

**UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN**  
**FACULTAD DE CIENCIAS BIOLÓGICAS**  
**INSTITUTO DE BIOTECNOLOGÍA**

RC-07-016  
Rev. 02-10/07



**Análisis del transcriptoma, basado en RNA-seq, de una levadura  
cervecera tipo lager en fermentación a nivel industrial y laboratorio**

POR:

**CÉSAR IGNACIO HERNÁNDEZ VÁSQUEZ**

COMO REQUISITO PARCIAL PARA OBTENER EL GRADO DE:

**MAESTRÍA EN CIENCIAS CON ORIENTACIÓN EN MICROBIOLOGÍA**

# UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

## FACULTAD DE CIENCIAS BIOLÓGICAS

### INSTITUTO DE BIOTECNOLOGÍA

“Análisis del transcriptoma, basado en RNA-seq, de una levadura cervecera tipo lager en fermentación a nivel industrial y laboratorio”



#### Comité de tesis

---

**Dr. Benito Pereyra Alférez**  
Director de tesis

---

**Dr. Jorge Hugo García García**  
Secretario

---

**Dr. Luis Jesús Galán Wong**  
1er. Vocal

---

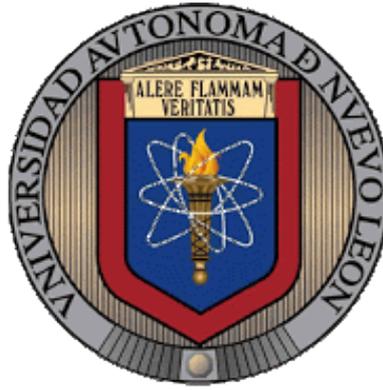
**Dr. José María Viader Salvadó**  
2do. Vocal

---

**Dr. Hamlet Avilés Arnaut**  
3er. Vocal

---

**MC. Esmeralda Pérez Ortega**  
Asesor Externo



Este trabajo se realizó en el laboratorio 4 (L4) del Instituto de Biotecnología de la Facultad de Ciencias Biológicas, Universidad Autónoma de Nuevo León, con la dirección del Dr. Benito Pereyra Alférez. Así como en el laboratorio de Microbiología de Cervecería Cuauhtémoc Moctezuma bajo la dirección de la M.C. Esmeralda Pérez Ortega

## Agradecimientos

A la **Vida** por brindarme un universo de oportunidades para hacer lo que más cultiva mi espíritu; la ciencia.

Al **CONACYT**, por tributar la beca para sustentar mis estudios de maestría.

Al **Dr. Benito Pereyra Aférez**, por todo su apoyo académico y espiritual. Muchas gracias por esas tertulias que abonaron tanto en esta investigación y en mi espíritu.

Al **Dr. Jorge Hugo García García** por compartir de forma extraordinaria todas las ideas y críticas a este trabajo. Muchas gracias por esas tardes de pláticas con profundo sentido científico que aportaron sustancialmente en mi pensamiento y espíritu.

Al **Dr. Luis J. Galán Wong** por el apoyo en el proceso de la investigación.

Al **Dr. José María Viader Salvadó** por sus buenos consejos en bioinformática y apoyo en el desarrollo de la investigación.

Al **Dr. Hamlet Avilés Arnaut** por su aporte y disponibilidad valiosa en esta investigación.

A la **MC. Esmeralda Pérez Ortega** por su valiosa tutoría y aportes críticos en el diseño, desarrollo y conclusión de este trabajo de investigación.

A la **Universidad Autónoma de Nuevo León** por acogerme en su fantástico ambiente académico. Además de proveer los medios para desarrollar e impulsar mi carrera profesional.

A **Cervecería Cuauhtémoc Moctezuma**, por todo el sustento necesario en el diseño, desarrollo y conclusión de este proyecto.

A mi esposa **Lorena** e hija **Luna**. Por brindarme una estabilidad emocional necesaria para enfrentar el camino de la investigación. Gracias por todo el apoyo incondicional. Son lo mejor que me ha pasado.

A mis **padres**, Ignacio Hernández y Escolástica Vásquez por todo el apoyo incondicional, amor y regaños a lo largo de mi vida. Sin ustedes prácticamente sería un ente viviente sin alma ni sueños. Gracias por ser uno de los motores que me impulsa a seguir.

A mis **hermanos**, Josué, Larissa y Eneida por formar parte de mi vida y compartir tantas experiencias juntos. Los amo ¡canijos!

A mis amigos de la universidad, Oscar, Mario y Cecy, del L4 (Clau, Hugo, Ramiro, Adrián, Eduardo, Carlos, Astrid, Pris, Viky, Fer, Jessica 1 y 2, y los que faltaron por nombrar). Gracias por formar parte en este bonito camino universitario y académico.

A mis **hermanos de otra madre**, Eduardo, Yair, Martín. Gracias, hermanos por todas las aventuras y experiencias.

## **Dedicatoria**

Para Lorena y Luna,  
Luz en mi camino.

Para mis padres, Ignacio y Escolástica,  
Cimientos de mi espíritu.

	<b>Página</b>
<b>Lista de tablas</b> .....	VI
<b>Lista de figuras</b> .....	VII
<b>Nomenclatura</b> .....	VIII
<b>Resumen</b> .....	IX
<b>Abstract</b> .....	X
<b>1. Introducción</b> .....	1
<b>2. Antecedentes</b> .....	3
2.1 Especies de <i>Saccharomyces</i> protagonistas en la elaboración de cerveza.....	3
2.2 Tecnologías de secuenciación de RNA y herramientas bioinformáticas para el análisis de datos.....	5
2.3 Estudios transcriptómicos en levaduras de interés industrial.....	7
2.4 El problema central de la biotecnología y estudios de fermentación a escala industrial.....	10
<b>3. Justificación</b> .....	12
<b>4. Hipótesis</b> .....	12
<b>5. Objetivos</b> .....	12
5.1 Objetivo General.....	12
5.2 Objetivos específicos.....	12
<b>6. Materiales y Métodos</b> .....	13
6.1 Material biológico y medios de cultivo.....	13
6.2 Fermentación industrial y diseño de fermentaciones de laboratorio.....	13
6.3 Extracción de RNA.....	15
6.4 Análisis estadístico de los datos analíticos de los días 2 y 5 de fermentación.....	15
6.5 Preprocesamiento, ensamble y análisis de expresión diferencial de los datos provenientes de RNA-seq .....	16
6.6 Análisis de los datos transcriptómicos de la cepa 820 mediante la utilización de genoma de referencia proveniente de la cepa <i>S. cerevisiae</i> 288C.....	16
<b>7. Resultados</b> .....	18
7.1 Muestreo de cocimientos.....	18
7.2 Muestreo de los F2 y F5 de las fermentaciones de laboratorio e industrial.....	18
7.3 Análisis fisicoquímico del mosto de los días 2 y 5 de la fermentación.....	19
7.4 Preprocesamiento de los datos transcriptómicos.....	22
7.5 Ensamble <i>de novo</i> del transcriptoma de la levadura 820 y anotación.....	25
7.6 Análisis exploratorio y de expresión diferencial de genes en fermentación industrial y de laboratorio.....	26
7.6.1 Análisis exploratorio.....	26
7.6.2 Análisis de expresión diferencial de genes.....	28

7.7 Análisis de enriquecimiento por ontología de genes y rutas metabólicas.....	31
7.7.1 Análisis de enriquecimiento por ontología de genes de F2 industrial.....	32
7.7.2 Análisis de enriquecimiento por rutas metabólicas de F2 industrial.....	34
7.7.3 Análisis de enriquecimiento por ontología de genes de F2 laboratorio.....	36
7.7.4 Análisis de enriquecimiento por rutas metabólicas de F2 industrial.....	38
7.7.5 Análisis de enriquecimiento por ontología de genes de F5 industrial y laboratorio	39
7.7.6 Análisis de enriquecimiento por rutas metabólicas de F5 industrial y de laboratorio.....	41
7.8 Análisis del transcriptoma de la cepa 820 con genoma de referencia.....	41
7.8.1 Análisis exploratorio de Los datos transcriptómicos de la cepa 820 con genoma de referencia.....	41
7.8.2 Análisis de expresión diferencial de genes obtenidos con un enfoque guiado por genoma de referencia.....	42
7.8.3 Análisis de enriquecimiento por ontología de genes GO de los datos transcriptómicos obtenidos mediante genoma de referencia.....	44
7.8.4 Análisis de enriquecimiento por ontología de genes del conjunto de F2 industrial	44
7.8.5 Análisis de enriquecimiento por ontología de genes de F2 laboratorio.....	47
7.8.6 Análisis de enriquecimiento de rutas metabólicas de F2 industrial.....	48
7.8.7 Análisis de enriquecimiento de rutas metabólicas de F2 laboratorio.....	49
7.8.8 El reducido número de genes diferencialmente expresados en F5 no sustenta un análisis de enriquecimiento GO y KEGG.....	50
<b>8. Discusión.....</b>	<b>51</b>
8.1 Utilización de maltosa y maltotriosa.....	51
8.2 Metabolismo de aminoácidos, compuestos volátiles y de azufre a nivel industrial y laboratorio.....	52
8.3 Observaciones sobre las diferencias moleculares en el día 2 de la fermentación a nivel industrial y laboratorio.....	54
8.4 Las diferencias observadas en el día 5 de la fermentación a nivel industrial y laboratorio indican respuesta a estrés y crecimiento celular, respectivamente.....	57
8.5 Los datos transcriptómicos obtenidos con un enfoque guiados por genoma de referencia sustentan las observaciones en las comparaciones de F2 obtenido mediante ensamblaje <i>de novo</i> .....	60
8.6 Las diferencias observadas en F5 sugieren un aporte minoritario por parte del subgenoma de <i>S. cerevisiae</i> .....	61
<b>9. Conclusiones.....</b>	<b>63</b>
<b>10. Literatura citada.....</b>	<b>65</b>

## Lista de tablas

Tabla	Descripción	Página
I.	Porcentaje de carbohidratos y aminoácidos en el mosto inicial.....	13
II.	Esquema de muestreo de mosto .....	18
III.	Esquema de muestreos terminado de las fermentaciones a nivel industrial y laboratorio.....	19
IV.	Captación de Maltosa y Maltotriosa por parte de la levadura.....	20
V.	Tabla comparativa del consumo de aminoácidos en fermentación industrial y laboratorio.....	21
VI.	Producción de compuestos volátiles y de azufre.....	21
VII.	Datos de secuenciación de ARN.....	22
VIII.	Resultados por número de secuencias de los datos transcriptómicos que pasan el valor E indicando identidad en algún tipo de base de datos de ARNr. También se muestra las lecturas que no cumplieron con el valor E indicando que están libres de contaminación con RNAr.....	23
IX	Resultados de la limpieza de ARNr. Se representa el porcentaje de coincidencia de los archivos de lecturas de secuenciación con la respectiva base de datos de ARNr.....	24

## Lista de figuras

Figura	Descripción	Pagina
1.	Representación esquemática de un experimento típico en secuenciación de ARN y análisis de perfil de expresión.....	7
2.	Vista general del ensamblaje Trinity y del flujo de trabajo.....	8
3.	Representación esquemática de los eventos secuenciales con posible efecto de estrés, para la levadura, en el proceso cervecero.....	11
4.	Esquema general de las fermentaciones de laboratorio e industrial.....	14
5.	Resultados representativos globales del preprocesamiento de los datos de secuenciación.....	25
6.	Análisis exploratorio de los datos de cuantificación de transcritos.....	29
7.	Mapa de calor y agrupamiento jerárquico de los transcritos diferencialmente expresados en el día 2 de la fermentación industrial y de laboratorio.....	30
8.	Mapa de calor y agrupamiento jerárquico de los transcritos diferencialmente expresados en el día 5 de la fermentación industrial y de laboratorio.....	31
9.	Análisis de enriquecimiento por ontología de genes (GO) de los GDE's en F2 industrial.....	33
10.	Representación esquemática de la ruta de Ehrlich.....	35
11.	Gráfico de dispersión del análisis de enriquecimiento de vías metabólicas con el uso de la base de datos KEGG de los GDE's de F2 industrial.....	36
12.	Análisis de enriquecimiento por ontología de genes (GO) de los GDE's...	37
13.	Gráfico de dispersión del análisis de enriquecimiento por rutas metabólicas (KEGG) para el conjunto de genes diferencialmente expresados en F2 laboratorio.....	38
14.	Histogramas del análisis de enriquecimiento por ontología de los conjuntos de genes diferencialmente expresados (GDE's) en F5 industrial y laboratorio.....	40
15.	Análisis multidimensional a escala de los datos transcriptómicos provenientes de un enfoque guiado por genoma de referencia (S.c S288C).....	42
16.	Análisis de expresión diferencial de los genes obtenidos mediante genoma de referencia.....	43
17.	Análisis de enriquecimiento por ontología de genes de los datos transcriptómicos de F2 industrial, obtenidos mediante genoma de referencia.....	46
18.	Análisis por ontología de genes de los datos transcriptómicos obtenidos con genoma de referencia en F2 laboratorio.....	48
19.	Gráfico de dispersión del análisis de enriquecimiento de vías metabólicas con el uso de la base de datos KEGG de los GDE's de F2 industrial.....	49
20.	Gráfico de dispersión del análisis de enriquecimiento de vías metabólicas con el uso de la base de datos KEGG de los GDE's de F2 laboratorio.....	50

## Nomenclatura

<b>RNA-seq</b>	Secuenciación de ARN (por sus siglas en ingles)
<b>PCA</b>	Análisis de componentes principales (por sus siglas en ingles)
<b>GDE's</b>	Genes diferencialmente expresados
<b>Mal-3</b>	Maltotriosa
<b>Mal-2</b>	Maltosa
<b>S.c</b>	<i>Saccharomyces cerevisiae</i>
<b>S. e</b>	<i>Saccharomyces eubayanus</i>
<b>S. p</b>	<i>Saccharomyces pastorianus</i>
<b>GO</b>	Ontología de genes
<b>KEGG</b>	Enciclopedia de Kioto de genes y genomas (por sus siglas en ingles)
<b>k-mer</b>	Tamaño específico de una secuencia utilizada por algoritmos de ensamblaje
<b>CCM</b>	Cervecería Cuauhtémoc Moctezuma
<b>ORF</b>	Marco de lectura abierto (por sus siglas en inglés)
<b>Contigs</b>	Segmentos de ADN superpuestos que juntos representan una región consenso de ADN
<b>Nx (ej. N50)</b>	Se define como la longitud de los contigs tal que usando contigs de igual o mayor tamaño produce la mitad de las bases del genoma
<b>Cepa 820</b>	Levadura tipo lager en estudio

## Resumen

La levadura juega un papel central en la elaboración de cerveza, esencialmente debido a su metabolismo de azúcares y aminoácidos el cual se relaciona estrechamente a la producción de etanol, compuestos de alto valor sensorial como esteroides y alcoholes superiores, así como compuestos volátiles de azufre. El entendimiento del proceso cervecero está limitado a las observaciones de datos analíticos que denotan importancia práctica. Estos datos experimentales se han obtenido de ensayos a nivel de laboratorio y, en base a éstos, se realiza el escalamiento a nivel industrial. Sin embargo, el estudio del metabolismo de la levadura trasciende desde entornos pragmáticos hasta escenarios puramente básicos. Información previa, sobre enfoques con uso del conocimiento de la secuencia nucleotídica del total del RNA mensajero (RNA-seq, por sus siglas en inglés), expone un apoyo importante en la elucidación de distintos aspectos de la fisiología de la levadura, como su respuesta al estrés, adaptación a cocultivos en fermentación, metabolismo de azúcares y aminoácidos. En este trabajo determinamos el transcriptoma de una levadura lager en condiciones de fermentación industrial y de laboratorio. Con el fin de buscar correlación y/o posibles diferencias, realizamos comparación entre los días 2 y 5 de fermentación (F2 y F5) de cada condición. La comparación de los genes diferencialmente expresados (GDE) de F2, mostró 286 y 480 GDE, para el nivel industrial y laboratorio, respectivamente. Mientras que F5, fueron solo 143 y 298 GDE para el fermentador a nivel industrial y de laboratorio, respectivamente. El análisis ontológico mostró diferencias genéticas en distintos procesos biológicos y metabólicos. El análisis de los GDE en F2 a nivel industrial, mostró un sensible enriquecimiento (FDR < 0.05) de procesos implicados en el metabolismo de carbohidratos (sce00051, sce00052). Mientras que a nivel laboratorio, los genes soportan procesos biológicos relacionados con el ciclo celular mitótico y meiótico (sce04111 y sce04113, respectivamente). En F5 a nivel industrial se observaron procesos de respuesta a droga (GO:0042493) y transporte de amonio (GO:0072488), mientras que a nivel laboratorio procesos como regulación del desarrollo (GO:0050793), morfología celular (GO:0022604) y crecimiento polarizado (GO:0030) fueron observados. Estos datos sugieren que, en las condiciones industriales, la distribución de los nutrientes es muy diferente a su homólogo en laboratorio, esto en F2. Lo observado en las comparaciones de F5 evoca un panorama de estrés a nivel industrial mientras que a nivel laboratorio los eventos se asocian principalmente al crecimiento y morfología celular. Estas observaciones proponen que la distribución de nutrientes entre las condiciones sucede de forma diferencial provocando diferencias en el metabolismo de la levadura. Además, las condiciones estresantes son más acentuadas a nivel industrial lo cual provoca respuesta genética diferencial. Sin embargo, los datos analíticos para los compuestos aromáticos indican que no sobrepasan los umbrales de detección sensorial reportados. El mismo enfoque fue empleado con la utilización de genoma de referencia de *S. cerevisiae* S288C con el objetivo de corroborar los resultados obtenidos con genoma de referencia. Lo obtenido en las comparaciones de F2 sustentan lo ocurrido con enfoque *de novo* mientras que en las comparaciones de F5 se observaron una cantidad reducida de genes que no soportaron análisis ontológicos. La interpretación de estos resultados puede indicar que el fenotipo expresado en F5 proviene del subgenoma de *S. eubayanus*.

## Abstract

Yeast plays a fundamental role in brewing due to its sugar and amino acid metabolism which is correlated with alcohol, esters, and fusel alcohol production. There are two types of beer: Ale and lager, being the latter very important in an economic context. The understanding of the brewing process is limited to analytical data that denotes practical relevance. However, figure out yeast's metabolism has transcendental scenarios such as pragmatic and basic science. Previous literature, with RNA-seq approach, support interesting knowledge about yeast's physiology such as stress response, co-culture adaptation, sugar, and amino acid metabolism among others. In this work, the yeast's transcriptome in lab and industrial scale was determined. The comparisons were set out between day two (F2) and day five (F5) of fermentation. To find out possible differences, statistical test between conditions was carried out. Results describe below; 286 differential expressed genes (DEG's) was observed at industrial scale and 480 DEG's at lab scale. On other hand, 143 DEG's at industrial scale F5 and 298 DEG's at lab scale F5 was observed. Through ontological analyzes, these genetic differences were translated into different biological and metabolic processes. Comparisons in F2 show that at the industrial level processes as carbohydrate metabolism (sce00051, sce00052) they are significantly enriched (FDR <0.05). While at the laboratory level the genes support biological processes related to the mitotic and meiotic cell cycle (sce04111 and sce04113, respectively). At the industrial level F5, drug response processes (GO: 0042493) and ammonium transport (GO: 0072488) were observed, while at the laboratory level F5 processes such as development regulation (GO: 0050793), cell morphology (GO: 0022604) and polarized growth (GO: 0030) was supported. These data suggest that in industrial conditions, the distribution of nutrients is very different from their laboratory counterpart, this in F2. The observed in the F5 comparisons evokes a panorama of stress at the industrial level while at the laboratory level the events are mainly associated with cell growth and cell morphology. These observations propose that the distribution of nutrients between the conditions occurs differentially causing differences in yeast metabolism. Besides, stressful conditions are more pronounced at the industrial level which causes a differential genetic response. However, analytical data for aromatic compounds indicate that they do not exceed the reported sensory detection thresholds. The same approach was used with the use of the *S. cerevisiae* S288C reference genome to corroborate the results obtained with de novo transcriptome assembly. The results obtained in the F2 comparisons support what happened with the de novo approach, while in the F5 comparisons a small number of genes were observed that did not support ontological analyzes. The interpretation of these results may indicate that the phenotype expressed in F5 comes from the subgenome of *S. eubayanus*.

## 1. Introducción

La elaboración de cerveza es un proceso antiguo datado entre 6000 y 8000 años (Hornsey, 2003) sin embargo, su entendimiento científico se abordó con los trabajos realizados por Luis Pasteur quién en 1857 describió por primera vez la fermentación como un proceso de cambios bioquímicos en un medio orgánico realizados por un microorganismo (Barnett, 2000). Hoy en día se sabe que los microorganismos responsables de la fermentación son levaduras del género *Saccharomyces*, específicamente las especies *Saccharomyces cerevisiae* (Ale) y *Saccharomyces pastorianus* (Lager) (Boulton y Quain, 2001). Fenómenos de selección artificial han permitido la adaptación de cepas de *Saccharomyces* a nichos de fermentación industrial e incluso se cree que las cepas tipo lager han surgido a partir de un evento de hibridación entre *S. cerevisiae* y *S. eubayanus* (Gallone *et al.*, 2018).

Dado que el metabolismo de la levadura afecta directamente las propiedades físicas y químicas de la cerveza (Ferreira y Guido, 2018; Pires *et al.*, 2014), es importante determinar, fundamentalmente, qué procesos metabólicos están participando activamente. Distintos abordajes existen en la caracterización del metabolismo de la levadura en fase fermentativa, siendo más frecuente lo relacionado con el metabolismo de carbohidratos y especialmente su captación por parte de la levadura (Brickwedde *et al.*, 2018; Cousseau *et al.*, 2013; Vidgren *et al.*, 2010). Por otro lado, existen estudios que abarcan mayor profundidad como lo es la caracterización funcional de genes involucrados en la fermentación, secuenciación del genoma completo, dinámica metabólica en proceso de fermentación e incluso análisis comparativo de proteoma y transcriptoma en proceso involucrados en la afectación del producto final; cerveza o vino (Akao *et al.*, 2011; Bolat *et al.*, 2013; Olesen *et al.*, 2002).

Sin embargo, la mayor parte de los datos reportados en la literatura sobre los perfiles metabólicos de levaduras cerveceras y no cerveceras están desarrollados a escala de laboratorio. En el laboratorio 4 del instituto de Biotecnología, UANL se han desarrollado distintos trabajos que comprenden el metabolismo y producción de compuestos azufrados (Leal-Guerra 2008), análisis genómicos y proteómicos involucrados en la regulación de la expresión génica en la levadura (Herrera-Gamboa. 2018) y otro estudio que aborda en la regulación genética de la floculación

de levadura cervecera (Damas Buenrostro. 2008) y actualmente se está desarrollando un estudio de transcriptómica en la levadura cervecera 790 en las condiciones de dos mostos con contenido de nitrógeno diferente para elucidar el efecto en la fisiología de la levadura (Elizondo-González, 2019, trabajo de tesis doctoral en desarrollo).

Estos trabajos tienen una constante; todos, al igual que en la literatura, se han realizado a escala de laboratorio y por tanto es importante abordar estos aspectos de la levadura en un escenario que represente la realidad; fermentación a nivel industrial. En contexto, sólo una investigación aborda cuestiones relacionadas con el metabolismo de carbohidratos y otros procesos metabólicos de la levadura a escala industrial y los compara con resultados a nivel laboratorio realizados en pretérito (Gibson *et al.*, 2008). Los aportes que surgen de investigaciones a nivel industrial son de vital importancia ya que emanan interpretaciones que rozan estrechamente las fronteras de la biotecnología aplicada.

Uno de los enfoques que nos permitirá evaluar el perfil transcripcional de la levadura cervecera es mediante análisis del transcriptoma basado en secuenciación de RNA. Dado que el análisis de la regulación de la expresión genética ha sido abordado desde diferentes puntos de vista, indiscutiblemente, el advenimiento de las tecnologías de secuenciación de nueva generación ha hecho que el análisis de la expresión sea mucho más confiable y de mayor cobertura, especialmente en estudios relacionados con el metabolismo de cualquier organismo (Wang *et al.*, 2009).

En este trabajo planteamos elucidar el comportamiento metabólico, en las condiciones a nivel de laboratorio y a nivel industrial, de una levadura cervecera tipo lager, a través del análisis transcriptómico generado con la tecnología de secuenciación de RNA (RNA-seq, por sus siglas en inglés). Esto brindará importante información sobre distintos procesos metabólicos de la levadura analizada ya que el panorama de la secuenciación abarca prácticamente todos los transcritos presentes en las condiciones específicas de escalamiento analizadas en un mismo medio de cultivo.

## 2. Antecedentes

### 2.1 Especies de *Saccharomyces* y la elaboración de cerveza

La relación ancestral de la humanidad y la producción de bebidas fermentadas es datada entre 6000 y 8000 años de antigüedad. En la historia humana diferentes civilizaciones comenzaron los senderos sistemáticos en la elaboración de bebidas fermentadas. Por ejemplo, el aguamiel o vino de miel fue producido en Asia durante el periodo Védico, hace ~1700-1100 a.C. Los griegos, celtas, sajones y vikingos también producían este tipo de bebidas. En Egipto, Babilonia, Roma y China se producía vino de uvas y cerveza a partir de cebada malteada (Alba-Lois, L. y Segal-Kischinevzky, C, 2010a). En Sudamérica se producía la chicha a partir de granos o frutas siendo más frecuente el uso de maíz, en América del Norte el octil, ahora conocido como pulque, era producido a partir de agave, un tipo de cactus (Godoy *et al.* 2003).

Sin embargo, a pesar de su antigua precedencia, la elaboración de bebidas fermentadas permanecía en un estatus esencialmente empírico. Los trabajos magistrales de diferentes científicos brindaron las bases para la comprensión de distintos procesos realizados por microorganismos. Como lo fueron las observaciones de Robert Hook y Anton van Leeuwenhoek, siendo el último en describir en primera instancia a las levaduras provenientes de ambientes fermentativos (Gerard J. Tortora *et al.*, 2019; Alba-Lois, L. y Segal-Kischinevzky, C, 2010). No obstante, no fue sino hasta los trabajos realizados por Luis Pasteur (1857) que el entendimiento sobre el proceso de fermentación era realizado principalmente por levaduras (Barnett, 2000).

Hoy en día se sabe que levaduras del género *Saccharomyces* son los principales caracteres en las fermentaciones de bebidas alcohólicas y en la industria panadera. (Legras *et al.*, 2007). El conjunto denominado *Saccharomyces sensu stricto* abarca desde la industrialmente ubicua *Saccharomyces cerevisiae* hasta especies confinadas a delimitados nichos geográficos naturales. Este clado se conforma por las especies; *S. cerevisiae*, *S. paradoxus*, *S. cariocanus*, *S. mikatae*, *S. kudriavzevii*, *S. arboricolus*, *S. eubayanus*, *S. ovarum*, *Naumovozyma castelli*, así como el híbrido ínter específico entre *S. cerevisiae* y *S. eubayanus*; *S. pastorianus* (Borneman y Pretorius, 2014).

Las especies mejor adaptadas a los nichos industriales de elaboración de cerveza son *S. cerevisiae* y *S. pastorianus*. Originalmente, la clasificación de estas levaduras se basaba en la capacidad de floculación: fermentación o floculación superior; levaduras tipo Ale y fermentación o floculación inferior; levaduras tipo lager. Actualmente parámetros como la temperatura de fermentación, producción de compuestos volátiles, utilización de carbohidratos complejos son utilizados para una mejor descripción de las levaduras cerveceras (Lodolo *et al.*, 2008).

*S. cerevisiae* es la levadura más utilizada en cervezas tipo Ale y los fenotipos intrínsecos de esta especie son la adaptación a ambientes ricos en carbohidratos, ambientes reducidos en oxígeno, alta tolerancia a etanol y algunos rasgos adquiridos por la interacción en nichos de fermentación: utilización eficiente de maltotriosa y la nula producción de vinil guayacol (2-metoxi-4-vinilfenol), el cual es un compuesto que puede modificar el perfil sensorial de la cerveza (Gallone *et al.*, 2018). En cuanto a *S. pastorianus*, representa el 90% del mercado cervecero y es por mucho el estilo de cerveza más popular del globo. Las cepas de levadura tipo lager son categorizadas en dos linajes distintos conocidos comúnmente como “Saaz”; tipo 1 y “Frohberg”; tipo 2. La hipótesis más aceptada del surgimiento de la especie *S. pastorianus* involucra dos eventos de hibridación independientes cada uno con diferentes cepas tipo Ale domesticadas, *S. cerevisiae*, y una especie silvestre, *S. eubayanus* (Baker *et al.*, 2015).

En cuanto a los fenotipos industrialmente interesantes, ha sido sugerido que parte del subgenoma de *S. pastorianus* perteneciente a *S. eubayanus* brinda la característica de tolerancia al frío, mientras que el subgenoma que corresponde a *S. cerevisiae* proporciona los fenotipos antes mencionados como la fermentación y la utilización eficientes de maltotriosa. Sin embargo, existe algo de discusión en esta tesis ya que se ha observado que transportadores de maltotriosa provenientes de *S. eubayanus* promueven el metabolismo de maltotriosa en cepas de levaduras lager tipo 1 (Magalhães *et al.*, 2016). Igualmente, cepas de *S. cerevisiae* tienen rasgos que les permiten tolerar bajas temperaturas lo que implica que el aporte de tolerancia al frío, no necesariamente provienen del subgenoma perteneciente a *S. eubayanus* (García-Ríos *et al.*, 2017).

Aunque recientemente se ha reportado que el rasgo involucrado en la adaptación de las temperaturas está fuertemente asociado con el contenido genético proveniente en el DNA mitocondrial (ADNmt) de la levadura, brindando

apoyo a la idea de que el aporte del ADNmt proveniente de *S. eubayanus* está involucrado en la adaptación a bajas temperatura y que el ADNmt de *S. cerevisiae* orquesta el crecimiento óptimo en altas temperaturas (Baker *et al.*, 2019)

## **2.2 Tecnologías de secuenciación de RNA y herramientas bioinformáticas para el análisis de datos.**

Distintos enfoques han surgido durante los últimos 40 años que abordan la elucidación del orden correcto de las bases nucleotídicas en una muestra de ADN (secuenciación). A lo largo de la historia se han desarrollado tecnologías que enfrentaron este tipo de problemas. Las tecnologías de primera generación que se basan, esencialmente, en reacciones de amplificación interrumpida de fragmentos de ADN generados a partir de tratamientos enzimáticos o físicos de la muestra inicial (metodología “Shot-gun”), con la utilización de di-deoxi nucleótidos (ddNTP, por sus siglas en inglés). Seguido de esto, las tecnologías de secuenciación de segunda generación, también nombradas a partir de este punto; tecnologías de secuenciación de nueva generación se basaban principalmente en la implementación de procesamiento en paralelo, el cual es su característica más importante. Posteriormente, las tecnologías de secuenciación de tercera generación se independizaron de metodologías involucradas con la amplificación, alejándose así de los posibles errores cometidos en la amplificación. Su enfoque principal es la utilización de nano poros por donde pasa una molécula de cadena sencilla de ADN y los cambios en el potencial eléctrico son específicos de cada nucleótido dado su tamaño y conformación (Shendure *et al.*, 2017).

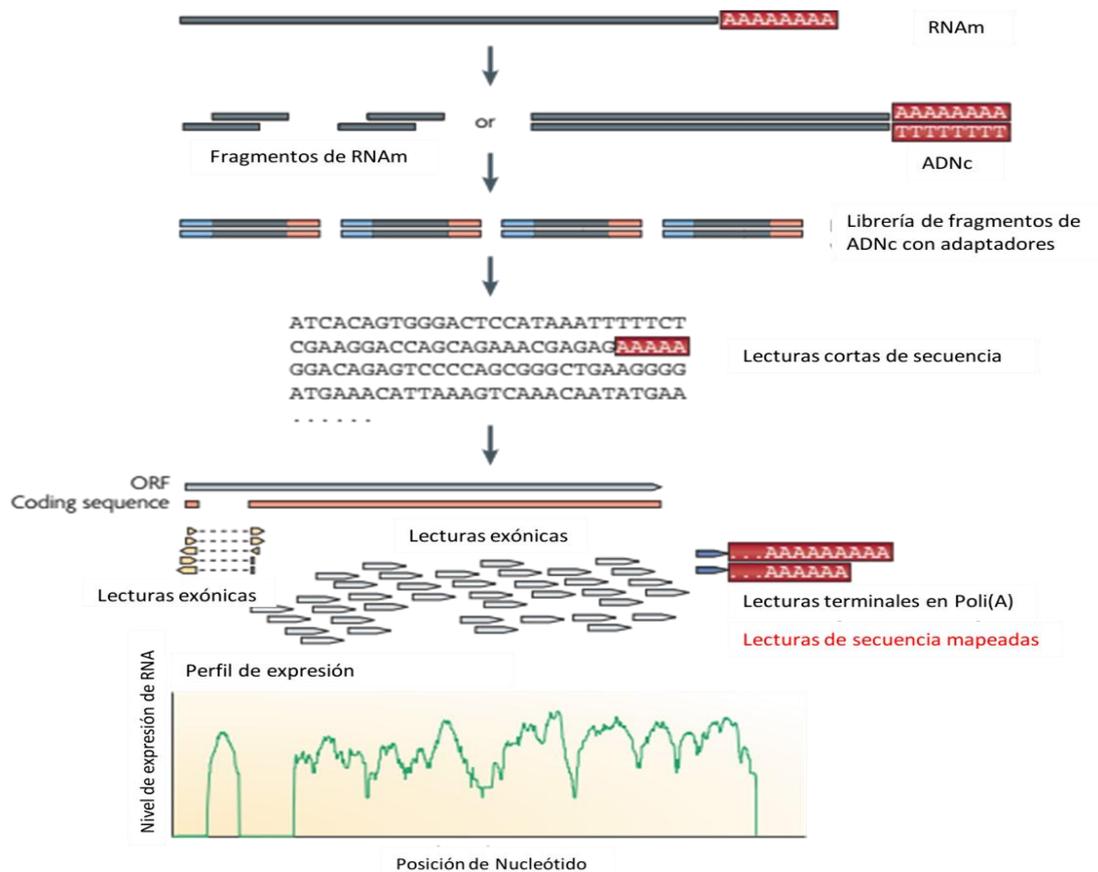
A partir de las tecnologías antes mencionadas ha surgido la tecnología de secuenciación de RNA la cual ayuda a abordar escenarios con un panorama mayoritariamente funcional, esto es el análisis de transcriptomas. El transcriptoma se define como el conjunto global de ARN mensajero (RANm) en una célula, en un estado específico de crecimiento o condición fisiológica. Y la importancia de su entendimiento radica esencialmente en la comprensión de la funcionalidad del genoma en fenotipos de interés como lo son las enfermedades, determinación de constituyentes moleculares en células y tejidos, entre otros aspectos importantes (Hrdlickova *et al.*, 2017). Un experimento típico de secuenciación de RNA-seq, se basa en que las muestras de ARN largos son convertidas primero a librerías ADN

complementario (ADNc) fragmentando, ya sea por la fragmentación directa de ARN o por la fragmentación de ADNc. Posteriormente son agregados los adaptadores a cada fragmento de ADNc y secuencias cortas (30-400 pb) son generadas por tecnologías de secuenciación de alto rendimiento. En seguida, las lecturas de secuencia resultantes son alineadas con un genoma de referencia o transcriptoma y son clasificadas en tres tipos; lecturas exónicas, lecturas de unión y lecturas terminales de poli(A). Estos tres tipos de lecturas son utilizadas para la construcción de un perfil de expresión de genes individuales (Figura 1) (Wang *et al.*, 2009).

Gracias a las nuevas tecnologías de secuenciación los costos de este tipo de proyectos se han disminuidos significativamente. Esto resulta en una producción de datos que evidencian la necesidad de herramientas bioinformáticas para su óptimo análisis. En el campo de RNA-seq no existe consenso entre las herramientas utilizadas para el análisis de datos. Si se cuenta con un genoma de referencia el enfoque se asimila a lo expuesto en la figura 1, donde las lecturas resultantes se alinean con un genoma de referencia para poder cuantificar las características del transcriptoma, seguido de la identificación propia de los transcritos y su expresión diferencial. Por otro lado, si no se cuenta con un genoma de referencia las lecturas resultantes del experimento RNA-seq necesitan ser ensambladas *de novo* para posteriormente alinear las lecturas con los transcritos ensamblados y seguir con los análisis posteriores (Conesa *et al.*, 2016).

Una revisión interesante expone algunas de las herramientas comúnmente utilizadas en el proceso de alineamiento y mapeo de las lecturas de secuencias, así como las herramientas involucradas en el análisis y determinación de genes diferencialmente expresados en los enfoques donde el genoma de referencia está debidamente anotado y normalizado. Además, se realiza una evaluación de rendimiento y comparación entre las herramientas utilizadas (Costa-Silva *et al.*, 2017). En contextos donde genoma o transcriptoma de referencia no está disponible el ensamble *de novo* de las lecturas de secuencia es el enfoque más utilizado. Para cumplir con este objetivo, el software Trinity es la herramienta que se expone en este trabajo. Los procesos del algoritmo se dividen en tres fases: i) Inchworm donde se construyen diccionarios de tamaño  $k$  (k-mer,  $k=25$  por defecto) de todas las secuencias y elimina los posibles errores en el diccionario. Se selecciona el k-mer más frecuente y se establece como semilla para la construcción del contig. ii) la fase de Crysalias se encarga de agrupar los contigs si existe una superposición de  $k-1$

bases, construye un grafo de Bruijn en cada conjunto para posteriormente determinar la ruta con mayor frecuencia y por último iii) la fase Butterfly que integra los grafos de Bruijn creados en Crysalius en conjunto con las lecturas originales para crear transcritos lineales de tamaño completo (Raychowdhury *et al.*, 2011). De esta manera se obtienen los transcritos ensamblados de forma eficiente y robusta para los análisis posteriores (Figura 2).



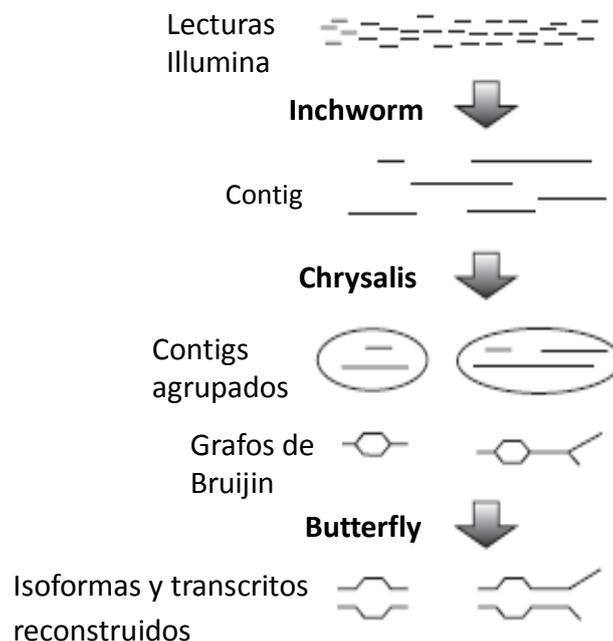
**Figura 1.** Representación esquemática de un experimento típico en secuenciación de ARN y análisis de perfil de expresión. Tomado y modificado de Wang y cols (2009).

### 2.3 Estudios transcriptómicos en levaduras de interés industrial

Los análisis transcriptómicos protagonizan un papel importante en la elucidación de los perfiles de expresión en especies económicamente importantes del género *Saccharomyces*. Distintos aspectos de la fisiología de *Saccharomyces cerevisiae* y otras levaduras de interés industrial han sido abordados mediante el uso de secuenciación de RNA. Por ejemplo, los mecanismos subyacentes en el estrés inducido por ácido acético en la levadura *S. cerevisiae* utilizada para la

producción de biocombustible (bioetanol) has sido eficazmente caracterizados, culminando con un conjunto de genes encargados en los mecanismos de respuesta. Esto permitió caracterizar la respuesta fisiológica de la célula de levadura lo cual es importante saber al momento de realizar fermentaciones para la producción de biocombustible (Dong *et al.*, 2017).

Otra investigación interesante muestra la utilidad de los enfoques de RNA-seq para generar conocimiento sobre la dinámica metabólica de tres cepas genéticamente distintas de *S. cerevisiae*. En dicha investigación se determinaron ciertos conjuntos de genes relacionados con la fermentación de harina de pan, siendo importantes los genes involucrados en el metabolismo de azúcares y aminoácidos, además de algunos genes involucrados en la respuesta fisiológica al estrés osmótico. Sin embargo, a pesar de los distintos fondos genéticos de las cepas utilizadas lograron determinar que el perfil metabólico era esencialmente el mismo entre las tres cepas (Aslankoochi *et al.*, 2013).



**Figura 2.** Vista general del ensamblaje Trinity y del flujo de trabajo. Se muestran los pasos clave en el algoritmo. Tomado y modificado de Raychowdhury y cols (2011).

Por otro lado, los experimentos de caracterización funcional de genes mediante deleciones (Knock out) o mutaciones es un método muy utilizado. Parte de

la caracterización funcional de dichos genes radica en el fondo transcripcional posterior a la mutación y dicha determinación es fácilmente abordada con enfoques de RNA-seq. Un ejemplo es lo reportado por Zhao y cols (2015) quienes mutaron una cepa de *S. cerevisiae* en el gen que codifica al factor de transcripción Taf25 y determinaron la perturbación global de dicha cepa mutante en la respuesta al estrés oxidativo. Esta investigación determinó que la vía de MAPK y la de cinasa A están íntimamente involucradas en la respuesta eficiente al estrés oxidativo en la cepa de levadura.

Otros enfoques, en RNA-seq, han sido empleados para ganar conocimiento en las fermentaciones mixtas que se usan en la producción del vino, el mapeo exhaustivo en el metabolismo de aminoácidos involucrado en fermentaciones de vino y otras interesantes investigaciones que abordan temas similares (Carvalho-Netto *et al.*, 2015; Cubillos *et al.*, 2019; Shekhawat *et al.*, 2019). La trascendencia pragmática de los estudios basados en RNA-seq saca a flote el potencial de su inmediata aplicación.

Por último y no menos importante están los resultados expuestos por Gibson y cols (2008) quienes idearon, en primera instancia, la hipótesis relacionada con la respuesta metabólica diferencial de una levadura frente en fermentaciones a nivel industrial y de laboratorio. Gibson y cols (2008) reportan que distintos genes relacionados con el metabolismo de carbohidratos están diferencialmente utilizados en los niveles de fermentación estudiados. A partir de estos resultados decidimos probar dicha hipótesis con el suceso probablemente homólogo que tienen lugar en Cervecería Cuauhtémoc Moctezuma.

La mayoría de los estudios recopilados aquí arrojan luz sobre varios problemas de biología fundamental que pueden ser utilizados para la extrapolación a contextos económicos. Esto es, creación de nuevas cepas de levadura más eficientes con fenotipos más resistentes a los distintos factores de estrés presentes en un ambiente de fermentación industrial.

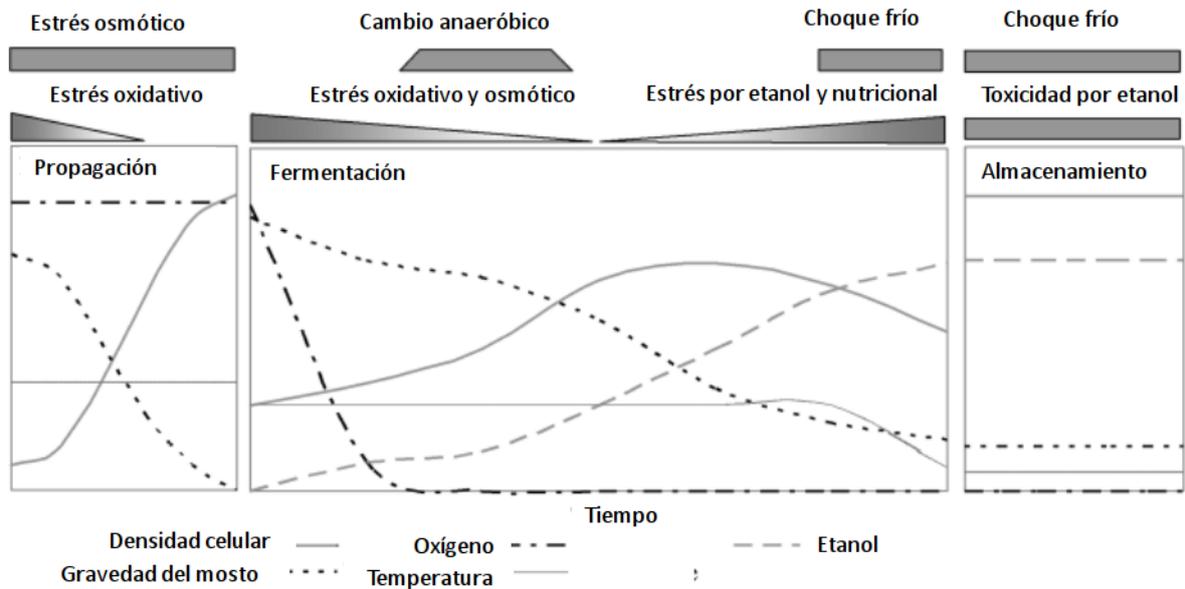
## **2.4 El problema central de la biotecnología y estudios de fermentación a escala industrial**

El escalamiento de procesos de fermentación como problema central de la biotecnología se planteó por primera vez durante la producción de penicilina en el año 1940. Como es evidente, el escalamiento es un proceso enfocado en la producción masiva de recursos biológicos. Sin embargo, distintos aspectos como la geometría y condiciones físicas del escalamiento pueden provocar efectos negativos en la fermentación culminando en bajos rendimientos del producto deseado (Schmidt, 2005). Específicamente en fermentaciones relacionadas con la producción de cerveza existe información que sustenta que la geometría y tamaño de los fermentadores tienen efectos en el rendimiento de la producción, afectando la distribución de los nutrientes (Casalta *et al.*, 2010; Speers y Stokes, 2009). Aunado a los efectos físicos del escalamiento también están presentes otros aspectos importantes que afectan la fisiología de la levadura. En el manejo industrial de la levadura cervecera se presentan distintas situaciones de estrés que afectan negativamente la fisiología de la célula. Estas condiciones abarcan las fluctuaciones en la concentración de oxígeno, potencial osmótico, pH, concentración de etanol, disponibilidad de nutrientes y temperatura. Estos factores de estrés para la célula están presentes en todo el proceso cervecero, desde la propagación hasta el almacenamiento de la cerveza (Figura 2) (Gibson *et al.*, 2007).

Aunado a los efectos mencionados anteriormente, es importante mencionar que existen otros aspectos como la heterogeneidad fenotípica en poblaciones clonales de microorganismos. Esta diferencia fenotípica está dada principalmente por efectos estocásticos. Estos efectos estocásticos surgen del movimiento Browniano de biomoléculas de bajo número de copias involucradas en la expresión génica (Takhaviev & Heinemann, 2018). Existen investigaciones que sustentan que la expresión estocástica de genes está involucrada en la heterogeneidad fenotípica, específicamente, a nivel de transcritos y proteínas (Elowitz *et al.*, 2002; Golding *et al.*, 2005; Raj y van Oudenaarden, 2008).

Debido a lo antes expuesto es necesaria la caracterización transcripcional y metabólica de la levadura cervecera en condiciones industriales para tener marcos de referencia más acertados y más cercanos a la realidad. Algunas investigaciones interesantes elucidan la dinámica metabólica de una levadura cervecera y el efecto que tienen la edad generacional en la calidad de la cerveza, en contextos puramente industriales (Kucharczyk *et al.*, 2018; Olesen *et al.*, 2002). No obstante, la información sobre comportamientos metabólicos de la levadura en fermentación

industrial es limitada y es necesaria la aportación de nuevos descubrimientos para tener un mayor control en procesos y por ende construir plataformas de mejoramiento del proceso.



**Figura 3.** Representación esquemática de los eventos secuenciales con posible efecto de estrés, para la levadura, en el proceso cervecero. Tomado y modificado de Gibson y cols (2007).

### **3. Justificación**

El entendimiento del metabolismo de levaduras de interés industrial está limitado a análisis fisicoquímicos. Sin embargo, el conocimiento básico del transcriptoma permite entender a profundidad lo que sucede a nivel molecular en las condiciones específicas de fermentación tanto a nivel industrial como laboratorio. Traduciendo este conocimiento básico en herramientas de posible aplicación en desarrollo y/o solución de problemas a nivel industrial además de contribuir en el enriquecimiento sobre el conocimiento de la bioquímica de la levadura en fermentación industrial y de laboratorio.

### **4. Hipótesis**

La levadura, en fermentaciones a nivel laboratorio e industrial, no posee la misma regulación de la expresión genética y/o una correlación lineal de otros eventos moleculares.

### **5. Objetivos**

#### **5.1 Objetivo General**

Analizar y comparar el transcriptoma de la cepa 820 en fermentación a nivel industrial y de laboratorio

#### **5.2 Objetivos Específicos**

1. Diseñar fermentaciones a nivel laboratorio con el mismo medio de cultivo que se utilizará en la fermentación industrial.
2. Extraer RNA total de levadura tomada en los días 2 y 5 de fermentación industrial y de laboratorio.
3. Ensamblar el transcriptoma *de novo* mediante la utilización de software Trinity-v2.8.4.
4. Obtener de los genes diferencialmente expresados en las condiciones mediante la utilización de software estadístico R versión 3.5.2 (2018-12-20).
5. Realizar análisis funcionales de los genes diferencialmente expresados mediante ontología de genes y rutas metabólicas.
6. Analizar los datos transcriptómicos con genoma de referencia

## 6. Materiales y Métodos

**6.1 Material biológico y medios de cultivo.** En este trabajo se empleó la cepa cervecera tipo lager 820. Las condiciones de manipulación y cultivo representan puntualmente el protocolo industrial. En breve: la cepa debidamente propagada, se recopiló a partir de los tanques de inoculación en las instalaciones de fermentación. Estas muestras de levadura fueron utilizadas para las fermentaciones que se diseñaron a escala laboratorio. El medio de cultivo utilizado constó de mosto empleado por la empresa en la elaboración de cerveza. La Tabla 1 muestra la composición del mosto utilizado.

Debido a que la fermentación industrial se replicó a escala de laboratorio, fue necesaria la obtención de mosto. Por esto, el muestreo de mostos se estableció de acuerdo con lo estipulado en el protocolo industrial.

**Tabla 1.** Porcentaje de carbohidratos y aminoácidos en el mosto inicial.

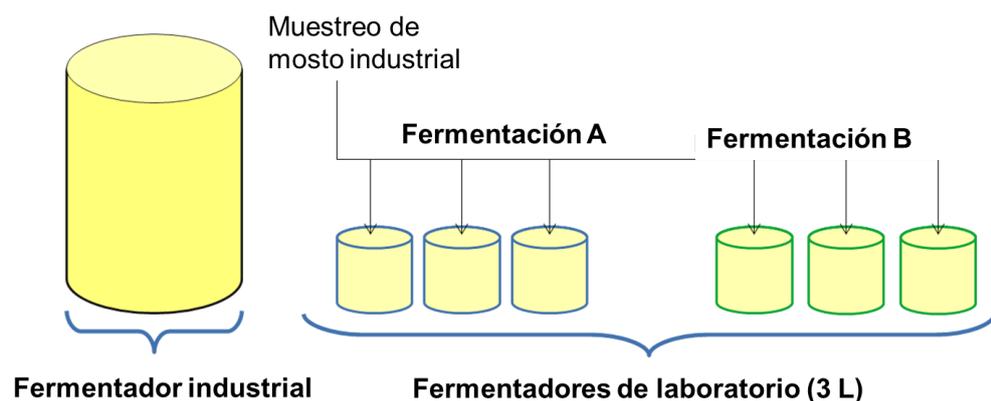
Carbohidratos (% p/p)							Aminoácidos (ppm)									
Glu	Fru	Mal-2	Mal-3	E	D	S	K	V	M	L	I	H	G	F	T	P
4.85	0.11	11.09	3.05	2.5	3.2	7.2	6.5	7.7	1.3	10.3	4.3	2.2	2.7	5.5	3.4	19.4

Glu: Glucosa, Fru: Fructosa, Mal: Maltosa y Mal-3: Maltotriosa. Los aminoácidos están representados en código de una letra.

**6.2 Fermentación industrial y diseño de fermentaciones de laboratorio.** En total tres conjuntos de fermentaciones fueron llevadas a cabo: i) fermentación industrial (6700 hL por parte CCM), ii) fermentación de laboratorio a la par de la fermentación industrial (réplica experimental) y iii) la réplica biológica. Ambas fermentaciones de laboratorio con un volumen de 2.5 L. En breve: la fermentación industrial se llevó a cabo estrictamente como lo indica el protocolo de la empresa. La elaboración y establecimiento de la fermentación fueron controladas por personal capacitado. Las directrices empleadas por el protocolo industrial sirvieron de base para la fermentación a escala laboratorio. El proceso industrial se resume en el llenado de un tanque con 8 cocimientos y su inoculación simultánea en el primer cocimiento. El inóculo utilizado fue el de la cepa 820 y las condiciones de inoculación y fermentación fueron establecidas de acuerdo con el protocolo industrial. Para las

fermentaciones a nivel de laboratorio es importante enfatizar que el mismo mosto utilizado a nivel industrial fue usado a nivel laboratorio, dejando como condición las dos diferentes escalas. Una vez obtenidos los mostos, fueron mezclados en matraces bola de 6 L, oxigenados por 5 min y posteriormente vertidos en los biorreactores de laboratorio, el proceso de vaciado se realizó bajo técnicas asépticas con mechero de uso industrial. La fermentación a escala laboratorio, representó puntualmente las mismas condiciones anteriormente descritas.

Los cálculos para escalar el inóculo fueron realizados por personal capacitado de la empresa. La inoculación y la fermentación fueron réplicas de lo antes mencionado. La única variación que se presentó en la réplica biológica fue la edad generacional de la cepa cervecera utilizada. Entiéndase como edad generacional al proceso de reutilización de levadura de una fermentación en una posterior. Los cálculos para el inóculo fueron de igual forma, realizados por personal de la empresa. Las dos fermentaciones de laboratorio se realizaron por triplicado necesario para enfoques estadísticos. Es importante destacar que debido a la dificultad de repetir fermentaciones a nivel industrial solo se realizó una vez y esta muestra se utilizó como referencia en las comparaciones posteriores. La Figura 4 expone, de forma resumida el diseño de las fermentaciones de laboratorio e industrial.



**Figura 4.** Esquema general de las fermentaciones de laboratorio e industrial. La fermentación **A** corresponde a la réplica experimental de lo sucedido a nivel industrial y la Fermentación **B** corresponde a la réplica biológica.

De cada conjunto de fermentaciones se tomaron muestras de 10 mL de levadura a los días 2 y 5 de la fermentación, así como otro conjunto de muestras

para los análisis fisicoquímicos, de carbohidratos, aminoácidos, producción de alcohol, compuestos volátiles y de azufre. Se recuperó la pastilla celular por medio de centrifugación, 3000 rpm por 5 min a 4°C, posteriormente la biomasa obtenida se homogenizó en amortiguador PBS 1X, seguido de una congelación instantánea en nitrógeno líquido y su almacenamiento hasta su uso en experimentos posteriores.

**6.3 Extracción de RNA.** Para la extracción de RNA el método de trizol (TRI reagent Sigma Aldrich) fue empleado y el procedimiento fue realizado de acuerdo con el proveedor y a Rio C. D y cols (2012) con algunas modificaciones que se mencionan a continuación. En breve; se recuperó la levadura por centrifugación a 3,000 rpm seguido de la adición de un volumen apropiado de TRI Reagent se incubó por 15 min, luego se adicionó 200 µL de cloroformo y se incubó por 10 min. Posteriormente se centrifugó a 12,000 x g por 15 min. La fase acuosa fue transferida a un tubo nuevo y se agregó 0.5 mL de 2-propanol por cada mL de TRI Reagent. Se incubó a temperatura ambiente de 5-10 min y fue centrifugada a 12,000 x g por 10 min a 4°C. La pastilla celular fue recuperada y lavada con 1 mL de etanol al 75% por cada mL de TRI Reagent, se agitó en vortex y se centrifugó a 7,500 x g por 5 min a 4°C. La pastilla de ARN fue secada por aireación, seguido de la adición de agua DEPC y mezcla por pipeteo. Las muestras de ARN total fueron sujetas a electroforesis en gel de agarosa al 1 % en buffer TAE para ver su correcta extracción, además de su verificación de calidad del RNA mediante el análisis de chips de ARN (Agilent Technologies). Una vez verificada la calidad del ARN se procedió al envío a GENEWIZ para su secuenciación. La secuenciación fue específica para ARN mensajeros (ARNm).

**6.4 Análisis estadístico de los datos analíticos de los días 2 y 5 de fermentación.** Con el fin de buscar diferencias entre el consumo, por parte de la levadura, de carbohidratos, aminoácidos y la producción de etanol, compuestos volátiles y de azufre se empleó una prueba de t-Student de una sola muestra tomando como media poblacional los datos obtenidos a nivel industrial, esto debido a la dificultad de realizar réplicas experimentales y biológicas a nivel industrial. Para cumplir con dicho análisis el software estadístico R versión 3.5.2 fue utilizado.

**6.5 Preprocesamiento, ensamble y análisis de expresión diferencial de los datos provenientes de RNA-seq.** Una vez obtenidas las lecturas de secuenciación, estas fueron sujetas a análisis de preprocesamiento con el fin de verificar la calidad de las secuencias, búsqueda de errores y/o contaminación, así como la remoción de adaptadores y lecturas con baja calidad que puedan afectar los análisis posteriores, esto para disminuir todo lo posible la variación técnica y dar más importancia a la variación biológica y experimental. Para cumplir con la verificación y control de calidad de las bases el software FASTQC versión 0.11.8 fue utilizado, el cual permitió ver distintas métricas relacionadas con la calidad de las lecturas de secuenciación, como lo es la distribución de tamaño de las lecturas, el valor Phred de las bases nucleotídicas en las lecturas de secuenciación, presencia de adaptadores entre otros (Simon Andrews, <https://www.bioinformatics.babraham.ac.uk/index.html>, revisado el día 15 de febrero del 2019).

Para efectuar el procesamiento antes del ensamble, las lecturas crudas sirvieron de entrada para el programa Trimmomatic versión 0.38, el cual realizó el filtrado de lecturas con bajos valores de calidad, adaptadores, entre otras opciones disponibles (Bolger et al., 2014). Para filtrar posible contaminación de ARN ribosomal el software SortMeRNA versión 2.1 fue utilizado (Kopylova et al., 2012). Posteriormente las lecturas filtradas fueron utilizadas en el programa Trinity versión 2.8.4 para el ensamble de las lecturas de secuenciación (Raychowdhury et al., 2011), seguido de una anotación funcional de dichos transcritos mediante el uso de Trinotate (Bryant et al., 2017). Al obtener los transcritos debidamente anotados se siguió con el análisis de genes diferencialmente expresados usando el software R versión 3.5.2 (2018-12-20), en especial el paquete Bioconductor-EdgeR versión 3.8 (Bioconductor, 2019).

**6.6 Análisis de los datos transcriptómicos de la cepa 820 mediante la utilización de genoma de referencia proveniente de la cepa *S. cerevisiae* 288C.** Con el fin de brindar soporte a los análisis realizados mediante ensamble *de novo* del transcriptoma de la cepa de estudio, se procedió a realizar un análisis con el mismo enfoque diferenciando sólo la utilización de genoma de referencia. Para cumplir con este objetivo se utilizó el flujo de trabajo de los programas STAR versión

2.7, HTSeq y edgeR/Bioconductor. El programa STAR (Dobin *et al.*, 2013) se utilizó para alinear las lecturas de secuenciación, ya procesadas, contra el genoma de referencia (S.c 288C). El genoma de referencia constó de la versión R64-1-1 de la cepa *Saccharomyces cerevisiae* S288C y su descarga fue desde el portal de buscador de genomas Ensembl genome browser 97 ([ftp://ftp.ensembl.org/pub/release-97/fasta/saccharomyces\\_cerevisiae/dna/](ftp://ftp.ensembl.org/pub/release-97/fasta/saccharomyces_cerevisiae/dna/)), de igual forma las anotaciones de dicho genoma se obtuvieron de la misma plataforma en formato gff3 ([ftp://ftp.ensembl.org/pub/release-97/gff3/saccharomyces\\_cerevisiae/](ftp://ftp.ensembl.org/pub/release-97/gff3/saccharomyces_cerevisiae/)). El resultado del programa STAR fueron archivos con la extensión SAM/BAM los cuales constaron de las alineaciones de las lecturas hacia el genoma de referencia. Estos archivos sirvieron de entrada para el programa HTSeq (Anders *et al.*, 2015).

Una vez alineadas las lecturas de secuenciación en el genoma de referencia, se efectuó un procesamiento con el programa HTSeq con el parámetro HTSeq-count, esto para contabilizar y brindar un valor numérico a la proporción de lecturas alineadas en cada gene del genoma de referencia. El archivo de salida obtenido fue un archivo txt en formato de tabla, en donde las filas representa los genes del genoma de referencia que presentaron al menos una lectura mapeada y las columnas representan la condición de nuestro experimento. Una vez obtenida la matriz de conteos de los genes se siguió con un análisis de expresión diferencial de genes mediante el programa edgeR siguiendo el fuljo de trabajo expuesto por Chen y cols (2016).

## 7. Resultados

### 7.1 Muestreo de cocimientos

Los muestreos del mosto estuvieron divididos en 8 cocimientos espaciados temporalmente por 3 horas en promedio cada uno. La Tabla 2 expone el esquema de muestreo de cocimientos concluido, incluyendo el día, hora y número de cocimiento. Cabe mencionar que al momento de los muestreos no se presentó ningún inconveniente que pudiera afectar los experimentos posteriores.

Tabla 2. Esquema de muestreo de mosto concluido

Muestra A	Muestra B (réplica)	No. mosto/cocimiento	Hora	Día
A1	B1	3999	02:30:00 pm*	
A2	B2	4001	04:51:00 pm	
A3	B3	4004	07:40:00 pm	1
A4	B4	4006	10:05:00 pm	
A5	B5	4007	01:50:00 am	2
A6	B6	4009	03:30:00 am	
A7	B7	4012	06:53:00 am	
A8	B8	4014	08:31:00 am	

*\*Hora de inoculación en fermentación industrial*

### 7.2 Muestreo de levadura en F2 y F5 de las fermentaciones de laboratorio e industrial

Las muestras de levadura se tomaron de F2 y F5 en cada escala. Es de importancia enfatizar que, debido a la hora de inoculación, tomada como hora cero de la fermentación industrial, sucedió en el primer cocimiento; 2:30 pm del día 1. El programa de fermentación en laboratorio se desfaso en tiempo ya que el último cocimiento se obtuvo hasta la hora 8:13 am del día 2 (Tabla 2). Esto resultó en el establecimiento de las fermentaciones de laboratorio hasta el día 2, tiempo en el que se obtuvieron todos los cocimientos necesarios para las fermentaciones a nivel laboratorio. En la Tabla 3 se puede observar el esquema de muestreo de las fermentaciones.

**Tabla 3.** Esquema de muestreos terminado de las fermentaciones a nivel industrial y laboratorio

Muestra	Descripción	Fecha
F2I	4 tubos de 10 mL* + 100 mL*	15/09/2018
F5I	4 tubos de 10 ml* + 4.5L***	18/09/2018
1-LF2		
2-LF2		
3-LF2	4 tubos de 10 ml* + 300 mL**	16/09/2018
1-LF5		
2-LF5		
3-LF5	4 tubos de 10 ml* + 4.5L**	19/09/2018
4-LF2		
5-LF2		
6-LF2	4 tubos de 10 ml* + 300 mL**	28/09/2018
4-LF5		
5-LF5		
6-LF5	4 tubos de 10 ml* + 4.5L**	01/10/2018

\*Muestras empleadas en RNA-seq; \*\*Análisis de fisicoquímicos en la empresa

\*\*\* Análisis sensorial por parte de cervecería

Las muestras listadas en la tabla 3 fueron sujetas a extracción de ARN y posteriormente enviadas a la empresa Genewiz (NJ, USA) para la secuenciación de ARN mensajero (ARNm).

### 7.3 Análisis fisicoquímico del mosto de los días 2 y 5 de la fermentación

Con el fin de proporcionar evidencia sobre el consumo de algunos componentes del mosto se emplearon herramientas analíticas para determinar la proporción de cada componente; los azúcares se midieron en el mosto inicial, F2 y F5 a nivel industrial y laboratorio. Los resultados obtenidos indican un patrón de consumo empezando por la glucosa y fructosa que en el mosto inicial representaron 4.85 y 0.11 %(p/p) respectivamente. En F2 dichos carbohidratos se manifestaron como no detectables dado el consumo primordial que impartió la levadura. Los eventos respecto a los carbohidratos complejos como la maltosa y la maltotriosa se presentaron de forma diferente al representar un porcentaje de carbohidrato residual; para la maltosa en F5 laboratorio se observó un 0.49 %(p/p) y un 1.44%(p/p) para la maltotriosa, mientras que en F5 industrial fue de 0.57 %(p/p) de maltosa y 1.7 %(p/p) de maltotriosa. Tomando en cuenta la proporción inicial en el mosto, este evento de carbohidrato residual representa un 4.43% y un 47.21% de

carbohidrato residual a nivel laboratorio, para maltosa y maltotriosa respectivamente. Para F5 industrial se presentó un 5.15% de maltosa residual y un 55.92% de maltotriosa. Por otro lado, el consumo de maltosa y maltotriosa en F2 fue significativamente mayor en laboratorio (Tabla 4. En el día cinco el consumo de maltosa se manifestó igual en ambas condiciones y el consumo de maltotriosa fue significativamente mayor en fermentación de laboratorio.

**Tabla 4.** Captación de Maltosa y Maltotriosa por parte de la levadura.

	Inicial	Ind F2	Lab F2	Ind F5	Lab F5
<b>Maltosa</b>	11.06	7.32	4.82 ± 0.24*	0.57	0.49 ± 0.061
<b>Maltotriosa</b>	3.05	2.52	2.16 ± 0.11*	1.7	1.44 ± 0.1*

*Ind: Fermentación industrial, Lab: Fermentación a nivel laboratorio. \* indica diferencia significativa ( $P < 0.05$ )*

Otro componente importante del mosto es el FAN (amino-nitrógeno libre), que representa la fuente de nitrógeno proveniente de aminoácidos o pequeños oligopéptidos que la levadura utiliza para su crecimiento. Los resultados indican que en el día 2 el consumo de glutamina, ácido aspártico, valina, leucina, isoleucina, fenilalanina y tirosina es significativamente mayor en las fermentaciones de laboratorio mientras que el consumo de metionina en el mismo día es mayor en la fermentación industrial. El consumo de aminoácidos en el día cinco representa un consumo significativamente mayor en las fermentaciones de laboratorio para los aminoácidos valina, lucina, tirosina y prolina. Algo interesante es que en el aminoácido Prolina se observó un aparente aumento desde el día dos al cinco en la fermentación industrial (Tabla 5).

Por último, la medición de compuestos aromáticos como ésteres, alcoholes de cadena larga y compuestos de azufre fue realizada solo en F5. Los resultados apuntan que la mayor producción de compuestos volátiles como acetato de etilo, acetato de isoamilo, hexanoato de etilo, propanol, isobutanol y alcohol isoamílico fue significativamente mayor su producción en las fermentaciones de laboratorio. En cuanto a los compuestos de azufre el sulfuro de hidrogeno, metanotiol, sulfuro de etileno y S-metiltioacetato fueron mayormente producidos en las fermentaciones de

laboratorio y el dimetilsulfuro fue significativamente mayor en la fermentación industrial (Tabla 6).

**Tabla 5.** Datos comparativos del consumo de aminoácidos en fermentación industrial y laboratorio. Los datos mostrados están en ppm.

AA	Inicial	Ind F2	Lab F2	Ind F5	Lab F5
<b>Glu</b>	2.5	2.1	1.117 ± 0.1327*	0.3	0.25 ± 0.067
<b>Asp</b>	3.2	0.8	0.31 ± 0.0862*	0.08	0.065 ± 0.011
<b>Ser</b>	7.2	0.3	0.357 ± 0.125	0.07	0.065 ± 0.011
<b>Lys</b>	6.5	0.5	0.8 ± 0.171	0.1	0.1 ± 0.0213
<b>Val</b>	7.7	2.3	0.135 ± 0.04*	0.05	0.033 ± 0.0055*
<b>Met</b>	1.3	0.03*	0.095 ± 0.005	0.2	0.183 ± 0.047
<b>Leu</b>	10.3	0.4	0.1 ± 0.033*	0.05	0.045 ± 0.0084
<b>Ile</b>	4.3	0.4	0.062 ± 0.012*	0.03	0.022 ± 0.003*
<b>His</b>	2.2	0.9	0.833 ± 0.154	0.1	0.09 ± 0.023
<b>Gly</b>	2.7	2.2	2.15 ± 0.41	0.5	0.46 ± 0.08
<b>Phe</b>	5.5	1.4	0.043 ± 0.013*	0.03	0.018 ± 0.0065
<b>Tyr</b>	3.4	1.6	0.121 ± 0.025*	0.1	0.065 ± 0.012*
<b>Pro</b>	19.4	14.7	13.467 ± 1.047	19	12.833 ± 1.67*

**Tabla 6.** Producción de compuestos volátiles y de azufre (ppm).

Compuestos volátiles	Industrial	Laboratorio
Acetaldehído	7.4	7.13 ± 0.29
Acetato de etilo	17.4	29.9 ± 2.056*
Propanol	9.7	13 ± 0.28 *
Isobutanol	9.9	15.516 ± 0.188*
Acetato de isoamilo	1.6	3.16 ± 0.25*
Alcohol Isoamílico	74.0	107.52 ± 2.02*
Hexanoato de etilo	0.2	0.33 ± 0.021*
Butanodiona	172.9	206.55 ± 35 ppb
Pentanodiona	100.5	141.42 ± 24.43 ppb
Compuestos de azufre		ppb
Sulfuro de hidrógeno	29.0	64.18 ± 5.38*
Metanotiol	1.6	2.3 ± 0.093*
Dimetilsulfuro	25.1	17.51 ± 1.3*
Sulfuro de etileno	0.4	0.55 ± 0.056*
S-metiltioacetato	8.3	12 ± 0.85*

\* Indica diferencia significativa ( $P < 0.05$ )

#### 7.4 Preprocesamiento de los datos transcriptómicos.

Los datos crudos de la secuenciación de ARN constaron de 28 archivos dos para cada muestra con un promedio de 22 millones de lecturas. Los 28 archivos corresponden a las 14 muestra de RNA obtenidas de los días dos y cinco de la fermentación industrial y laboratorio. Dada la plataforma de secuenciación se obtuvieron dos archivos representando un conjunto de lecturas forward y reverse para cada muestra. La nomenclatura de los archivos siguió una progresión numérica desde 1 al 14, representando el código de cada muestra, seguido de R1 para las lecturas forward o R2 para las lecturas reverse (Tabla 7). La longitud de las secuencias fue de 150 bases y el contenido GC se manifestó en 43% para todos los archivos analizados.

**Tabla 7.** Datos de secuenciación de ARN.

<b>Archivo</b>	<b>Muestra</b>	<b>No. secuencias</b>
1_R1_001.fastq.gz	F2 Ind forward	16684598
1_R2_001.fastq.gz	F2 Ind reverse	
2_R1_001.fastq.gz	F5 Ind forward	15844645
2_R2_001.fastq.gz	F5 Ind reverse	
3_R1_001.fastq.gz	F2 Lab-1 forward	19251976
3_R2_001.fastq.gz	F2 Lab-1 reverse	
4_R1_001.fastq.gz	F2 Lab-2 forward	19936414
4_R2_001.fastq.gz	F2 Lab-2 reverse	
5_R1_001.fastq.gz	F2 Lab-3 forward	25353412
5_R2_001.fastq.gz	F2 Lab-3 reverse	
6_R1_001.fastq.gz	F5 Lab-1 forward	21343668
6_R2_001.fastq.gz	F5 Lab-1 reverse	
7_R1_001.fastq.gz	F5 Lab-2 forward	21468599
7_R2_001.fastq.gz	F5 Lab-2 reverse	
8_R1_001.fastq.gz	F5 Lab-3 forward	19756854
8_R2_001.fastq.gz	F5 Lab-3 reverse	
9_R1_001.fastq.gz	F2 Lab-4 forward	21446884
9_R2_001.fastq.gz	F2 Lab-4 reverse	
10_R1_001.fastq.gz	F2 Lab-5 forward	19260248
10_R2_001.fastq.gz	F2 Lab-5 reverse	
11_R1_001.fastq.gz	F2 Lab-6 forward	25984135
11_R2_001.fastq.gz	F2 Lab-6 reverse	
12_R1_001.fastq.gz	F5 Lab-4 forward	19650081
12_R2_001.fastq.gz	F5 Lab-4 reverse	
13_R1_001.fastq.gz	F5 Lab-5 forward	23607240
13_R2_001.fastq.gz	F5 Lab-5 reverse	
14_R1_001.fastq.gz	F5 Lab-6 forward	25283113
14_R2_001.fastq.gz	F5 Lab-6 reverse	

Para cumplir con el preprocesamiento de las lecturas de secuenciación de ARN se utilizó el software SortMeRNA versión 2.1 (Kopylova *et al.*, 2012) el cual permitió la eliminación de posibles contaminaciones de ARN ribosomal. Para esto, se construyó una base de datos correspondiente a secuencias de RNA ribosomal 16S y 23S para organismos procariotas y 18S, 28S para organismos eucariotas. Una vez construida la base de datos de ARNr se analizaron los archivos crudos de las lecturas de secuenciación. Debido a que se obtuvieron dos archivos por muestra, secuencias forward y reverse, se concatenaron ambos archivos en uno solo para cumplir con el criterio de comparación con la base de datos de ARNr. Los resultados de la limpieza de ARNr por número de secuencias que coinciden con la base de datos de ARNr se muestran en la Tabla 8 y los resultados por base de datos de ARNr se muestra en la Tabla 9.

**Tabla 8.** Resultados por número de secuencias de los datos transcriptómicos que pasan el valor E indicando identidad en algún tipo de base de datos de ARNr. También se muestra las lecturas que no cumplieron con el valor E indicando que están libres de contaminación con RNAr

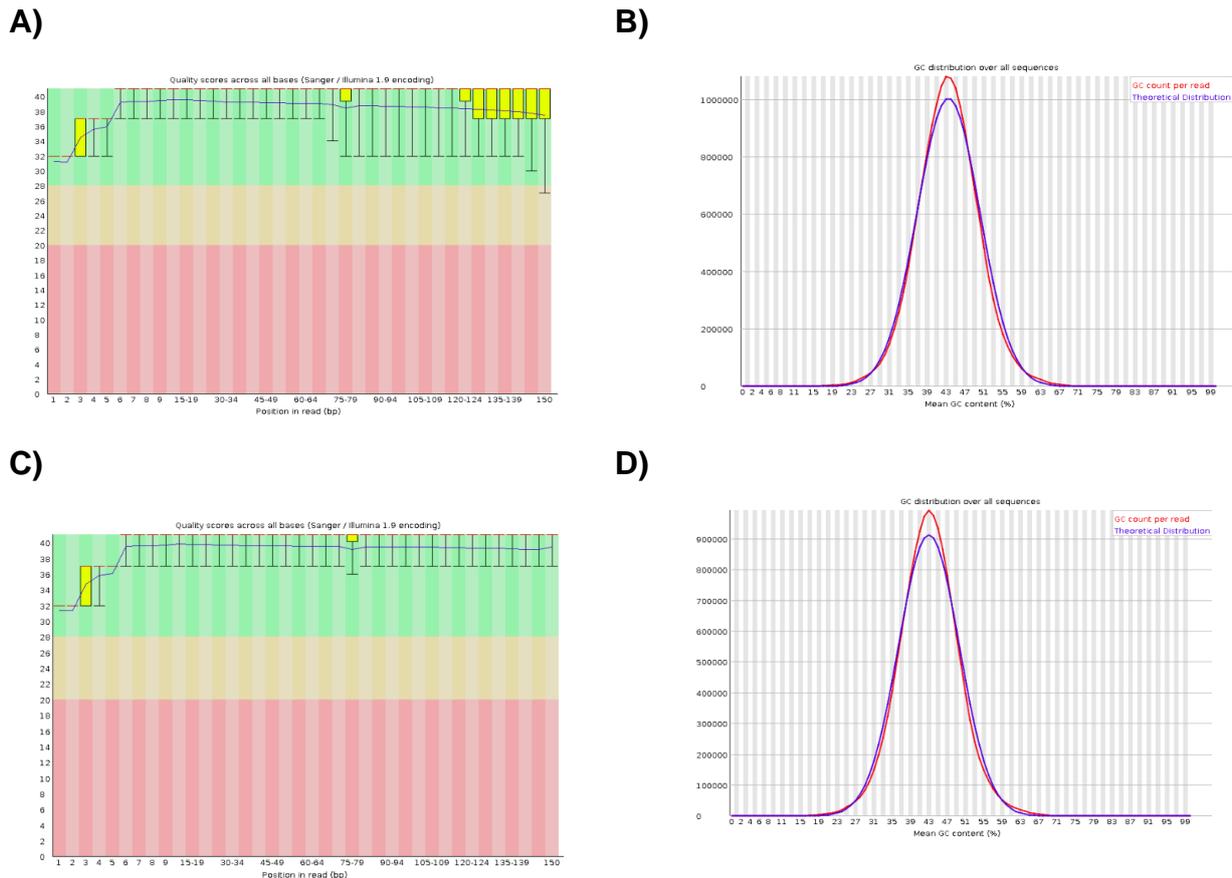
Archivo	Muestra	Total de secuencias analizadas	% de lecturas que cumple con el valor E	% de lecturas que no cumple con el valor E
1_001_merge.fastq	F2 Ind	33369196	1245648 (3.73%)	32123548 (96.27%)
2_001_merge.fastq	F5 Ind	31689290	885685 (2.79%)	30803605 (97.21%)
3_001_merge.fastq	F2 Lab 1	38503952	902470 (2.34%)	37601482 (97.66%)
4_001_merge.fastq	F2 Lab 2	39872828	1159190 (2.91%)	38713638 (97.09%)
5_001_merge.fastq	F2 Lab 3	50706824	2013924 (3.97%)	48692900 (96.03%)
6_001_merge.fastq	F5 Lab 1	42687336	1236621 (2.90%)	41450715 (97.10%)
7_001_merge.fastq	F5 Lab 2	42937198	2737876 (6.38%)	40199322 (93.62%)
8_001_merge.fastq	F5 Lab 3	39513708	1059438 (2.68%)	38454270 (97.32%)
9_001_merge.fastq	F2 Lab 4	42893768	1247665 (2.91%)	41646103 (97.09%)
10_001_merge.fastq	F2 Lab 5	38520496	1527530 (3.97%)	36992966 (96.03%)
11_001_merge.fastq	F2 Lab 6	51968270	2096536 (4.03%)	49871734 (95.97%)
12_001_merge.fastq	F5 Lab 4	39300162	2600795 (6.62%)	36699367 (93.38%)
13_001_merge.fastq	F5 Lab 5	47214480	2380679 (5.04%)	44833801 (94.96%)
14_001_merge.fastq	F5 Lab 6	50566226	1243578 (2.46%)	49322648 (97.54%)

Una vez cumplido el primer paso de limpieza de las lecturas de secuenciación se procedió a analizar la calidad de las secuencias con el software Trimmomatic versión 0.38 (Bolger *et al.*, 2014). El cual permitió eliminar las secuencias con baja calidad, debajo de 33 Phred, y los adaptadores empleados en la secuenciación. Los resultados integrados que comprenden la eliminación de ARNr, lecturas de mala

calidad y los adaptadores de secuenciación se visualizaron con la herramienta bioinformática FASTQC versión 0.11.8 (Simon Andrews, <https://www.bioinformatics.babraham.ac.uk/index.html>, revisado el día 15 de febrero del 2019). La Figura 5 muestra un resultado representativo global del procesamiento de las lecturas de secuenciación.

**Tabla 9.** Resultados de la limpieza de ARNr. Se representa el porcentaje de coincidencia de los archivos de lecturas de secuenciación con la respectiva base de datos de ARNr.

Archivo	Base de datos de ARN ribosomal (% de coincidencia)			
	Bac-16S	Bac-23S	Euc-18S	Euc-28S
1_001_merge.fastq (F2 Ind)	0.01	0.00	1.05	2.67
2_001_merge.fastq (F5 Ind)	0.01	0.00	0.94	1.85
3_001_merge.fastq (F2 Lab 1)	0.01	0.00	0.69	1.65
4_001_merge.fastq (F2 Lab 2)	0.01	0.00	0.90	2.00
5_001_merge.fastq (F2 Lab 3)	0.01	0.00	1.28	2.68
6_001_merge.fastq (F5 Lab 1)	0.01	0.00	0.93	1.95
7_001_merge.fastq (F5 Lab 2)	0.01	0.01	2.11	4.24
8_001_merge.fastq (F5 Lab 3)	0.01	0.00	0.89	1.78
9_001_merge.fastq (F2 Lab 4)	0.01	0.00	0.87	2.02
10_001_merge.fastq (F2 Lab 5)	0.01	0.01	1.25	2.70
11_001_merge.fastq (F2 Lab 6)	0.01	0.00	1.28	2.74
12_001_merge.fastq (F5 Lab 4)	0.14	0.24	2.03	4.21
13_001_merge.fastq (F5 Lab 5)	0.1	0.1	1.67	3.34
14_001_merge.fastq (F5 Lab 6)	0.01	0.00	0.77	1.68



**Figura 5.** Resultados representativos globales del preprocesamiento, eliminación de secuencias de baja calidad, adaptadores y secuencias que coinciden con las bases de datos de ARNr. **A).** Distribución de la calidad de cada base a lo largo de las lecturas de secuenciación sin procesar. **B).** Distribución del contenido GC respecto al número de lecturas de secuenciación sin procesar. **C)** Distribución de la calidad de las bases de las lecturas de secuenciación después del procesamiento. Nótese el cambio en el aumento de calidad de las bases. **D)** Distribución del contenido GC respecto al número de lecturas de secuenciación después del procesamiento. Se observó un decremento en el contenido GC de 43 a 42%.

## 7.5 Ensamble de *de novo* del transcriptoma de la levadura y anotación

Una vez cumplido con el control de calidad de las lecturas de secuenciación, las lecturas preprocesadas sirvieron de entrada para el programa Trinity versión 2.8.4. Para cumplir con los parámetros necesarios del programa todas las lecturas fueron concatenadas en dos archivos generales que contenían todas las lecturas forward y otro con todas las lecturas reverse. El número de CPU se estableció a 10 que corresponde a 20 tareas paralelizadas, una memoria de 45 GB y el tipo de librería a RF. Transcurrido el tiempo de computación se obtuvo un archivo denominado Trinity.fasta el cual contiene todos los genes ensamblados *de novo* a partir de las lecturas de secuenciación. Se obtuvieron un total de 24321 transcritos Trinity, un total de 16946 genes Trinity y un contenido GC global de 39.56%. En

cuanto a todos los contigs ensamblados se observó una media de longitud de 1206 bases, un promedio de 3303.78 bases y un total de bases ensambladas de 80351264. Las estadísticas basadas en todos los contigs exponen contig N10 de 19301 en longitud, N20 de 14504, N30 de 12187, N40 de 10021 y contig N50 de 8295.

Una vez ensamblado el transcriptoma se procedió con la anotación de dichos transcritos. Para esto, el software Trinotate (Bryant et al., 2017) fue utilizado. En primera instancia se hizo una búsqueda de transcritos Trinity presentes en bases de datos como UNIPROT y SWISPROT. Además, se implementó el programa Transdecoder v 5.5.0 para extraer a partir del archivo Trinity.fasta las proteínas candidatas con los ORF más largos, el archivo resultante de este análisis fue transdecoder.pep el cual contiene las proteínas predichas con el ORF más grande presentes en el archivo Trinity.fasta. Obtenidos las proteínas con los ORF más largos se procedió a la búsqueda en las mismas bases de datos mencionadas anteriormente.

Además de esto, se construyeron bases de datos correspondientes a los softwares HAMMER el cual identifica dominios estructural-funcionales de proteínas, SignalP el cual predice péptidos señales en secuencias de proteínas, tmHMM que predice dominios transmembrana. El archivo transdecoder.pep sirvió de entrada, para todos los programas mencionados anteriormente. Los archivos generados fueron cargados en el programa SQLite con el fin de producir un reporte unificado de todas las anotaciones funcionales de los transcritos. Con la utilidad *Trinotate.sqlite report* se creó un reporte con todas las anotaciones funcionales correspondientes a cada transcrito. Dicho archivo se nombró *trinotate\_annotations\_reports.xls*.

## **7.6 Análisis exploratorio y de expresión diferencial de genes en fermentación industrial y de laboratorio.**

### **7.6.1 Análisis exploratorio**

Para abordar el análisis exploratorio de expresión diferencial de genes se utilizaron utilidades del software Trinity que dependen del software RSEM (Li y Dewey, 2011). RSEM a su vez hace uso del programa Bowtie 2 el cual es un alineador de secuencias. Al obtener los archivos BAM/SAM resultado de la

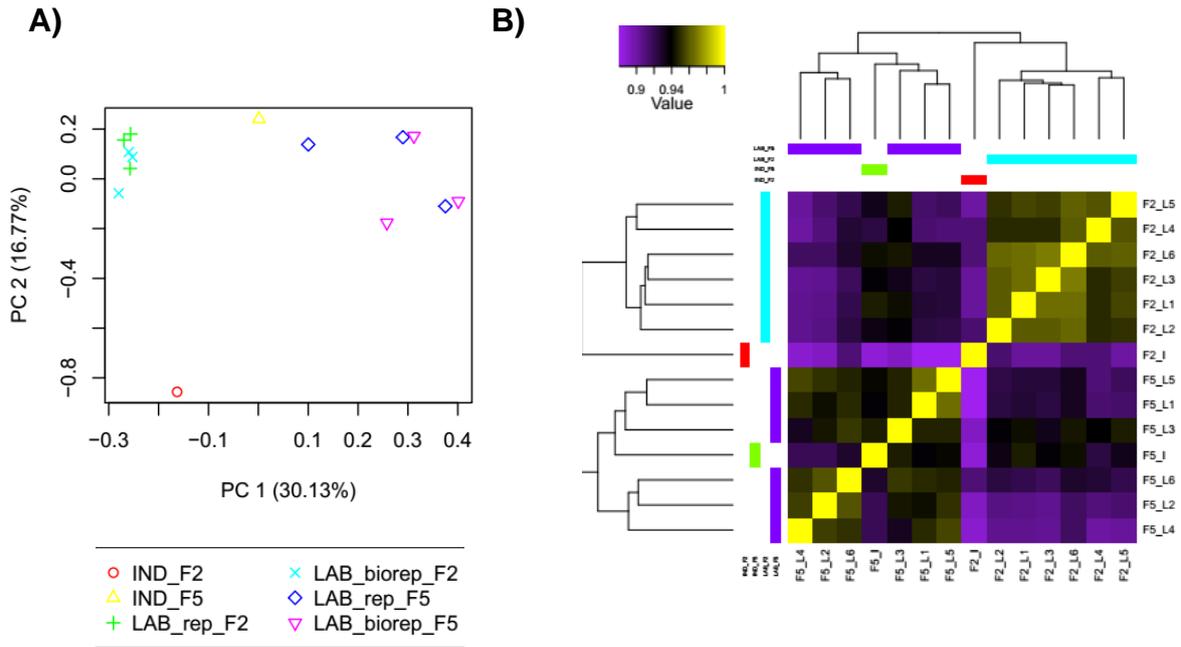
alineación de las lecturas de secuenciación, el programa RSEM los utiliza de entrada para cuantificar la proporción de lecturas alineadas para cada transcrito. En esencia el archivo resultante es una matriz de datos donde las filas representan a cada transcrito y las columnas cada muestra o condición. Los datos obtenidos en esta matriz representan la cuantificación de los transcritos.

Para verificar la calidad de muestreo, de las réplicas biológicas se realizó un análisis de componentes principales (PCA, por sus siglas en inglés) y un agrupamiento jerárquico empleando un mapa de calor como recurso gráfico que nos indica de forma visual si existe correlación entre las réplicas, además indicando una posible relación de las muestras (Figura 6A). Los resultados del análisis exploratorio indican que las muestras de la réplica experimental y biológica del día dos de la fermentación a nivel de laboratorio son muy similares, respecto al conteo de transcritos, al verse agrupadas en un solo conjunto demostrando que contienen aproximadamente la misma magnitud en su variación. Por otro lado, se puede observar que la muestra correspondiente a F2 de la fermentación industrial no se agrupa con las muestras de la fermentación de laboratorios del mismo día. Esto indica que la variación representada en F2 industrial contiene una magnitud diferente en la variación observada en su respectiva réplica experimental y biológica a nivel laboratorio. Representando así una diferencia en la cuantificación de los transcritos en el día dos de la fermentación a nivel industrial (Figura 6A). De igual forma la agrupación de las muestras correspondientes a la réplica experimental y biológica del día 5 de la fermentación a nivel laboratorio indican una mayor diferencia entre dichas muestras ya que la distancia entre cada muestra es mayor a la observada el conjunto del día 2 a nivel laboratorio. No obstante, la muestra correspondiente a F5 industrial manifiesta una variación relativamente parecida al conjunto de F5 laboratorio. Por tanto, la cuantificación de transcritos en F5 es relativamente similar tanto a nivel industrial como a nivel laboratorio. Es importante mencionar que las pequeñas distancias, proyectadas tanto en el componente principal 1 y 2, que existen entre las muestras del día 5 pueden representar distinta cuantificación de algunos transcritos (Figura 6A). Los resultados antes mencionados se corroboran con un agrupamiento jerárquico de las muestras antes mencionadas. En la figura 6B podemos observar que la agrupación de todas las muestras de F2 laboratorio se efectuó en un solo conjunto (Dendograma de la Figura 6B). Lo mismo

sucede con las muestras de F5 laboratorio. El mapa de calor indica una correlación directa entre cada muestra con su conjunto símil. Nótese como la muestra F2 industrial muestra un valor bajo de correlación con las respectivas muestras de la réplica experimental y biológica a nivel laboratorio confirmando lo observado en el gráfico de componentes principales.

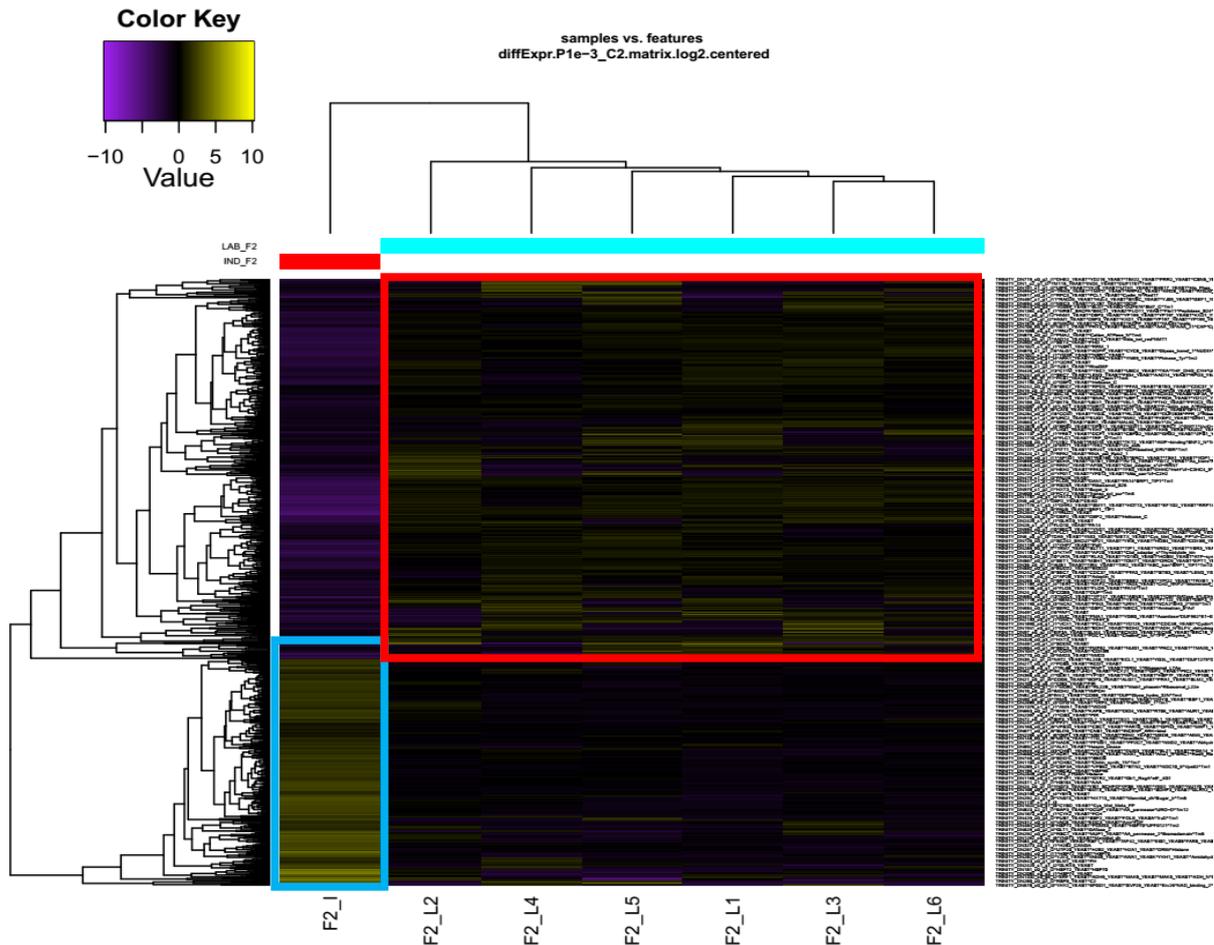
### **7.6.2 Análisis de expresión diferencial de genes**

Una vez completado el análisis exploratorio se procedió con el análisis de expresión diferencial de genes comparando la muestra de F2 industrial contra todas las muestras de F2 laboratorio, es decir 6 muestras de F2 laboratorio. De la misma forma para el día 5 las comparaciones se realizaron entre la muestra F5 industrial contra las 6 muestras de F5 laboratorio. Las comparaciones en F2 de la fermentación resultaron en 286 transcritos Trinity a nivel industrial y 480 transcritos Trinity a nivel laboratorio diferencialmente expresados, para F5 industrial se obtuvieron 143 Transcritos Trinity y a nivel laboratorio se obtuvieron 298 transcritos Trinity. Es importante mencionar que al momento de las anotaciones de las matrices de expresión un mismo gen coincidió con distintas isoformas resultantes del ensamble por Trinity, esto significa que distintos transcritos Trinity mapean con un solo gen manifestando redundancia en dichas matrices de expresión. Estos resultados serán filtrados en los análisis de enriquecimiento por ontología de genes y vías metabólicas.



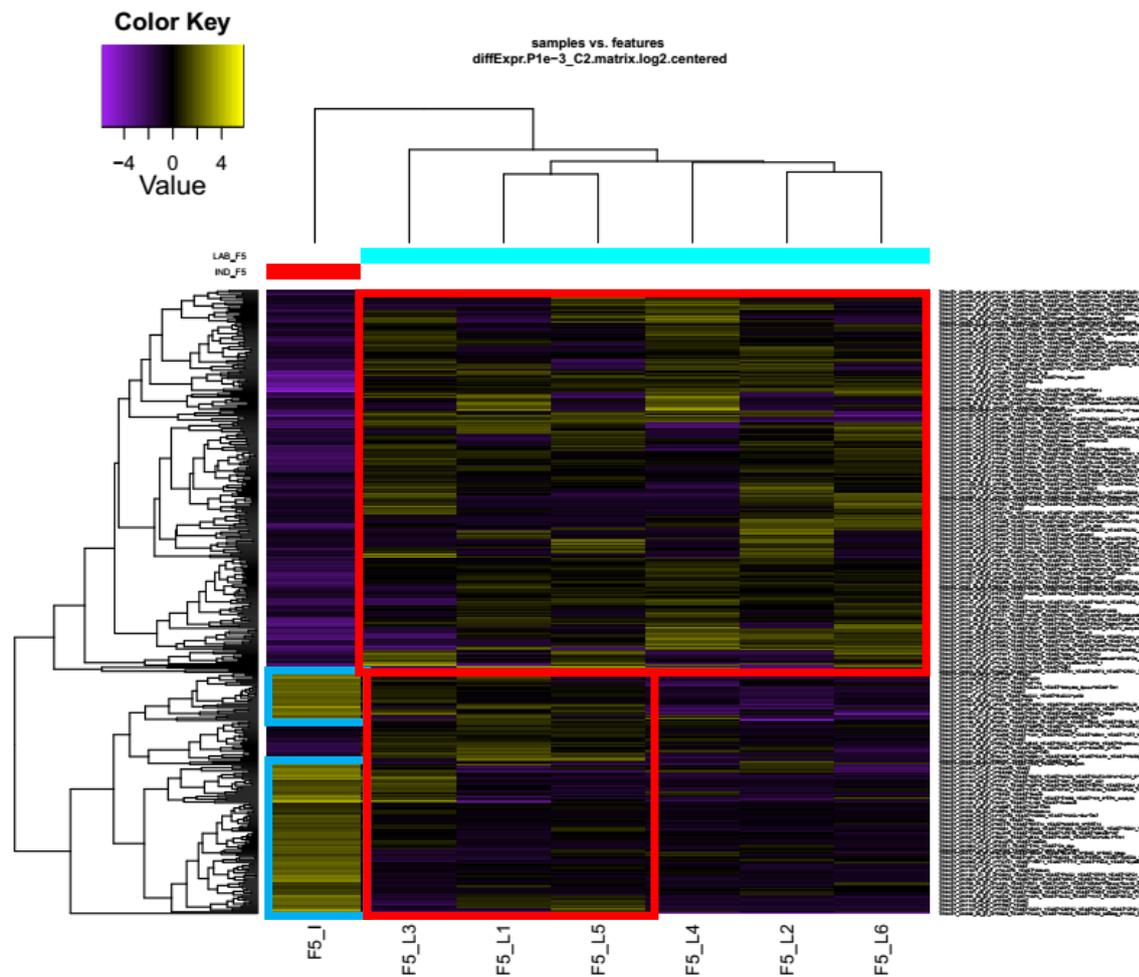
**Figura 6.** Análisis exploratorio de los datos de cuantificación de transcritos. **A)** Gráfico de análisis de componentes principales que indica la correlación de cada réplica y muestra relativo a su similitud en el conteo de transcritos. **B)** Agrupamiento jerárquico y correlación de muestras relativo al conteo de transcritos.

Una representación visual de los resultados obtenidos en las comparaciones en F2 se expone en la Figura 7 donde es posible ver un mapa de calor con agrupamiento jerárquico de los distintos transcritos diferencialmente expresados en cada condición. Es posible observar que los transcritos a nivel industrial de F2 se agrupan de forma importante, (Dendograma representado en el costado izquierdo de la figura 6) rectángulo resaltado en azul. Mientras que los transcritos expresados a nivel laboratorio de forma general se agrupan en el recuadro rojo. Sin embargo, es posible ver algunos transcritos dispersos a lo largo del mapa de calor. Por otro lado, el agrupamiento de las muestras dado su patrón de expresión indica un conjunto representado por las réplicas experimentales y biológicas y otro conjunto con solo una muestra representada por la muestra de F2 industrial. Evidenciando así diferencia en el patrón de transcripción entre la fermentación a nivel industrial y de laboratorio del día 2.



**Figura 7.** Mapa de calor y agrupamiento jerárquico de los transcritos diferencialmente expresados en el día 2 de la fermentación industrial y de laboratorio, en el costado izquierdo y la agrupación de las muestras respecto al patrón de expresión de dichos transcritos en la parte superior. Explicación en el texto principal.

De igual forma, la representación visual de las comparaciones realizadas en el día 5 de la fermentación se expone en la Figura 8 donde se emplea la misma interpretación mencionada para las comparaciones de F2. Igualmente, se observa el agrupamiento de los transcritos en los recuadros representados en color azul para la condición industrial y en rojo para la condición laboratorio.



**Figura 8.** Mapa de calor y agrupamiento jerárquico de los transcritos diferencialmente expresados en el día 5 de la fermentación industrial y de laboratorio en el costado izquierdo. La agrupación de las muestras respecto al patrón de expresión de dichos transcritos se observa en la parte superior.

### 7.7 Análisis de enriquecimiento por ontología de genes y rutas metabólicas.

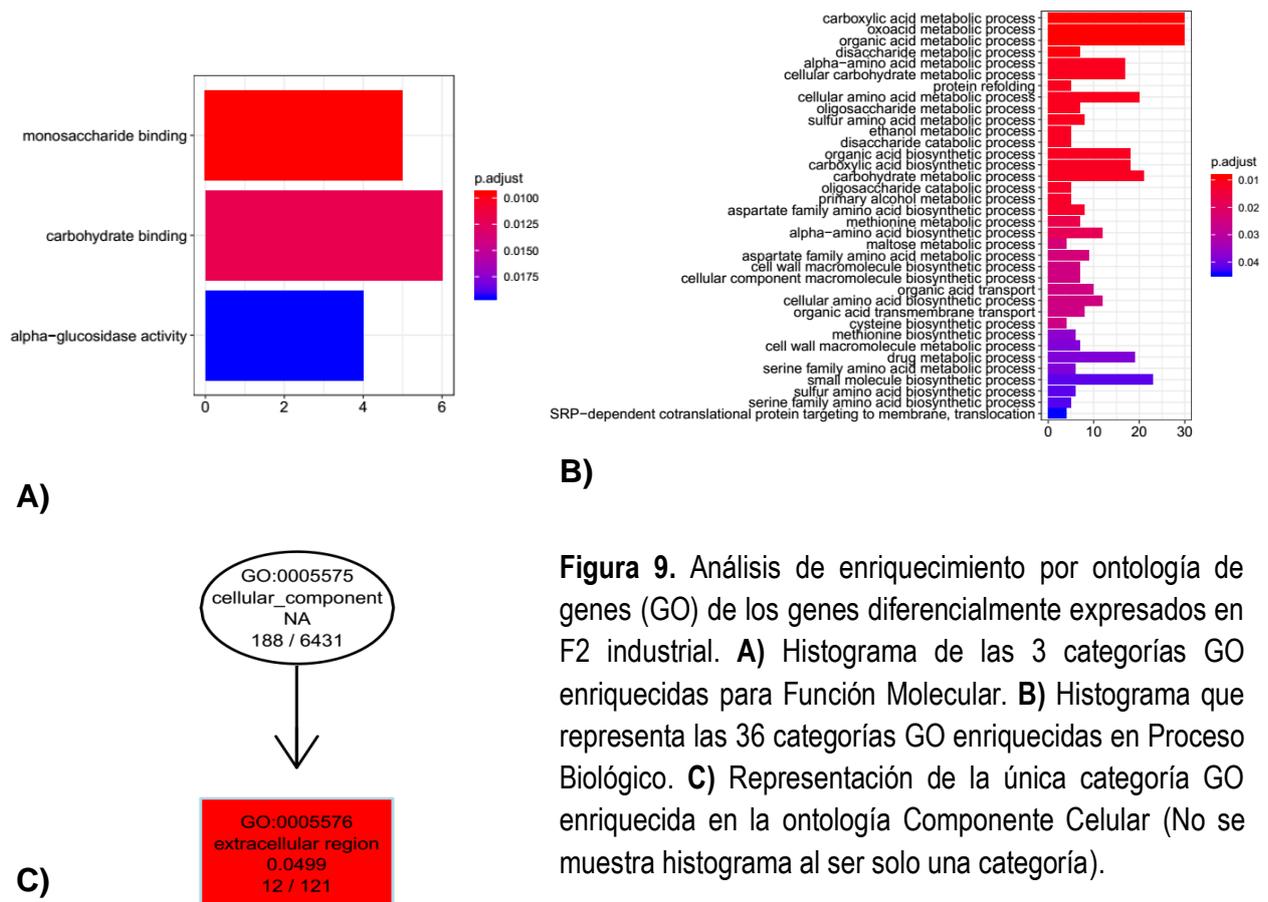
Para obtener información sobre el panorama general de los genes diferencialmente expresados (GDE) en las fermentaciones a nivel industrial y de laboratorio se empleó un enfoque de agrupamiento por medio de ontología (GO) de genes y rutas metabólicas. Para la clasificación de los GDE's un análisis de enriquecimiento GO, por rutas metabólicas (KEGG) y el paquete clusterProfile versión 3.12.0 (Yu *et al.*, 2012) fueron utilizados. La lista de consulta de fondo constó de la base de datos de anotación de genoma completo de la levadura *Saccharomyces cerevisiae* contenida en el paquete org.Sc.sgd.db versión 3.8.2 (Carlson, 2019). Los paquetes antes mencionados forman parte del proyecto Biocconductor (Morgan, 2018).

### 7.7.1 Análisis de enriquecimiento por ontología de genes de F2 industrial

El análisis de enriquecimiento GO para los 286 GDE's anotados en F2 industrial resultó en 3 categorías GO significativamente sobrerrepresentadas (FDR < 0.05), para Función Molecular 36 subcategorías GO y para componente celular sólo una categoría GO resultó significativamente sobrerrepresentada (Figura 9). Dentro de la ontología Función Molecular, los GDE's se asociaron mayormente a unión a monosacáridos (GO:0048029), unión a carbohidratos (GO:0030246) y actividad alfa-glucosidasa (GO:0090599). Para la categoría de Proceso Biológico se demuestra que los GDE's se asocian principalmente a procesos metabólicos de ácidos carboxílicos (GO:0019752), procesos metabólicos de oxoácidos (GO:0043436), metabolismo de ácidos orgánicos (GO:0006082), metabolismo de disacáridos (GO:0005984), metabolismo de alfa aminoácidos (GO:1901605), metabolismo de carbohidratos celulares (GO:0044262), plegamiento de proteínas (GO:0042026), metabolismo de aminoácidos celulares (GO:0006520), metabolismo de oligosacáridos (GO:0009311), por mencionar algunos (Figura 9B). En la categoría Componente Celular sólo un término, región extracelular (GO:0005576), fue significativamente enriquecido (FDR < 0.05).

Es importante mencionar que en la mayoría de los términos relacionados con metabolismo de carbohidratos celular (GO:0044262) en la categoría de Proceso Biológico se componen de genes como *DSF1* manitol deshidrogenasa, metabolismo del manitol (Jordan *et al.*, 2016), *GID7* Proteína 7 de degradación inducida por glucosa, regulación negativa de la gluconeogénesis (Regelmann *et al.*, 2003), *CHS1* y 2 sintasa 1 y 2 de quitina, involucrada en la formación del septo y reparación bajo condiciones adversas. Mantenimiento de la morfología de la célula (Leal *et al.*, 1994; Gaudet *et al.*, 2011), *EMI2* Glucoquinasa-2 putativa, relacionada con la fosforilación de aldohexosas y captación de glucosa. También involucrada en esporulación. Además, existe evidencia indicando que activa *IME1* inductor de la meiosis (Enyenihi y Saunders, 2003), *MAL32* Alfa glucosidasa involucrada en la hidrólisis de las uniones no reductoras alfa (1→4) de residuos alfa-D-Glucosa liberando alfa-D-Glucosa (Brown *et al.*, 2010). En esta misma categoría también se encontró el gen *MAL11* (*AGT1*) que codifica para una permeasa, simportador,

general para alfa-glicósidos como maltosa, turanosa, isomaltosa, alfa-metilglucosa, maltotriosa, palatinosa, trealosa, melezitosa y glucosa, transportando de forma concomitante protones dentro de la célula (Day *et al.*, 2002), *SUC2* que codifica para una invertasa involucrada en el metabolismo de la sacarosa, hidrolizando la sacarosa en fructosa y glucosa (Naumov y Naumova 2010).



**B)**

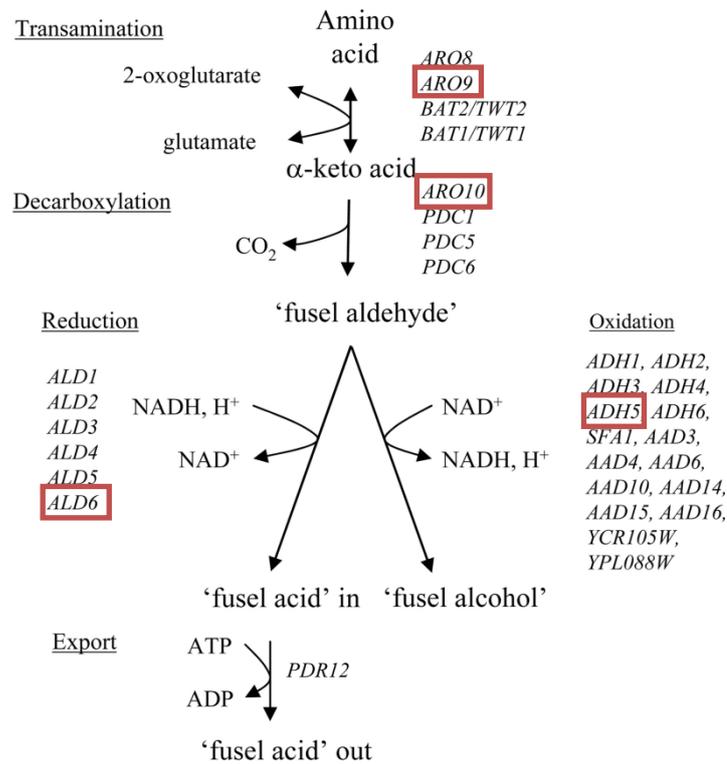
**Figura 9.** Análisis de enriquecimiento por ontología de genes (GO) de los genes diferencialmente expresados en F2 industrial. **A)** Histograma de las 3 categorías GO enriquecidas para Función Molecular. **B)** Histograma que representa las 36 categorías GO enriquecidas en Proceso Biológico. **C)** Representación de la única categoría GO enriquecida en la ontología Componente Celular (No se muestra histograma al ser solo una categoría).

Otros genes interesantes son *H XK1* que codifica para una hexoquinasa-1 que regula el paso inicial en la glicólisis al catalizar la fosforilación de la D-glucosa (Rodríguez *et al.*, 2001) y el gen *IMA5* que codifica una oligo-1,6-glicosidasa que se encarga de la hidrólisis de las uniones glicosídicas alfa 1→6 de algunos oligosacáridos producidos por el tratamiento de almidón y glucógeno con alfa amilasa, y en isomaltosa. Presenta afinidad hacia isomaltosa, maltosa y palatinosa (Teste y Parrou, 2010).

En cuanto a la subcategoría de Proceso Biológico; metabolismo de aminoácidos celular (GO:0006520) se encontró enriquecida con 20 genes de los cuales algunos interesantes para la producción de compuestos aromáticos fueron los genes *ARO9* que codifica para una transaminasa general para aminoácidos aromáticos involucrada en el metabolismo de triptófano (Iraqi *et al.*, 1998), *ARO10* que codifica una descarboxilasa que cataliza la descarboxilación de aminoácidos (Kneen *et al.*, 2011) y el gen *ADH5* que codifica a una alcohol deshidrogenasa que cataliza la remoción de un protón de alcoholes primarios para su reducción a aldehídos (Smith *et al.*, 2004). En la subcategoría de procesos metabólicos de ácidos carboxílicos (GO:0019752) se encuentra el gen *ALD6* que codifica para aldehído deshidrogenasa citosólica la cual realiza la conversión de acetaldehído a ácido acético (Saint *et al.*, 2004). En resumen, se tienen los tres pasos involucrados en la vía de Ehrlich la cual se encarga de la producción de compuestos aromáticos como alcoholes y ácidos de cadena larga. En la figura 10 se muestra la ruta de Ehrlich con los distintos genes involucrados y resaltado en recuadros los genes diferencialmente expresados en F2 industrial.

### **7.7.2 Análisis de enriquecimiento por rutas metabólicas de F2 industrial**

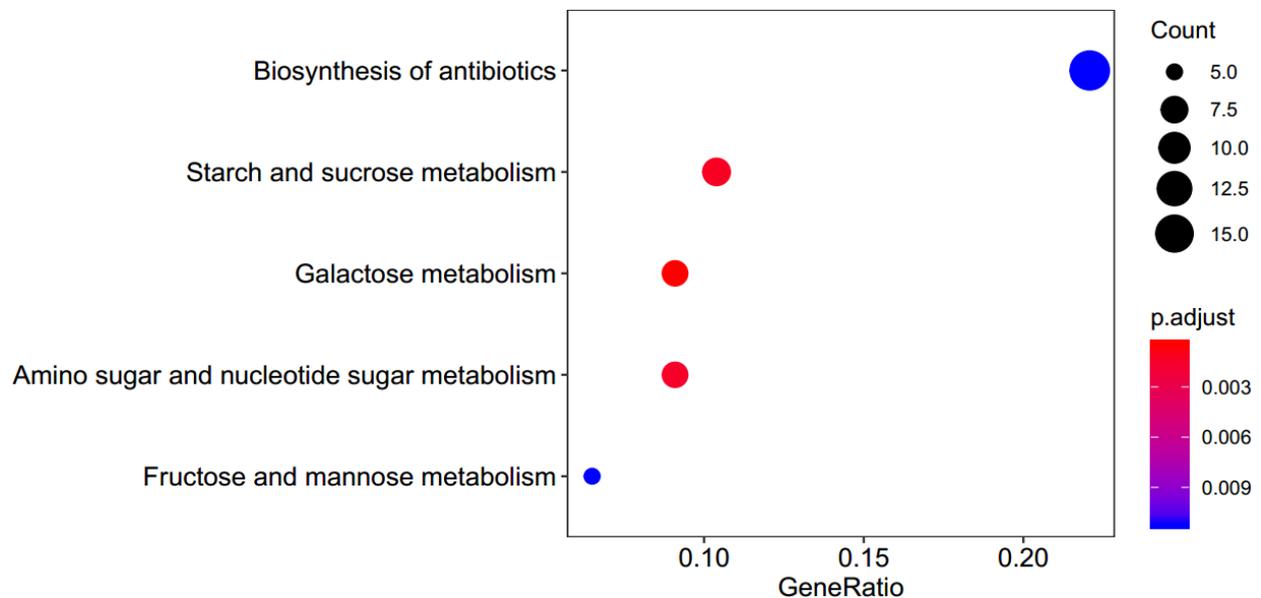
Para obtener conocimiento sobre la funcionalidad de los genes diferencialmente expresados en F2 industrial se realizó un análisis de enriquecimiento de rutas metabólicas utilizando la base de datos de la enciclopedia de Kioto de genes y genomas (KEGG por sus siglas en inglés). El análisis de enriquecimiento para rutas metabólicas evidenció que la síntesis de antibióticos (sce01130, con 17), metabolismo de almidón y sacarosa (sce00500, con 8 genes), metabolismo de amino azúcares y nucleótidos (sce00520, con 7 genes), metabolismo de galactosa (sce00052, con 7 genes) y el metabolismo de fructosa y manosa (sce00051, con 5 genes) fueron significativamente enriquecidos (Figura 11).



**Figura 10.** Representación esquemática de la ruta de Ehrlich involucrada en el catabolismo de aminoácidos como leucina, valina, isoleucina, fenilalanina, tirosina, triptófano y metionina con la consecuente producción de compuestos aromáticos como alcoholes y ácidos de fusel. En recuadro rojo se encuentran los genes diferencialmente expresados en F2 industrial. Tomado y modificado de Hazelwood y cols (2008).

Indicando que los procesos metabólicos involucrados en el metabolismo de carbohidratos y aminoácidos está positivamente regulado en F2 industrial. En el metabolismo de almidón y sacarosa observamos genes como *EMI2*, *SUC2*, *UGP1*, *MAL32*, *HXK1*, *IMA1* y 5, *PGI1* los cuales están involucrados en el catabolismo de carbohidratos. Los genes que enriquecen la categoría KEGG del metabolismo de galactosa son prácticamente los mismos que se observan en el metabolismo de almidón y sacarosa a excepción de *UGP1* el cual no se encuentra en la categoría de metabolismo de galactosa. Los genes presentes en la categoría del metabolismo de fructosa y manosa son *DSF1*, *YNR073C*, que codifican para una manitol deshidrogenasa, *EMI2*, *HXK1* y *PSA1*. La categoría de metabolismo de amino azucres y nucleótidos son esencialmente los mismos para las categorías anteriormente descritas sumando los genes *CHS1* y 2 que codifican para una sintasa de quitina. Por último, la categoría KEGG mayormente enriquecida;

Biosíntesis de antibióticos contiene genes que codifican mayormente deshidrogenasas, transferasas, sintasas y cinasas.

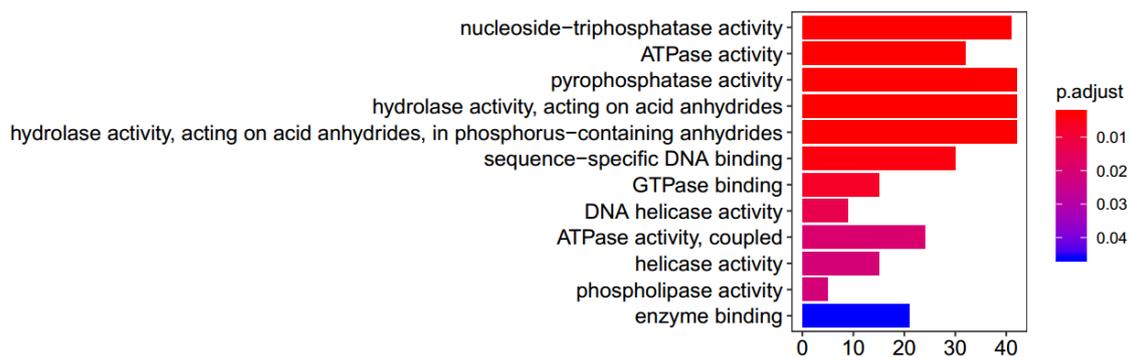


**Figura 11.** Gráfico de dispersión del análisis de enriquecimiento de vías metabólicas con el uso de la base de datos KEGG de los GDE's de F2 industrial. El eje-x indica la proporción de genes observados frente a la base de datos KEGG. El eje-y indica el nombre de la ruta metabólica KEGG. El tamaño del punto indica el número de genes que enriquecen la ruta KEGG y los colores de acuerdo al valor p ajustado. Se muestran las únicas rutas metabólicas enriquecidas en F2 industrial con un valor p ajustado (FDR) menor a 0.05.

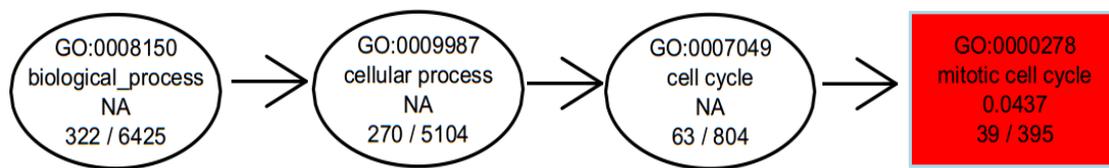
### 7.7.3 Análisis de enriquecimiento por ontología de genes de F2 laboratorio

Del mismo modo mencionado anteriormente para F2 industrial, se emplearon análisis de enriquecimiento por ontología de genes para F2 laboratorio, así como para F5 industrial y laboratorio. Los resultados para los 480 genes anotados en F2 laboratorio indican que sólo los términos de Función Molecular y Proceso Biológico fueron significativamente enriquecidos (FDR < 0.05) mientras que para Componente Celular el valor p ajustado fue mayor a 0.05 (FDR=0.051). El término Función Molecular fue enriquecido en 12 categorías GO y el término Proceso Biológico fue enriquecido solo en una categoría GO. En Función Molecular los GDE's se asociaron a actividad nucleótido trifosfatasa (GO:0017111), actividad ATPasa (GO:0016887), actividad pirofosfatasa (GO:0016462), actividad hidrolasa en ácidos

anhídridos y en anhídridos que contienen fosforo (GO:0016818), unión a secuencias específicas de ADN (GO:0043565), unión a GTPasa (GO:0051020), actividad de helicasa de ADN (GO:0003678), actividad ATPasa (GO:0042623), actividad helicasa (GO:0004386), actividad fosfolipasa (GO:0004620) y unión a enzimas (GO:0019899). Para Proceso Biológico solo la categoría ciclo celular mitótico (GO:0000278) fue enriquecida (FDR=0.04) (Figura 12).



**A)**



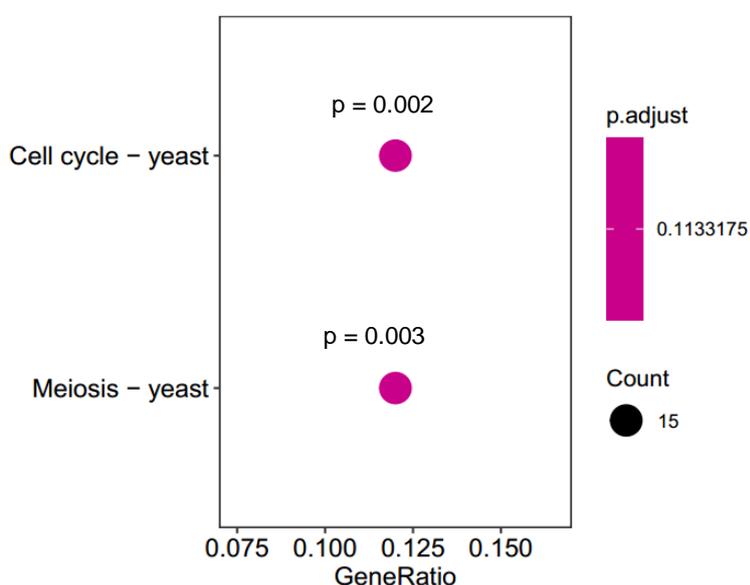
**B)**

**Figura 12.** Análisis de enriquecimiento por ontología de genes (GO) de los genes diferencialmente expresados en F2 laboratorio. **A)** Histograma de las 12 categorías GO enriquecidas para Función Molecular. **B)** Representación de la única categoría GO enriquecida en la ontología Componente Celular (No se muestra histograma al ser solo una categoría).

En resumen, los niveles de expresión de la levadura en F2 laboratorio están centrados en la regulación del ciclo celular mitótico. Algunos de los genes codifican para proteínas de unión a ADN, algunas cinasas involucradas en la regulación del ciclo celular, ATPasas y GTPasa involucradas en procesos regulatorios del ciclo celular.

#### 7.7.4 Análisis de enriquecimiento por rutas metabólicas de F2 laboratorio

Para visualizar la funcionalidad de los GDE's en F2 laboratorio se procedió con el mismo análisis de enriquecimiento para rutas metabólicas con la utilización de la base de datos KEGG. Sin embargo, los resultados indican que ninguna ruta metabólica esta significativamente enriquecida en el conjunto de genes diferencialmente expresados en F2 laboratorio. En este conjunto de genes se obtuvo un valor P de 0.002 y 0.003 para las rutas metabólicas de ciclo celular en levadura (sce04111) y meiosis en levadura (sce04113), respectivamente. Sin embargo, los valores p ajustados (FDR) no pasaron el punto de corte  $FDR < 0.05$  al obtener un FDR de 0.11 para ambas rutas metabólicas (Figura 13).



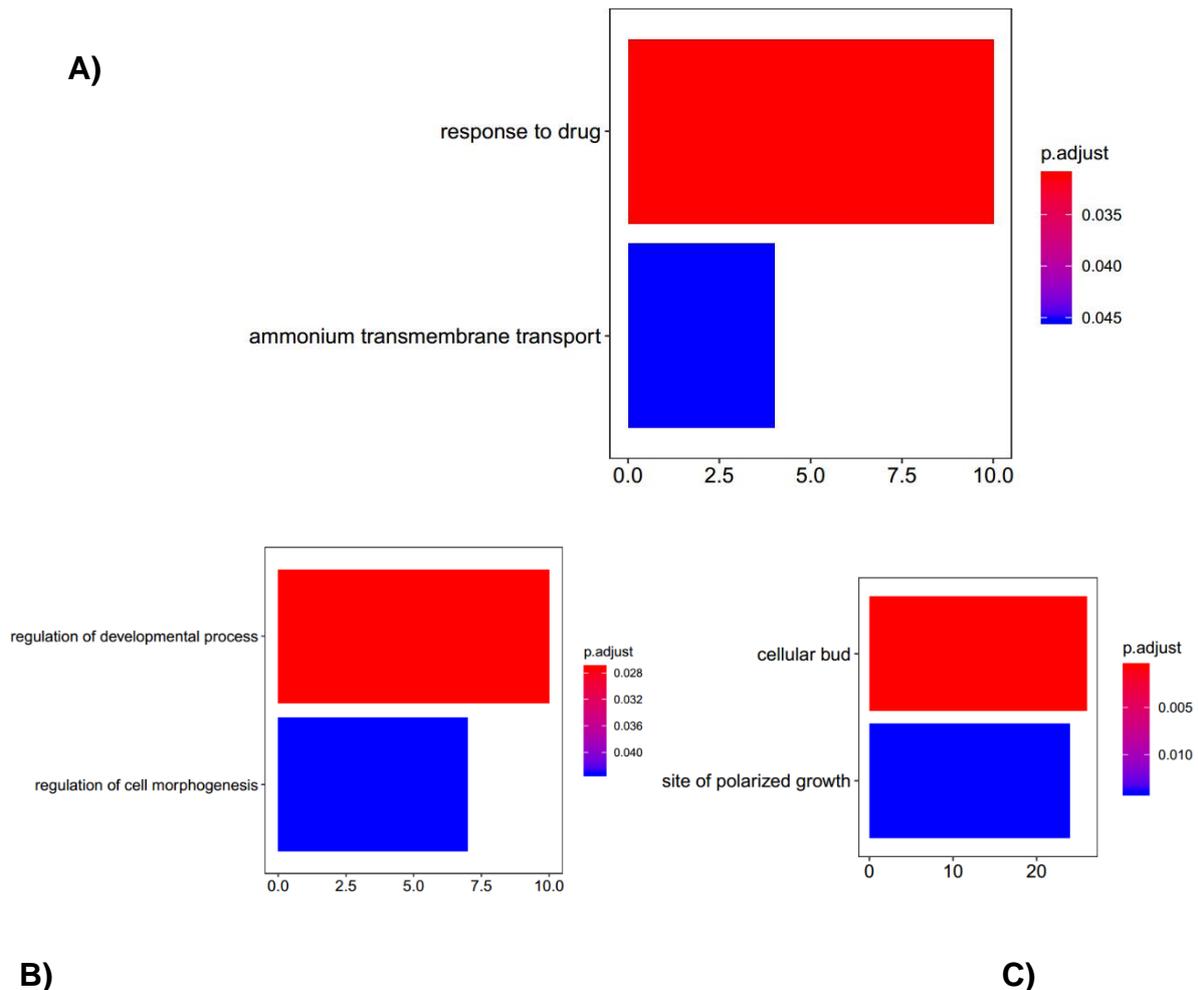
**Figura 13.** Gráfico de dispersión del análisis de enriquecimiento por rutas metabólicas (KEGG) para el conjunto de genes diferencialmente expresados en F2 laboratorio. Como se puede observar, los valores p ajustados (FDR) no cumplen con el criterio estadístico  $FDR < 0.05$ .

Los resultados anteriormente descritos, F2 laboratorio e industrial, se pueden interpretar como genes regulados positivamente en una condición y regulados a la baja en la condición contraria. Es decir, los genes que se observaron diferencialmente expresados en F2 industrial son genes regulados negativamente en F2 laboratorio. Partiendo de esta interpretación se puede resumir que las rutas metabólicas significativamente enriquecidas en F2 industrial están negativamente reguladas a nivel laboratorio.

### **7.7.5 Análisis de enriquecimiento por ontología de genes de F5 industrial y laboratorio**

El análisis de enriquecimiento por ontología de genes asociado al conjunto de los 143 genes diferencialmente expresados en F5 industrial resultó en dos subcategorías GO dentro del término Proceso Biológico (FDR < 0.05). La subcategoría respuesta a fármacos (GO:0042493) resultó significativamente con 10 GDE's, mientras que la subcategoría transporte transmembranal de amonio (GO:0072488) resultó significativamente enriquecida con 4 genes. En respuesta a fármacos genes como *SAM3*, *HNM1*, *AQR1*, *AGP1* y *ATO3* codifican para transportadores de nitrógeno, en forma de poliaminas. Otros genes tales como *PDR3* y *RSC9* presentan roles reguladores. Genes como *PDR16*, *VPS27* y *DOA4*, tienen funciones como remodelación lipídica de la membrana plasmática, tráfico vacuolar de proteínas y esterasa involucrada en procesos de ubiquitinación, respectivamente. La subcategoría de transporte de amonio resultó enriquecida con los mismos genes presentes en respuesta a fármacos; *SAM3*, *HNM1*, *AQR1*, *AGP1* y *ATO3* (Figura 14A).

Para el conjunto de los 298 GDE's en F5 laboratorio, el análisis de enriquecimiento por ontología presentó significancia estadística en las subcategorías de proceso de regulación del desarrollo (GO:0050793) con 10 genes y regulación de la morfología celular (GO:0022604) con 7 genes, esto dentro del término Proceso Biológico. También se observó enriquecimiento significativo dentro del término Componente Celular en las subcategorías de gema celular (GO:0005933) con 26 genes y sitio de crecimiento polarizado (GO:0030427) con 24 genes.



**Figura 14.** Histogramas del análisis de enriquecimiento por ontología de los conjuntos de genes diferencialmente expresados (GDE's) en F5 industrial y laboratorio. **A)** Categorías enriquecidas en Proceso Biológico con los GDE's en F5 industrial. **B)** Categorías significativamente enriquecidas en Proceso Biológico con los GDE's observados en F5 laboratorio. **C)** Categorías enriquecidas dentro de Componente Celular en F5 laboratorio.

Los genes que enriquecen significativamente la regulación del desarrollo, en F5 laboratorio, se relaciona con factores de transcripción como *MSN4* y *2*, una cinasa *YCK1*, un transportador de oligopéptidos *OPT2*, un importante regulador de la transcripción *SNF2*, entre otros. En cuanto a la regulación de la morfogénesis celular se encuentran nuevamente genes como *OPT2*, *YCK1*. Una cinasa importante en la regulación de la división celular *CDC7*, un gen involucrado en un apareamiento eficiente *FIG2*, una beta-glucano sintasa involucrada en la síntesis de beta glucanos, entre otros (Figura 14B y C).

### **7.7.6 Análisis de enriquecimiento por rutas metabólicas de F5 industrial y de laboratorio**

Al momento de realizar el análisis de enriquecimiento para ambas condiciones no se observaron rutas metabólicas significativamente enriquecidas (FDR > 0.05) indicando que en las comparaciones de dichas condiciones no existen diferencias respecto a las rutas metabólicas. Sin embargo, las categorías GO significativamente enriquecidas brindan luz sobre algunos procesos diferencialmente regulados en las condiciones.

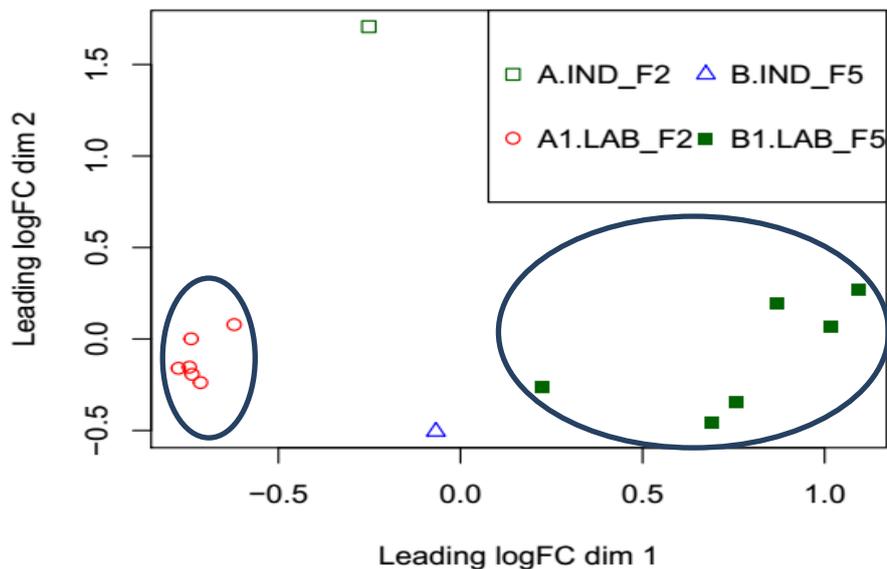
## **7.8 ANÁLISIS DEL TRANSCRIPTOMA DE LA CEPA 820 CON GENOMA DE REFERENCIA**

Para brindar más evidencia en los resultados obtenidos en el análisis por ensamble *de novo* se realizó un análisis guiado por genoma de referencia. Para esto el genoma de la cepa *S. cerevisiae* 288C y sus anotaciones respectivas fueron utilizados.

### **7.8.1 Análisis exploratorio de los datos transcriptómicos de la cepa 820 con genoma de referencia**

Con el fin de verificar posibles diferencias entre las muestras, análisis igualmente realizado en el enfoque *de novo*, se procedió a inspeccionar la relación de cada muestra tomando en cuenta el conteo de las lecturas mapeadas en los genes dentro del genoma de referencia. Para esto, se efectuó una visualización por medio de gráficos multidimensionales a escala (MDS plot), el cual tiene el mismo principio que el análisis de componentes principales empleado en la sección anterior. De hecho, podemos ver que el comportamiento de las muestras tiene aproximaciones considerables con los resultados obtenidos mediante ensamble *de novo* (Figura 15). Las muestras de F2 laboratorio se agrupan en un solo conjunto, apartándose en variación respecto a la muestra F2 industrial. De igual forma que en la sección anterior, esta distancia en variación de las muestras de F2 industrial y laboratorio se traducen en distinto patrón de expresión. Por otro lado, podemos observar que las muestras de F5 laboratorio, a pesar de que presenta más variación entre sí, se agrupan en un solo conjunto y la muestra de F5 industrial presenta

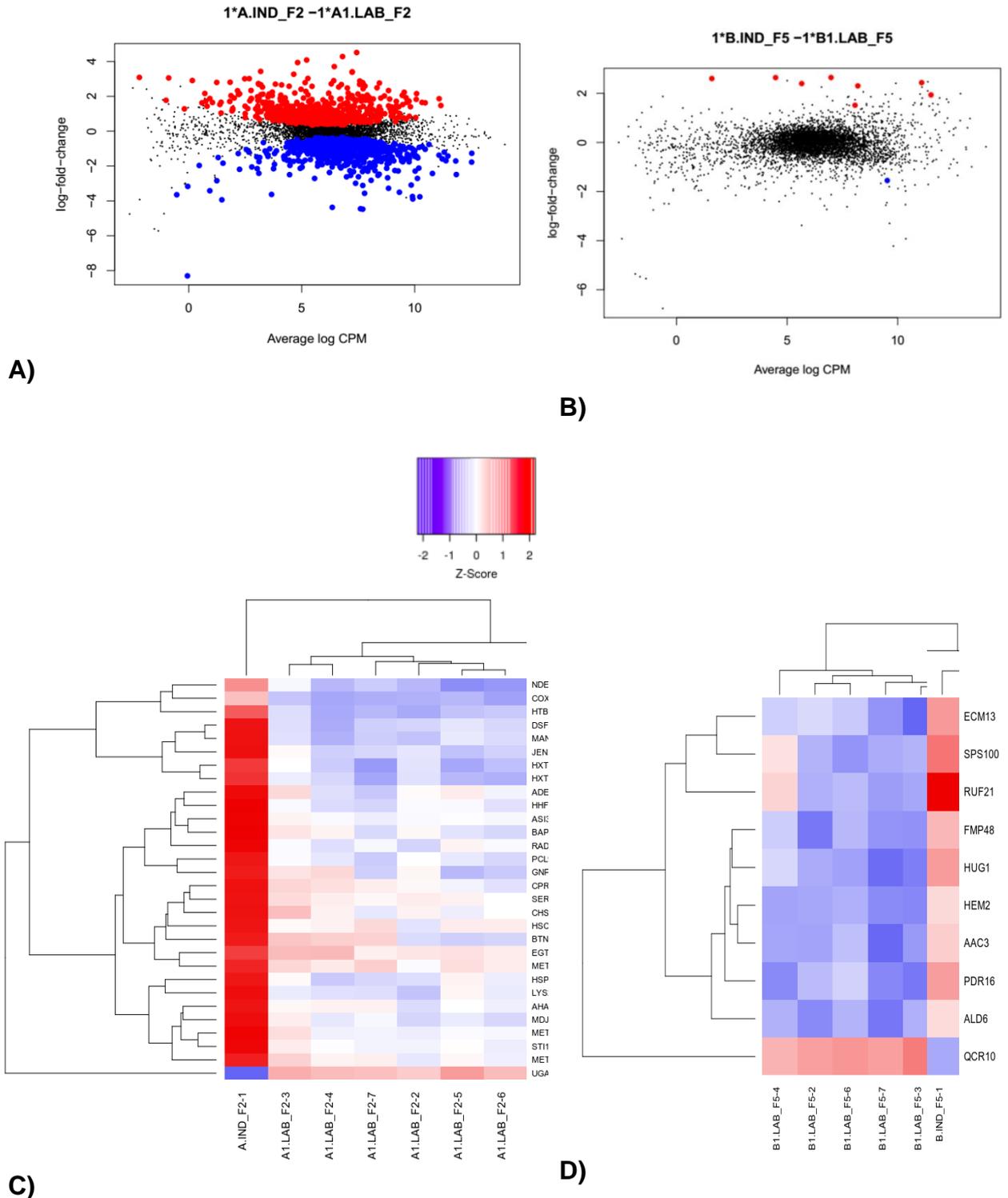
aproximadamente la misma magnitud de variación. Este mismo comportamiento sustenta la validez conductual de los datos obtenidos en el enfoque *de novo*.



**Figura 15.** Análisis multidimensional a escala de los datos transcryptómicos provenientes de un enfoque guiado por genoma de referencia (S.c S288C). Se puede observar las agrupaciones de las muestras muy similar a lo observado en el enfoque *de novo*.

### 7.8.2 Análisis de expresión diferencial de genes obtenidos con un enfoque guiado por genoma de referencia

Al igual que en el enfoque *de novo*, el software edgeR fue utilizado para cumplir con el análisis de expresión diferencial de los genes obtenidos con un enfoque guiado por referencia. La variación observada en los análisis exploratorios de F2 resultó en 200 genes diferencialmente expresados en condiciones industriales y 159 genes a nivel laboratorio. En cuanto a F5, la reducida variación observada en los análisis multidimensionales exploratorios resultó en tan solo 8 genes diferencialmente expresados a nivel industrial y solamente 1 gen a nivel laboratorio (Figura 16).



**Figura 16.** Análisis de expresión diferencial de los genes obtenidos mediante genoma de referencia. **A)** Gráfico MD que muestra los 200 GDE's en F2 industrial (puntos rojos) y los 159 GDE's en F2 laboratorio (puntos azules). **B)** Gráfico MD que muestra los 8 GDE's en F5 industrial (puntos rojos) y solo un GDE en F5 laboratorio (punto azul). **C)** Mapa frío que muestra el top 30 de los genes mayormente significativos en F2 industrial y laboratorio. **D)** Mapa frío que muestra los únicos 10 GDE's obtenidos en F5 industrial y laboratorio.

Las evidentes diferencias en las comparaciones de F2 industrial y laboratorio se pueden observar en las gráficas multidimensionales que grafican la tasa de cambio o fold change contra la abundancia relativa de cada gen, average log CPM. Con la misma óptica se puede resaltar las pequeñas diferencias observadas en las comparaciones de F5 en donde valores positivos en el logaritmo base 2 de la tasa de cambio (fold change) indican genes diferencialmente expresados en condiciones industriales y valores negativos indican genes diferencialmente expresados en condiciones de laboratorio. Este mismo enfoque aplica para las comparaciones de F2.

Es importante poner énfasis en que las diferencias observadas en las comparaciones por pares, están solamente explicadas por el subgenoma de *S. cerevisiae* constituyente genómico de la cepa híbrida de estudio (cepa 820, *S. cerevisiae* x *S. eubayanus*).

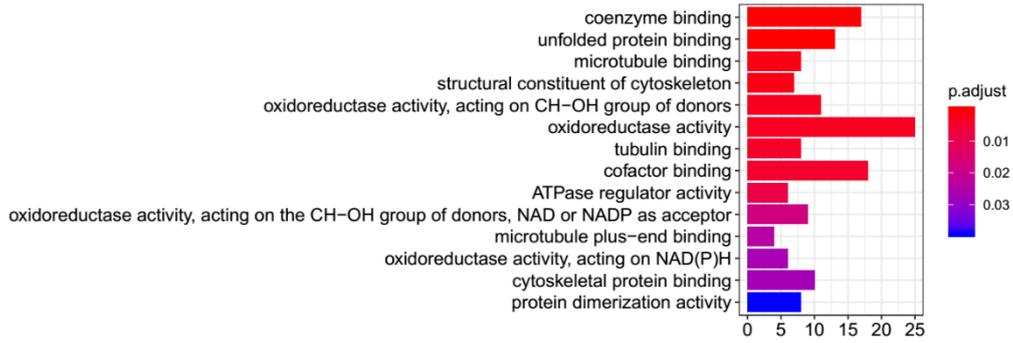
### **7.8.3 Análisis de enriquecimiento por ontología de genes GO de los datos transcriptómicos obtenidos mediante genoma de referencia**

Al igual que el procesamiento de los datos transcriptómicos obtenidos *de novo*, se optó por un análisis de enriquecimiento de ontología de genes. Cada conjunto de genes diferencialmente expresados en las condiciones obtenidos mediante genoma de referencia fueron sometidos a la agrupación funcional.

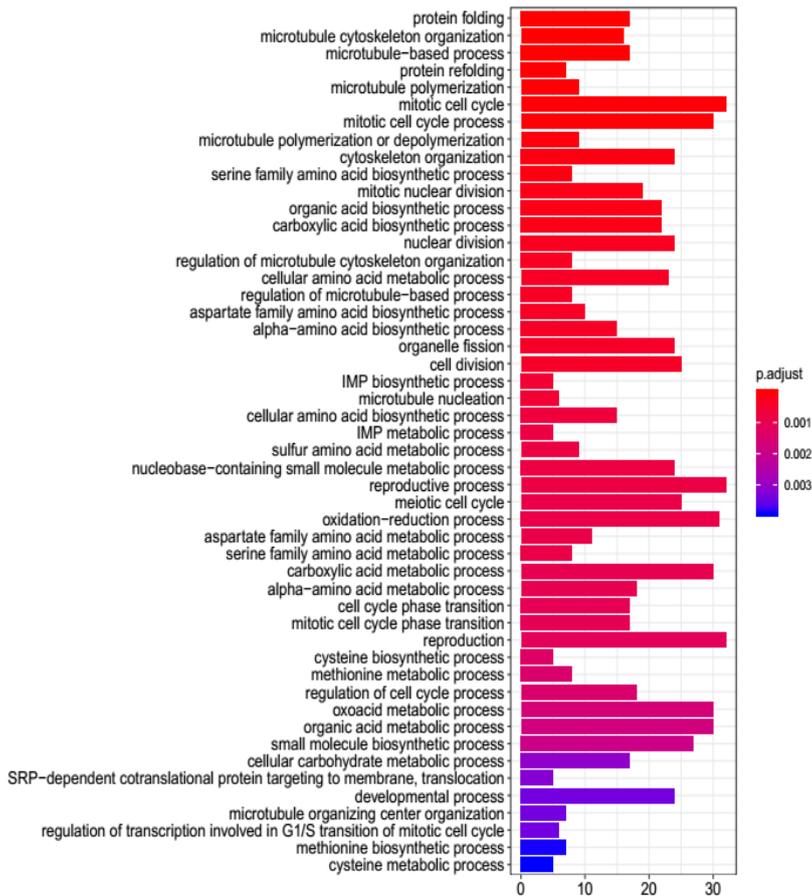
### **7.8.4 Análisis de enriquecimiento por ontología de genes del conjunto de F2 industrial**

El análisis del conjunto de los 200 genes diferencialmente expresados en la condición de F2 industrial resultaron en 119 categorías significativamente enriquecidas dentro del término Proceso Biológico, 14 categorías dentro el termino Función Molecular y 32 categorías en Componente Celular. A diferencia del análisis de enriquecimiento de los datos obtenidos *de novo*, en esta parte del análisis se obtuvieron mayor número de categorías significativamente enriquecidas en cada término.

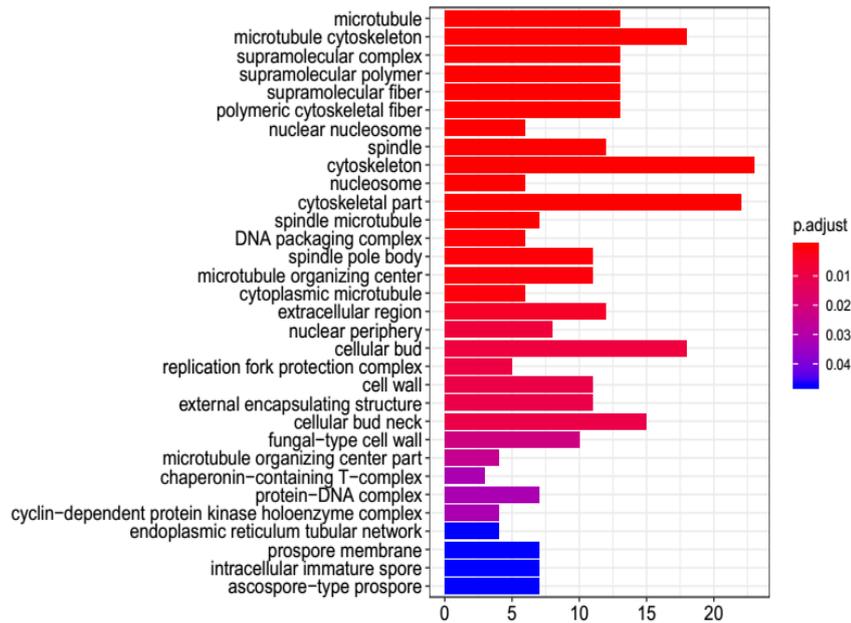
A)



B)



C)

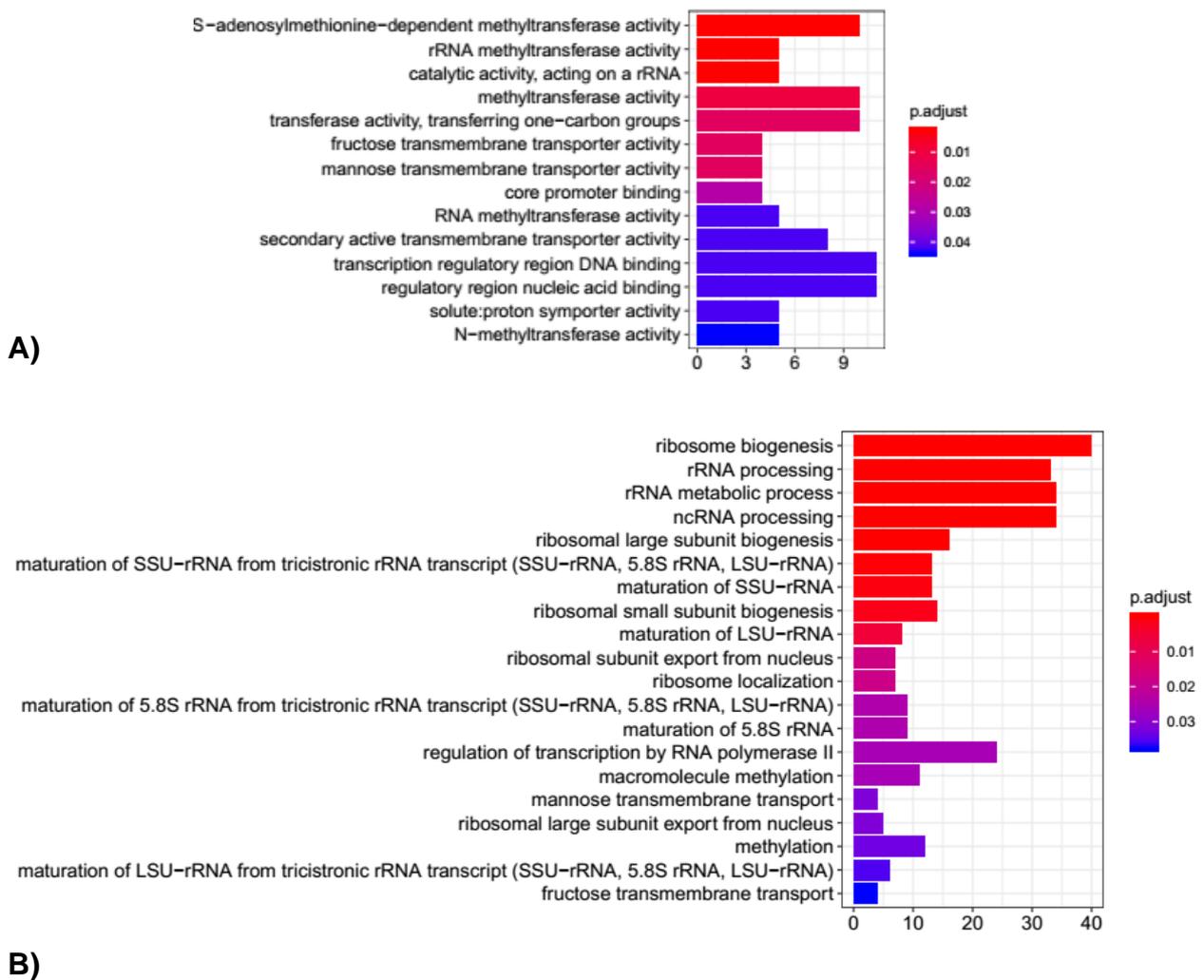


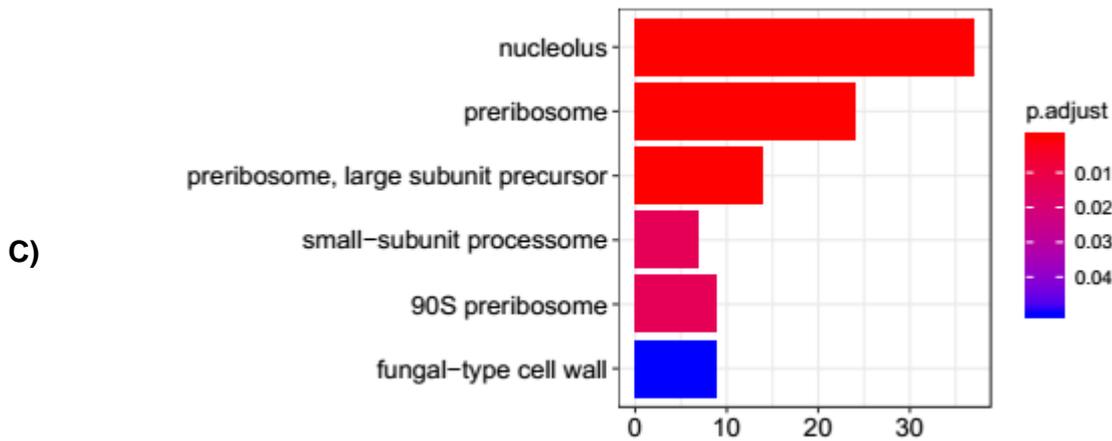
**Figura 17.** Análisis de enriquecimiento por ontología de genes de los datos transcriptómicos de F2 industrial, obtenidos mediante genoma de referencia. **A)** Categorías significativamente enriquecidas dentro del término Función Molecular. **B)** Categorías enriquecidas que cumplen los criterios estadísticos dentro de conjunto Proceso Biológico) Categorías que se ajustan al umbral estadístico dentro de la agrupación Componente Celular.

Cada categoría enriquecida en los análisis de los datos obtenidos *de novo* están dentro de lo observado en los análisis con genoma de referencia. Sin embargo, en el conjunto de genes correspondiente a F2 industrial *de novo* no se observaron categorías enriquecidas con el ciclo celular, mientras que en F2 industrial obtenido con genoma de referencia se observan categorías como proceso celular mitótico (GO:1903047), división nuclear mitótica (GO:0140014), división celular (GO:0051301), ciclo reproductivo (GO:0022414), ciclo celular meiótico (GO:0051321), transición de fase del ciclo celular mitótico (GO:0044772), reproducción (GO:0000003), regulación del ciclo celular (GO:0010564), regulación de la transcripción involucrada en la transición de G1/S del ciclo celular mitótico (GO:0000083), reproducción sexual (GO:0019953), esporulación sexual (GO:0034293), salida de la mitosis (GO:0010458) entre otros Procesos Biológicos relacionados con el ciclo celular mitótico y meiótico (Figura 17).

### 7.8.5 Análisis de enriquecimiento por ontología de genes de F2 laboratorio

El análisis de enriquecimiento GO para los 159 GDE's en F2 laboratorio evidenciaron significancia estadística en 14 categorías a nivel Función Molecular, 26 categorías en Proceso Biológico y 6 en Componente Celular. De forma discordante con los datos obtenidos con un enfoque *de novo* obtuvimos categorías ontológicas completamente diferentes a las observadas en el enfoque *de novo*.





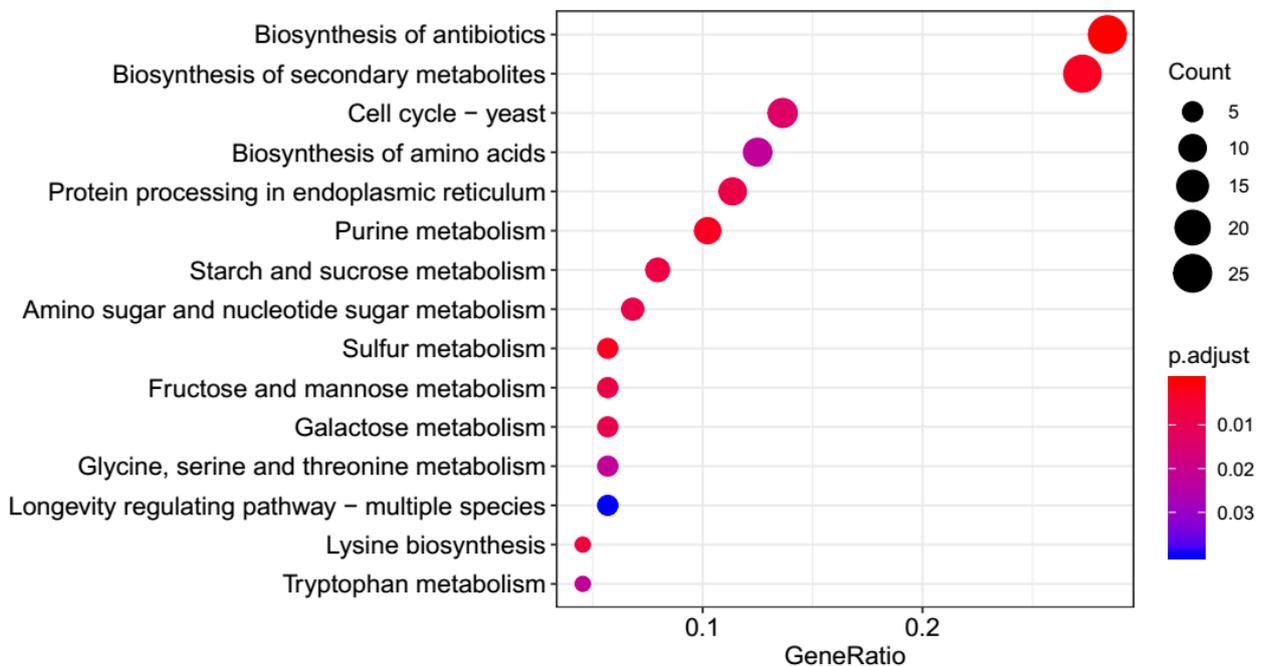
**Figura 18.** Análisis por ontología de genes de los datos transcriptómicos obtenidos con genoma de referencia en F2 laboratorio. **A)** Categorías ontológicas enriquecidas dentro de Función Molecular. **B)** Categorías que se ajustan a los criterios estadísticos dentro del término Proceso Biológico. **C)** Categorías ontológicas que cumplen el umbral estadístico dentro de Componente Celular.

Las diferentes categorías observadas en el enfoque con genoma de referencia hacen énfasis principalmente en Procesos Biológicos como biogénesis del ribosoma (GO:0042254), proceso metabólico de ARN ribosomal (GO:0006364), biogénesis de la subunidad grande del ribosoma (GO:0042273), biogénesis de la subunidad pequeña del ribosoma (GO:0042274), entre otras categorías relacionadas con el metabolismo de ARN ribosomal y biogénesis de ribosomas, además encontramos que las categorías transporte de manosa y fructosa se encontraron significativamente enriquecidas con genes *HXT1*, 3, 4 y 7 (GO:0015761 y GO:0015755, respectivamente) (Figura 18).

#### 7.8.6 Análisis de enriquecimiento de rutas metabólicas de F2 industrial

Con el fin de hacer inferencias en la funcionalidad de los conjuntos de los GDE's en F2 industrial se realizaron análisis de enriquecimiento de rutas metabólicas utilizando la base de datos KEGG. Al igual que los GDE's en F2 industrial obtenidos con ensamble *de novo*, las rutas metabólicas significativamente enriquecidas aquí fueron las mismas; biosíntesis de antibióticos (sce01130), metabolismo de almidón y sacarosa (sce00500), metabolismo de galactosa (sce00052), metabolismo de fructosa y manosa (sce00051) y metabolismo de amino azúcares y azúcar de nucleótidos (sce00520). Estos resultados sustentan lo obtenido mediante ensamblaje del transcriptoma *de novo*. No obstante, otras rutas

metabólicas fueron enriquecidas con los datos obtenidos con genoma de referencia. Biosíntesis de aminoácidos (sce01230), ciclo celular en levaduras (sce04111), biosíntesis de metabolitos secundarios (sce01110), son algunas de las categorías KEGG significativamente enriquecidas bajo el contexto de análisis guiado por genoma de referencia (Figura 19). Esto podría significar que el proceso de ensamblaje *de novo* presentó algunas problemáticas en la reconstrucción de los transcritos basándose solo en las lecturas de secuenciación. Además, en el análisis mediante genoma de referencia no implica ensamblaje de transcriptoma sino solo el mapeo de las lecturas de secuenciación hacia el genoma de referencia indicando mayor exactitud en el análisis de expresión diferencial.

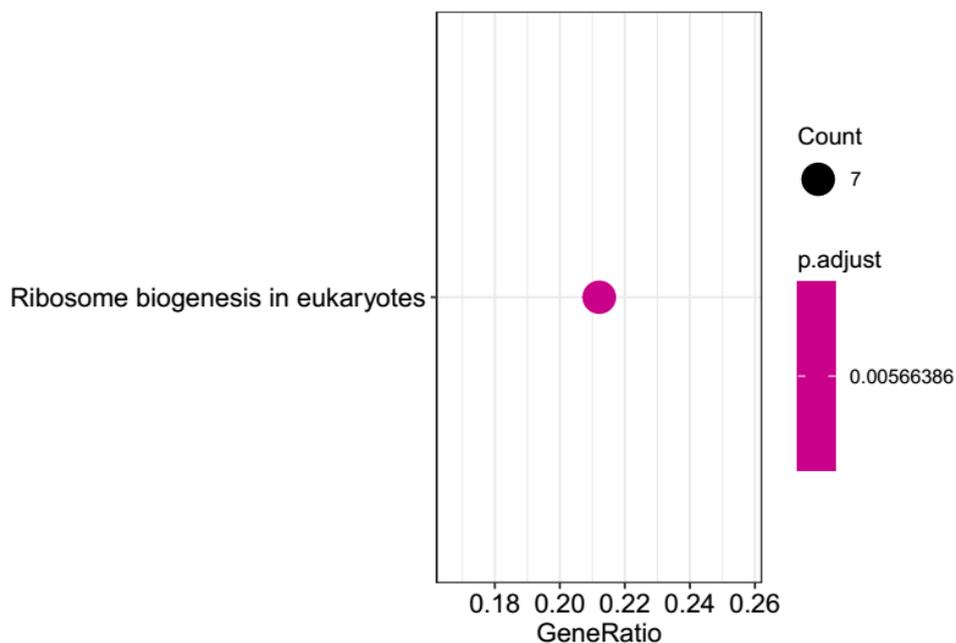


**Figura 19.** Gráfico de dispersión del análisis de enriquecimiento de vías metabólicas con el uso de la base de datos KEGG de los GDE's de F2 industrial. Datos obtenidos mediante ensamblaje *de novo* del transcriptoma. El eje x indica la proporción de genes observados frente a la base de datos KEGG. El eje y indica el nombre de la ruta metabólica KEGG. El tamaño del punto indica el número de genes que enriquecen la ruta KEGG y los colores de acuerdo al valor p ajustado. Se muestran las únicas rutas metabólicas enriquecidas en F2 industrial (FDR < 0.05).

### 7.8.7 Análisis de enriquecimiento de rutas metabólicas de F2 laboratorio

El análisis ontológico realizado para F2 laboratorio se tradujo en tan solo una categoría KEGG significativamente enriquecida; biogénesis de ribosoma en

eucariotas (sce03008). Dado que la significancia estadística en las rutas metabólicas enriquecidas en F2 laboratorios, ensamble *de novo*, es dudosa (Figura 9), estos resultados podrían sustentar un crecimiento mitótico por parte de la levadura de estudio. Debido a que los genes obtenidos con genoma de referencia enriquecen significativamente la biogénesis de ribosomas (Figura 19), probablemente sea un indicador de un estado biosintético de la levadura que sustenta muchos procesos biológicos entre ellos la división celular.



**Figura 20.** Gráfico de dispersión del análisis de enriquecimiento de vías metabólicas con el uso de la base de datos KEGG de los GDE's de F2 laboratorio.

### 7.8.8 El reducido número de genes diferencialmente expresados en F5 no sustenta un análisis de enriquecimiento GO y KEGG

Debido al reducido número de GDE's obtenidos en las comparaciones de F5 industrial y laboratorio, no es posible obtener categoría GO significativamente enriquecidas y mucho menos rutas metabólicas de acuerdo con la base de datos KEGG. Esto contrasta con los resultados obtenidos en el ensamblaje *de novo* ya que se observaron 143 GDE's en F5 industrial y 298 en F5 laboratorio, los cuales alimentaron, aunque de forma reducida, categorías GO específicas.

## 8. Discusión

### 8.1 Utilización de maltosa y maltotriosa

Los resultados reportados para la glucosa y fructuosa se manifestaron como no detectables por lo que se omitió su presentación. Esto es lo esperado ya que mientras se está en crecimiento, condiciones aerobias, el consumo de fructosa y de glucosa es prioritario para la célula seguida del consumo de maltosa y maltotriosa. Este patrón de consumo está determinado por la represión por glucosa o por catabolito (Meneses *et al.*, 2002). Tomando en cuenta el porcentaje inicial de maltotriosa (3.05% p/p) y de maltosa (11.06% p/p) el azúcar residual en la fermentación industrial fue del 55.74% y 5.16% respectivamente. Para la fermentación de laboratorio que se corrió a la par de la industrial manifestó 45.43% de maltotriosa y 3.5% de maltosa residual, ligeramente por debajo de lo observado en la fermentación industrial. Los datos de la fermentación de la réplica biológica (fermentadores 1R-LF5 hasta 3R-LF5) sustentan 49.19% de maltotriosa y 5.58% de maltosa residual. Partiendo de estos datos, se puede concluir que la levadura empleada en la fermentación posee un pobre sistema de captación de maltotriosa dado que consume solo es 50% de la maltotriosa representando un problema ya que la maltotriosa es el segundo carbohidrato más abundante en el mosto. El gen *AGT1* es el responsable para el transporte de maltotriosa en cepas de *Sacharomyces cerevisiae* (Orikasa *et al.*, 2018). Sin embargo, un homólogo, *MTY1*, también se encuentran en el subgenoma perteneciente a *S. eubayanus* en la cepa cervecera *S. pastorianus* lo que permite un uso eficiente de maltotriosa y por tanto eficiencia en la producción (Brickwedde *et al.*, 2018; Cousseau *et al.*, 2013). Dado los resultados podemos inferir que dicha cepa puede pertenecer al grupo I (Zaas) aunque existe reportes que apoyan la idea que el grupo I es muy heterogéneo en cuanto a la utilización de maltotriosa y que la variación en dicha captación es principalmente debido a la ausencia o presencia de transportadores de membrana específicos, en especial *MTY1* en cepas tipo lager (Magalhães *et al.*, 2016).

Una investigación interesante sugiere que a la ausencia de transportadores para maltotriosa, cepas de *S. cerevisiae* recurren la secreción de enzimas extra celulares que hidrolizan la maltotriosa a unidades de glucosa y así permitir el

crecimiento, aunque el costo en la aptitud biológica se manifiesta con un crecimiento retardado (Alves *et al.*, 2018). En investigaciones previas en nuestro laboratorio (Elizondo González, 2019, trabajo de tesis doctoral en desarrollo), con el mismo enfoque, pero con la cepa industrial 790, se obtuvo que la eficiencia del metabolismo de maltosa y maltotriosa es muy alta al dar resultados no detectables para dichos carbohidratos. Integrando los datos de consumo de carbohidratos, la cepa 820 puede presentar los transportadores, pero con muy baja afinidad o eficiencia, ya que se observa un consumo del 50% o bien puede presentar enzimas extracelulares que degraden la maltotriosa a unidades de glucosa para su óptima captación. Cualquiera de estos panoramas requiere más investigación y un enfoque importante es el de observar el perfil transcripcional de dicha cepa. Dado que las empresas cerveceras han adaptado el hábito de disminuir el número de veces en la reutilización la cepa de levadura hasta un determinado número fermentaciones, los fenómenos de adaptación y domesticación han disminuido su frecuencia. Por lo que la tarea de obtener nuevas y mejores cepas adaptadas a nichos industriales diversos ha caído en laboratorios de investigación y desarrollo. Hay evidencia que indica que la evolución adaptativa de cepas de *S. pastorianus* surge cuando se enfrenta a condiciones de selección, como lo son medios ricos en maltotriosa. Mejorando hasta cinco veces más la eficiencia de captación de maltotriosa o más resistencia a ambientes de altas concentraciones de alcohol (Brickwedde *et al.*, 2017; Krogerus y Holmström, 2018).

## **8.2 Metabolismo de aminoácidos, compuestos volátiles y de azufre a nivel industrial y laboratorio**

El consumo de aminoácidos por parte de la levadura fue medido en los días 2 y 5 de la fermentación teniendo diferencia significativa ( $P < 0.01$ ) en el día 2 para los aminoácidos Glutamato, Aspartato, Valina, Metionina, Leucina, Isoleucina, Fenilalanina y Tirosina, siendo el mayor consumo en las fermentaciones a escala de laboratorio. Las diferencias ( $P < 0.01$ ) para el día 5 se redujeron solo a los aminoácidos Valina, Isoleucina, Tirosina donde el mayor consumo se presentó a nivel laboratorio. En el día 5 la Prolina pareció aumentar su concentración en las fermentaciones a nivel laboratorio, una explicación plausible de este acontecimiento puede ser la actividad de algunas proteasas provenientes de células lisadas o bien

provenientes de la malta, las cuales tiene efecto en proteínas y así aumentando ligeramente la concentración de prolina. Estas diferencias en el consumo de aminoácidos a nivel industrial y de laboratorio pueden estar dadas por las diferencias en los volúmenes utilizados en las fermentaciones e incluso la geometría de los fermentadores y la edad generacional de la levadura (Speers y Stokes, 2009). Esto toma sentido ya que para la obtención de la media muestral de consumo de aminoácidos se partió de un experimento donde la levadura tiene la misma edad generacional que la utilizada a nivel industrial y otro experimento donde se cambió la edad generacional de la levadura (replica biológica). Entonces nuestra media muestral de consumo de aminoácidos contiene parte de la variabilidad biológica debido a la edad generacional de la levadura. De forma paralela, las diferencias observadas en la producción de compuestos volátiles están estrechamente relacionadas con el metabolismo de aminoácidos. Las diferencias en el consumo de aminoácidos que se mantienen en el día 5 repercuten directamente en la cerveza ya que se ha demostrado que la Valina tiene efectos en la producción de compuestos volátiles promoviendo la producción de isobutil alcohol y 2-metil-propanol (Kobayashi *et al.*, 2008). Una investigación confirma que la valina está involucrada en el enriquecimiento de compuestos volátiles como el diacetilo, propanol e isobutanol (Yin *et al.*, 2018). Esto concuerda con nuestros resultados, ya que el mayor consumo de Valina en fermentaciones de laboratorio se correlaciona con la mayor producción de isobutanol, propanol y acetato de etilo también a nivel de laboratorio. Por otra parte, la isoleucina también tiene correlación directa con algunos de los compuestos volátiles ya que en fermentaciones de vino se ha demostrado que la adición de isoleucina promueve la producción de acetato de amilo, acetato de hexilo, acetato de etilo y amil alcohol (Trinh *et al.*, 2010). En nuestros resultados observamos un incremento en la producción de acetato de isoamilo, alcohol isoamílico y hexanoato de etilo. Estos datos sugieren que además de la suplementación específica, que exponen las investigaciones anteriores, el consumo diferencial en escala laboratorio puede producir cerveza no comparable con la producida a nivel industrial. No obstante, los niveles de detección sensorial expuestos por Olaniran y cols (2017) están muy por encima de los observados en nuestros resultados indicado que la apreciación sensorial tal vez no sea significativa. Procopio y cols, (2014) exponen un patrón de consumo por parte de *S. pastorianus*, en fermentaciones a escala laboratorio, muy similar a lo observado en nuestros

resultados. Sin embargo, el perfil de compuestos volátiles difiere sustancialmente. Esto tienen sentido ya que se ha establecido que el patrón de consumo de aminoácidos por parte de especies de *Saccharomyces* es un proceso conservado y fundamentalmente igual entre las especies analizadas (Crépin *et al.*, 2012). Por otro lado, la producción de compuestos volátiles está condicionada por la temperatura y por la abundancia de los aminoácidos específicos anteriormente descritos (Olaniran *et al.*, 2011). En cuanto a la producción de compuestos de azufre se observó diferencia significativa siendo a nivel de laboratorio la mayor producción de sulfuro de hidrogeno, metanotiol, sulfuro de etileno y S-metil-tioacetato. En nuestro laboratorio ya se ha caracterizado que la cepa 820 es una fuerte productora de compuestos volátiles y que en mostos reducidos en compuestos nitrogenados esta levadura tiende a la producción de dichos compuestos (Leal. 2011). Los resultados nos indican que la distribución de los compuestos nitrogenados es menos homogénea a nivel laboratorio permitiendo crear un ambiente aparentemente bajo en dichos compuestos por lo cual se observa una producción mayor de compuestos de azufre a nivel de laboratorio que a nivel industrial. Sin embargo, como se observa en el día 2 a nivel laboratorio el consumo de aminoácidos es mayor en casi todos los aminoácidos, pero en el día 5 el consumo de dichos aminoácidos prácticamente es el mismo. Los datos obtenidos hasta el momento muestran diferencias estadísticamente significativas e incluso se especula que dichas diferencias persistan a nivel molecular. Sin embargo, como sucede con los compuestos volátiles los cuales están por debajo de los límites de detección sensorial, tal vez estas diferencias no representen un significado práctico en contextos industriales.

### **8.3 Observaciones sobre las diferencias moleculares en el día 2 de la fermentación a nivel industrial y laboratorio**

Diferencias moleculares fueron determinadas en los análisis de expresión diferencial de genes a la hora de las comparaciones en parejas entre F2 industrial y F2 laboratorio. Se determinaron 286 genes regulados positivamente en condiciones industriales y 480 genes regulados positivamente en condiciones de laboratorio. Otra óptica de interpretación puede realizarse de tal forma que los genes regulados positivamente en F2 industrial son genes negativamente regulados en F2 laboratorio y viceversa. El análisis de agrupación de genes por su ontología estableció que, a nivel industrial, parte del metabolismo de carbohidratos esta positivamente regulada,

al observar genes como *EMI2*, *MAL32*, *MAL11* (*AGT1*) y *SUC2*. Es interesante destacar que el gen *MAL11*, sinónimo de *AGT1*, está diferencialmente expresado en condiciones industriales. El transcrito TRINITY\_DN738\_c0\_g1\_i3 de longitud 7217 nt correspondiente al gen *MAL11*, contiene un marco de lectura en el rango de 1366 a 3192 nt. Este marco de lectura abierto codifica a una proteína que presenta una identidad del 92% (561aa / 610 aa) con la cepa IFO 2010 reportada por Ogura y cols 2008 y un 93% (556 aa / 596 aa) con la cepa de referencia S288C. Esta permeasa de alfa glucósidos está involucrada en el transporte de maltosa, sacarosa, trealosa, maltotriosa, entre otros. Esto, mediante un mecanismo de transporte por simportador acoplado a protones (Henderson y Poolman, 2017) □. Dado que este gen ha sido reportado que proviene del subgenoma de *Saccharomyces cerevisiae*, algunos investigadores han reportado que cepas híbridas como *S. pastorianus* presentan un codón de paro en el nucleótido 1183 mientras que cepas tipo Ale codifican una proteína totalmente funcional (Vidgren *et al.*, 2009; Vidgren *et al.*, 2005). El gen encontrado en esta investigación no presenta dicho codón de paro por lo que probablemente codifique a una permeasa funcional. Esto pone en evidencia un problema fundamental; si el transportador conservado proveniente de la cepa parental *S. eubayanus* es funcional, como explicar la reducida utilización de maltotriosa en la cepa de estudio. Nuestros datos analíticos exponen que cerca del 50% de la maltotriosa no es captada por la levadura implicando algún defecto en su utilización.

Una investigación interesante prueba la hipótesis de que la interacción cruzada de los cromosomas parentales de levaduras híbridas juega un papel importante en el metabolismo de la maltotriosa (Brouwers *et al.*, 2019). Es decir, en cepas híbridas con mutantes nulos para ScAGT1, SeAGT1 complementa la funcionalidad mientras que el subgenoma de Sc brinda la parte regulatoria que activa los transportadores en el subgenoma de *S. eubayanus*. Partiendo de esta idea, podría haber fallas en la regulación cruzada de dichos genes lo que resulta en un pobre o nulo escenario de utilización de maltotriosa. Sin embargo, la cuestión fundamental permanece intrigante ya que existe un 50% de consumo de maltotriosa. Dentro de las categorías del metabolismo de carbohidratos que fueron enriquecidas significativamente, se encuentran dos genes interesantes que podrían explicar este consumo de maltotriosa. El gen *IMA1* y *5* son alfa glucosidasas que realizan

hidrólisis de algunos alfa glucósidos como la maltosa, isomaltosa y maltotriosa (Teste *et al.*, 2010). El gen *IMA5* ha sido responsabilizado de la hidrólisis extracelular de maltotriosa en cepas de *S. cerevisiae* mutantes para *AGT1*. El fenotipo observado fue de sobreexpresión de *IMA5*, *MAL12* y transportadores de glucosa de alta afinidad como *HXT4*, 6 y 7 (Alves *et al.*, 2018). En nuestra investigación observamos que el gen *IMA5* y un transportador de glucosa *HXT3* tuvieron una expresión diferencial a nivel industrial lo que trae consigo una posible explicación del fenotipo de captación de maltotriosa por la levadura de estudio.

Por otro lado, en la parte de metabolismo de aminoácidos se pudo observar que genes importantes en la ruta de Ehrlich como *ARO9* y *10*, *ADH5* y *ALD6* están positivamente regulados a nivel industrial indicando que gran parte de la ruta metabólica implicada en el metabolismo de aminoácidos como valina, isoleucina, fenilalanina, tirosina y metionina con la subsecuente producción de compuestos aromáticos (Hazelwood *et al.*, 2008) esta positivamente regulada en condiciones industriales pero no a nivel laboratorio. Estos resultados representan una brecha en el conocimiento del metabolismo de la levadura en ambas condiciones industriales y de laboratorio ya que nuestros resultados analíticos indican que, para casi todos los aminoácidos involucrados en la ruta de Ehrlich, mencionados anteriormente, están significativamente reducidos a nivel laboratorio y la producción de compuestos aromáticos como, ésteres y alcoholes de cadena larga se ven mayormente producidos a nivel de laboratorio. Dejando en total discrepancia los resultados analíticos con los datos transcriptómicos. Sin embargo, dichas diferencias en la producción de compuestos aromáticos no tienen impacto económico al no sobrepasar los valores umbral de detección sensorial expuestos en la discusión 7.2.

Es importante resaltar que el consumo de la mayoría de los aminoácidos fue significativamente mayor a nivel laboratorio. Esto concuerda con las diferencias moleculares observadas a nivel laboratorio ya que el análisis de enriquecimiento por ontología de los 480 genes regulados positivamente en F2 laboratorio denotan una sobrerrepresentación de Funciones Moleculares relacionadas con regulación como los son factores de transcripción, actividad ATPasa, actividad helicasa, unión a enzimas, entre otros. Este tipo de subcategorías enriquecidas en Función Molecular se traducen en el enriquecimiento de Procesos Biológicos como el ciclo celular

mitótico. Dado que el amino nitrógeno libre del mosto cervecero es el principal inductor de crecimiento celular en la levadura además de conferir mayor resistencia a estrés inducido por fermentación (Lei *et al.*, 2019; Lei *et al.*, 2013). En forma de especulación, podríamos involucrar la variante del inóculo ya que a nivel industrial se implementa suficiente levadura para reducir el tiempo de reproducción de la levadura y así reducir el tiempo de fermentación. En el caso de las condiciones de laboratorio pudo haber existido algunas discrepancias en el escalamiento del inóculo promoviendo que la levadura se dedique a alimentar su ciclo reproductivo.

Por otro lado, en la ruta metabólica Biosíntesis de antibióticos los genes agrupados en esta ruta metabólica tienen que ver más con la biosíntesis de aminoácidos como Lisina, cisteína, glutamato, entre otros. Este perfil de genes pudo aumentar las concentraciones de aminoácidos necesarios para el mantenimiento del estado reproductivo que depende puramente de fenómenos biosintéticos (Polymenis y Aramayo, 2015). Un apoyo a esta idea es que el análisis de enriquecimiento de rutas metabólicas indica que el ciclo celular y meiosis en la levadura están informalmente enriquecidas. Sin embargo, al momento del ajuste de los valores p se observa que este enriquecimiento representa un falso positivo al no cumplir con el criterio estadístico. Por lo que se debe tener cuidado a la hora de tomar estos datos en las interpretaciones fisiológicas de la levadura.

#### **8.4 Las diferencias observadas en el día 5 de la fermentación a nivel industrial y laboratorio indican respuesta a estrés y crecimiento celular, respectivamente.**

Las diferencias moleculares observadas en el día 5 de la fermentación entre las comparaciones industrial contra laboratorio indican que los 143 genes regulados positivamente a nivel industrial sobre representaron las subcategorías respuesta a fármaco y transporte transmembranal de amonio dentro de Proceso Biológico a nivel industrial. Los genes presentes en la categoría respuesta a fármaco sustentan la idea de que a nivel industrial existen condiciones más estresantes que a nivel laboratorio. Genes como *AQR1* codifican para un probable transportador que confiere resistencia a ácidos monocarboxílicos de cadena corta, además involucrado en la multifármaco resistencia (Tenreiro *et al.*, 2002). Otros transportadores como *HNM1*, *AGP2*, *ATO3* y *SAM3*. El primero involucrado en el transporte hacia dentro de la célula de compuestos nitrogenados y betaína durante

estrés hipersalino (Kiewietdejonge *et al.*, 2006). *AGP2* está involucrado en la respuesta al estrés oxidativo y existe evidencia que indica que regula el transporte de poliaminas y carnitina no por transporte sino por una actividad tipo sensor que transduce señales medioambientales que activan la transcripción de genes como *SAM3* y *HNM1* (Aouida *et al.*, 2013; Lee *et al.*, 2002). Está bien estudiado que las poliaminas juegan papeles importantes en la respuesta estrés medioambiental, como estrés oxidativo, estrés osmótico, por pH, entre otros (Valdés y Herrera, 2013). Por otro lado, el gen *ATO3* está involucrado en el estrés inducido por pH, además de estar involucrado en la desintoxicación por amonio dentro de la célula (Palková *et al.*, 2002). otros genes como *PDR16* presentan actividad de transferencia de fosfatidilinositol la cual está involucrada en la composición de fosfolípidos de membranas endoplasmáticas y plasmáticas lo cual altera la composición de membranas dando un posible mecanismo de multifármaco resistencia (van den Hazel *et al.*, 1999). El gen *PDR3* confirma aún más la situación estresante que enfrenta la levadura a nivel industrial ya codifica para un factor transcripcional que activa la red pleiotrópica de fármaco resistencia. Activa positivamente a *SNQ2* y *PDR5* los cuales codifican para una transportador tipo ABC involucrado en el transporte de fármacos. El gen *PDR3* se involucra también en la respuesta a solventes orgánicos (Vanacloig *et al.*, 2019). Con estos resultados exponemos que las condiciones a nivel industrial se presentan pronunciadamente hostiles hacia la levadura de tal forma que inducen significativamente la maquinaria bioquímica y genética para combatir dicho estrés. En cuanto al transporte de amonio, los genes presentes en esta categoría son los mismos presentes en respuesta a fármaco; *SAM3*, *HNM1*, *AQR1*, *AGP1* y *ATO3*.

Es interesante notar que solo pocos genes relacionados con el estrés están positivamente regulados en las condiciones de laboratorio, en contraste, mayor número de genes relacionados con el desarrollo de la levadura están positivamente regulados. Los dos únicos genes involucrados a la respuesta a estrés presentes en las condiciones de laboratorio son dos factores transcripcionales *MSN2* y *5* los cuales reconocen y se unen a la secuencia denominada elemento de respuesta a estrés (STRE, por sus siglas en inglés) activando la transcripción de genes como *CTT1*, *DDR2* y *HSP12*. Donde *CTT1* es una catalasa que ayuda a la protección frente al estrés oxidativo por peróxido de hidrogeno, *DDR2* es una proteína de

respuesta a multiestrés la cual es activada por varios xenobióticos, estrés fisiológico y medioambiental. *HSP12* codifica para una proteína de membrana de choque térmico involucrada en el mantenimiento de la organización de la membrana frente a estrés por calor, estrés oxidativo, osmótico, y baja de nutrientes como glucosa, oleato y alcohol (Estruch *et al.*, 2000). A pesar de la expresión de estos factores transcripcionales involucrados en la respuesta a estrés, la mayoría de los genes están directamente involucrados en la regulación del crecimiento celular. Tal como el gen *CDC7* que codifica una cinasa de treonina/serina necesaria para el inicio de la síntesis de ADN durante la meiosis además de regular el paso inicial de la replicación en la fase S del ciclo celular (Rossbach *et al.*, 2017; Valentin *et al.*, 2006; Wan *et al.*, 2006). Esto es interesante ya que al existir evidencia de que *CDC7* está involucrado en la meiosis, esto conduce a ser posibles eventos meióticos. Dado que uno de los principales fenotipos cerveceros es la escasa o nula reproducción sexual (Gallone *et al.*, 2018) esto resulta en que la cepa de estudio podría estar en una etapa no consagrada de adaptación a nicho cervecero, implicando dos ideas importantes; por un lado, esta cepa podría tener un potencial desconocido en cuanto a procesos de fermentación ya que aún falta camino evolutivo por recorrer en el nicho industrial. No obstante, si se quiere mantener un producto constante con la cepa de estudio es importante verificar la estabilidad genética ya que podría estar frente a eventos importantes de adaptación y así cambiar procesos metabólicos que afecte el proceso cervecero.

Otro gen interesante es *FIG2* que se requiere para un apareamiento eficiente en levaduras. También juega papeles importantes en el mantenimiento de la integridad de la pared celular durante el apareamiento (Guo *et al.*, 2000). En general junto con la categoría de la regulación de morfología celular y las subcategorías de gema celular y crecimiento polarizado, dentro del término Componente Celular, los genes se involucran en el mantenimiento de la morfogénesis y ciclo celular de la levadura.

## **8.5 Los datos transcriptómicos obtenidos con un enfoque guiados por genoma de referencia sustentan las observaciones en las comparaciones de F2 obtenido mediante ensamblaje *de novo*.**

Los resultados obtenidos para las comparaciones por pares en F2 industrial y laboratorio, obtenidos con genoma de referencia, indican un sustento importante frente a los obtenidos con ensamblaje *de novo*. Además de esto, ampliaron las observaciones ya que rutas metabólicas adicionales fueron significativamente enriquecidas. Este escenario podría indicar algunos problemas inherentes en el ensamblaje *de novo* que no permiten una reconstrucción completa del transcriptoma. Está bien demostrados que el ensamblaje *de novo*, con la utilización del software Trinity, se efectúan descubrimientos en nuevas isoformas de algunos transcritos ampliando el panorama frente a estudios guiados por genoma de referencia (Lu *et al.*, 2013; Haas *et al.*, 2013). Dado que la cepa de estudio se especula que es una levadura híbrida (*S. cerevisiae* x *S. eubayanus*) es relevante los datos obtenidos con el genoma de referencia (*S.c* S288C).

Los resultados indican que las rutas metabólicas significativamente enriquecidas presentes en el ensamblaje *de novo* de F2 están presentes en los resultados obtenidos con genoma de referencia brindando soporte a dichos resultados. Sin embargo, otras rutas metabólicas como el ciclo celular mitótico y metabolismo de algunos aminoácidos se vieron significativamente enriquecidas. Dado que en algunas cepas lager el aporte genético del subgenoma de *S. cerevisiae* (haploide) es más reducido que el de su contraparte de *S. eubayanus* (diploide) (Walther *et al.*, 2014) podría pensarse que este aporte genético proviene mayoritariamente de *S. eubayanus*. Sin embargo, está demostrado que los genomas de cepas de *S. eubayanus* y *S. cerevisiae* manifiestan una sintenia relativamente importante compartiendo no solo el mismo contenido genético sino también el orden específico de la mayoría de los loci (Baker *et al.*, 2015). Partiendo de esta idea, es posible que las lecturas de secuenciación provenientes de RNA mensajero de la cepa de estudio puedan mapear libremente en el genoma de ambas cepas parentales del híbrido en cuestión. Esto explica los resultados discrepantes entre el análisis guiado por genoma de referencia y *de novo*.

Por otro lado, el análisis mediante genoma de referencia, apoya la idea de que la levadura en cuestión presenta ciclo celular sexual al ver categorías ontológicas enriquecidas significativamente relacionadas con el ciclo celular meiótico, evidenciando así la importancia de mantener monitoreado el cultivo

principal con el fin de mantener un producto homogéneo. Ya que es bien reportado que el ciclo celular meiótico representa la materia prima para la evolución de cualquier organismo y por ende cambios en el genotipo y fenotipo (Owens *et al.*, 2018). En cuanto al ciclo celular mitótico, es posible observar que no sólo en F2 laboratorio, ensamblaje *de novo*, está presente este fenotipo sino también a nivel industrial en las comparaciones hechas con los datos obtenidos mediante genoma de referencia. Las rutas biosintéticas de aminoácidos, metabolismo de azúcares y categorías ontológicas relacionadas con el citoesqueleto de la levadura indican un estado evidentemente de crecimiento celular. Es importante especular, con la evidencia recabada hasta el momento, que la levadura de estudio utiliza la materia prima, esqueleto de carbono proveniente de aminoácidos y azúcares, para sustentar su exacerbado crecimiento celular, no solo a nivel laboratorio sino también a nivel industrial. Otro aspecto que evidencia el crecimiento celular es lo obtenido en F2 laboratorio, con datos obtenidos mediante genoma de referencia, es que la única ruta metabólica significativamente enriquecida fue la relacionada con la biogénesis del ribosoma en eucariotas. Este importante complejo macromolecular es de suma importancia en estados biosintéticos, como el crecimiento celular (Dever *et al.*, 2016).

#### **8.6 Las diferencias observadas en F5 sugieren un aporte minoritario por parte del subgenoma de *S. cerevisiae*.**

Los resultados obtenidos para las comparaciones de F5 fueron evidentemente reducidos. Esto se observó en el análisis exploratorio mediante gráfico multidimensional, que expuso similitud de variación en las muestras correspondientes a F5. El escaso número de 8 GDE's a nivel industrial y tan solo 1 GDE a nivel laboratorio, sustentan la idea de que a nivel en el día 5 de la fermentación existe reducida diferencia. Retomando la idea de las cepas lager tipo 1 presentan normalmente mayor contenido de subgenoma de *S. eubayanus* (diploide) (Walther *et al.*, 2014), esto puede explicar que, bajo el contexto del genoma de *S. cerevisiae*, no existen diferencias moleculares en el día 5 de la fermentación. Esto es económicamente importante ya que en el último tiempo de la fermentación la levadura produce todo el conjunto de moléculas sensoriales, alcohol y otros componentes importantes para el producto final (Boulton y Quain, 2001). Bajo el esquema de la misma idea es importante realizar el análisis con genoma de

referencia perteneciente a *S. eubayanus* para evidenciar y sustentar las diferencias encontradas en los análisis transcriptómicos por medio de ensamblaje *de novo*.

## 9. Conclusiones

Bajo las condiciones establecidas en los experimentos realizados en esta investigación, las conclusiones procedentes de los datos obtenidos son las siguientes:

1. Los datos transcriptómicos procedentes de las comparaciones del día 2 de la fermentación exponen diferencias moleculares que resultan en genes involucrados en el metabolismo de carbohidratos positivamente regulados a nivel industrial. A nivel laboratorio se observan genes involucrados en la regulación del ciclo celular mitótico de la levadura. Las rutas metabólicas positivamente reguladas en el día 2 de la fermentación a nivel industrial confirman los eventos moleculares relacionados con el metabolismo de carbohidratos y biosíntesis de aminoácidos. También, en F2 industrial se vio positivamente activada la mayor parte de la ruta de Ehrlich involucrada en la producción de compuestos aromáticos. Esto es genes como *ARO9*, *ARO10*, *ADH5* y *ALD6* fueron positivamente regulados en F2 industrial
2. Las diferencias moleculares para las comparaciones de F5 indican que a nivel industrial existe un ambiente con mayor efecto estresante que a nivel laboratorio, al verse enriquecidas categorías como respuesta a fármaco y transporte de amonio dentro de Procesos Biológicos. En cuanto a F2 laboratorio se confirma la idea de que los procesos de desarrollo celular y regulación de la morfogénesis están positivamente regulados. Además de ver Componentes Celulares como gema celular y crecimiento polarizado significativamente enriquecidos.
3. Los datos transcriptómicos apuntan hacia la idea de que el proceso evolutivo de la levadura de estudio está en proceso, ya que presenta genes relacionados con la meiosis y reducida utilización de la maltotriosa.
4. Los datos obtenidos mediante genoma de referencia, sustentan las diferencias observadas para las comparaciones en F2 obtenidas mediante ensamblaje *de novo*. Implicando rutas metabólicas relacionadas con el procesamiento de carbohidratos positivamente reguladas a nivel industrial. Además, amplía el panorama genético y metabólico al observar que en el F2 industrial están presentes rutas metabólicas relacionadas con el ciclo celular mitótico y estados biosintéticos de la célula. También se vieron enriquecidas categorías ontológicas

relacionadas con reproducción sexual en levadura, sustentando la idea de posibles eventos meióticos en la levadura de estudio. Rutas metabólicas como biogénesis de ribosomas en eucariotas en F2 laboratorio, apoyando la idea de que la levadura desvía su fuente de carbono hacia eventos biosintéticos como ciclos celulares mitóticos y meióticos.

5. Bajo el contexto del genoma de *S. cerevisiae* no se observaron diferencias moleculares en las comparaciones del día 5 de la fermentación. Contrastando lo obtenido con enfoque *de novo*. Esto posiblemente evidencie el aporte mayoritario del subgenoma de *S. eubayanus* en estas diferencias.

## 10. Literatura Citada

- Akao, T., Yashiro, I., Hosoyama, A., Kitagaki, H., Horikawa, H., Watanabe, D., Shimoi, H. (2011). Whole-genome sequencing of sake yeast *Saccharomyces cerevisiae* Kyokai no. 7. *DNA Research*, 18(6), 423–434. <https://doi.org/10.1093/dnares/dsr029>
- Alba-Lois, L. y Segal-Kischinevzky, C. (2010) Beer & Wine Makers. *Nature Education* 3(9):17
- Alves, S. L., Thevelein, J. M., & Stambuk, B. U. (2018). Extracellular maltotriose hydrolysis by *Saccharomyces cerevisiae* cells lacking the AGT1 permease. *Letters in Applied Microbiology*, 67(4), 377–383. <https://doi.org/10.1111/lam.13048>
- Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics (Oxford, England)*, 31(2), 166–169. <https://doi.org/10.1093/bioinformatics/btu638>
- Aouida, M., Teixeira, M. R., Thevelein, J. M., Poulin, R., Ramotar, D., & Ramotar, D. (2013). Agp2, a Member of the Yeast Amino Acid Permease Family, Positively Regulates Polyamine Transport at the Transcriptional Level. *PLoS ONE*, 8(6), e65717. <https://doi.org/10.1371/journal.pone.0065717>
- Aslankoochi, E., Zhu, B., Rezaei, M. N., Voordeckers, K., Maeyer, D. De, Marchal, K., Verstrepen, K. J. (2013). Dynamics of the *Saccharomyces cerevisiae* transcriptome during bread dough fermentation. *Applied and Environmental Microbiology*, 79(23), 7325–7333. <https://doi.org/10.1128/AEM.02649-13>
- Baker, E. C. P., Peris, D., Moriarty, R. V., Li, X. C., Fay, J. C., & Hittinger, C. T. (2019). Mitochondrial DNA and temperature tolerance in lager yeasts. *Science Advances*, 5(1), 1–8. <https://doi.org/10.1126/sciadv.aav1869>
- Baker, E. C., Wang, B., Bellora, N., Peris, D., Hulfachor, A. B., Koshalek, J. A., Hittinger, C. T. (2015). The genome sequence of *saccharomyces eubayanus* and the domestication of lager-brewing yeasts. *Molecular Biology and Evolution*, 32(11), 2818–2831. <https://doi.org/10.1093/molbev/msv168>
- Barnett, J. A. (2000). A history of research on yeasts 2: Louis Pasteur and his contemporaries, 1850-1880. *Yeast*, 16(8), 755–771. [https://doi.org/10.1002/1097-0061\(20000615\)16:8<755::AID-YEA587>3.0.CO;2-4](https://doi.org/10.1002/1097-0061(20000615)16:8<755::AID-YEA587>3.0.CO;2-4)
- Bioconductor. (2019). Bioconductor - edgeR. Retrieved March 16, 2019, from <https://bioconductor.org/packages/release/bioc/html/edgeR.html>
- Bolat, I., Romagnoli, G., Zhu, F., Pronk, J. T., & Daran, J. M. (2013). Functional analysis and transcriptional regulation of two orthologs of ARO10, encoding broad-substrate-specificity 2-oxo-acid decarboxylases, in the brewing yeast *Saccharomyces pastorianus* CBS1483. *FEMS Yeast Research*, 13(6), 505–517. <https://doi.org/10.1111/1567-1364.12051>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Borneman, A. R., & Pretorius, I. S. (2014). Genomic insights into the *Saccharomyces sensu stricto* complex. *Genetics*, 199(2), 281–291. <https://doi.org/10.1534/genetics.114.173633>

- Boulton, C. Christopher M. ., & Quain, D. (2001). *Brewing yeast and fermentation*. Blackwell Science. Retrieved from <https://www.wiley.com/en-us/Brewing+Yeast+and+Fermentation-p-9781405152686>
- Brickwedde, A., Brouwers, N., Broek, M. van den, Gallego Murillo, J. S., Fraiture, J. L., Pronk, J. T., & Daran, J. M. G. (2018). Structural, physiological and regulatory analysis of maltose transporter genes in *Saccharomyces eubayanus* CBS 12357T. *Frontiers in Microbiology*, 9(AUG), 1–18. <https://doi.org/10.3389/fmicb.2018.01786>
- Brickwedde, A., van den Broek, M., Geertman, J. M. A., Magalhães, F., Kuijpers, N. G. A., Gibson, B., ... Daran, J. M. G. (2017). Evolutionary engineering in chemostat cultures for improved maltotriose fermentation kinetics in *saccharomyces pastorianus* lager brewing yeast. *Frontiers in Microbiology*, 8(SEP), 1–15. <https://doi.org/10.3389/fmicb.2017.01690>
- Brouwers, N., Brickwedde, A., Vries, A. R. G. de, Broek, M. van den, Weening, S. M., Eijnden, L. van den, ... Daran, J.-M. G. (2019). Maltotriose consumption by hybrid *Saccharomyces pastorianus* is heterotic and results from regulatory cross-talk between parental sub-genomes. *BioRxiv*, 679563. <https://doi.org/10.1101/679563>
- Brown, C. A., Murray, A. W., & Verstrepen, K. J. (2010). Rapid Expansion and Functional Divergence of Subtelomeric Gene Families in Yeasts. *Curr Biol.*, 20(10), 895–903. <https://doi.org/10.1016/j.cub.2010.04.027>
- Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., Whited, J. L. (2017). A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. *Cell Reports*, 18(3), 762–776. <https://doi.org/10.1016/j.celrep.2016.12.063>
- Carlson M (2019). *org.Sc.sgd.db: Genome wide annotation for Yeast*. R package version 3.8.2.
- Carvalho-Netto, O. V, Pereira, G. A., Calderón, L. AL, Argueso, J. L., Carazzolle, M. F., Mofatto, L. S., ... Noronha, M. F. (2015). *Saccharomyces cerevisiae* transcriptional reprogramming due to bacterial contamination during industrial scale bioethanol production. *Microbial Cell Factories*, 14(1), 1–13. <https://doi.org/10.1186/s12934-015-0196-6>
- Casalta, E., Aguera, E., Picou, C., Rodriguez-Bencomo, J. J., Salmon, J. M., & Sablayrolles, J. M. (2010). A comparison of laboratory and pilot-scale fermentations in winemaking conditions. *Applied Microbiology and Biotechnology*, 87(5), 1665–1673. <https://doi.org/10.1007/s00253-010-2652-x>
- Chen, Y., Lun, A. T. L., & Smyth, G. K. (2016). From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research*, 5, 1438. <https://doi.org/10.12688/f1000research.8987.2>
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., ... Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, 17(1), 1–19. <https://doi.org/10.1186/s13059-016-0881-8>
- Costa-Silva, J., Domingues, D., & Lopes, F. M. (2017). RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS ONE*, 12(12), 1–18. <https://doi.org/10.1371/journal.pone.0190152>
- Cousseau, F. E. M., Alves, S. L., Trichez, D., & Stambuk, B. U. (2013). Characterization of maltotriose transporters from the *Saccharomyces eubayanus* subgenome of the hybrid *Saccharomyces pastorianus* lager brewing

- yeast strain Weihenstephan 34/70. *Letters in Applied Microbiology*, 56(1), 21–29. <https://doi.org/10.1111/lam.12011>
- Crépin, L., Nidelet, T., Sanchez, I., Dequin, S., & Camarasa, C. (2012). Sequential use of nitrogen compounds by *Saccharomyces cerevisiae* during wine fermentation: A model based on kinetic and regulation characteristics of nitrogen permeases. *Applied and Environmental Microbiology*, 78(22), 8102–8111. <https://doi.org/10.1128/AEM.02294-12>
- Cubillos, F. A., Tisné, S., García, V., Molinet, J., Liti, G., Abarca, V., ... Brice, C. (2019). Identification of Nitrogen Consumption Genetic Variants in Yeast Through QTL Mapping and Bulk Segregant RNA-Seq Analyses. *G3 & Genes/Genomes/Genetics*, 7(6), 1693–1705. <https://doi.org/10.1534/g3.117.042127>
- Damas Buenrostro Luis Cástulo (2008). Evaluación y análisis de la expresión genética de *Saccharomyces cerevisiae* bajo condiciones de fermentación que afectan la floculación. Tesis de doctorado con especialidad en Biotecnología. Facultad de Ciencias Biológicas, UANL.
- Day, R. E., Higgins, V. J., Rogers, P. J., & Dawes, I. W. (2002). Characterization of the putative maltose transporters encoded by YDL247w and YJR160c. *Yeast*, 19(12), 1015–1027. <https://doi.org/10.1002/yea.894>
- Dever, T. E., Kinzy, T. G., & Pavitt, G. D. (2016). Mechanism and Regulation of Protein Synthesis in *Saccharomyces cerevisiae*. *Genetics*, 203(1), 65–107. <https://doi.org/10.1534/genetics.115.186221>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Dong, Y., Hu, J., Fan, L., & Chen, Q. (2017). RNA-Seq-based transcriptomic and metabolomic analysis reveal stress responses and programmed cell death induced by acetic acid in *Saccharomyces cerevisiae*. *Scientific Reports*, 7(June 2016), 1–16. <https://doi.org/10.1038/srep42659>
- Elizondo Gonzales, Ramiro. 2020. Tesis Doctoral en proceso. Tesis de Doctorado En Biotecnología. Facultad de Ciencias Biológicas, UANL.
- Elowitz, M. B., Levine, A. J., Siggia, E. D., & Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science*, 297(5584), 1183–1186. <https://doi.org/10.1126/science.1070919>
- Enyenihi, A. H., & Saunders, W. S. (2003). Large-Scale Functional Genomic Analysis of Sporulation and Meiosis in *Saccharomyces cerevisiae*, 54(January), 47–54.
- Estruch, F., Boy-Marcotte, E., Jacquet, M., Garreau, H., Hasan, R. N., & Renault, G. (2000). Hyperphosphorylation of Msn2p and Msn4p in response to heat shock and the diauxic shift is inhibited by cAMP in *Saccharomyces cerevisiae*. *Microbiology*, 146(9), 2113–2120. <https://doi.org/10.1099/00221287-146-9-2113>
- Ferreira, M., & Guido, F. (2018). Impact of Wort Amino Acids on Beer Flavour: A Review. *Fermentation*, 4(2), 23. <https://doi.org/10.3390/fermentation4020023>
- Gallone, B., Mertens, S., Gordon, J. L., Maere, S., Verstrepen, K. J., & Steensels, J. (2018). Origins, evolution, domestication and diversity of *Saccharomyces* beer yeasts. *Current Opinion in Biotechnology*, 49, 148–155. <https://doi.org/10.1016/j.copbio.2017.08.005>
- García-Ríos, E., Morard, M., Parts, L., Liti, G., & Guillamón, J. M. (2017). The genetic architecture of low-temperature adaptation in the wine yeast

- Saccharomyces cerevisiae. *BMC Genomics*, 18(1), 1–13. <https://doi.org/10.1186/s12864-017-3572-2>
- Gaudet, P., Livstone, M. S., Lewis, S. E., & Thomas, P. D. (2011). Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium, 12(5). <https://doi.org/10.1093/bib/bbr042>
- Gerard J. Tortora BRF& CLC. Microbiology: an introduction. 13a ed. Pearson Education I, editor. 2019.
- Gibson, B. R., Boulton, C. A., Box, W. G., Graham, N. S., Lawrence, S. J., Linforth, R. S. T., & Smart, K. A. (2008). Carbohydrate utilization and the lager yeast transcriptome during brewery fermentation. *Yeast*, 25(8), 549–562. <https://doi.org/10.1002/yea.1609>
- Gibson, B. R., Lawrence, S. J., Leclaire, J. P. R., Powell, C. D., & Smart, K. A. (2007). Yeast responses to stresses associated with industrial brewery handling. *FEMS Microbiology Reviews* (Vol. 31). <https://doi.org/10.1111/j.1574-6976.2007.00076.x>
- Godoy, A., Herrera, T. & Ulloa, M. (2003) *Más allá del pulque y el tepache: Las bebidas alcohólicas no destiladas indígenas de México*. Mexico: UNAM, Instituto de Investigaciones Antropológicas.
- Golding, I., Paulsson, J., Zawilski, S. M., & Cox, E. C. (2005). Real-time kinetics of gene activity in individual bacteria. *Cell*, 123(6), 1025–1036. <https://doi.org/10.1016/j.cell.2005.09.031>
- Guo, B., Styles, C. A., Feng, Q., & Fink, G. R. (2000). A Saccharomyces gene family involved in invasive growth, cell-cell adhesion, and mating. *Proceedings of the National Academy of Sciences*, 97(22), 12158–12163. <https://doi.org/10.1073/pnas.220420397>
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., ... Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8(8), 1494–1512. <https://doi.org/10.1038/nprot.2013.084>
- Hazelwood, L. A., Daran, J.-M., van Maris, A. J. A., Pronk, J. T., & Dickinson, J. R. (2008). The Ehrlich pathway for fusel alcohol production: a century of research on Saccharomyces cerevisiae metabolism. *Applied and Environmental Microbiology*, 74(8), 2259–2266. <https://doi.org/10.1128/AEM.02625-07>
- Henderson, R., & Poolman, B. (2017). Proton-solute coupling mechanism of the maltose transporter from Saccharomyces cerevisiae. *Scientific Reports*, 7(1), 14375. <https://doi.org/10.1038/s41598-017-14438-1>
- Herrera Gamboa Jessica Giselle (2018). Análisis genómico funcional de Saccharomyces sp. relacionado con el proteoma de la malta. Tesis de Doctorado con Especialidad en Biotecnología. Facultad de Ciencias Biológicas, UANL.
- Hornsey, I. S. (2003). *History of Beer and Brewing*. Cambridge: Royal Society of Chemistry. <https://doi.org/10.1039/9781847550026>
- Hrdlickova, R., Toloue, M., & Tian, B. (2017). RNA-Seq methods for transcriptome analysis. *Wiley Interdisciplinary Reviews. RNA*, 8(1). <https://doi.org/10.1002/wrna.1364>
- Iraqui, I., Vissers, S., Cartiaux, M., & Urrestarazu, A. (1998). Characterisation of Saccharomyces cerevisiae ARO8 and ARO9 genes encoding aromatic aminotransferases I and II reveals a new aminotransferase subfamily. *Molecular and General Genetics MGG*, 257(2), 238–248. <https://doi.org/10.1007/s004380050644>

- Jordan, P., Choe, J. Y., Boles, E., & Oreb, M. (2016). Hxt13, Hxt15, Hxt16 and Hxt17 from *Saccharomyces cerevisiae* represent a novel type of polyol transporters. *Scientific reports*, 6, 23502. doi:10.1038/srep23502
- Kiewietdejonge, A., Pitts, M., Cabuhat, L., Sherman, C., Kladwang, W., Miramontes, G., ... Ramirez, R. M. (2006). Hypersaline stress induces the turnover of phosphatidylcholine and results in the synthesis of the renal osmoprotectant glycerophosphocholine in *Saccharomyces cerevisiae*. *FEMS Yeast Research*, 6(2), 205–217. <https://doi.org/10.1111/j.1567-1364.2006.00030.x>
- Kneen, M. M., Stan, R., Yep, A., Tyler, R. P., Saehuan, C., & McLeish, M. J. (2011). Characterization of a thiamin diphosphate-dependent phenylpyruvate decarboxylase from *Saccharomyces cerevisiae*. *FEBS Journal*, 278(11), 1842–1853. <https://doi.org/10.1111/j.1742-4658.2011.08103.x>
- Kobayashi, M., Shimizu, H., & Shioya, S. (2008). Beer Volatile Compounds and Their Application to Low-Malt Beer Fermentation. *Journal of Bioscience and Bioengineering*, 106(4), 317–323. <https://doi.org/10.1263/jbb.106.317>
- Kopylova, E., Noé, L., & Touzet, H. (2012). SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, 28(24), 3211–3217. <https://doi.org/10.1093/bioinformatics/bts611>
- Krogerus, K., & Holmström, S. (2018). Enhanced Wort Fermentation with De Novo Lager Hybrids. *Applied and Environmental Microbiology*, 84(4), 1–20. <https://doi.org/10.1128/AEM.02302-17>
- Kucharczyk, K., Tuszyński, T., & Żyła, K. (2018). Effect of yeast harvest moment on a brewing process in beer produced on an industrial scale. *Czech Journal of Food Sciences*, 36(No. 5), 365–371. <https://doi.org/10.17221/157/2017-cjfs>
- Leal Guerra Clara Susana (2011). Regulación de la expresión de genes involucrados en la producción de compuestos de azufre en *Saccharomyces cerevisiae*. Tesis de Maestría con Especialidad en Microbiología. Facultad de Ciencias Biológicas, UANL.
- Leal-Morales, C. A., Bracker, C. E., & Bartnicki-, S. (1994). Subcellular localization, abundance and stability of chitin synthetases 1 and 2 from *Saccharomyces cerevisiae*. *Microbiology*. 10.1099/13500872-140-9-2207
- Lee, J., Lee, B., Shin, D., Kwak, S.-S., Bahk, J. D., Lim, C. O., & Yun, D.-J. (2002). Carnitine uptake by AGP2 in yeast *Saccharomyces cerevisiae* is dependent on Hog1 MAP kinase pathway. *Molecules and Cells*, 13(3), 407–412. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12132580>
- Legras, J. L., Merdinoglu, D., Cornuet, J. M., & Karst, F. (2007). Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. *Molecular Ecology*, 16(10), 2091–2102. <https://doi.org/10.1111/j.1365-294X.2007.03266.x>
- Lei, H., Feng, L., Peng, F., & Xu, H. (2019). Amino Acid Supplementations Enhance the Stress Resistance and Fermentation Performance of Lager Yeast During High Gravity Fermentation. *Applied Biochemistry and Biotechnology*, 187(2), 540–555. <https://doi.org/10.1007/s12010-018-2840-1>
- Lei, H., Zheng, L., Wang, C., Zhao, H., & Zhao, M. (2013). Effects of worts treated with proteases on the assimilation of free amino acids and fermentation performance of lager yeast. *International Journal of Food Microbiology*, 161(2), 76–83. <https://doi.org/10.1016/j.ijfoodmicro.2012.11.024>
- Lodolo, E. J., Kock, J. L. F., Axcell, B. C., & Brooks, M. (2008). The yeast *Saccharomyces cerevisiae* - The main character in beer brewing. *FEMS Yeast Research*, 8(7), 1018–1036. <https://doi.org/10.1111/j.1567-1364.2008.00433.x>

- Lu, B., Zeng, Z., & Shi, T. (2013). Comparative study of de novo assembly and genome-guided assembly strategies for transcriptome reconstruction based on RNA-Seq. *Science China Life Sciences*, 56(2), 143–155. <https://doi.org/10.1007/s11427-013-4442-z>
- Magalhães, F., Vidgren, V., Ruohonen, L., & Gibson, B. (2016). Maltose and maltotriose utilisation by group I strains of the hybrid lager yeast *Saccharomyces pastorianus*. *FEMS Yeast Research*, 16(5), 1–11. <https://doi.org/10.1093/femsyr/fow053>
- Martin Morgan (2018). BiocManager: Access the Bioconductor Project Package Repository. R package version 1.30.4. <https://CRAN.R-project.org/package=BiocManager>
- Meneses, F. J., Henschke, P. A., & Jiranek, V. (2002). A Survey of Industrial Strains of *Saccharomyces cerevisiae* Reveals Numerous Altered Patterns of Maltose and Sucrose Utilisation. *Journal of the Institute of Brewing*, 108(3), 310–321. <https://doi.org/10.1002/j.2050-0416.2002.tb00556.x>
- Naumov, G. I., & Naumova, E. S. (2010). [Comparative genetics of yeasts: a novel beta-fructosidase gene SUC8 in *Saccharomyces cerevisiae*]. *Genetika*, 46(3), 364–372. Recuperado de <http://www.ncbi.nlm.nih.gov/pubmed/20391781>
- Ogura, K., Nakamura, S., Kadokura, T., & Nakazato, A. (2008). MAL genes of industrial yeasts. *Journal of the brewing society of Japan*, 103(8), 629–637. <https://doi.org/10.6013/jbrewsocjapan1988.103.629>
- Olaniran, A. O., Hiralal, L., Mokoena, M. P., & Pillay, B. (2017). Flavour-active volatile compounds in beer: production, regulation and control. *Journal of the Institute of Brewing*, 123(1), 13–23. <https://doi.org/10.1002/jib.389>
- Olesen, K., Felding, T., Gjermansen, C., & Hansen, J. (2002). The dynamics of the *Saccharomyces carlsbergensis* brewing yeast transcriptome during a production-scale lager beer fermentation. *FEMS Yeast Research*, 2(4), 563–573. [https://doi.org/10.1016/S1567-1356\(02\)00155-1](https://doi.org/10.1016/S1567-1356(02)00155-1)
- Orikasa, Y., Mikumo, D., & Ohwada, T. (2018). A 2-Deoxyglucose-Resistant Mutant of *Saccharomyces cerevisiae* Shows Enhanced Maltose Fermentative Ability by the Activation of MAL Genes. *Foods (Basel, Switzerland)*, 7(4), 52. <https://doi.org/10.3390/foods7040052>
- Owens, S., Tang, S., & Hunter, N. (2018). Monitoring Recombination During Meiosis in Budding Yeast. *Methods in Enzymology*, 601, 275–307. <https://doi.org/10.1016/bs.mie.2017.12.005>
- Palková, Z., Devaux, F., Ričicová, M., Mináriková, L., Le Crom, S., & Jacq, C. (2002). Ammonia Pulses and Metabolic Oscillations Guide Yeast Colony Development. *Molecular Biology of the Cell*, 13(11), 3901–3914. <https://doi.org/10.1091/mbc.e01-12-0149>
- Pires, E. J., Teixeira, J. A., Brányik, T., & Vicente, A. A. (2014). Yeast: The soul of beer's aroma - A review of flavour-active esters and higher alcohols produced by the brewing yeast. *Applied Microbiology and Biotechnology*, 98(5), 1937–1949. <https://doi.org/10.1007/s00253-013-5470-0>
- Polymenis, M., & Aramayo, R. (2015). Translate to divide: control of the cell cycle by protein synthesis. *Microbial Cell*, 2(4), 94–104. <https://doi.org/10.15698/mic2015.04.198>
- Procopio, S., Brunner, M., & Becker, T. (2014). Differential transcribed yeast genes involved in flavour formation and its associated amino acid metabolism during brewery fermentation. *European Food Research and Technology*, 239(3), 421–439. <https://doi.org/10.1007/s00217-014-2236-6>

- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Raj, A., & van Oudenaarden, A. (2008). Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell*, 135(2), 216–226. <https://doi.org/10.1016/j.cell.2008.09.050>
- Raychowdhury, R., Gnirke, A., Fan, L., Yassour, M., Regev, A., di Palma, F., ... Adiconis, X. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), 644–652. <https://doi.org/10.1038/nbt.1883>
- Regelmann, J., Schu, T., Josupeit, F. S., Horak, J., Rose, M., Entian, K., ... Goethe-universita, W. (2003). Catabolite Degradation of Fructose-1, 6-bisphosphatase in the Yeast *Saccharomyces cerevisiae* : A Genome-wide Screen Identifies Eight Novel GID Genes and Indicates the Existence of Two Degradation Pathways, 14(April), 1652–1663. <https://doi.org/10.1091/mbc.E02>
- Rio, D. C., Ares, M., Hannon, G. J., & Nilsen, T. W. (2010). Purification of RNA using TRIzol (TRI reagent). *Cold Spring Harbor Protocols*, 2010(6), pdb.prot5439. <https://doi.org/10.1101/pdb.prot5439>
- Rodríguez, A., Cera, T. de la, Herrero, P., & Moreno, F. (2001). The hexokinase 2 protein regulates the expression of the GLK1, HXK1 and HXK2 genes of *Saccharomyces cerevisiae*. *Biochemical Journal*, 355(3), 625–631. <https://doi.org/10.1042/bj3550625>
- Rosbach, D., Bryan, D. S., Hesselberth, J. R., & Sclafani, R. (2017). Localization of Cdc7 Protein Kinase During DNA Replication in *Saccharomyces cerevisiae*. *G3 (Bethesda, Md.)*, 7(11), 3757–3774. <https://doi.org/10.1534/g3.117.300223>
- Saint-Prix, F., Bönquist, L., & Dequin, S. (2004). Functional analysis of the ALD gene family of *Saccharomyces cerevisiae* during anaerobic growth on glucose: the NADP<sup>+</sup>-dependent Ald6p and Ald5p isoforms play a major role in acetate formation. *Microbiology*, 150(7), 2209–2220. <https://doi.org/10.1099/mic.0.26999-0>
- Schmidt, F. R. (2005). Optimization and scale up of industrial fermentation processes. *Applied Microbiology and Biotechnology*, 68(4), 425–435. <https://doi.org/10.1007/s00253-005-0003-0>
- Shekhawat, K., Patterton, H., Bauer, F. F., & Setati, M. E. (2019). RNA-seq based transcriptional analysis of *Saccharomyces cerevisiae* and *Lachancea thermotolerans* in mixed-culture fermentations under anaerobic conditions, 1–15.
- Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., & Waterston, R. H. (2017). DNA sequencing at 40: Past, present and future. *Nature*, 550(7676), 345–353. <https://doi.org/10.1038/nature24286>
- Simon Andrews. (n.d.). Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. Retrieved Feb 15, 2019, from <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Smith, M. G., Des Etages, S. G., & Snyder, M. (2004). Microbial synergy via an ethanol-triggered pathway. *Molecular and Cellular Biology*, 24(9), 3874–3884. <https://doi.org/10.1128/mcb.24.9.3874-3884.2004>
- Speers, R. A., & Stokes, S. (2009). Effects of vessel geometry, fermenting volume and yeast repitching on fermenting beer. *Journal of the Institute of Brewing*, 115(2), 148–150. <https://doi.org/10.1002/j.2050-0416.2009.tb00360.x>

- Takhaveev, V., & Heinemann, M. (2018). Metabolic heterogeneity in clonal microbial populations. *Current Opinion in Microbiology*, 45, 30–38. <https://doi.org/10.1016/j.mib.2018.02.004>
- Tenreiro, S., Nunes, P. A., Viegas, C. A., Neves, M. S., Teixeira, M. C., Cabral, M. G., & Sá-Correia, I. (2002). AQR1 Gene (ORF YNL065w) Encodes a Plasma Membrane Transporter of the Major Facilitator Superfamily That Confers Resistance to Short-Chain Monocarboxylic Acids and Quinidine in *Saccharomyces cerevisiae*. *Biochemical and Biophysical Research Communications*, 292(3), 741–748. <https://doi.org/10.1006/bbrc.2002.6703>
- Teste, M.-A., François, J. M., & Parrou, J.-L. (2010). Characterization of a New Multigene Family Encoding Isomaltases in the Yeast *Saccharomyces cerevisiae*, the IMA Family. *Journal of Biological Chemistry*, 285(35), 26815–26824. <https://doi.org/10.1074/jbc.M110.145946>
- Trinh, T. T. T., Woon, W. Y., Yu, B., Curran, P., & Liu, S. Q. (2010). Effect of L-isoleucine and L-phenylalanine addition on aroma compound formation during longan juice fermentation by a co-culture of *Saccharomyces cerevisiae* and *Williopsis saturnus*. *South African Journal of Enology and Viticulture*, 31(2), 116–124.
- Valdés-Santiago, L., & Ruiz-Herrera, J. (2013). Stress and polyamine metabolism in fungi. *Frontiers in Chemistry*, 1, 42. <https://doi.org/10.3389/fchem.2013.00042>
- Valentin, G., Schwob, E., & Della Seta, F. (2006). Dual role of the Cdc7-regulatory protein Dbf4 during yeast meiosis. *The Journal of Biological Chemistry*, 281(5), 2828–2834. <https://doi.org/10.1074/jbc.M510626200>
- Van den Hazel, H. B., Pichler, H., Matta, M. A. do V., Leitner, E., Goffeau, A., & Daum, G. (1999). *PDR16* and *PDR17*, Two Homologous Genes of *Saccharomyces cerevisiae*, Affect Lipid Biosynthesis and Resistance to Multiple Drugs. *Journal of Biological Chemistry*, 274(4), 1934–1941. <https://doi.org/10.1074/jbc.274.4.1934>
- Vanacloig-Pedros, E., Lozano-Pérez, C., Alarcón, B., Pascual-Ahuir, A., & Proft, M. (2019). Live-cell assays reveal selectivity and sensitivity of the multidrug response in budding yeast. *The Journal of Biological Chemistry*, jbc.RA119.009291. <https://doi.org/10.1074/jbc.RA119.009291>
- Vidgren, V., Huuskonen, A., Virtanen, H., Ruohonen, L., & Londesborough, J. (2009). Improved fermentation performance of a lager yeast after repair of its AGT1 maltose and maltotriose transporter genes. *Applied and Environmental Microbiology*, 75(8), 2333–2345. <https://doi.org/10.1128/AEM.01558-08>
- Vidgren, V., Multanen, J. P., Ruohonen, L., & Londesborough, J. (2010). The temperature dependence of maltose transport in ale and lager strains of brewer's yeast. *FEMS Yeast Research*, 10(4), 402–411. <https://doi.org/10.1111/j.1567-1364.2010.00627.x>
- Vidgren, V., Ruohonen, L., & Londesborough, J. (2005). Characterization and functional analysis of the MAL and MPH Loci for maltose utilization in some ale and lager yeast strains. *Applied and Environmental Microbiology*, 71(12), 7846–7857. <https://doi.org/10.1128/AEM.71.12.7846-7857.2005>
- Walther, A., Hesselbart, A., & Wendland, J. (2014). Genome sequence of *Saccharomyces carlsbergensis*, the world's first pure culture lager yeast. *G3 (Bethesda, Md.)*, 4(5), 783–793. <https://doi.org/10.1534/g3.113.010090>
- Wan, L., Zhang, C., Shokat, K. M., & Hollingsworth, N. M. (2006). Chemical Inactivation of Cdc7 Kinase in Budding Yeast Results in a Reversible Arrest

- That Allows Efficient Cell Synchronization Prior to Meiotic Recombination. *Genetics*, 174(4), 1767–1774. <https://doi.org/10.1534/genetics.106.064303>
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1), 57–63. <https://doi.org/10.1038/nrg2484>
- Yin, H., He, Y., Dong, J., & Lu, J. (2018). Transcriptional profiling of amino acid supplementation and impact on aroma production in a lager yeast fermentation. *Journal of the Institute of Brewing*, 124(4), 425–433. <https://doi.org/10.1002/jib.508>
- Yu G, Wang L, Han Y, He Q (2012). “clusterProfiler: an R package for comparing biological themes among gene clusters.” *OMICS: A Journal of Integrative Biology*, 16(5), 284-287. doi: [10.1089/omi.2011.0118](https://doi.org/10.1089/omi.2011.0118).
- Zhao, H., Chen, J., Liu, J., & Han, B. (2015). Transcriptome analysis reveals the oxidative stress response in *Saccharomyces cerevisiae*. *RSC Advances*, 5(29), 22923–22934. <https://doi.org/10.1039/c4ra14600j>