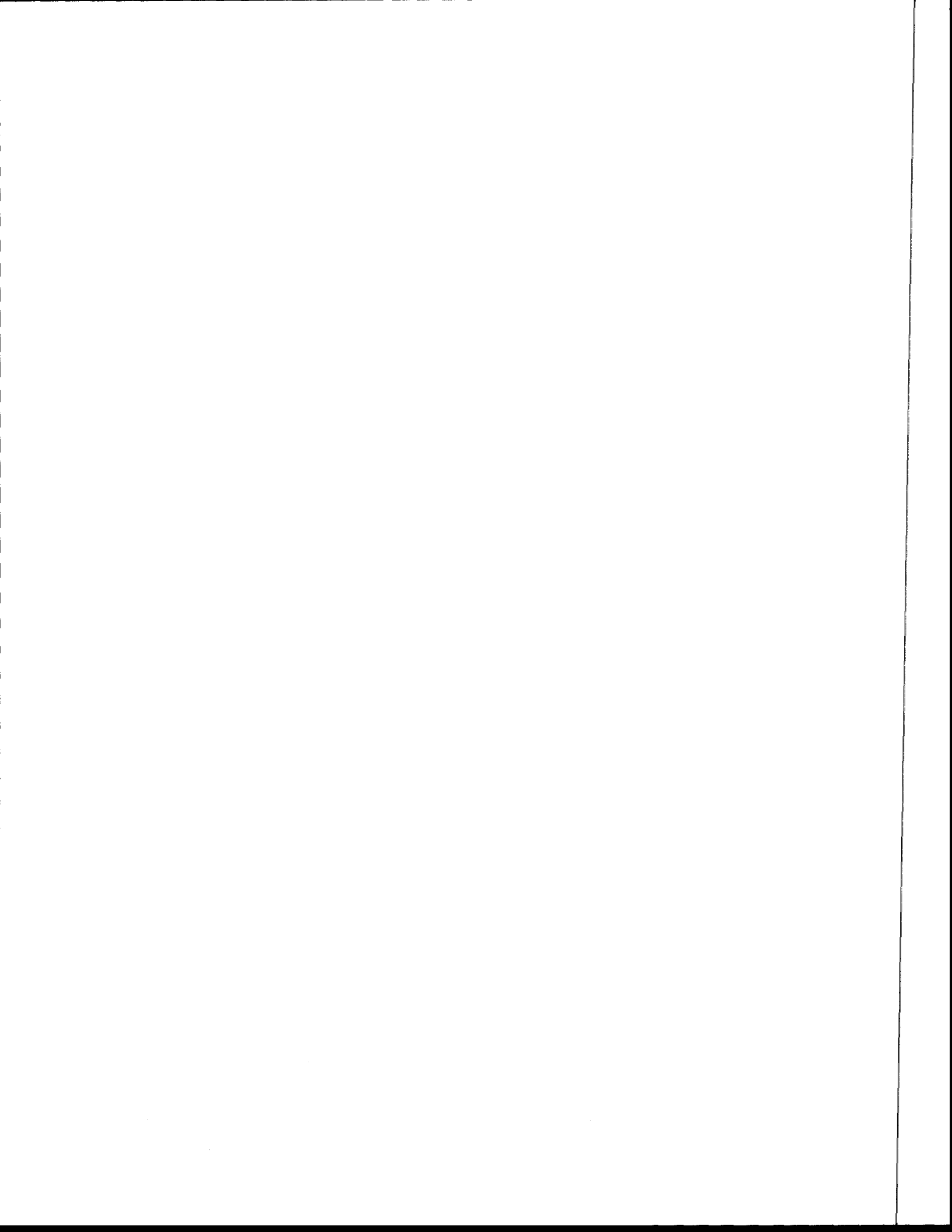


TABLA DE CONTENIDOS

| | |
|--|------|
| TABLA DE CONTENIDOS | I |
| ÍNDICE DE FIGURAS | IV |
| ÍNDICE DE TABLAS | V |
| RESUMEN | VI |
| ABSTRACT | VIII |
| 1. MARCO TEÓRICO | 1 |
| 1.1 Acuicultura Chilena | 1 |
| 1.2 El salmón en Chile | 1 |
| 1.3 Enfermedades del salmón registradas en Chile | 2 |
| 1.3.1 Enfermedad Bacteriana del Riñón (BKD)..... | 2 |
| 1.3.2 Anemia Infecciosa del Salmón (Virus ISA)..... | 3 |
| 1.3.3 Virus de Necrosis Pancreática Infecciosa (IPN)..... | 3 |
| 1.3.4 <i>Piscirickettsia salmonis</i> | 3 |
| 1.3.5 Parásitos | 4 |
| 1.4 <i>Caligus rogercresseyi</i> | 4 |
| 1.4.1 Ciclo de vida de <i>Caligus rogercresseyi</i> | 5 |
| 1.4.2 Problemas producidos por <i>Caligus</i> en Chile..... | 6 |
| 1.4.3 Principales fármacos utilizados contra <i>Caligus rogercresseyi</i> en Chile | 7 |
| 1.4.3.1 Resistencia farmacológica | 8 |
| 1.5 Bioinformática y Blancos terapéuticos..... | 10 |
| 1.6 Análisis de transcriptomas. | 11 |
| 1.6.1 Relevancia de la bioinformática en el análisis de datos transcriptómicos..... | 12 |
| 1.7 Iniciativa del análisis del transcriptoma del <i>Caligus rogercresseyi</i> | 12 |
| 2. HIPÓTESIS | 14 |
| 3. OBJETIVOS | 15 |
| 3.1 Objetivo general | 15 |
| 3.2 Objetivos específicos..... | 15 |
| 4. METODOLOGÍA..... | 16 |
| 4.1 Estrategia de trabajo | 16 |



| | |
|---|----|
| 4.2 Detalle de la metodología implementada..... | 18 |
| 4.2.1 Análisis de calidad de las secuencias..... | 18 |
| 4.2.2 Corte y filtrado de secuencias..... | 19 |
| 4.2.3 Unificación de los set de datos..... | 20 |
| 4.2.4 Limpieza de agentes contaminantes..... | 20 |
| 4.2.4.1 Limpieza de agentes contaminantes del salmón..... | 20 |
| 4.2.4.2 Limpieza de agentes contaminantes contra base de datos nt..... | 22 |
| 4.2.5 Ensamble de las lecturas..... | 24 |
| 4.2.6 Anotación y Caracterización funcional..... | 25 |
| 6.2.7 Identificación de blancos terapéuticos putativos..... | 27 |
| 4.2.8 Expresión diferencial..... | 28 |
| 4.2.8.1 Generación de matriz de estimación de abundancia de transcripción..... | 28 |
| 4.2.8.2 Identificación de transcritos expresados diferencialmente..... | 28 |
| 4.2.8.3 Analizando los transcritos expresados diferencialmente..... | 28 |
| 4.2.9 Pipeline de automatización del proceso..... | 29 |
| 5. RESULTADOS..... | 30 |
| 5.1 Lecturas trabajadas..... | 30 |
| 5.2 Ensamble <i>de novo</i> | 31 |
| 5.3 Anotación..... | 32 |
| 5.4 Análisis de blancos terapéuticos putativos..... | 39 |
| 5.4.1 Análisis de blancos terapéuticos putativos a partir de componentes de membrana..... | 40 |
| 5.4.2 Análisis de blancos terapéuticos putativos a partir de genes esenciales..... | 43 |
| 5.5 Expresión Diferencial..... | 46 |
| 5.6 Scripts desarrollados..... | 50 |
| 6. DISCUSIÓN..... | 52 |
| 8.1 De la problemática económica y ecológica..... | 52 |
| 8.2 De los diseños experimentales..... | 52 |
| 8.3 De la problemática filosófica de la bioinformática en el día de hoy..... | 53 |
| 8.4 Del análisis de los datos..... | 53 |
| 8.5 De la importancia de la bioinformática en estudios genómicos..... | 54 |
| 8.6 De la calidad y la limpieza de los datos obtenidos de la secuenciación..... | 55 |
| 8.7 De las bioinformática y las bases de datos públicas..... | 55 |

| | |
|---|----|
| 8.8 Del “divide y vencerás” | 56 |
| 8.9 De la validación de las herramientas públicas | 56 |
| 8.10 De Google y la bioinformática | 57 |
| 8.11 De los resultados obtenidos | 57 |
| 7. CONCLUSIÓN | 59 |
| BIBLIOGRAFÍA | 61 |
| ANEXOS | 69 |
| Anexo 1 | 69 |
| Anexo 2 | 70 |
| Anexo 3 | 71 |
| Anexo 4 | 72 |

ÍNDICE DE FIGURAS

| | |
|---|----|
| Figura 1. 1 <i>Caligus rogercresseyi</i> | 5 |
| Figura 1. 2 Ciclo de vida de <i>Caligus rogercresseyi</i> | 6 |
| | |
| Figura 4. 1 Flujo de trabajo planteado | 17 |
| Figura 4. 2 Flujo de limpieza de contaminantes..... | 22 |
| Figura 4. 3 Flujo de limpieza de contaminantes por taxonomía..... | 24 |
| | |
| Figura 5. 1 Gráfico de pérdida de lecturas | 30 |
| Figura 5. 2 Gráfico de anotaciones | 33 |
| Figura 5. 3 Gráfico de anotaciones por base de datos..... | 34 |
| Figura 5. 4 Gráfico de funciones moleculares..... | 36 |
| Figura 5. 5 Gráfico de procesos biológicos..... | 37 |
| Figura 5. 6 Gráfico de componentes celulares..... | 39 |
| Figura 5. 7 Gráfico de anotacion contra TMHMM | 40 |
| Figura 5. 8 Diagrama de proceso de GO y TMHMM | 40 |
| Figura 5. 9 Gráfico de dominios transmembranales GO..... | 41 |
| Figura 5. 10 Gráfico de cantidad de dominios transmembranales GO | 42 |
| Figura 5. 11 Gráfico de anotacion contra DEG..... | 43 |
| Figura 5. 12 Diagrama de proceso de DEG y TMHMM..... | 44 |
| Figura 5. 13 Gráfico de dominios transmembranales DEG | 45 |
| Figura 5. 14 Gráfico de cantidad de dominios transmembranales DEG | 45 |
| Figura 5. 15 Resultados caso 1 análisis de expresión diferencial..... | 47 |
| Figura 5. 16 Resultados caso 2 análisis de expresión diferencial..... | 47 |
| Figura 5. 17 Resultados caso 3 análisis de expresión diferencial..... | 48 |
| Figura 5. 18 Resultados caso 4 análisis de expresión diferencial..... | 48 |
| Figura 5. 19 Resultados caso 5 análisis de expresión diferencial..... | 49 |
| Figura 5. 20 Diagrama de scripts realizados..... | 50 |

ÍNDICE DE TABLAS

| | |
|---|----|
| Tabla 5. 1: Detalles de las lecturas perdidas en cada etapa realizada hasta la formación del ensamble final. | 30 |
| Tabla 5. 2: Muestra los valores asociadas al total de genes, total de transcritos, etc., para cada uno de los ensamblajes descritos. | 31 |
| Tabla 5. 3: Muestra los estadísticos asociados a los contigs de los transcritos, para cada uno de los ensamblajes descritos. | 31 |
| Tabla 5. 4: Muestra los estadísticos asociados a las isoformas, para cada uno de los ensamblajes descritos. | 32 |
| Tabla 5. 5: Recuento de número de transcritos y porcentajes de los mismos según en la cantidad de bases de datos en las que fueron anotados. | 34 |
| Tabla 5. 6: Segundo nivel de casificación para Funciones Moleculares. | 35 |
| Tabla 5. 7 : Tercer nivel de casificación para Funciones Moleculares. | 35 |
| Tabla 5. 8 Segundo nivel de casificación para Procesos Biológicos. | 36 |
| Tabla 5. 9: Tercer nivel de casificación para Procesos Biológicos. | 37 |
| Tabla 5. 10: Segundo nivel de casificación para Componentes Celulares. | 38 |
| Tabla 5. 11: Tercer nivel de casificación para Componentes Celulares. | 38 |
| Tabla 5. 12: Cantidad de dominios transmembrana encontrados en el análisis de TMHMM y GO. | 42 |
| Tabla 5. 13: Cantidad de anotaciones de dominios de membrana predichos por base de datos. | 43 |
| Tabla 5. 14: Cantidad de dominios transmembrana encontrados en el análisis de TMHMM y DEG. | 46 |
| Tabla 5. 15: Cantidad de anotaciones por base de datos. | 46 |
| Tabla 5. 16: Scripts pertenecientes a etapa de limpieza de datos. | 51 |
| Tabla 5. 17: Scripts perteneciente a formato de anotaciones. | 51 |