

## INDICE DE CONTENIDOS

INDICE DE CONTENIDOS .....	1
INDICE DE TABLAS .....	5
INDICE DE FIGURAS .....	7
RESUMEN .....	8
ABSTRACT .....	10
<b>CAPITULO 1 - INTRODUCCION AL PROBLEMA Y OBJETIVOS .....</b>	<b>12</b>
1.1 Motivación.....	12
1.2 Objetivos.....	16
1.2.1 Objetivo general .....	16
1.2.2 Objetivos específicos .....	16
1.3 Diseño experimental computacional .....	17
1.3.1 Selección de genomas microbianos y virales.....	17
1.3.2 Construcción de metagenomas conceptuales.....	17
1.3.3 Recopilación y clasificación de <i>softwares</i> de <i>binning</i> .....	17
1.3.4 Implementación de <i>softwares</i> funcionales.....	18
1.3.5 Evaluación y ranking de <i>softwares</i> de <i>binning</i> .....	18
<b>CAPITULO 2 - METAGENOMICA .....</b>	<b>20</b>
2.1 Introducción .....	20
2.1.1 Metagenómica.....	20
2.1.2 Metagenomas acidofilos.....	23
2.1.3 Genomas acidófilos.....	24
2.1.4 Virus acidófilos .....	24
2.1.5 Metagenoma conceptual .....	24
2.1.6 Tecnologías de secuenciación masiva.....	25
2.1.7 Simuladores de datos NGS.....	25

2.2 Metodología .....	27
2.3 Resultados .....	29
2.3.3 Metagenoma conceptual construido .....	31
2.4 Conclusión .....	33
<b>CAPITULO 3 – BINNING Y SOFTWARES DE BINNING .....</b>	<b>34</b>
3.1 Introducción .....	34
3.1.1 Binning .....	34
3.1.2 Softwares de binning .....	36
3.2 Metodología .....	37
3.3 Resultados .....	38
3.3.1 Diversidad .....	38
3.3.2 Clasificación .....	38
3.3.3 Categorización .....	43
3.4 Conclusión .....	47
<b>CAPITULO 4 - IMPLEMENTACIÓN DE SOFTWARES DE BINNING .....</b>	<b>48</b>
4.1 Introducción .....	48
4.1.1 Implementación de softwares funcionales .....	48
4.1.2 Sistemas operativos .....	48
4.1.3 Lenguajes de programación .....	49
4.2. Metodología .....	50
4.3 Resultados .....	51
4.3.1 Ejecución .....	51
4.3.2 Archivos de entrada .....	53
4.3.3 Archivos de salida .....	56
4.4 Conclusión .....	57

<b>CAPITULO 5 – BINNING DE LOS METAGENOMAS SIMULADOS Y EVALUACIÓN COMPARATIVA DE LOS METODOS FUNCIONALES .....</b>	<b>58</b>
5.1. Introducción .....	58
5.1.1 Evaluación comparativa o benchmarking.....	58
5.2. Metodología .....	60
5.2.1 Porcentaje de bases clasificadas .....	61
5.2.2 Porcentaje de bases correctamente clasificadas .....	61
5.2.3 Puntaje taxonómico.....	63
5.2.4 Índice de Variación de la Información .....	64
5.3. Resultados.....	65
5.3.1 Porcentaje de bases clasificadas .....	65
5.3.2 Porcentaje de bases correctamente clasificadas .....	66
5.3.3 Puntaje taxonómico.....	69
5.3.4 Distancia de variación de la información .....	70
5.3.5 Ranking integrativo.....	71
5.4 Conclusión .....	73
<b>CAPITULO 6 – DISCUSIÓN Y CONCLUSIONES .....</b>	<b>74</b>
6.1 Hitos.....	74
6.2 Discusión .....	76
6.2.1 Metagenoma Conceptual .....	76
6.2.2 Evaluación.....	76
6.2.3 Fracción de interés.....	77
<b>CAPITULO 7 – PRODUCTOS DE LA TESIS.....</b>	<b>79</b>
7.1 Publicaciones en preparación.....	79
7.2 Congresos .....	80
<b>REFERENCIAS.....</b>	<b>83</b>
<b>ANEXOS .....</b>	<b>i</b>

Tabla microorganismos acidófilos..... i

Tabla calidad de ensamble .....ii

Recopilación Softwares de Binning .....iv

Tabla de salidas de softwares de binning.....vi

Tabla de bases clasificadas..... viii

Tabla de bases correctamente clasificadas en la fracción viral .....ix

Tabla de bases correctamente clasificadas en la fracción microbiana ..... x

Tabla de bases correctamente clasificadas en la fracción comunitaria .....xi

Tabla de puntajes taxonómicos .....xii

## INDICE DE TABLAS

<b>Tabla 1.</b> Metagenomas secuenciados. ....	23
<b>Tabla 2.</b> Principales plataformas NGS. ....	25
<b>Tabla 3.</b> Simuladores NGS. ....	26
<b>Tabla 4.</b> Parámetros utilizados en la simulación con GRINDER. ....	28
<b>Tabla 5.</b> Microorganismos utilizados en la composición del metagenoma conceptual. ....	29
<b>Tabla 6.</b> Genomas virales de ambientes acidófilos en bases de datos públicas y privadas. ....	30
<b>Tabla 7.</b> Genomas y su representación en el metagenoma. ....	32
<b>Tabla 8.</b> Bases de datos. ....	37
<b>Tabla 9.</b> Diversidad de softwares de binning. ....	38
<b>Tabla 10.</b> Softwares de binning basados en alineamiento. ....	39
<b>Tabla 11.</b> Softwares de binning basados en composición. ....	40
<b>Tabla 12.</b> Softwares de binning híbridos. ....	41
<b>Tabla 13.</b> Softwares de binning basados en la filogenia. ....	42
<b>Tabla 14.</b> Softwares de binning no basados en taxonomía. ....	42
<b>Tabla 15.</b> Softwares de binning en la categoría obsoletos. ....	43
<b>Tabla 16.</b> Softwares de binning en la categoría limitados. ....	44
<b>Tabla 17.</b> Softwares de binning en la categoría funcionales. ....	45
<b>Tabla 18.</b> Softwares de binning ejecutados. ....	51
<b>Tabla 19.</b> Softwares de binning seleccionados y archivos de entrada. ....	53
<b>Tabla 20.</b> Salidas de los softwares de binning. ....	56
<b>Tabla 21.</b> Niveles taxonómicos y su puntaje. ....	63
<b>Tabla 22.</b> Resumen criterios de evaluación. ....	72

<b>Tabla 23.</b> Puntajes y ranking.....	72
<b>Tabla 24.</b> Ranking.....	73
<b>Tabla 25.</b> Microorganismos acidófilos secuenciados depositados en bases de datos públicas y privadas, modificado de (Valdes et al., 2010). ....	i
<b>Tabla 26.</b> Tabla calidad ensamble. ....	ii
<b>Tabla 27.</b> Tabla calidad ensamble, continuación. ....	iii
<b>Tabla 28.</b> Recopilación de softwares de binning.....	iv
<b>Tabla 29.</b> Recopilación de softwares de binning, continuación.....	v
<b>Tabla 30.</b> Salidas de softwares de binning. ....	vi
<b>Tabla 31.</b> Salidas de softwares de binning, continuación. ....	vii
<b>Tabla 32.</b> Porcentaje de bases clasificadas. ....	viii
<b>Tabla 33.</b> Porcentaje de bases correctamente clasificadas en la fracción viral. ....	ix
<b>Tabla 34.</b> Porcentaje de bases correctamente clasificadas en la fracción microbiana.....	x
<b>Tabla 35.</b> Porcentaje de bases correctamente clasificadas en la fracción comunitaria.....	xi
<b>Tabla 36.</b> Puntajes taxonómicos. ....	xii
<b>Tabla 37.</b> Puntajes taxonómicos, continuación.....	xiii

## INDICE DE FIGURAS

<b>Figura 1.</b> Diferencias entre el análisis genómico y metagenómico. ....	13
<b>Figura 2.</b> Diagrama diseño experimental. ....	19
<b>Figura 3.</b> Metagenomas secuenciados. ....	21
<b>Figura 4.</b> Evolución de proyectos genómicos y metagenómicos en GOLD. ....	22
<b>Figura 5.</b> Calidad por base de las lecturas. ....	31
<b>Figura 6.</b> Proceso de binning. ....	34
<b>Figura 7.</b> Clasificación de softwares de binning. ....	35
<b>Figura 8.</b> Distribución de los softwares de binning en los últimos años. ....	36
<b>Figura 9.</b> Softwares de binning. ....	46
<b>Figura 10.</b> Porcentaje de bases correctas por contig. ....	54
<b>Figura 11.</b> Set de contigs de referencia. ....	60
<b>Figura 12.</b> Porcentaje de bases correctamente clasificadas. ....	62
<b>Figura 14.</b> Variación de la información y cantidades relativas. ....	64
<b>Figura 15.</b> Porcentaje de bases clasificadas. ....	65
<b>Figura 16.</b> Porcentaje de bases correctamente clasificadas en la fracción viral. ...	66
<b>Figura 17.</b> Porcentaje de bases correctamente clasificadas en la fracción microbiana. ....	67
<b>Figura 18.</b> Porcentaje de bases correctamente clasificadas en la fracción comunitaria. ....	68
<b>Figura 19.</b> Gráfico de distribución de puntajes taxonómicos. ....	69
<b>Figura 20.</b> Árbol neighbor joining de variación de la información. ....	70
<b>Figura 21.</b> Poster SoiBio. ....	81