 16	Parámetros para la ejecución del software Cuffdiff	34
Figura 17	Gráfico de barras, que expresa la cantidad de Pics para las condiciones Low CO ₂ y High CO ₂ .	35
Figura 18	Posicionamiento de Pics respecto al Transcrito que los contiene	36
Figura 19	Tipos de Transcritos que contiene Pics en la condición Low CO ₂	37
Figura 20	Tipos de Transcritos que contiene Pics en la condición High CO ₂	38
Figura 21	Porcentaje de Pics coincidentes con al menos un ARNnc de la base de datos NONCODE	39
Figura 22	Gráfico de dispersión, que relaciona el contenido de GC y la Cobertura en Bajo CO ₂ .	40
Figura 23	Gráfico de dispersión, que relaciona el contenido de GC y la Cobertura en Alto CO ₂ .	41

Figura 24	Reducción del tamaño del set de lecturas Illumina, posterior a la clusterización realizada mediante CD-HIT para Bajo CO ₂	43
Figura 25	Reducción del tamaño del set de lecturas Illumina, posterior a la clusterización realizada mediante CD-HIT. Alto CO ₂	43
Figura 26	Reducción en la cantidad de Pics luego de la clusterización de lecturas duplicadas, para la condición de Bajo CO ₂ .	45
Figura 27	Reducción en la cantidad de Pics luego de la clusterización de lecturas duplicadas, para la condición de Alto CO ₂ .	45