



UNIVERSIDAD NACIONAL DE COLOMBIA

Secuenciación y caracterización del genoma de la cepa *Saccharomyces cerevisiae* 202-3 con potencial para la producción de etanol de segunda generación.

Yina Alejandra Cifuentes Triana

Universidad Nacional de Colombia
Facultad de Ciencias, Instituto de Biotecnología de la Universidad Nacional
Bogotá, Colombia
2016

Secuenciación y caracterización del genoma de la cepa *Saccharomyces cerevisiae* 202-3 con potencial para la producción de etanol de segunda generación.

Yina Alejandra Cifuentes Triana

Tesis presentada como requisito parcial para optar al título de:
Magister en Ciencias- Microbiología

Director (a):

Ph.D., Mario Enrique Velásquez Lozano

Codirector (a):

Ph.D., Andrés Mauricio Pinzón Velasco

Línea de Investigación:

Bioprocesos

Grupo de Investigación:

Grupo de Investigación en Procesos Químicos y Bioquímicos

Grupo de Investigación en Bioinformática y Biología de Sistemas

Universidad Nacional de Colombia

Facultad de Ciencias, Instituto de Biotecnología Universidad Nacional (IBUN)

Bogotá, Colombia

2016

Este trabajo lo dedico a mi familia por su apoyo incondicional.

Agradecimientos

Como primera instancia agradezco a las personas que me han acompañado durante toda mi formación académica y como persona, mi familia. Agradezco a mis tutores los Doctores Mario Enrique Velásquez y Andrés Mauricio Pinzón, pues gracias a su apoyo y ánimo pude desarrollar este proyecto de investigación.

Agradezco al grupo de investigación en GeorgiaTech liderado por el Profesor Jordan King, por brindarme las herramientas necesarias para realizar este trabajo. Así como el apoyo financiero de Ecopetrol y COLCIENCIAS.

Agradezco a todo el Grupo de Investigación en Procesos Químicos y Bioquímicos, en especial a mis compañeros de trabajo Juan Pablo Ortiz, Diana Morales, Paola Morales y Richard Ruiz, quienes apoyaron mi investigación. Así como el Grupo de investigación de Bioinformática y Biología Computacional y a todo el equipo que conforma el Instituto de Biotecnología IBUN.

Agradezco al Grupo de Investigación GERMINA, por permitirme usar sus instalaciones así como el laboratorio de biotecnología en el departamento de Agronomía.

Finalmente agradezco a todos mis amigos, Laura Jutinico, Luisa Dorado, Brian Castro, Sandra Montaña, Sergio Latorre, Jorge Charry, Rodrigo Pérez y muchos más por darme ánimo en los momentos más críticos de esta fase de mi vida.

Resumen

En este trabajo se presenta la secuenciación y caracterización del genoma de un aislado natural de levadura *Saccharomyces cerevisiae* (cepa 202-3), identificada como una cepa con potencial para la producción de etanol de segunda generación. En estudios previos llevados a cabo por el Grupo de Investigación de Procesos Químicos y Bioquímicos enfocados a la producción de etanol de segunda generación a partir de hidrolizados de bagazo de caña, se identificó de un grupo de cepas aisladas de *S. cerevisiae* un consumo de xilosa (azúcar presente en hidrolizados) del 2-5% para la cepa 202-3, siendo esta una ventaja para la fermentación de hidrolizados, ya que muchas cepas de *S. cerevisiae* son incapaces de procesarlo. Esta ventaja y otros aspectos positivos destacados durante los perfiles de fermentación evaluados en hidrolizados de bagazo de caña, hicieron de la cepa 202-3 una cepa candidata para mejorar la producción de etanol de segunda generación, por lo que se propuso como primer paso estudiar la cepa a nivel genómico.

La caracterización molecular se llevó a cabo mediante la secuenciación del genoma con la plataforma Hiseq Illumina 2000 finales pareados, el ensamblaje se realizó con diferentes programas, eligiendo finalmente el ensamblador ABySS con un kmer de 89. La predicción de genes se desarrolló con la aproximación de modelos ocultos de Markov con Augustus. Los genes identificados fueron anotados basados en similaridad con bases de datos públicas de nucleótidos y proteínas. Los registros fueron organizados a partir de funciones ontológicas a diferentes niveles jerárquicos, que permitieron identificar funciones y roles metabólicos centrales de la cepa de *S. cerevisiae* 202-3, destacando la presencia de cuatro posibles proteínas nuevas, dos de ellas asociadas probablemente al consumo positivo de xilosa de la cepa *S. cerevisiae* 202-3.

Palabras clave: etanol celulósico, *S. cerevisiae*, secuenciación genómica, consumo xilosa.

Abstract

In this work the sequencing and genome characterization of a natural isolate of *Saccharomyces cerevisiae* yeast (strain 202-3), identified with potential for the production of second generation ethanol is presented. In previous studies conducted by the Research Group of Chemical and Biochemical Processes focused on the production of second-generation ethanol from sugarcane bagasse hydrolysates, we identified a group of isolates of *S. cerevisiae* with a consumption of xylose (sugar present in hydrolysates) of 2-5% for strain 202-3, this being an advantage for the fermentation of hydrolysates, since many strains of *S. cerevisiae* are incapable of processing. This advantage and other prominent positive aspects during fermentation profiles evaluated bagasse hydrolysates made the strain 202-3 a candidate strain to improve the production of second-generation ethanol, which was proposed as a first step to study the strain at the genomic level.

The molecular characterization was carried out by genome sequencing with the Illumina HiSeq 2000 platform paired end; the assembly was performed with different programs, finally choosing the assembler ABYSS with kmer 89. Gene prediction was developed with the approach of hidden Markov models with Augustus. The genes identified were scored based on similarity with public databases of nucleotide and protein. Records were organized from ontological functions at different hierarchical levels, which identified central metabolic functions and roles of the *S. cerevisiae* strain 202-3, highlighting the presence of four possible new proteins, two of them probably associated with the positive consumption of xylose.

Keywords: cellulosic ethanol, *S. cerevisiae*, genome sequencing, xylose consumption.

Contenido

Agradecimientos	IV
Resumen	V
Abstract	VI
Lista de figuras	IX
Lista de tablas	XI
Lista de Símbolos y abreviaturas	XII
Introducción	XIII
Objetivos	XIV
1. Estado del arte	15
1.1 Producción de etanol de segunda generación.....	15
1.1.2. Retos del proceso para obtener etanol de segunda generación	16
1.2 Métodos analíticos en genómica	24
1.2.1 Secuenciación de cepas de <i>Saccharomyces cerevisiae</i> y su aplicación	24
1.2.2 Estrategias de ensamblaje para lecturas cortas.....	25
1.2.3 Anotación genómica	27
2. Materiales y métodos	29
2.1 Crecimiento y mantenimiento de la cepa de <i>S. cerevisiae</i>	29
2.2 Extracción de ADN y estrategia de secuenciación genómica	29
2.3 Análisis Computacional.....	29
2.3.1 Preprocesamiento de los datos.....	30
2.3.2 Ensamblaje del genoma : Aproximación <i>de novo</i>	31
2.3.3 Ensamblaje del genoma : Aproximación con genoma de referencia	31
2.3.4 Anotación genómica: Anotación estructural	32
2.3.5 Anotación genómica: Anotación funcional.....	32
3. Resultados y Discusión	33
3.1 Preprocesamiento de los datos	33
3.2 Ensamblaje	35
3.2.1 Ensamblaje <i>de novo</i>	35
3.3 Anotación genómica.....	36
3.3.1 Predicción de CDS	36
3.3.2 Predicción de ARN ribosomal y regiones ITS	36
3.3.3 Predicción de ARN de transferencia	37
3.3.4 Anotación funcional	37
3.3.5 Ensamblaje con genoma de referencia.....	40
3.4 Identificación de rutas metabólicas	47
3.5 Caracterización de proteínas no identificadas	49
4. Conclusiones y Recomendaciones	51

Anexo 1: Consumo de xilosa en medios sintéticos y amplificación del gen XDH1 ..	53
Anexo 2: Análisis de calidad del ADN sometido a secuenciación	55
Anexo 3: Métricas de los ensamblajes de novo.	56
Anexo 4: ARN de transferencia en <i>S. cerevisiae</i> 202-3	60
Anexo 5: Reportes obtenidos durante la anotación con términos GO	62
Anexo 6: Asociación de genes identificados en ensamblaje <i>de novo</i> a rutas metabólicas.	66
Anexo 7: Mutaciones encontradas con MUDI para la cepa 202-3 respecto a la cepa S288c.	124
Anexo 8: Filogenia para algunas cepas de <i>S. cerevisiae</i>.....	150
Anexo 9: Publicaciones y participaciones en eventos.....	151
Bibliografía	158

Lista de figuras

<i>Figura 1-1. Productos obtenidos a partir de materias lignocelulosicas. En este esquema se incluyen los azúcares fermentables del proceso, como glucosa, xilosa, arabinosa, manosa y galactosa. Así como algunos inhibidores formados durante la etapa de pretratamiento e hidrólisis, como el 2-furaldehído, 5-hidroximetilfurfural, ácido levúlinico, ácido fórmico, compuestos fenólicos y ácido acético. Adaptado de Ibraheem et al., 2013.</i>	16
<i>Figura 1-2. Resumen de los retos más destacados que son motivo de investigación para implementar la producción de etanol lignocelulósico a nivel industrial. Adaptado de Chen et al., 2016.</i>	17
<i>Figura 1-3. Rutas principales de bioconversión dependientes de NAD(P)H de algunos inhibidores aldehídos por <i>Saccharomyces cerevisiae</i>. Adaptado de Jayakody et al., 2013.</i>	18
<i>Figura 1-4. Rutas metabólicas propuestas para obtener etanol a partir de xilosa en cepas de <i>S. cerevisiae</i>. La ruta representada al costado izquierdo corresponde a la cepa SR8 y la otra ruta corresponde a la cepa SXA-R2P-E, para más detalle leer texto. Adaptado de Li et al., 2016.</i>	19
<i>Figura 1-5. Ruta de la xilosa, metabolismo de reducción-oxidación (en fungi) y el de isomerización (presente en fungi y bacterias). Los genes resaltados en azul corresponden a los genes putativos reportados para <i>S. cerevisiae</i> empleados en la ruta de la xilosa. Adaptado de Wenger, et al., 2010.</i>	21
<i>Figura 1-6. Perfiles de fermentación durante 48 horas en hidrolizados de bagazo de caña para el grupo de cepas nativas y la levadura industrial Red fermentis. Consumo de glucosa (línea continúa). Producción de etanol (línea punteada). Tomado de Morales et al., 2014.</i>	22
<i>Figura 1-7. Amplificación del gen XDH1 reportado por Wenger et al., 2010 para la cepa nativa 202-3. Se emplean como control positivo la cepa Lavin EC1118 y como control negativo la cepa BY474-2. Ver anexo 1.</i>	23
<i>Figura 1-8. Diagrama general de trabajo para realizar el ensamblaje y anotación de un genoma. Adaptado de Ekblom et al., 2014.</i>	26
<i>Figura 1-9. Esquema del trabajo realizado en Mudi. Adaptado de Iida et al., 2014.</i>	27
<i>Figura 2-1. Descripción general del procedimiento computacional implementado para el análisis de las lecturas obtenidas con la tecnología Hiseq Illumina 2000 con finales pareados de la cepa <i>S. cerevisiae</i> 202-3.</i>	30
<i>Figura 3-1. Calidad de las lecturas por base, antes (derecha) y después (izquierda) de la evaluación de la calidad y pre-procesamiento.</i>	34
<i>Figura 3-2. Distribución de los términos GO mapeado por nivel. Proceso biológico (verde), función molecular (azul) y componente celular (amarillo).</i>	39
<i>Figura 3-3. Tipos de proteínas identificadas con mutaciones.</i>	41

Figura 3-5. Esquema general de los genes identificados en el consumo de xilosa para la cepa 202-3, teniendo en cuenta los resultados del anexo 6. La imagen fue adaptada de Nijland et al., 2014.48

Figura 3-6. Alineamiento estructural para el gen A. G4.t1 con la subunidad de hemoglobina de Homo sapiens, con un TM-score de 0,992 y B. G3707.t1 con la metiltransferasa asociada a Aspergillus fumigata, con un TM-score de 0,910. Las proteínas modeladas para los genes desconocidos se representan con una línea más delgada en morado.49

Figura 3-7. Diagrama propuesto a partir de datos en literatura del rol de GAL4 en el metabolismo de xilosa. La propuesta se hace en base a los dominios que se encuentran en los genes G3810.t1 y G3706.t1 no reportados para la cepa de referencia S288c, que podrían estar asociados al metabolismo de xilosa al igual que GAL4. A continuación se mencionan las citas empleadas para la estructuración del diagrama: [1] Nijland et al., 2014; [2] Matsushika et al., 2014; [3] Traven et al., 2006; [4] Ren et al., 2000 y [5] Angermayr et al., 2003.....50

Lista de tablas

<i>Tabla 1-1. Desempeño de fermentación en hidrolizados de bagazo de caña para cada levadura nativa. Tomado de Morales et al., 2014.</i>	23
<i>Tabla 1-2. Evaluación del consumo de azúcares (xilosa y glucosa) para la cepa nativa 202-3 y la cepa comercial Lavin EC1118 en medios sintéticos. Ver anexo 1.</i>	23
<i>Tabla 1-3. Reportes del secuenciamiento del genoma para cepas de <i>S. cerevisiae</i> de uso industrial.</i>	24
<i>Tabla 3-1. Número de lecturas y su longitud en las etapas de pre-procesamiento</i>	33
<i>Tabla 3-2. Resumen de las métricas para los mejores ensamblajes de contigs obtenidos con diferentes ensambladores.</i>	35
<i>Tabla 3-3. Resumen de las métricas para los mejores ensamblajes de scaffolds obtenidos con diferentes ensambladores.</i>	35
<i>Tabla 3-4. Predicción de regiones codificantes con las herramientas GeneMark-ES y Augustus para el ensamblaje realizado con ABySS con el k-mer 89.</i>	36
<i>Tabla 3-5. Proteínas que no presentaron ningún resultado al realizar el Blastp con la base de datos RefSeq. Para estas proteínas se realizó nuevamente la búsqueda en la base de datos no redundante (nr) de proteínas.</i>	38
<i>Tabla 3-6. Resumen de la anotación con Blast2GO (B2G) más InterProScan (IPS).</i>	39
<i>Tabla 3-7. Mutaciones identificadas con MUDI para el genoma de la cepa <i>S. cerevisiae</i> 202-3 respecto a la cepa S288c.</i>	40
<i>Tabla 3-8. Número de mutaciones encontradas para los genes asociados al metabolismo de xilosa y otros genes de interés.</i>	42
<i>Tabla 3-9. Genes predichos con Augustus para las regiones no mapeadas con BWA para la cepa <i>S. cerevisiae</i> 202-3 y comparación con los genes obtenidos en el ensamblaje de novo con ABySS.</i>	44

Lista de Símbolos y abreviaturas

Abreviatura	Termino
GO	Ontología de genes –(gene ontology)
%GC	Porcentaje del contenido de guanina-citosina
ADN	Ácido desoxirribonucleico
ARN	Ácido ribonucleico
ARNr	Ácido ribonucleico ribosomal
ARNt	Ácido ribonucleico de transferencia
k-mer	Subcadenas de longitud k
p.b	Pares de bases nitrogenadas
Blast	Herramienta de búsqueda de alineamientos locales (Basic Alignment Search Tool, por sus siglas en inglés)
CDS	Secuencias de ADN codificantes (coding DNA sequences, por sus siglas en inglés)
KEGG	Enciclopedia de genes y genomas de Kyoto (Kyoto Encyclopedia of Genes and Genomes, por sus siglas en inglés)
E.C.	Código de la comisión enzimática (Enzyme Commission numbers, por sus siglas en inglés)
BWA	Burrows-Wheeler Aligner
G###.t1	Nomenclatura de genes obtenidos mediante la aproximación del ensamblaje <i>de novo</i> , siendo # el número del gen.
g###.t1	Nomenclatura de genes no mapeados de la aproximación del ensamblaje con genoma de referencia, siendo # el número del gen.
NAD⁺/NADH⁺	Nicotinamida adenina dinucleótido oxidado/reducido
NADP⁺/NADPH⁺	Nicotinamida adenina dinucleótido fosfato oxidado/reducido
XR	Xilosa reductasa
XD	Xilosa deshidrogenasa
XK	Xiluloquinasa
HXT	Familia de genes de transportadores de azúcares (HXT1, HXT2...)
GAL	Familia de genes de transportadores de azúcares (GAL1, GAL2...)
IPS	InterProScan

Introducción

En la actualidad existe un creciente interés por la búsqueda de alternativas al uso de combustibles fósiles, como es el caso del uso de materiales lignocelulósicos para la obtención final de bioetanol. El uso de estos materiales es un reto pues requieren de un proceso de degradación especial (pretratamiento con ácido, hidrólisis enzimática, entre otras alternativas) que promueve la formación de algunos inhibidores (hidroximetilfurfural, furfural, ácido acético, entre otros), para poder acceder finalmente a los azúcares presentes en la celulosa y hemicelulosa, correspondientes a xilosa y glucosa mayoritariamente (Ibraheem et al., 2013).

S. cerevisiae es la levadura que por excelencia se ha usado para varios procesos industriales como la obtención de cerveza, vino, entre otros productos, dada su robustez y rápida habilidad para fermentar azúcares como glucosa. Teniendo en cuenta esta habilidad y su alta resistencia a altas concentraciones de etanol y crecimiento en presencia de inhibidores, resulta ser el microorganismo que por excelencia se emplea para obtener etanol de los azúcares presentes en los hidrolizados lignocelulósicos (Jayakody et al., 2013). El mayor obstáculo que presenta el proceso de fermentación es la no utilización de xilosa por parte de *S. cerevisiae*, por lo que el desarrollo de una levadura que combine las cualidades para realizar fermentaciones a escala industrial con la habilidad de fermentar pentosas es lo ideal para hacer el proceso factible a grandes escalas (Chen et al., 2016).

La mayoría de estrategias planteadas, enmarcadas en el campo de la ingeniería metabólica proponen el uso de genes heterólogos de bacterias y hongos que tienen incorporado en su ruta metabólica la fermentación de xilosa, rutas que básicamente consisten en el sistema xilosa reductasa/xilosa deshidrogenasa y la xilosa isomerasa (Chen et al., 2016). Todos estos esfuerzos han sido considerados pues dada la diversidad genética que posee *S. cerevisiae*, se han encontrado algunas cepas que pueden consumir xilosa como única fuente de carbono (Wenger et al., 2010). La presencia de estos genes endógenos abre una puerta para evitar el uso de genes heterólogos para la fermentación de xilosa.

Por este motivo se evaluó el consumo de xilosa y glucosa en hidrolizados de bagazo de caña por un grupo de cepas aisladas previamente y caracterizadas como *S. cerevisiae*, estos estudios mostraron un consumo de xilosa por parte de la levadura *S. cerevisiae* 202-3. Por lo anteriormente mencionado, con el fin de caracterizar la base molecular que justifique dicho consumo frente a cepas que no lo hacen (como la cepa de referencia S288c), se planteó la secuenciación del genoma de la cepa 202-3, para evaluar posibles genes candidatos y futuras modificaciones genéticas que mejoren el consumo y posterior fermentación del azúcar xilosa.

Objetivos

Objetivo general:

El objetivo principal del presente trabajo fue caracterizar el genoma de la cepa *S. cerevisiae* 202-3, con el fin de identificar posibles indicios de la capacidad de consumo de xilosa durante la producción de etanol de segunda generación.

Objetivos específicos:

- 1) Establecer las características genómicas de la cepa *S. cerevisiae* 202-3, mediante los procedimientos de ensamblaje y anotación.
- 2) Inferir, a partir del genoma caracterizado y del análisis comparativo con cepas *S. cerevisiae* con capacidad o no de consumir xilosa, los genes candidatos asociados a este proceso.

1. Estado del arte

1.1 Producción de etanol de segunda generación

La rápida industrialización ha resultado en una creciente demanda por combustibles fósiles. Esta alta demanda ha provocado un incremento continuo en el precio de los combustibles, una alta contribución a las emisiones de gas y al calentamiento global. Por lo que surge la necesidad de desarrollar alternativas ambientalmente sostenibles para satisfacer esta necesidad creciente. Una de las opciones amigables con el medio ambiente es el bioetanol, dada su baja emisión neta de gases de efecto invernadero (CO₂ y SO₂) en comparación con la gasolina, diesel o queroseno. Con esta alternativa en marcha las leyes que protegen el medio ambiente en muchos países, exigen el uso de bioetanol como un aditivo en los combustibles, hecho que incrementa las necesidades en la producción de bioetanol (Ibraheem et al., 2013; Chen et al., 2016).

En la actualidad, la mayor producción de bioetanol proviene de jugos de caña de azúcar y maíz (primera generación). Sin embargo, el uso de estas materias primas genera controversia pues son parte fundamental de la cadena alimenticia de la humanidad. El resultado es un producto inconveniente para la opinión pública y de costos de producción muy elevados. En Colombia la producción de etanol es de 1.200.000 litros por día en seis refinерías, basado en mayor medida en la producción del valle del río Cauca con caña de azúcar, con lo cual es posible reemplazar 8,5% de las gasolinas que se consumen en el país. Para el año 2020, la meta es alcanzar mezclas de etanol-gasolina del 20% (Fedebioenergías, 2007). Por este motivo explorar alternativas, como el etanol de segunda generación, producido a partir de fuentes mucho más económicas, como la lignocelulosa, promete ser una alternativa más viable y sostenible (Ibraheem et al., 2013; Chen et al., 2016).

La producción de etanol de segunda generación involucra una serie de subprocesos, donde se incluye la etapa de pretratamiento o extracción de los sustratos celulósicos del material lignocelulósico crudo, como el bagazo de caña. En seguida la celulosa se somete a un proceso de hidrólisis donde se obtienen azúcares fermentables en forma de pentosas y hexosas. Los azúcares obtenidos son fermentados por microorganismos etanologénicos, incluidos bacterias, levaduras y fungi, en la figura 1-1 se presenta un resumen de los sustratos obtenidos durante el proceso, incluyendo algunos inhibidores (Ibraheem et al., 2013).

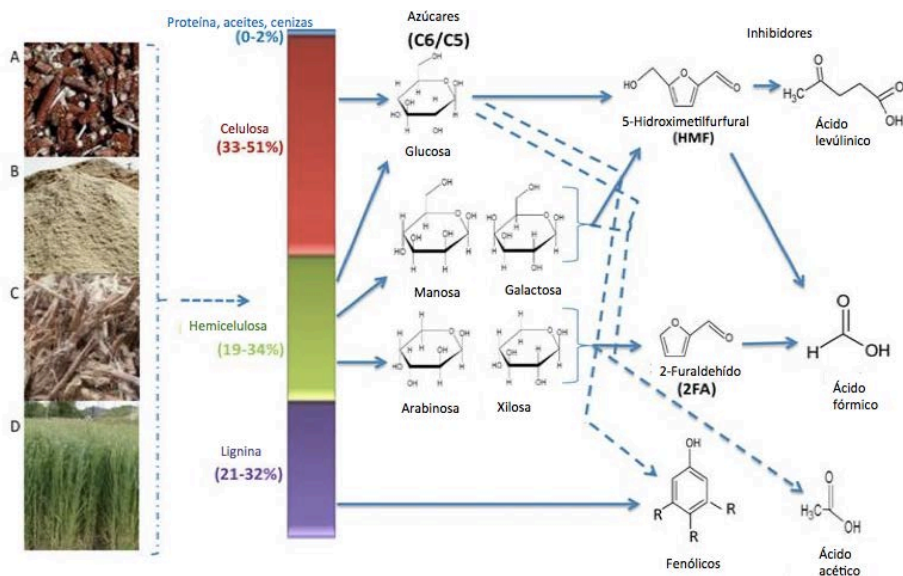


Figura 1-1. Productos obtenidos a partir de materias lignocelulósicas. En este esquema se incluyen los azúcares fermentables del proceso, como glucosa, xilosa, arabinosa, manosa y galactosa. Así como algunos inhibidores formados durante la etapa de pretratamiento e hidrólisis, como el 2-furaldehído, 5-hidroximetilfurfural, ácido levúlnico, ácido fórmico, compuestos fenólicos y ácido acético. Adaptado de Ibraheem et al., 2013.

1.1.2. Retos del proceso para obtener etanol de segunda generación

Actualmente la tecnología para producir etanol celulósico, ha pasado por varias etapas de optimización. Sin embargo estos cambios aún no hacen sostenible el proceso en términos de costos y competitividad de producción (Chen et al., 2016). En la figura 1-2, se presentan las dificultades actuales que aún tiene el proceso, allí se destacan las etapas de proceso principales: pretratamiento, hidrólisis enzimática y fermentación. Durante el pretratamiento sobresalen como obstáculos la heterogeneidad de los materiales lignocelulósicos, la producción de inhibidores (listados en la figura 1-1) durante la etapa de pretratamiento, para el proceso de hidrólisis y fermentación se evidencia en común una problemática sea la opción de sacarificación y fermentación en simultáneo (SSF por sus siglas en inglés) y proceso de hidrólisis y fermentación por separado (SHF por sus siglas en inglés), la fermentación eficiente de xilosa.

Los inhibidores, que se mencionan en las figuras 1-1 y 1-2, con grupos aldehído reducen significativamente la capacidad de fermentación para la mayoría de microorganismos, incluida *S. cerevisiae*. Incluso a bajas concentraciones, estos compuestos pueden llegar a inhibir el crecimiento, reducir el rendimiento final en la producción de etanol, reducir la actividad enzimática, provocar daños en membrana y en la pared celular, inhibición de síntesis de proteínas y daño de ADN (Jayakody et al., 2013). Para solucionar este obstáculo se ha identificado que uno de los mejores métodos para detoxificar el medio, es el uso de levaduras mediante el aprovechamiento de la capacidad enzimática que tiene *S. cerevisiae* para disminuir la toxicidad del medio, identificando los genes que están

envueltos en este proceso acoplados con NADH y NADPH, figura 1-3 (Jayakody et al., 2013).

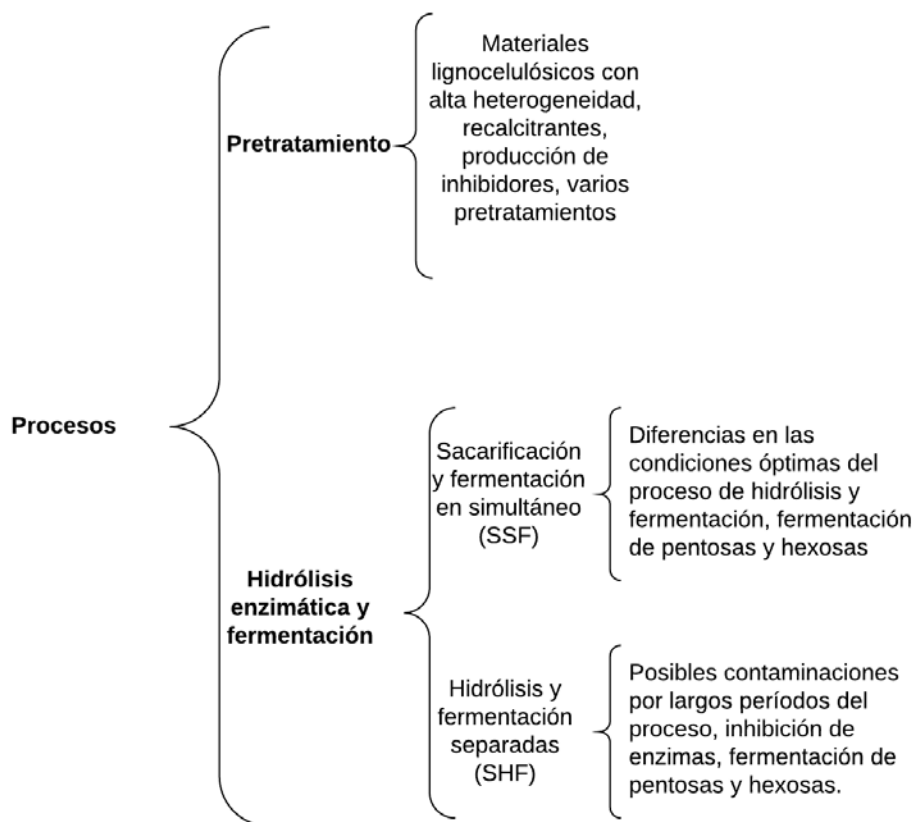


Figura 1-2. Resumen de los retos más destacados que son motivo de investigación para implementar la producción de etanol lignocelulósico a nivel industrial. Adaptado de Chen et al., 2016.

Aunque el mecanismo propuesto en la figura 1-3, corresponde a *S. cerevisiae*, es importante resaltar que esta capacidad de detoxificar es dependiente de la cepa, por eso algunos estudios se concentran en encontrar cepas con alta resistencia (Sánchez et al., 2012), en otros casos los estudios se concentran en realizar mejoramiento genético mediante la sobreexpresión de las enzimas dependientes de NAD(P)H, codificadas por los genes representados en la figura 1-3 o incrementando el flujo del ciclo de las pentosas fosfato (Passoth et al., 2014).

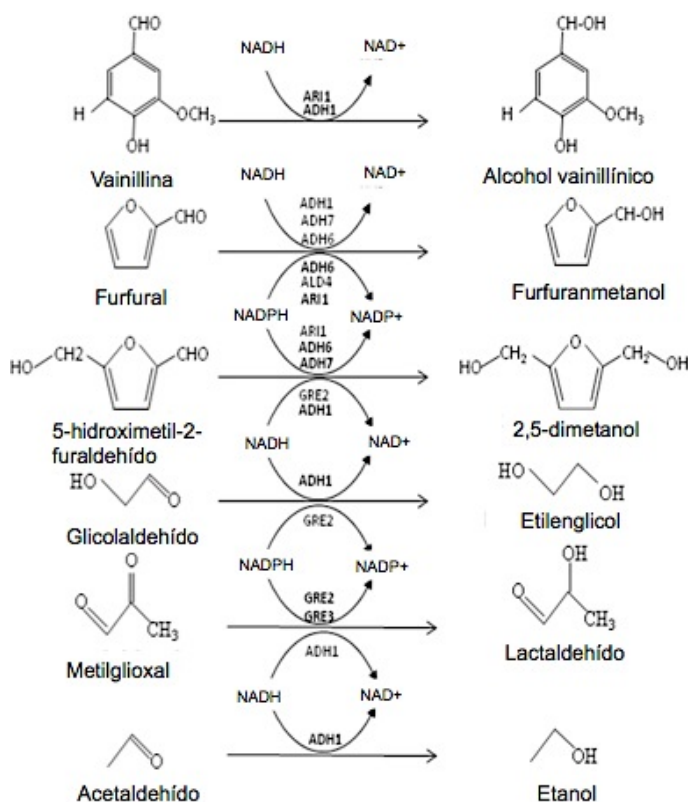


Figura 1-3. Rutas principales de bioconversión dependientes de NAD(P)H de algunos inhibidores aldehídos por *Saccharomyces cerevisiae*. Adaptado de Jayakody et al., 2013.

Para el caso de la hidrólisis y la fermentación, las dos estrategias propuestas, SSF y SHF, presentan la misma dificultad relacionada con el aprovechamiento de las hexosas y pentosas de los sustratos, como se observa en la figura 1-2.

El uso de las pentosas en especial la xilosa, resulta ser un tema de gran importancia pues muchos microorganismos pueden tener diferentes capacidades en el uso de azúcares, para el crecimiento, pero en general la mayoría prefiere usar los azúcares de 6 carbonos como la glucosa, la cual pueden convertir en piruvato y finalmente oxidarla hasta dióxido de carbono y agua con la producción de ATP. Gracias a este proceso tan eficiente muchas células suprimen el metabolismo de otros azúcares en presencia de glucosa (Harcus et al., 2013, Chen R et al., 2016).

Ya ha sido establecida la expresión heteróloga de rutas metabólicas de xilosa en microorganismos etalógicos como *S. cerevisiae* y *Zymomonas mobilis* que no usan naturalmente xilosa u otros azúcares de 5 carbonos. Realizar estas modificaciones es una tarea que todavía está en desarrollo, pues sigue siendo menos eficiente frente al uso de glucosa (Chen R et al., 2016). Para el caso de bacterias las restricciones en el catabolismo de xilosa se presentan por causa a la toxicidad del xilitol y la competencia por los transportadores de glucosa, por lo que a pesar de la búsqueda de microorganismos capaces de utilizar eficientemente xilosa de manera natural aún sigue predominando el uso de *S. cerevisiae*, dada su robustez y su potencial como hospedero para biología sintética para modificaciones de *de novo* e ingeniería de rutas metabólicas complejas (Kavšček et al., 2015, Apel et al. 2016). Adicionalmente, para optimizar el proceso en bacterias se requiere de la expresión de al menos 4 enzimas heterólogas lo cual requiere un trabajo de múltiples pasos, aumentando la tasa de error en las modificaciones

genéticas (Chen R et al., 2016).

En levaduras y en otras cepas de hongos, el xilitol es un intermediario de la ruta de la xilosa deshidrogenasa por lo que el xilitol parece no tener un impacto tóxico (Chen R et al., 2016). Sin embargo se presentan desequilibrios redox debido a los requerimientos del cofactor NADPH para la xilosa deshidrogenasa, por lo que los casos que se consideran de mayor éxito en levaduras han sido los que emplean la expresión de la ruta metabólica de las isomerasa de bacterias en combinación o no de la ruta de la xilosa deshidrogenasa. Los casos reportados corresponden a cepas de *S. cerevisiae*, cepas que han sido sometidas a adaptación o expresión de NADH oxidasa (Chen R et al., 2016), en la figura 1-4 se presentan las dos estrategias que principalmente se emplean para *S. cerevisiae*.

En estudios recientes se realiza una comparación de las transformaciones de las rutas metabólicas en *S. cerevisiae* que más exitosas han resultado para la fermentación de xilosa (cepas disponibles para uso público) (Li et al., 2016). Estas modificaciones corresponden a las dos estrategias ilustradas en la figura 1-4 para las cepas SR8 y SXA-R2P-E. La ruta oxido-reductasa, expresada para la cepa SR8, consiste en la expresión de dos o más copias de los genes XYL1 (XR), XYL2 (XDH) y XYL3 (XK) de la levadura *Scheffersomyces stipitis* (levadura capaz de fermentar xilosa), y la delección del gen acetaldehído deshidrogenasa, ALD6. La segunda ruta propuesta, corresponde a la xilosa isomerasa, expresada en la cepa SXA-R2P-E, con dos copias de los genes XYLA3 de *Piromyces sp.*, el gen TAL1 de *S. stipitis*, sobre-expresión endógena de XKS1 (XK) y GRE3 y la delección de PHO13. El estudio concluye que la expresión de la ruta oxido-reductasa para la cepa SR8 resulta tener mejores desempeños frente al consumo de xilosa y producción de etanol (Li et al., 2016, Sánchez et al., 2016).

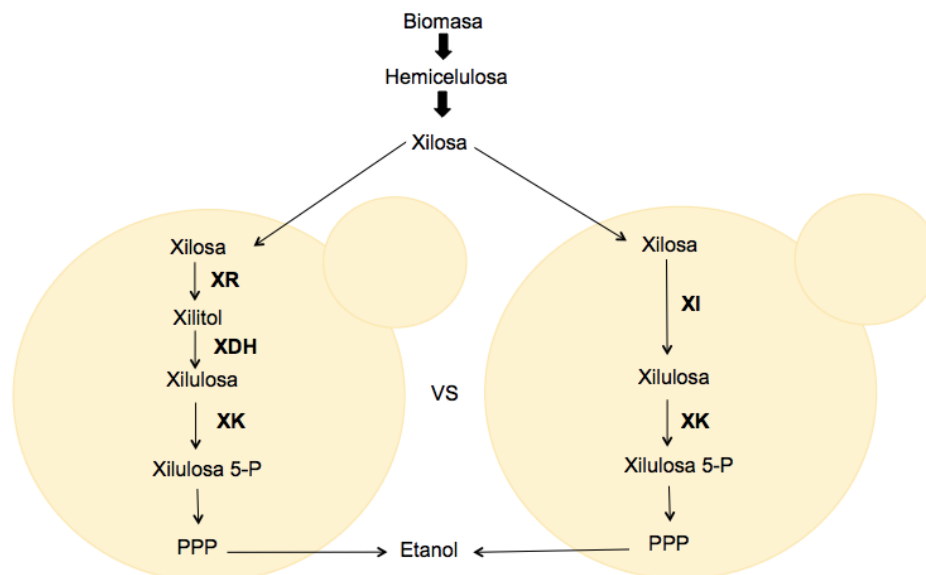


Figura 1-4. Rutas metabólicas propuestas para obtener etanol a partir de xilosa en cepas de *S. cerevisiae*. La ruta representada al costado izquierdo corresponde a la cepa SR8 y la otra ruta corresponde a la cepa SXA-R2P-E, para más detalle leer texto. Adaptado de Li et al., 2016.

Siendo *S. cerevisiae* un microorganismo tan estudiado, la disponibilidad de herramientas moleculares para realizar cambios es considerable, pero la modificación genética que se

ha realizado en su mayoría corresponde a cepas de laboratorio como es el caso de la cepa S228c y la serie de cepas CEN.PK y no a cepas industriales. Aunque estas cepas pueden ser modificadas fácilmente, para el caso de consumo de pentosas, su nivel de tolerancia a inhibidores es inferior frente a las cepas industriales, su productividad es más baja y la producción de otros metabolitos como glicerol es mayor (Oreb et al., 2012). Otra de las ventajas al usar cepas industriales es que suelen ser estables, a pesar de ser sometidas a secado y largos periodos de almacenamiento. Las dificultades que se presentan a la hora de modificar cepas industriales se debe al nivel de ploidia y al hecho de no conocer el contexto genómico de cada cepa (Oreb et al., 2012).

En escalas grandes de fermentación el uso de plásmidos no se considera apropiado, pues su mantenimiento depende de los marcadores de selección (Oreb et al., 2012). Por lo que las modificaciones genéticas realizadas deben ser planeadas para que tengan una integración estable en el genoma (Oreb et al., 2012). Como consecuencia la inserción de genes o delección debe ser realizada en todos los alelos para obtener un genotipo estable (Oreb et al., 2012). Así mismo el número de pasos empleados debe ser pequeño para disminuir el número de mutaciones negativas (Oreb et al., 2012). Por este motivo superar este obstáculo del aprovechamiento de pentosas sigue siendo objeto de estudio, desde el desarrollo de nuevas estrategias en la modificación de genes como en la forma en que se realiza la transformación como tal.

Recientes estudios optan por la estrategia de mejorar los transportadores transmembranales de azúcares, aumentando la afinidad por la xilosa, pues los transportadores reportados hasta el momento tienen una capacidad muy baja (Apel et al., 2016, Wang et al., 2016). El uso de nuevas herramientas de modificación genética como el sistema CRISPR/Cas9, permitirían lograr un acercamiento más certero para el mejoramiento de cepas industriales, como lo revelan estudios recientes (Stovicek et al., 2015, Sung et al., 2015, Jakočiūnas et al., 2016).

1.1.3. Ruta metabólica para la degradación de xilosa en *S. cerevisiae*

Como bien se mencionó anteriormente aún es un reto industrial lograr que cepas de *S. cerevisiae* fermenten xilosa de manera eficiente. Aunque varios autores reportan como negativo el consumo de xilosa en *S. cerevisiae* sin modificación genética (Wang et al., 2016), algunos estudios demuestran que aislados naturales de *S. cerevisiae* pueden consumir xilosa sin modificación genética alguna (Attfield et al., 2006; Wenger et al., 2010; Nogueira, 2016). Explotar esta capacidad endógena, resulta de gran importancia teniendo en cuenta que el número de modificaciones genéticas puede disminuir, siendo esto de gran interés para las cepas de uso industrial. Algunos autores aprovechando esta capacidad endógena que tiene *S. cerevisiae* de degradar pentosas, han demostrado que el isómero de la xilosa, xilulosa, puede ser fermentado con mayor facilidad por cepas sin modificación genética (Eliasson et al., 2000; Tamari et al., 2014).

La ruta metabólica que ha sido reportada en *S. cerevisiae* para la degradación de xilosa se encuentra representada en la figura 1-5. Teniendo en cuenta estos genes endógenos que posee *S. cerevisiae*, se han realizado algunas pruebas donde se demuestra que la mayor contribución para la actividad xilosa reductasa es atribuible a GRE3, YPR1 y YJR096W, también se ha demostrado que la sobreexpresión de los genes GRE3 y XYL2,

es suficiente para provocar el fenotipo positivo en el consumo de xilosa (Traff et al., 2002; Toivari et al., 2004.).

Dentro de los estudios más recientes se ha reportado un nuevo gen en *S. cerevisiae* denominado XDH1 (con actividad xilosa deshidrogenasa), el cual no está presente en todas la cepas, como es el caso de la levadura S288c. La presencia de este gen es suficiente para considerar positivo o no el consumo natural de xilosa en *Saccharomyces* (Wenger et al., 2010).

Algunos autores sugieren como paso limitante el transporte de xilosa a a célula (Kahar et al., 2011), pues como previamente se ha reportado *S. cerevisiae* presenta el potencial para degradar xilosa con los genes mencionados en la figura 1-5. Por esta razón en la actualidad se evidencia un creciente interés en mejorar los transportadores de azúcares en *S. cerevisiae*, por lo que se toma en consideración la familia de genes HXT y GAL (Kahar et al., 2011; Young et al., 2014; Apel et al., 2016).



Figura 1-5. Ruta de la xilosa, metabolismo de reducción-oxidación (en fungi) y el de isomerización (presente en fungi y bacterias). Los genes resaltados corresponden a los genes putativos reportados para *S. cerevisiae* empleados en la ruta de la xilosa. *Gen asociado a levaduras de *S. cerevisiae* con fenotipo positivo en el consumo de xilosa. Adaptado de Wenger, et al., 2010.

1.1.4. Antecedentes del grupo de investigación de procesos químicos y bioquímicos en la producción de etanol de segunda generación con aislados naturales de *S. cerevisiae*

El Grupo de Investigación en Procesos Químicos y Bioquímicos de la Facultad de Ingeniería de la Universidad Nacional de Colombia, desde años anteriores ha trabajado y evaluado los perfiles de fermentación en mieles de caña de azúcar para un grupo de aislados naturales de levaduras, enfocando su esfuerzo a la especie *S. cerevisiae*. Recientemente con el creciente interés y futuro en la producción de etanol de segunda generación, estos aislados de levaduras nativas de *S. cerevisiae* fueron empleados para evaluar el potencial de estas cepas para la producción de etanol de segunda generación,

teniendo en cuenta los retos con los que todavía cuenta el proceso a nivel mundial.

Los perfiles de fermentación fueron evaluados en hidrolizados de bagazo de caña y contrastados con el perfil de fermentación de la cepa industrial *Ethanol Red fermentis*, dado su potencial como cepa para producción de etanol de segunda generación (Oreb et al., 2012). Los perfiles de fermentación evaluados se encuentran en la figura 1-6, de este grupo de cepas resaltó el desempeño de la levadura 202-3. En la tabla 1-1, se presenta en detalle el consumo de azúcares encontrado para el grupo de cepas nativas evaluado. Como se observa, para la cepa 202-3 a pesar de ser la segunda cepa con mejor rendimiento, es la cepa que presenta un mayor consumo de xilosa frente a las demás cepas, por lo que fue seleccionada como la mejor cepa candidata para la producción de etanol de segunda generación.

Con el fin de evaluar el consumo de xilosa por parte de la cepa 202-3 y contribuir al estudio de la variación natural dentro de las levaduras *Saccharomyces* respecto al uso de xilosa (Wenger, et al., 2010), se estudió el consumo de xilosa como única fuente de carbono empleando como control positivo la cepa Lavin EC1118, caracterizada por poseer el gen XDH1 que le da la capacidad de ser una cepa xilosa positiva frente a otras cepas de *S. cerevisiae* (Wenger et al., 2010). Los perfiles encontrados de consumo de xilosa como única fuente de carbono y en presencia de glucosa se encuentran en la tabla 1-2. Allí se comprueba que respecto a la cepa Lavin EC1118, la cepa 202-3 tiene un consumo considerable de xilosa y un mejor perfil de consumo de glucosa.

Como primera aproximación para justificar el consumo de xilosa de la cepa 202-3, se amplificó el gen XDH1, presente en la cepa Lavin EC1118. Como se observa en la figura 1-7, el gen no se encuentra en la cepa 202-3. Por este motivo el presente trabajo tiene como finalidad emplear como primera aproximación la caracterización del genoma de la cepa 202-3, mediante el secuenciamiento del mismo con el fin de identificar los genes involucrados en el consumo de xilosa y establecer posibles diferencias con cepas reportadas en literatura de la misma especie para explotar el potencial de esta cepa en la producción de etanol de segunda generación.

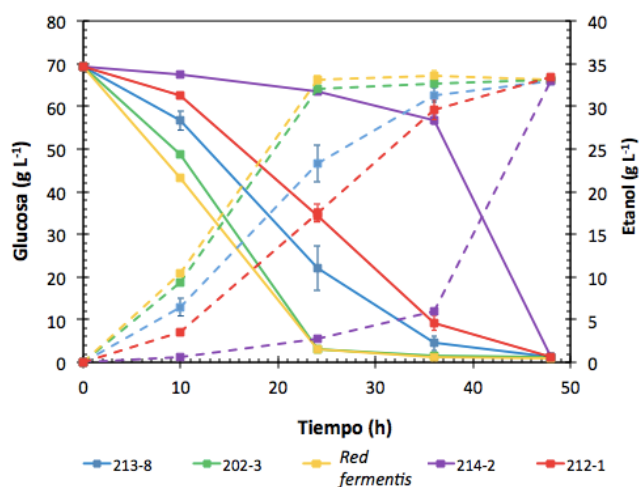


Figura 1-6. Perfiles de fermentación durante 48 horas en hidrolizados de bagazo de caña para el grupo de cepas nativas y la levadura industrial *Red fermentis*. Consumo de glucosa (línea continua). Producción de etanol (línea punteada). Tomado de Morales et al., 2014.

Tabla 1-1. Desempeño de fermentación en hidrolizados de bagazo de caña para cada levadura nativa. Tomado de Morales et al., 2014.

Cepa	Etanol (g/L)	Rendimiento etanol (g _{EtOH} /g _{Glu})	Consumo de glucosa (%)	Consumo de xilosa (%)	Tiempo de producción máxima de etanol (h)	Productividad (gEtOH/L.h)
213-8	32.908	0.474	98.176	4.923	28	1.175
202-3	32.093	0.477	98.271	6.470	24	1.337
214-2	32.693	0.475	97.995	3.215	40	0.817
212-1	33.352	0.481	98.139	3.172	28	1.191

Tabla 1-2. Evaluación del consumo de azúcares (xilosa y glucosa) para la cepa nativa 202-3 y la cepa comercial Lavin EC1118 en medios sintéticos. Ver anexo 1.

	Xilosa como unica fuente de carbono		Xilosa y glucosa como fuentes de carbono			
	%Consumo de xilosa		%Consumo de xilosa		%Consumo de glucosa	
Tiempo (h)	202-3	EC1118	202-3	EC1118	202-3	EC1118
24	0.968	2.746	1.477	3.842	59.284	49.919
72	7.493	6.110	6.051	6.757	100.000	96.981
120	8.962	11.457	10.587	10.803	100.000	100.000

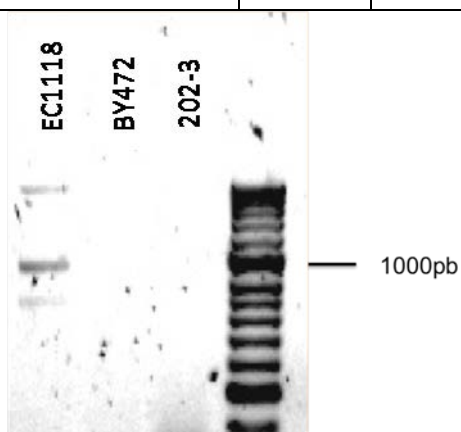


Figura 1-7. Amplificación del gen XDH1 (tamaño aproximado 1000 pb) reportado por Wenger et al., 2010 para la cepa nativa 202-3. Se emplean como control positivo la cepa Lavin EC1118 y como control negativo la cepa BY474-2. Ver anexo 1.

1.2 Métodos analíticos en genómica

1.2.1 Secuenciación de cepas de *Saccharomyces cerevisiae* y su aplicación

El genoma de *S. cerevisiae* se ha convertido en un importante modelo para el análisis comparativo, por este motivo en los últimos años para caracterizar cepas de uso industrial la secuenciación del genoma ha sido de gran interés (Treu et al., 2014), en la tabla 1-3 se mencionan algunos ejemplos. Los estudios realizados han demostrado que la tecnología humana y las condiciones hostiles (alta concentración de azúcar, altos niveles de etanol, bajo pH) han influenciado la evolución del género *Saccharomyces*, proporcionando una distinción entre las cepas de uso industrial (Akao et al., 2011; Treu et al., 2014). Aunque *S. cerevisiae* sea uno de los microorganismos más estudiados, aún hace falta información genómica en relación con el gran número de cepas que se han adaptado a diferentes ambientes (Treu et al., 2014; Gash et al., 2016). Mediante la comparación de los genomas de cepas de *S. cerevisiae*, algunos autores han identificado genes específicos de cada cepa (Wenger et al., 2010; Borneman et al., 2011; Treu et al., 2014), por lo que siempre es importante considerar que la información que ha sido producida no debe ser extrapolada a otras cepas de levadura dada su diversidad genómica y fenotípica (Zheng et al., 2012).

Tabla 1-3. Reportes del secuenciamiento del genoma para cepas de *S. cerevisiae* de uso industrial.

Nombre de la cepa	Número de ORFs	Aplicación	Referencia
AWRI1631	5687	Vino	Borneman et al., 2008..
Kyokai no. 7	5815	Sake	Akao et al., 2011.
CAT-1	6652	Bioetanol	Babrzadeh, 2012.
YJS329	5602	Bioetanol	Zheng et al., 2012.
R103	6364	Vino	Treu et al., 2014.
P283	6325	Vino	Treu et al., 2014.
IR-2	5628	Bioetanol	Sahara et al., 2014 (a).
NAM34-4C	5696	Bioetanol	Sahara et al., 2014 (b).
NCIM3186	5347	Bioetanol	Sravanthi et al., 2015.
S288c	6604	Referencia	SGD, 2016.

Una de las aplicaciones de los datos de secuenciación genómica, consiste en el uso de esta información para la elaboración de modelos a escala genómica, con el fin de generar hipótesis en el diseño de rutas metabólicas (Sánchez et al., 2015; Blais et al., 2013; Otero et al., 2010). Por lo que el entendimiento a nivel genómico para correlacionar genotipo a fenotipo ha aumentado, en el caso de *S. cerevisiae* uno de los primeros modelos propuestos fue el iFF708, para luego evolucionar en modelos como el iND750, iLL672, iIND750, iMM904, entre otros. Esta disponibilidad de modelos refleja la importancia de conocer más datos biológicos con el fin de generar un modelo que se acerque mucho más a la realidad (Sánchez et al., 2015).

1.2.2 Estrategias de ensamblaje para lecturas cortas

La investigación biológica ha cambiado rápidamente desde que surgieron las tecnologías de secuenciación, conocidas como secuenciación de nueva generación (NGS, por sus siglas en inglés). Estos secuenciadores producen lecturas de alta calidad de cortos tamaños a un precio moderado, acelerando así los campos de investigación en genómica, transcriptómica, metagenómica, proteómica, entre otros (Metwally et al., 2013). Con todos estos datos surge el gran inconveniente del ensamblaje de estas cortas lecturas, pues es imposible hasta la fecha secuenciar todo el genoma directamente en una sola lectura. El método de secuenciación aleatoria rompe todo el genoma al azar y secuencia cada fragmento independientemente. El proceso de reconstrucción de todo el genoma mediante la unión de las lecturas hasta el nivel de cromosoma, se conoce como ensamblaje del genoma (Metwally et al., 2013).

El primer paso para realizar el ensamblaje del genoma es tener un diagrama de trabajo, como se ilustra en la figura 1-8. El diagrama de trabajo, incluye las siguientes etapas fundamentales: la etapa de laboratorio donde se obtiene el ADN; la etapa de preprocesamiento (remoción de adaptadores, remoción de lecturas con baja calidad); ensamblaje *de novo* (o como más adelante se menciona, con genoma de referencia) y finalmente el proceso de anotación.

Antes de realizar la limpieza de las lecturas, se procede a realizar un análisis de calidad, llevado a cabo comúnmente por herramientas como FastQC (Andrews et al., 2015), donde se obtiene un reporte de diferentes parámetros, tales como contenido de homopolímeros de ADN, calidad de las secuencias (en el puntaje de la escala de Phred), contenido de guanina-citosina (%GC), número de N por cada lectura, entre otros. Una vez evaluados estos parámetros, se lleva a cabo la limpieza de las lecturas. Con el fin de asegurar la eliminación de los adaptadores empleados durante la preparación de las librerías se utilizan diferentes herramientas como Cutadapt (Martin et al., 2011) y como uso adicional algunos programas ofrecen la opción de remover otras lecturas según otros criterios de calidad, como es el caso de Prinseq (Schmieder et al., 2011), que también puede llevar a cabo la remoción de adaptadores.

Con las lecturas preprocesadas se realiza el ensamblaje, procedimiento mediante el cual se busca unir las lecturas obtenidas en el genoma sometido a secuenciación, según sea la disponibilidad de datos, el ensamblaje se puede realizar de dos maneras: **ensamblaje *de novo*** y con **genoma de referencia (resecuenciación)**. Cuando se habla de ensamblaje *de novo* se hace referencia al proceso mediante el cual se unen las lecturas para formar secuencias contiguas (denominadas como contigs) que comparten las mismas secuencias nucleotídicas según el ADN plantilla de donde derivaron las secuencias (Paszkiwicz et al., 2010); mientras que en el ensamblaje con genoma de referencia por lo general se habla de una resecuenciación, esto hace referencia a la disponibilidad previa de un genoma secuenciado de la misma especie, esta metodología suele implementarse con el fin de detectar mutaciones (SNPs) u otras variaciones genéticas entre individuos de la misma especie (Paszkiwicz et al., 2010).

En el caso del **ensamblaje de novo** cuando ya se obtienen los *contigs*, estos deben ser unidos en secuencias mucho más largas, conocidos como *scaffolds*, finalmente los scaffolds pueden ser unidos para formar los cromosomas, sin embargo este último paso requiere información detallada del mapa genético, por lo que no suele reportarse hasta este punto en la mayoría de los casos para los análisis realizados (Ekblom et al., 2014).

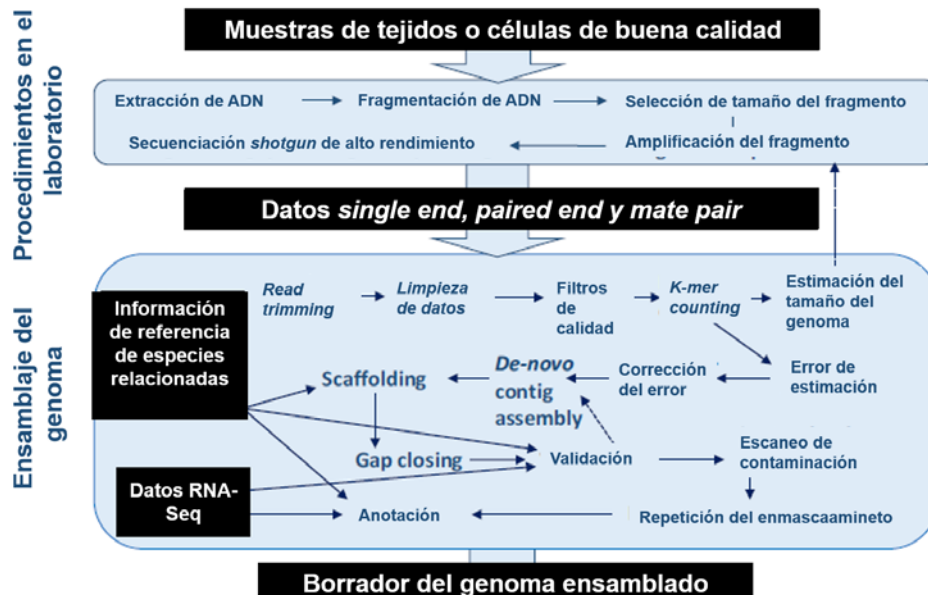


Figura 1-8. Diagrama general de trabajo para realizar el ensamblaje y anotación de un genoma. Adaptado de Ekblom et al., 2014.

Las herramientas que se emplean para realizar el ensamblaje del genoma *de novo* varían en términos de velocidad, escalabilidad y la calidad del ensamblaje. Aunque algunos métodos de ensamblaje superan claramente otros, aún es difícil predecir cual debe ser la herramienta adecuada (Ekblom et al., 2014).

La estrategia empleada por la mayoría de ensambladores *de novo* para secuencias cortas se divide en dos clases principalmente: métodos basados en extensión y los grafos De Bruijn (o Eulerian) (Ekblom et al., 2014). Los ensambladores más usados basados en extensión son SSAKE y JR-Assembler, son computacionalmente eficientes, pero son muy sensibles a los errores de secuenciación, regiones repetitivas y altos niveles de polimorfismo (Ekblom et al., 2014). Por otro lado los ensambladores basados en los grafos De Bruijn, suelen ser los más usados para el ensamblaje de lecturas cortas, donde las lecturas son particionadas en k-mers (subconjunto de las lecturas de longitud k) para luego formar los nodos del grafo y son unidas cuando comparten un k-1 mer; ejemplos de algunos softwares son SOAPdenovo, ABySS, Velvet, SPAdes, entre otros (Ekblom et al., 2014).

Para la identificación de mutaciones y variaciones genómicas, el esquema general que se debe seguir es: mapear las secuencias de las lecturas a un genoma de referencia, detectar los alelos mutados basados en comparación a la secuencia de referencia y filtración de los alelos mutados. La anotación de estas mutaciones puede ser de utilidad pues nos brinda una idea de la función o efecto de la mutación (Iida et al., 2014). En este

trabajo la búsqueda de mutaciones se realiza con la herramienta Mudi (Mutation discovery) mediante el flujo de trabajo que se observa en la figura 1-9.

En el caso de Mudi, BWA es la herramienta que realiza el alineamiento, sin embargo existen otros alineadores como MAQ, BOAT, PerM, SOAPv2, Bowtie, entre otros (Zhang et al., 2014). En general los algoritmos que se emplean para los alineadores son dos los basados en tabla hash y los que usan la transformada de Burrows-Wheeler (Zhang et al., 2014).

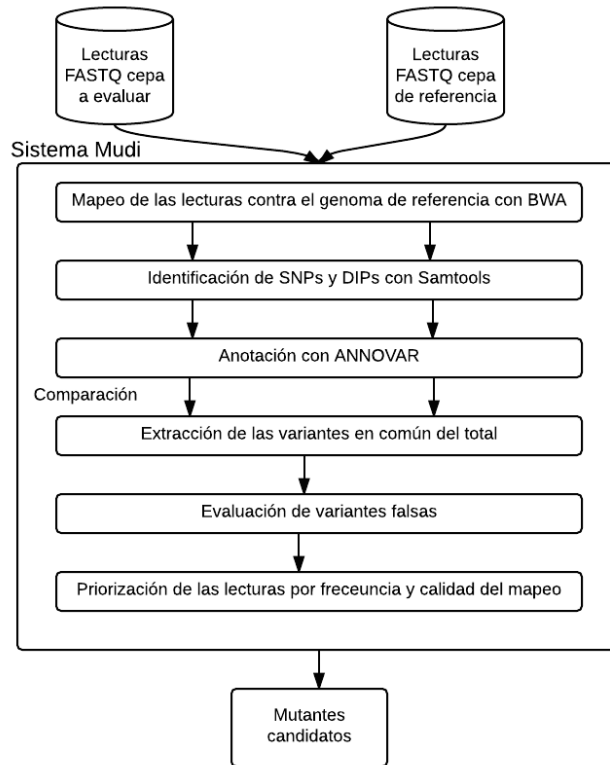


Figura 1-9. Esquema del trabajo realizado en Mudi. Adaptado de Iida et al., 2014.

1.2.3 Anotación genómica

Para explotar todo el potencial de las secuencias del genoma, surge la necesidad de anotarlo con información biológica que puede incluir modelos de genes e información funcional, como términos del “Gene Ontology Consortium” (GO) o de rutas metabólicas del “Kyoto encyclopedia of genes and genomes” (KEGG), información de modificaciones epigenéticas (proyecto ENCODE) y microRNA, enriquecen el ensamblaje obtenido (Ekblom et al., 2014).

El primer paso para realizar la anotación, o caracterización funcional, de un genoma es identificar las regiones del genoma que posiblemente codifican para distintos tipos de ARN. Este proceso se lleva a cabo mediante el uso de algoritmos específicos, que en términos generales son llamados “gene callers”. Este proceso se lleva a cabo de manera distinta en genomas eucariotas y en genomas procariontes. En el caso de genomas eucariotas, uno de los grandes inconvenientes es la presencia de intrones, por lo que predecir la estructura de un gen puede llegar a ser complejo (Stanke et al., 2005). Una de

las herramientas ampliamente validada y usada por la comunidad científica para llevar a cabo estas predicciones en genes eucariotas es Augustus y Genemark-ES. Estos predictores *ab initio*, usan estructuras génicas (motivos, secuencias cortas, codones de inicio, etc.) con un entrenamiento automático, en este caso, para eucariotas (Lomsadze et al., 2005), aproximándose al 100% de éxito dado su alto nivel de sensibilidad (Yandell et al., 2012). Estas herramientas se basan en modelos ocultos de Markov, un método estadístico en el que se define una distribución de probabilidad para varias secciones de la secuencia del genoma. Intrones, exones, regiones intergénicas, etc., corresponden a estados en el modelo y cada estado crea una secuencia de ADN con ciertas probabilidades de emisión pre-definida (Stanke et al., 2005).

Finalmente, una vez se ha logrado una identificación de las regiones génicas en el genoma, es necesario caracterizarlos funcionalmente, es decir que es necesario describir los posibles roles funcionales que estos genes tienen. Una de las aproximaciones comúnmente usada para este propósito es la de la asignación funcional mediante la comparación de la secuencia de estos genes con bases de datos de genes conocidos, y los cuales se encuentran clasificados de acuerdo a un lenguaje estandarizado conocido como Ontología. Con este fin el proyecto GO (*gene ontology*, por sus siglas en inglés) ha desarrollado anotaciones estandarizadas para integrar descriptores para la clasificación de las secuencias y sus atributos funcionales (Ashburner et al., 2000; Blake et al., 2013). Las ontologías para cualquier producto génico presentan relaciones con otros términos más generales y/o específicos y han sido agrupadas en tres grandes términos:

1. Función molecular
2. Proceso biológico
3. Componente celular.

La categoría **función molecular** se asocia con la actividad bioquímica del producto génico. El **proceso biológico** describe el objetivo de un producto génico. Por otro lado, el término **componente celular** incluye términos que describen el lugar donde el producto génico lleva a cabo su función (Gaudet et al., 2009). Una de las herramientas que permite la asociación de las secuencias génicas con términos GO es Blast2GO mediante búsquedas de similitud a través del uso de Blast e InterProScan, finalmente el programa permite asociar estos términos con términos EC (Enzymatic Commission numbers, por sus siglas en inglés) y mapearlos en la base de datos de KEGG (Conesa et al., 2008; Conesa et al., 2005).

2. Materiales y métodos

2.1 Crecimiento y mantenimiento de la cepa de *S. cerevisiae*

La levadura 202-3 corresponde al cepario aislado y caracterizado molecularmente del proyecto “Aislamiento, caracterización y evaluación de cepas nativas del Municipio de Puerto López (Meta) para la producción de etanol”.

Para su respectivo uso las levaduras fueron activadas y almacenadas en medio YPG (3g/L extracto de levadura, 20 g/L glucosa, 3g/L maltosa y 20 g/L agar) y almacenadas a 4°C en cajas Petri (medio modificado de Maqueda et al., 2010).

Para la preparación del inóculo y su posterior uso, la activación se llevó a cabo en erlenmeyer de 250 mL con 50 mL de medio líquido con triptosa (3.5 g/L), extracto de levadura (30 g/L), glucosa (10 g/L), fosfato diácido de potasio (2 g/L) y sulfato de amonio (1 g/L) (medio modificado de Atiyeh & Duvnjak, 2002). La incubación se llevó a cabo por un periodo de 16-24 horas a 30°C con agitación de 120 rpm.

2.2 Extracción de ADN y estrategia de secuenciación genómica

La extracción de ADN se realizó con el kit YeaStar columns (Zymo Research) de acuerdo a las recomendaciones del fabricante (ver anexo 1).

Se empleó el método de secuenciación aleatoria usando la plataforma *Illumina HiSeq 2000* con finales pareados.

2.3 Análisis Computacional

El método usado para el análisis computacional se representa de manera general en la figura 2-1, donde destacan 3 etapas: 1. Preprocesamiento, 2. Ensamblaje y 3. Asignación de genes a rutas metabólicas y posterior análisis. En los numerales 2.3.1-2.3.5 se habla con más detalle de las herramientas y métodos usados en cada paso.

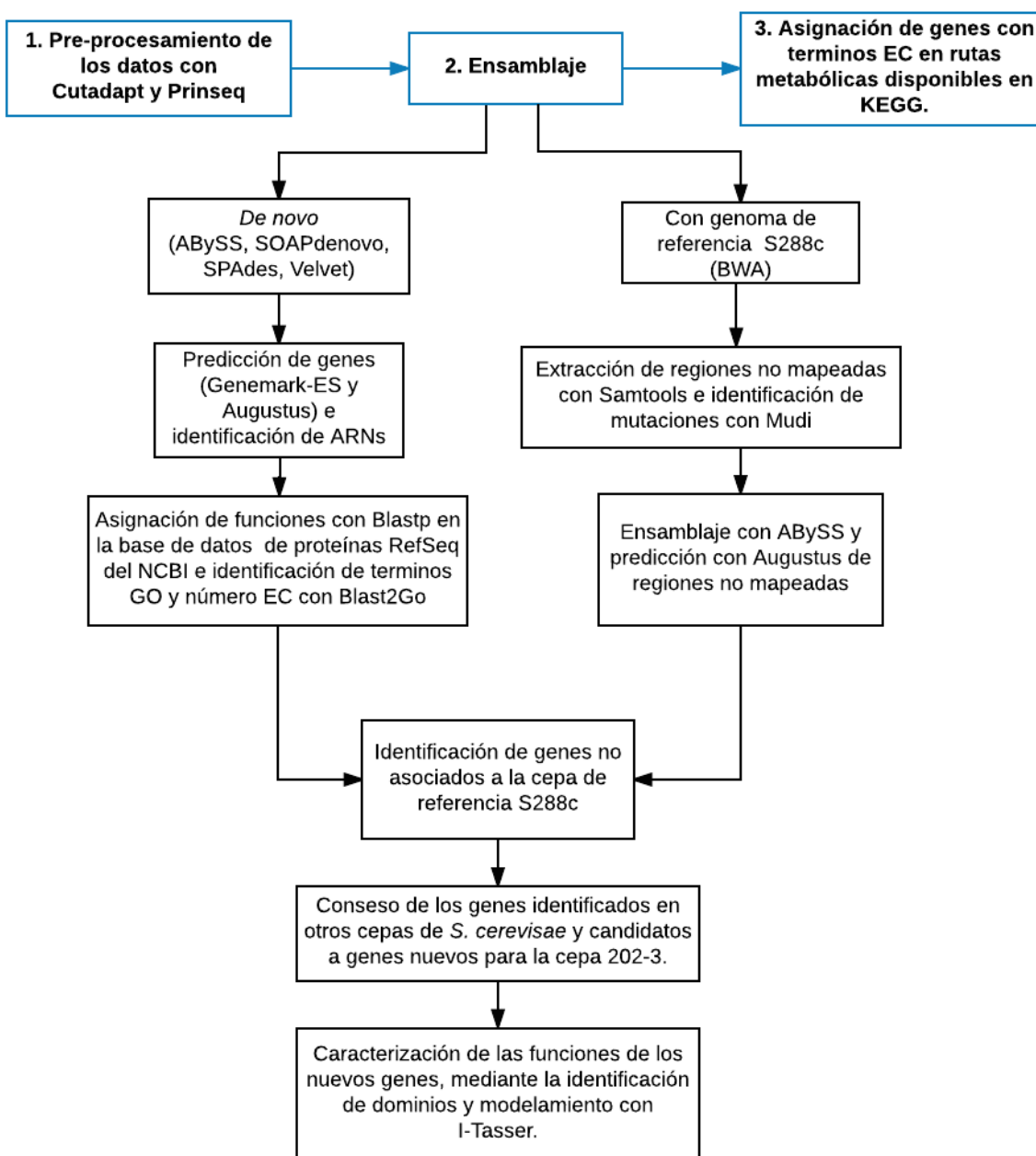


Figura 2-1. Descripción general del procedimiento computacional implementado para el análisis de las lecturas obtenidas con la tecnología Hiseq Illumina 2000 con finales pareados de la cepa *S. cerevisiae* 202-3.

2.3.1 Preprocesamiento de los datos

La calidad de las lecturas primarias se evaluó con el programa FastQC Version 0.11.4 (Andrews et al., 2015). Para la remoción de adaptadores, se empleó el software Cutadapt (Martin et al., 2011) y con Prinseq (Schmieder et al., 2011) se seleccionaron las lecturas que tenían un valor de calidad mayor a 24 Phred.

2.3.2 Ensamblaje del genoma: Aproximación *de novo*

Con el fin de establecer la mejor estrategia para el ensamblaje *de novo*, con la cual se pudiera minimizar el número de contigs y maximizar la cobertura del genoma, se implementaron los programas Velvet (Zerbino et al., 2008), SPAdes (Bankevich et al., 2012), ABySS (Simpson et al., 2009) y SOAPdenovo2 (Luo et al., 2012), con k-mers impares desde 21 hasta 91 y demás parámetros por defecto para lecturas de finales pareados y corrección de errores. Para todos los ensamblajes se evaluó el número de contigs, la longitud del contig más largo, el N50, el N75 y el número de N's por cada 100 Kpb, como medidas para seleccionar un ensamblaje óptimo con la herramienta Quast (Yandell et al., 2012; Gurevich et al., 2013).

Finalmente, para el ensamblaje seleccionado, los gaps fueron cerrados con la herramienta GapCloser paquete del software SOAPdenovo2.

Los scaffolds obtenidos de este numeral fueron anotados estructural y funcionalmente, según el procedimiento descrito el numeral 2.5.

2.3.3 Ensamblaje del genoma: Aproximación con genoma de referencia

Para realizar el ensamblaje con genoma de referencia se empleo el programa BWA (Li et al., 2009-a), con el genoma de las cepa de referencia *S. cerevisiae* S288c, cepa que ampliamente se emplea como genoma de referencia ya que fue la primera levadura en ser secuenciada y presenta un amplio trabajo a nivel de anotación y marcos de lectura caracterizados (Engel, 2014; Fisk, 2006).

Una vez realizado el ensamblaje por mapeo se obtuvo un conjunto de lecturas que no pudieron ser mapeadas contra dicho genoma. De esta manera se procedió a extraerlas con la herramienta samtools (Li et al., 2009-b) ensamblarlas *de novo* con el programa ABySS, eligiendo el mejor ensamblaje según parámetros mencionados en el numeral 2.3.2. Los scaffolds obtenidos de este numeral fueron anotados estructural y funcionalmente, según el procedimiento descrito el numeral 2.3.4 y 2.3.5 y comparados con los resultados obtenidos con la aproximación del ensamblaje *de novo*.

Dada la diversidad genómica que poseen las cepas de *S. cerevisiae* (Borneman et al., 2011), la identificación de mutaciones se considera un paso importante, pues permite justificar algunas diferencias fenotípicas respecto a otras cepas de la misma especie (Akao et al., 2011). Con el fin de encontrar estas posibles mutaciones en el genoma de la cepa *S. cerevisiae* 202-3 respecto a la cepa de referencia S288c, se implementó la herramienta Mudi (Iida et al., 2014).

Uno de los pasos limitantes a la hora de identificar mutaciones es el nivel de ploidia que presentan las levaduras, en el caso de levaduras que son más que haploides, la identificación de mutaciones se hace más compleja pues algunos genes pueden ser heterocigotos y pasan a ser detectados como mutaciones de un solo gen (Akao et al.,

2011), por tal motivo para evidenciar si la cepa 202-3 es más que haploide se amplificaron los genes MAT α y MAT α con los primers MAT α , MAT α y MAT reportados previamente (Huxley et al., 1990).

2.3.4 Anotación genómica: Anotación estructural

La predicción de secuencias de ADN codificante (CDS, por sus siglas en inglés) se llevó a cabo con Augustus (Stanke et al., 2005), empleando como plantilla la cepa *S. cerevisiae* S288c y con el programa GeneMark-ES (Lomsadze et al., 2005). Para evaluar la calidad de los genes obtenidos se implementó la herramienta BUSCO (Simao et al., 2015) como sugieren algunos reportes (Drozdova et al., 2016). Adicionalmente se realizó la predicción de ARN ribosomal y de transferencia con Barrnap (VBC, 2016) y ARAGORN (Laslett, 2004) respectivamente. Para complementar estos resultados se anotaron las regiones asociadas a elementos móviles y pequeños ARNs con la herramienta RepeatMasker (Graovac et al., 2009).

Con el fin de corroborar los datos experimentales obtenidos anteriormente se extrajeron las regiones ITS1 e ITS2 con ITSx (Bengtsson-Palme et al., 2013), para identificar la especie de la levadura con EzFungi (Jongsik Chun Lab et al., 2015).

2.3.5 Anotación genómica: Anotación funcional

Los CDS obtenidos anteriormente en formato GFF, fueron extraídos y se tradujeron a secuencias peptídicas con el *script* gff2fasta.pl (Imelfort et al., 2013) y se alinearon mediante la herramienta Blastp (Altschu et al., 1990), contra la base de datos Refseq protein del NCBI (Pruitt et al., 2005).

La obtención de las anotaciones textuales, términos GO, números E.C y rutas metabólicas de KEGG, se realizó con la herramienta Blast2GO (Conesa et al., 2005).

Finalmente para las secuencias que no fueron identificadas en la base de datos Refseq, se realizó la búsqueda en la base de datos no redundante del NCBI (Pruitt et al., 2005). Las proteínas cuyas funciones no fueron asociadas a la especie de *S. cerevisiae* fueron modeladas mediante la herramienta I-TASSER, pues esta herramienta permite asignar mediante la predicción de la estructura terciaria y algunos alineamientos estructurales las posibles funciones de la proteína (Yang et al., 2015).

3. Resultados y Discusión

Con el fin de desarrollar el objetivo principal de este proyecto como primera etapa se propuso establecer las características genómicas de la cepa *S. cerevisiae* 202-3, mediante los procedimientos de ensamblaje y anotación, resultados que se presentan en los numerales 3.1, 3.2 y 3.3. Finalmente para concluir el propósito de este trabajo en el numerales 3.4 y 3.5 se presentan los resultados asociados a la identificación de los genes candidatos asociados al proceso de consumo de xilosa.

3.1 Preprocesamiento de los datos

El ADN genómico de la cepa 202-3 *S. cerevisiae* fue secuenciado por la metodología Illumina/Solexa Hiseq 2000 de finales pareados, la calidad de las lecturas crudas y preprocesadas se presenta en la tabla 3-1 y figura 3-1 y anexo 2.

Tabla 3-1. Número de lecturas y su longitud en las etapas de pre-procesamiento

	Lecturas crudas (pe1)	Preprocesadas
<i>Número de secuencias</i>	11.938.533	10.867.914
<i>Secuencias de baja calidad</i>	0	0
<i>Longitud de las Secuencias</i>	101	80-101
<i>% GC</i>	38	38

	Lecturas crudas (pe2)	Preprocesadas
<i>Número de secuencias</i>	11.938.533	10.867.914
<i>Secuencias de baja calidad</i>	0	0
<i>Longitud de las Secuencias</i>	101	80-101
<i>%GC</i>	38	38

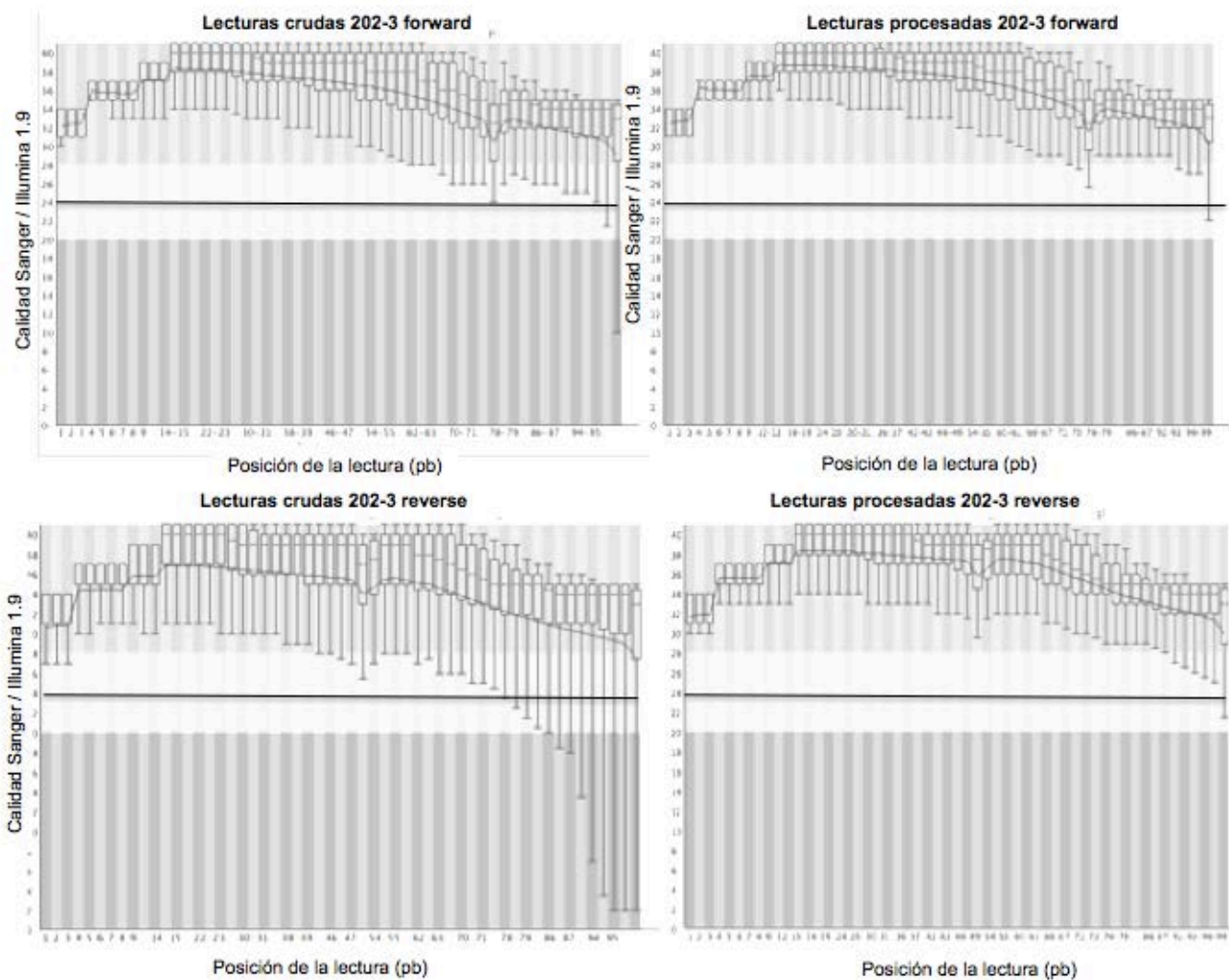


Figura 3-1. Calidad de las lecturas por base, antes (derecha) y después (izquierda) de la evaluación de la calidad y pre-procesamiento.

Para las lecturas obtenidas, se observa que la calidad para el conjunto forward (pe1) la calidad en la asignación de bases, medida con la unidad Phred, presenta mejores valores de calidad, sin embargo según el reporte de calidad proporcionado por FastQC, la calidad de los conjuntos de lecturas es bueno, a pesar de la dispersión en las posiciones finales, hecho que se presenta con frecuencia al usar esta tecnología de secuenciación (Dohm et al., 2008).

Teniendo en cuenta que no se encontraron parámetros de baja calidad según reporte por FastQC, se procedió a realizar la remoción de adaptadores con Cutadapt de las secuencias específicas de la librería empleada, una vez retirados los adaptadores, con Prinseq se procedió a eliminar las lecturas que no fueran mayores a 80 pb y para disminuir la dispersión en las posiciones finales de las lecturas se removieron las lecturas con valores de Phred medios menores a 24, igualmente se descartaron las lecturas que perdieron su par correspondiente después de emplear los filtros anteriores (ver tabla 3-1 y figura 3-1).

3.2 Ensamblaje

3.2.1 Ensamblaje *de novo*

En las tablas 3-2 y 3-3, se presentan las métricas que fueron empleadas para evaluar el desempeño de cada ensamblador, los parámetros que se tuvieron en cuenta corresponden a los reportados por Yandell et al., 2012, donde destacan el N50 de los contigs y scaffolds obtenidos, el número de N's en el ensamblaje y el porcentaje de cobertura (cobertura del genoma). Teniendo en cuenta estos parámetros se observa que para el ensamblaje de los contigs Velvet resulta ser una posible opción, sin embargo el número de N's es mayor frente a los demás ensamblajes y el tamaño esperado del genoma no corresponde al más próximo (12,1 Mb, SGD, 2014). Teniendo en cuenta la importancia de estos dos parámetros, el ensamblaje de Velvet fue descartado.

Tabla 3-2. Resumen de las métricas para los mejores ensamblajes de contigs obtenidos con diferentes ensambladores.

Ensamblador	<i>k</i> -mer	# Contigs	Contig mayor longitud (pb)	Tamaño total (pb)	NG50	NG75	# N's por 100 kbp
Velvet	81	147	706827	11770565	330754	234645	587.77
SPAdes	85	193	704005	11693852	301623	156040	0
ABYSS	89	184	548475	11856708	248053	132812	58.05
SOAPdenovo	81	842	110115	11590510	28815	16252	0

Con los tres ensamblajes restantes se procedió a realizar la comparación de los scaffolds obtenidos (tabla 3-3). Al comparar los scaffolds, resalta el ensamblaje realizado con ABYSS, presentando el menor número de contigs, el contig más largo, tamaño del genoma más aproximado (SGD, 2014) y con un valor intermedio del número de N's presentes una vez tratado el ensamblaje con Gapcloser, siendo este último parámetro bastante cercano al valor mínimo reportado por SPAdes.

Tabla 3-3. Resumen de las métricas para los mejores ensamblajes de scaffolds obtenidos con diferentes ensambladores.

Ensamblador	<i>k</i> -mer	#Contigs	Contig mayor longitud (pb)	Tamaño total (pb)	NG50	NG75	# N's por 100 kbp	
							Antes gapcloser	Después gapcloser
SPAdes	85	179	704005	11690365	301655	170567	3.37	1.86
ABYSS	89	146	711362	11861222	301461	180439	81.67	3.49
SOAPdenovo	81	159	703675	11682269	310026	232600	263.66	263.66

Teniendo en cuenta los parámetros mencionados anteriormente, el ensamblaje seleccionado fue el realizado con el ensamblador ABYSS con un *k*-mer de 89, para ver con más detalle todas las métricas reportadas de los *k*-mers probados ver Anexo 3.

3.3 Anotación genómica

3.3.1 Predicción de CDS

Teniendo en cuenta los diferentes reportes bibliográficos de secuenciamiento y predicción de genes, tabla 1-3, no hay un valor consenso para el número de genes totales en *S. cerevisiae*, sin embargo teniendo en cuenta la aproximación realizada por Lin et al., 2013, se esperan alrededor de 6091 proteínas. Según este valor de genes codificantes el predictor de genes que logra un mayor número de proteínas es el llevado a cabo por Augustus, por otro lado otros valores como la mediana, promedio, valor mínimo y máximo son mayores con este predictor, como se observa en la tabla 3-4. Sin embargo a pesar de tener estos valores, no es posible afirmar que predictor podría ser mejor pues para este proyecto no se cuenta con datos de expresión, para demostrar la correcta presencia o no de un modelo génico. Para fines de este proyecto la elección para los posteriores pasos se realizó con Augustus, por los datos presentados en la tabla 3-4 y los reportes bibliográficos de la tabla 1-3, adicionalmente la herramienta BUSCO corroboró que las proteínas predichas mapean más del 98% de los genes de la base de datos establecida en BUSCO para eucariotas y adicionalmente estas secuencias se encuentran en su mayoría completas y no fragmentadas. Aún así para contrastar estos resultados obtenidos se realiza una comparación con el ensamblaje con genoma de referencia y evidenciar la posible existencia de genes nuevos o no en la cepa *S. cerevisiae* 202-3.

Tabla 3-4. Predicción de regiones codificantes con las herramientas GeneMark-ES y Augustus para el ensamblaje realizado con ABySS con el k-mer 89.

Herramienta	Número de genes	Longitud de CDS (pb)			
		Mediana	Min	Promedio	Max
Augustus	5497	1269	39	1522	14730
GeneMark-ES	5162	1247	2	1464	9946

3.3.2 Predicción de ARN ribosomal y regiones ITS

Mediante la herramienta Barnap (VBC, 2016) se predijeron 5 genes ribosomales (ARNr), identificados como 5S, 18S, 28S, 5.8S, reportados en *S. cerevisiae* (Venema et al., 1999) y 12S, este último ARN mitocondrial fue identificado parcialmente (68%), esta dificultad para encontrar la secuencia completa de ARN ribosomal podría explicarse por el hecho de la diferencia entre la evolución del ADN nuclear y el ADN mitocondrial, la replicación del ADN mitocondrial no está ligada al ciclo celular, por lo que a este se atribuye una mayor tasa de mutación (Wolters et al., 2015), de allí la diferencia del gen ribosomal 12S predicho frente a lo reportados para las demás cepas de esta especie.

Para la identificación de las regiones ITS1 e ITS2, presentes en eucariotas se empleó la herramienta ITSx (Bengtsson-Palme et al., 2013). Con el fin de confirmar la identidad de la cepa estudiada se realizó una búsqueda en EzFungi (Jongsik Chun Lab, 2015) con los genes ITS1 e ITS2 identificados, allí se corroboró que la cepa corresponde a la especie *S. cerevisiae*, presentando el mayor porcentaje de identidad con la cepa CBS1171 (cepa para la producción de cerveza) y la cepa Kyokai N°7 (cepa para la producción de Sake).

3.3.3 Predicción de ARN de transferencia

La herramienta ARAGORN (Laslett et al., 2004) predijo 246 ARN de transferencia (ARNt), representados en el anexo 4, normalmente para *S. cerevisiae* se suelen encontrar aproximadamente 275 ARN de transferencia (Iben et al., 2012; Hani et al., 1998; Chan et al., 2009). Allí se observa que el ARNt de mayor frecuencia es el correspondiente al transporte de ácido aspártico (GTC), lo cual concuerda con lo reportado para *S. cerevisiae*, siendo este uno de los de mayor frecuencia junto con glicina (GCC) (Iben et al., 2012).

Para este proyecto el uso de la herramienta RepeatMasker se implementó como complemento de los resultados reportados en los numerales 3.3.2 y 3.3.3, mediante la anotación de las regiones repetitivas y de baja complejidad (presencia de transposones, pequeños ARNs, entre otros). No se empleó la opción de enmascarar estas regiones, teniendo en cuenta reportes previos de ensamblajes de levadura (Akao et al., 2011; Borneman et al., 2011) y recomendaciones del desarrollador del programa (Graovac et al., 2009).

Según RepeatMasker el genoma cuenta con cero regiones de baja complejidad, satélites y repeticiones simples, posee 165 pequeños ARNs representando el 0.13% de las secuencias, 16 SINEs (short interspersed elements, por sus siglas en inglés) siendo el 0.01% de las secuencias, 43 LINEs (long interspersed elements, por sus siglas en inglés) con un 0.02% de las secuencias y 15 de otros elementos móviles de ADN correspondiente a un 0.01% de las secuencias. En contraste con otras cepas de la misma especie, se encuentra una gran variedad de reportes, pues en el caso de los elementos móviles, la variación de estos permite incluso clasificar las levaduras por debajo del nivel de especie dada su diversidad (Legras et al., 2003).

3.3.4 Anotación funcional

De las 5497 proteínas, 8 de ellas no presentaron ningún resultado al realizar Blastp con la base de datos RefSeq, por lo que con estas secuencias se procedió a realizar la búsqueda en la base de datos no redundante de proteínas del NCBI al igual que para las secuencias que presentaron resultados con proteínas de otras especies. Dado que la predicción de modelos génicos puede llevar a concluir falsos resultados (Denton et al., 2014), durante el desarrollo de este proyecto se llevaron a cabo dos aproximaciones como se mencionó anteriormente, ensamblaje *de novo* y por mapeo con genoma de

referencia de la cepa S288c, de las regiones que no fueron mapeadas en el ensamblaje con genoma de referencia se contrastaron con las obtenidas en el ensamblaje *de novo*. De las 8 proteínas que no fueron asociadas a ningún resultado en RefSeq protein, se encontró que todas ellas están reportadas para otras cepas de *S. cerevisiae*, como se observa en la tabla 3-5. De este grupo de proteínas se evidencia la presencia de varios genes PAU asociados a los genes G2.t1, G149.t1 y G2676.t1 identificados en la cepa 202-3, esto se debe a la dificultad de identificar genes en las regiones subteloméricas, como lo son los genes de la familia PAU, y la diversidad que presentan estos genes entre la misma especie (Akao et al., 2011). Para el gen G262.t1, se encuentra que si está reportado para la cepa S288c, sin embargo este gen corresponde una proteína no caracterizada. Los genes G2759.t1 y G5333.t1, no reportan función como tal, pues no se identifican dominios, por lo que sus estado permanece como proteína hipotética.

Las proteínas que fueron identificadas como genes de otras especies, fueron analizados con más detalle y comparados con los resultados obtenidos en el ensamblaje por mapeo en la sección 3.3.5.

Tabla 3-5. Proteínas que no presentaron ningún resultado al realizar el Blastp con la base de datos RefSeq. Para estas proteínas se realizó nuevamente la búsqueda en la base de datos no redundante (nr) de proteínas.

Identificador del gen en <i>S. cerevisiae</i> 202-3	Tamaño (aa)	Resultado Blastp contra nr de proteínas	%Identidad	E-value
G2.t1	13	Pau18p, <i>S. cerevisiae</i> YJM1615 (parcial)	100	3E-4
G149.t1	14	Pau18p, <i>S. cerevisiae</i> YJM1615 (parcial)	100	2E-5
G262.t1	163	YHR180W Proteína Putativa no caracterizada, <i>S. cerevisiae</i> S288c	100	5E-115
G1926.t1	131	P301_P10001 Proteína hipotética, <i>S. cerevisiae</i> P301	100	1E-88
G2269.t1	131	P301_P10001 Proteína hipotética, <i>S. cerevisiae</i> P301	98%	1E-88
G2676.t1	32	Pau23p, <i>S. cerevisiae</i> YJM1078	75	4E-4
G2759.t1	74	CENPK1137D_4762 Proteína hipotética, <i>S. cerevisiae</i> CEN.PK113-7D	99	2E-46
G5333.t1	110	C1Q_05663 Proteína hipotética, <i>S. cerevisiae</i> JAY291	96	5E-49

El siguiente paso asociado al enriquecimiento de los resultados obtenidos con Blastp fue la asignación de los términos GO, mediante la herramienta Blast2GO, como se observa en la tabla 3-6, la mayoría de genes fueron mapeados con terminos GO, lo cual facilita su análisis para la búsqueda de funciones.

Tabla 3-6. Resumen de la anotación con Blast2GO (B2G) más InterProScan (IPS).

	B2G
Genes anotados con términos GO	5383
Genes con IPS	5284
GO+IPS	3887
Sin IPS	213

3.3.4.1. Análisis cuantitativo de las anotaciones funcionales: Ontología de Genes

La distribución de las anotaciones GO por niveles, se muestra en la figura 3-2. Allí se observa que la mayoría de secuencias cuenta con anotaciones de nivel 8 para componente celular y procesos biológicos, mientras que para las anotaciones de función molecular la mayoría son de nivel 6. A medida que aumenta el nivel de especificidad (niveles mayores a 10), los términos disminuyen.

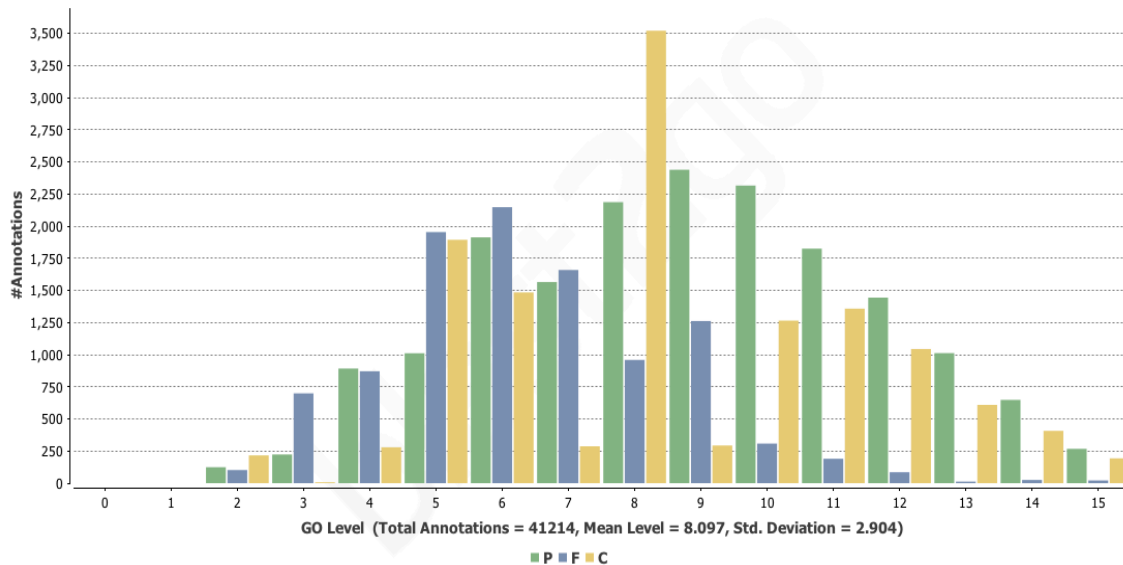


Figura 3-2. Distribución de los términos GO mapeado por nivel. Proceso biológico (verde), función molecular (azul) y componente celular (amarillo).

Con el fin de explorar con más detalle los términos GO asignados y contrastarlos con reportes para *S. cerevisiae* según la base de datos SGD, 2016 se obtuvieron las figuras 1-3, en el anexo 5. Allí se observa que para los procesos biológicos, la mayoría de anotaciones corresponde a procesos celulares, como la organización de organelos, regulación de procesos celulares, entre otros, esta distribución corresponde a lo observado en la base de datos SGD, 2016 para *S. cerevisiae*.

En el caso de las funciones moleculares, destacan los terminos de unión, atribuidos en su gran mayoría a uniones de ATP, metales, DNA, RNA y proteína. Finalmente para los componentes celulares hay una mayor representación de terminos asociados a organelos intracelulares, al igual que se reporta en la base de datos. Cabe mencionar que en general para *S. cerevisiae* aún hay un gran número de genes anotados como

desconocidos, para el caso de procesos biológicos hay 1822 genes anotados como desconocidos, para función molecular 2648 y finalmente componente celular 1412 (SGD, 2016), por lo que identificar la totalidad de terminos para las proteínas estudiadas aún está en desarrollo.

La identificación de las rutas metabólicas potenciales en la cepa *S. cerevisiae* 202-3, se realizó con el software Blast2GO con la implementación de las bases de datos de Kegg (Aoki-Kinoshita et al., 2007). Allí se asignaron 2906 secuencias a 115 rutas metabólicas diferentes, como se observa en la figura 4 del anexo 5 y anexo 6. Cabe resaltar que para algunas proteínas con actividad enzimática, Blast2GO no asigna número EC y esto sucede con frecuencia para las enzimas cuyos números no se encuentran completos, ocasionando errores en la anotación de los genes (Green et al., 2005), p.ej. la aldo ceto reductasa codificada con por el gen YJR096W con número EC 1.-.-.-., por este motivo la asignación de número EC fue sometida a revisión manual para los genes con funciones relevantes a la investigación. El análisis de las rutas identificadas se hará en el numeral 3.4.

3.3.5 Ensamblaje con genoma de referencia

Con el fin de corroborar los resultados obtenidos con el ensamblaje *de novo* y encontrar posibles mutaciones en algunos genes de interés se desarrolló la metodología del ensamblaje con genoma de referencia empleando como plantilla el genoma de la cepa S288c. Como se observa en la tabla 3-7 se obtienen diferentes porcentajes de cobertura para cada cromosoma, siendo el menor el correspondiente al ADN mitocondrial, lo que concuerda con el bajo porcentaje identificado en el gen 12S del ensamblaje *de novo*. En general el porcentaje total de mapeo, 97,7%, es relativamente alto frente a lo reportados para otras cepas de la misma especie, cuyos porcentajes oscilan entre el 76-94% (Drozdova et al., 2016).

Tabla 3-7. Mutaciones identificadas con Mudi para el genoma de la cepa *S. cerevisiae* 202-3 respecto a la cepa S288c.

# Cromosoma	Tamaño (pb)	% Cobertura
chr01	230218	94.60%
chr02	813184	98.80%
chr03	316620	99.70%
chr04	1531933	97.90%
chr05	576874	96.30%
chr06	270161	93.40%
chr07	1090940	98%
chr08	562643	97.40%
chr09	439888	96.30%
chr10	745751	98%
chr11	666816	99.80%
chr12	1078177	95.20%

chr13	924431	98.50%
chr14	784333	98.80%
chr15	1091291	98.90%
chr16	948066	97.50%
ADN Mitocondrial	85779	83.70%
%Total mapeado	97.7%	
Profundidad del mapeo	174	

Con el fin de identificar las funciones de las proteínas con mayor acumulación de mutaciones la lista de genes fue anotada en la plataforma PANTHER (Huaiyu et al., 2010), en la figura 3-3 se observa que la mayoría de mutaciones se encuentra asociada a proteínas oxidorreductasas, proteínas de citoesqueleto, proteínas de unión a calcio y chaperonas. el detalle de las mutaciones encontradas se encuentra en el Anexo 7.

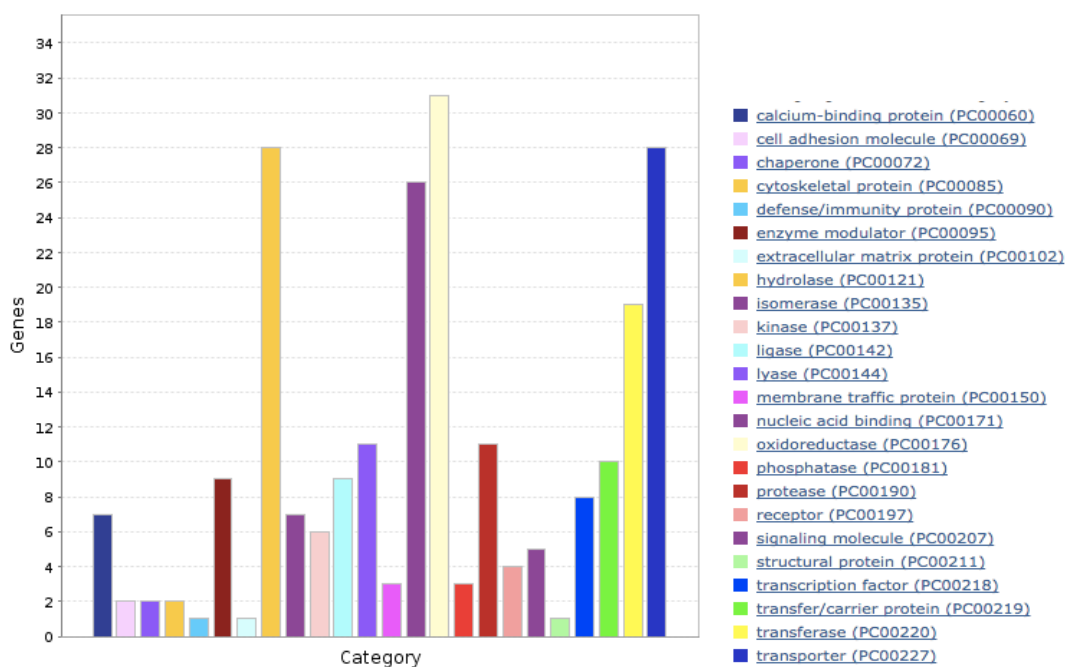


Figura 3-3. Tipos de proteínas identificadas con mutaciones.

Es necesario resaltar el gran número de mutaciones identificadas, 76298, así como los altos valores de cobertura, hasta 7529 lecturas, lo cual puede justificarse porque la cepa 202-3 es más que haploide como se observa en la figura 3-4. Este mismo caso se presenta para la cepa Kyokai N°7, cuyas secuencias corresponden a una cepa diploide y logran identificar 1347 sitios heterocigotos, caso que podría parecerse a la cepa 202-3, dada su cercanía filogenética (Ver anexo 8).

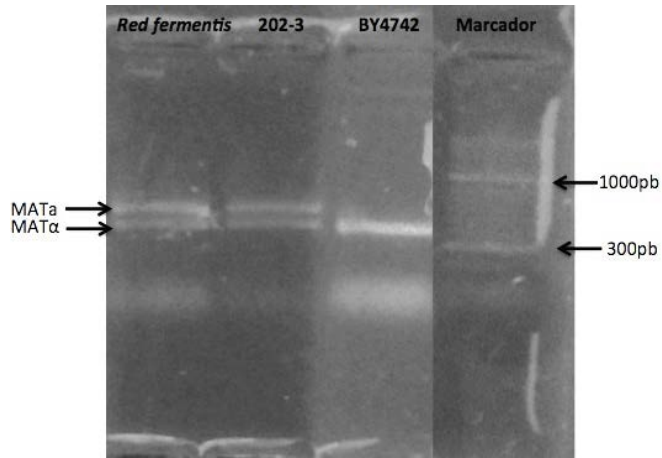


Figura 3-4. Identificación de los genes MATa y MATα, genes asociados a cepas que no son haploides. El estudio se realizó para la cepa *S. cerevisiae* 202-3, empleando como patrón la cepa haploide BY4742, MATα y la cepa industrial *Red fermentis*.

Las mutaciones de los genes reportados previamente y asociados al consumo de xilosa se encuentran en la tabla 3-8, los detalles de estas mutaciones se encuentran en el anexo 7. Como se observa la mayoría de genes tiene asociada una mutación a excepción del gen SOR2 y XDH1, este último no se encuentra en la cepa S288c ni 202-3. Las mutaciones que se reportan en este trabajo deben ser analizadas con detalle, pues muchas de ellas son mutaciones que afectan el marco de lectura, como el caso del gen HXT12.

Tabla 3-8. Número de mutaciones encontradas para los genes asociados al metabolismo de xilosa y otros genes de interés.

# Mutaciones	Gen (Nombre sistemático)
3	GRE3 (YHR104w)
2	GCY1 (YOR120W)
6	YPR1 (YDR368W)
4	YJR096W
21	YDL124W
9	XYL2(YLR070C)
4	SOR1 (YJR159W)
--	SOR2 (YDL246C)
--	XDH1
13	XKS1 (YGR194C)
9	TKL1 (YPR074C)
6	TAL1 (YLR354C)
18	HXT1(YHR094C)
8	HXT2 (YMR011W)
13	HXT4(YHR092C)
11	HXT5 (YHR096C)
32	HXT7 (YDR342C)

10	GAL2(YLR081W)
Otros genes asociados a la ruta de las petosas y de importancia para la producción de etanol	
1	GND1 (YHR183W)
7	ZWF1 (YNL241C)
4	RPE1 (YJL121C)

En la tabla 3-9 se presenta con la nomenclatura g###.t1 (#=número del gen) los genes no mapeados en el ensamblaje con genoma de referencia, en total se identifican 31 genes de los cuáles mediante Blastn se asocian en su mayoría a genes que previamente no fueron anotados con funciones pertenecientes a la especie *S. cerevisiae* del genoma ensamblado *de novo*. Al corroborar la descripción de estos genes en la base de datos no redundante de proteínas, estos genes se asocian a otras cepas de *S. cerevisiae* (Drozdova et al., 2016; Borneman et al., 2011; Argueso et al., 2009; Akao et al., 2011) a excepción de dos genes pertenecientes a *Homo sapiens* (G4.t1) y *Lachancea lanzarotensis* (G3707.t1).

Respecto a los 8 genes listados en la tabla 3-5, no se encuentra el par correspondiente a excepción de los genes G2269.t1 y G1926.t1, sin embargo, como se mencionó anteriormente esto se justifica dada la naturaleza de dichos genes. Es importante mencionar el caso del gen G3818.t1 del ensamblaje *de novo* que no se evidencio en el conjunto de 31 proteínas de la tabla 3-9, cuyo resultado indica un alto grado de identidad con la proteína K7_05679p, de la cepa Kyokai N°7, esta misma proteína que no presentó un ORF cercano a la cepa S288c, se considera como una proteína que reversamente se sobrelapa con la proteína K7_05679p un ortólogo del gen YMR317W (Akao et al., 2011), representado con una identidad del 97% en el gen G3817.t1 de la cepa *S. cerevisiae* 202-3.

Respecto a las proteínas cuyos pares fueron reportados es importante identificar posibles funciones que previamente no han sido descritas, pues nos pueden dar un indicio de la naturaleza del fenotipo positivo en el consumo de xilosa para la cepa 202-3. Por este motivo se menciona a continuación una breve descripción de los genes reportados en la tabla 3-9.

En el caso de la proteína G278.t1, esta fue identificada como el gen YPS6 en la cepa S288c con un 97% de identidad, al mapear este gen contra otras cepas de la misma especie, se encuentra una mayor identidad con el gen K7_06167, identificado previamente en el ensamblaje por mapeo con una menor longitud de la proteína. Este gen ha sido identificado en cepas como Kyokai N°7, EC1118 y RM11-1a (Akao et al., 2011). Igualmente pasa con el gen G3809.t1 presente en la cepa JAY291.

Cabe resaltar la importancia de la presencia de algunas proteínas que aunque no estén relacionadas directamente con la producción de etanol, pueden llegar a ser productos industriales de gran interés, como el caso de la proteína codificada por el gen G1407.t1 identificada en la cepa Kyokai N°7 previamente, cuyos dominios representan enzimas de

la familia epoxido hidrólidas (Botes et al., 1999). Otros genes de interés que no han sido reportados en la cepa S288c, corresponden a los genes G280.t1, este asociado a la función GPR1 (identificada por los dominios presentes con la herramienta Blastp), que podría ser un gen importante en la regulación del crecimiento (Won et al., 1997).

En la tabla 3-9, se presentan un grupo de proteínas que aunque tienen un gen asociado según reportes de otras cepas su función aún no ha sido identificada, estos casos fueron comprobados al buscar posibles dominios y no encontrar asignación alguna. Tal es el caso para los genes G2269.t1, G1409.t1, G2658, en el caso de la proteína G3810.t1, su función fue asignada recientemente mediante el programa AGAPE (Song et al., 2015), sin evidencia experimental, su función corresponde a un factor de transcripción. Al realizar una búsqueda más detallada del gen G3810.t1 y estudiar sus dominios (Bauer et al., 2015), se observa un dominio asociado a factores de transcripción que contienen un N-terminal como es el caso de GAL4 en *S. cerevisiae* (gen identificado en la cepa *S. cerevisiae* 202-3 con el gen G1994.t1), así mismo se encuentra el dominio Fungal_trans, que si bien también se encuentra en el gen GAL4 se ha reportado igualmente en activadores transcripcionales como el xlnR presente en *Aspergillus niger*, donde regulan la expresión de genes asociados a xilosa reductasa (Hasper et al., 2000). Confirmar la presencia de este gen que no está presente en la cepa de referencia S288c y si en cepas productoras de bioetanol (JAY291), resulta ser de gran importancia dado que estudios de expresión diferencial demuestran que los genes GAL se sobreexpresan al tener xilosa como fuente de carbono, siendo el gen GAL4 uno de los genes con altos niveles de expresión (Matsushika et al., 2014), por lo que podría estar implicado en el fenotipo positivo de consumo de xilosa en la cepa *S. cerevisiae* 202-3.

El gen identificado como G3811.t1 fue reportado previamente en la cepa de producción de bioetanol JAY291, gen C1Q_05665 (Argueso et al., 2009), cuya función no es específica pero al realizar una búsqueda de dominios (Bauer et al., 2015) se encuentra la superfamilia MFS que se encuentra ligada al transporte de azúcares fosfatados, iones, drogas, aminoácidos y otros sustratos, es interesante resaltar que en cepas cuyo consumo de xilosa se considera positivo, como la cepa EC1118, este gen no se encuentra, a pesar de estar asociado posiblemente al transporte de azúcares.

Tabla 3-9. Genes predichos con Augustus para las regiones no mapeadas con BWA para la cepa *S. cerevisiae* 202-3 y comparación con los genes obtenidos en el ensamblaje *de novo* con ABYSS.

Identificador Regiones no mapeadas	Tamaño (aa)	Identificador ensamblaje <i>de novo</i>	Tamaño (aa)	% Identidad	Gen asociado y descripción
g1.t1	128	G3132.t1	1180	100	VIN13_0549 Proteína hipotética, <i>S. cerevisiae</i> Vin13
g2.t1	229	G3133.t1	229	100	SCRG_03463 Proteína hipotética, <i>S. cerevisiae</i> RM11-1a
g3.t1	51	G5227.t1	892	100	YIL169C Proteína hipotética, <i>S. cerevisiae</i> S288c
g4.t1	424	G278.t1	532	100	K7_06169p, <i>S. cerevisiae</i> Kyokai no. 7
g5.t1	557	G4861.t1	557	100	AWRI796_3907 Proteína hipotética, <i>S. cerevisiae</i> AWRI796

g6.t1	143	G4.t1	143	100	Subunidad alfa de hemoglobina, <i>Homo sapiens</i>
g7.t1	172	G3132.t1	1180	100	Hpf1p, <i>S. cerevisiae</i> S288c
g8.t1	356	G2003.t1	356	100	C1Q_05652 Proteína hipotética, <i>S. cerevisiae</i> JAY291
g9.t1	118	G278.t1	532	100	Yps6p, <i>S. cerevisiae</i> S288c
g10.t1	594	G279.t1	594	100	Proteína hipotética conservada, <i>Saccharomyces cerevisiae</i> RM11-1a
g11.t1	275	G280.t1	275	100	K7_06167p, <i>S. cerevisiae</i> Kyokai no. 7
g12.t1	149	G4.t1	143	44.14	Subunidad alfa de hemoglobina, <i>Homo sapiens</i>
g13.t1	574	G3132.t1	229	82.81	Hpf1p <i>S. cerevisiae</i> FostersB
g14.t1	117	G3709.t1	400	90.7	Irc7p, <i>Saccharomyces cerevisiae</i>
g15.t1	517	G3708.t1	517	100	Seo1p, <i>S. cerevisiae</i> YJM1389
g16.t1	245	G3707.t1	245	100	LALA0S15e01596g1_1, <i>Lachancea lanzarotensis</i>
g17.t1	562	G3706.t1	562	100	Yrm1p, <i>Saccharomyces cerevisiae</i> YJM1450
g18.t1	123	G2436.t1	589	100	Oligo-1,6-glucosidasa IMA1 [<i>Saccharomyces cerevisiae</i> S288c], identificado con un 34% de identidad.
g19.t1	182	G2437.t1	463	100	WN66_02735 Proteína hipotética, <i>S. cerevisiae</i> , Factor de transcripción Mal33
g20.t1	692	G3808.t1	1184	100	WN66_01991 Proteína hipotética, <i>S. cerevisiae</i> , Factor de transcripción Mal33
g21.t1	544	G3809.t1	544	100	C1Q_05667 Proteína hipotética, <i>S. cerevisiae</i> JAY291
g22.t1	475	G3810.t1	475	100	C1Q_05666 Proteína hipotética <i>S. cerevisiae</i> JAY291
g23.t1	636	G3811.t1	636	100	C1Q_05665 Proteína hipotética, <i>S. cerevisiae</i> JAY291, transportador iones, azúcares fosfatados, drogas.
g24.t1	377	G1407.t1	377	100	K7_Ehl2p, <i>S. cerevisiae</i> Kyokai no. 7
g25.t1	265	G3808.t1	1184	99.53	WN66_01991 Proteína hipotética, <i>S. cerevisiae</i>

g26.t1	249	G1409.t1	174	100	K7_11205p, <i>S. cerevisiae</i> Kyokai no. 7, función no asignada
g27.t1	103	G2658.t1	103	99.03	CENPK1137D_1957 Proteína hipotética, <i>S. cerevisiae</i> CEN.PK113-7D, no se identifica función.
g28.t1	131	G2269.t1, G1926.t1	131	100	P301_P10001 Proteína hipotética, <i>S. cerevisiae</i> P301, no se identifica función.
g29.t1	515	G1041.t1	523	97.71	Agp3p, <i>S. cerevisiae</i> S288c
g30.t1	172	G2652.t1	787	100	Vth1p, <i>S. cerevisiae</i> S288c
g31.t1	402	G2439.t1	410	97.07	MAL33 <i>S. cerevisiae</i> S288c

Para los genes identificados como G3808.t1 y G2437.t1, solo se encuentra la secuencia parcial en en las regiones no mapeadas en la cepa S288c en el gen g20.t1 y g19.t1, respectivamente. Estos resultados contrastados con el número de mutaciones encontradas en la familia Mal3 (123 mutaciones) puede justificarse ya que son genes que se consideran un complejo de multigenes (SGD, 2016-b) , por lo que concluir un posible nuevo gen de esta familia no es muy claro. Igualmente sucede con el gen G2436.t1 asociado a IMA1, donde solo se encuentra una secuencia parcial en el gen g18.t1.

En el conjunto de proteínas que no mapearon contra la cepa S288c, tenemos los genes G279.t1 y G3133.t1, reportados para la cepa RM11-1a. El gen G279.t1, tiene la superfamilia amidasa y para el gen G3133.t1 se identifica la familia acetiltransferasa (Bauer et al., 2015). Finalmente el gen identificado como G4861.t1, presente en cepas AWRI796_3907 y RM11-1a (SCRG_02226), tiene una región conservada perteneciente a la función amino permeasa (tirosina).

El gen G3706.t1 se encuentra también en las regiones que no mapearon con el genoma de referencia S288c y se identifica como g17.t1, al realizar Blastp, uno de los primeros resultados es el gen Yrm1, sin embargo este tan solo presenta un 34% de identidad, siendo el gen G1555.t1 quien presenta un 99% de identidad con el gen Yrm1. Dados estos resultados se podría pensar que este gen es exclusivo para la cepa *S. cerevisiae* 202-3, cuya secuencia presenta un dominio de unión de ADN GAL4-like Zn2Cys6 (Bauer et al., 2015), que no se encuentra en Yrm1 y si en factores transcripcionales como GAL4. Igualmente sucede con el gen G3707.t1, cuya secuencia par para las regiones que no mapearon es la g16.t1, esta secuencia presenta un 37% de identidad con el gen LALA0S15e0159g1_1 de la levadura *Lachancea thermotolerans*, este gen no se encuentra presente en ningún otra cepa de levadura *S. cerevisiae*. Según resultados obtenidos con Blastp, allí se identifica un dominio metiltransferasa y a la superfamilia AdoMet_MTases, metiltransferasas asociadas a la biosíntesis de menaquinona y ubiquinona (Bauer et al., 2015).

Otro de los genes considerados como candidatos a genes únicos en la cepa *S. cerevisiae* 202-3, es el gen G3708.t1, representado también en el gen g15.t1 de las regiones sin mapear, este gen tiene una identidad del 57% con el gen SEO1 presente en *S. cerevisiae*, gen que también se encuentra representado en G1928.t1 con un 100% de

identidad. Fue identificada la super familia MFS, de transportadores secundarios y la función de transportador de D-galactonato, actividad que puede ser asociada a más de un dominio y que no corresponde a ninguna familia o dominio (TIGR00893: 2A0114).

Para los genes G3709.t1 asociado con g14.t1 y G3132.t1 a g13.t1, g7.t1 y g1.t1, no son concluyentes pues son genes cuya función fue descrita con claridad para la caracterización de ORFs del ensamblaje *de novo*, además de las diferencias en los tamaños de cada proteína al comparar las secuencias asignadas. Igualmente para el gen G4.t1 asociado con g12.t1 y G278.t1 con g9.t1, aunque este último presente un alto porcentaje de identidad la longitud de la secuencia no corresponde a la misma. Esto puede deberse al error que está asociado a la predicción de genes con ensamblajes de genomas borradores (Denton et al., 2014).

Finalmente para el gen G4.t1, cuyo equivalente en el ensamblaje por mapeo corresponde al gen g6.t1, no se encuentra para ninguna otra cepa de *S. cerevisiae* y se asocia con una subunidad alfa de hemoglobina en humanos, realizar una afirmación respecto a esta proteína dado que su mayor porcentaje de identidad corresponde al de una proteína humana y teniendo en cuenta que durante los experimentos de secuenciación pueden existir casos de contaminaciones cruzadas de especies (Merchan et al., 2014), se recomienda afirmar la presencia de esta proteína mediante otros ensayos experimentales.

Al igual que para el proyecto de secuenciación de la cepa diploide Kyokai N°7, se destaca que a pesar de conocer la existencia de los genes MATa y MAT α , durante el proceso de ensamblaje y anotación solo logra identificarse el gen MAT α (Akao et al., 2011), posiblemente al igual que la cepa Kyokai N°7 se deba a la presencia de regiones heterocigotas.

3.4 Identificación de rutas metabólicas

En la figura 3-5, se presenta el resumen de los genes encontrados en la cepa 202-3, una vez analizadas las rutas metabólicas identificadas en KEGG (ver anexo 6), incluyendo los transportadores mencionados previamente. Se plantea que para la cepa 202-3, la ruta metabólica consiste en los mismos genes presentes para la cepa S288c, cuyo fenotipo respecto al consumo de xilosa se considera negativo. Durante la identificación de los genes asociados al metabolismo de xilosa no se evidenció el gen SOR2 dado su alto porcentaje de identidad con el gen SOR1, encontrando dos copias de este mismo gen. A pesar de encontrar los genes asociados a la cepa S288c, nuevamente cabe destacar las mutaciones encontradas para estos genes, por lo que no se debe descartar el efecto de estas en el fenotipo positivo frente al consumo de xilosa para la cepa 202-3.

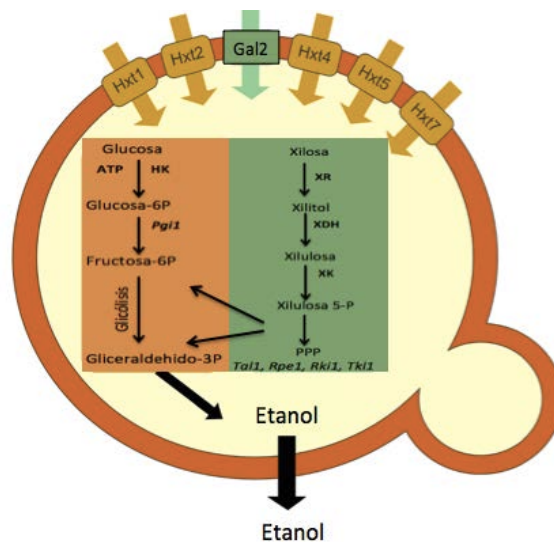


Figura 3-5. Esquema general de los genes identificados en el consumo de xilosa para la cepa 202-3, teniendo en cuenta los resultados del anexo 6. La imagen fue adaptada de Nijland et al., 2014.

Como parte de la discusión y del propósito de esta investigación, se destacan los genes mencionados en el apartado 3.3, los cuáles presentan el dominio de unión de ADN GAL4-like Zn2Cys6, para el gen G3810.t1, reportado también para la cepa JAY291 y el gen G3706.t1, candidato a gen exclusivo de la cepa *S. cerevisiae* 202-3. Estos genes se consideran candidatos en respuesta al consumo de xilosa para el aislado natural de *S. cerevisiae* 202-3, teniendo en cuenta los reportes que indican que los factores de transcripción asociados a la familia GAL tienen altos niveles de expresión ante la presencia de xilosa y el desarrollo de la patente WO 2012/143513 A2 (Klassen et al., 2012), donde mencionan la importancia de la mutación encontrada en el gen CEP3, gen que posee el mismo dominio que los genes reportados en este estudio.

Respecto a las mutaciones encontradas para los genes de interés, relacionados con xilosa, producción de etanol y genes de resistencia a inhibidores, se encuentra que aunque esta cepa tiene rendimientos destacados en la producción de etanol frente a otras cepas industriales como *Red fermentis*, solo se encuentra presente el gen SUC2 con una sola mutación, algunos autores reportan que si una cepa industrial posee más de un gen SUCX, su capacidad invertasa será mejor, destacando su rendimiento frente a azúcares como la sacarosa (Argueso et al., 2009), sin embargo cepas como la JAY291, empleada para la producción de bioetanol no posee dicho gen y destaca por su producción (Argueso et al., 2009). Igualmente el gen RTM1 que codifica una enzima transportadora de lípidos, que siempre se encuentra asociado con los genes SUCX, no fue encontrado, al igual que en la cepa JAY291 (Argueso et al., 2009).

Para la familia de genes FLOX (X=1,2...), se observan en el ensamblaje *de novo* los genes FLO1, FLO5 y FLO9, sin embargo cada uno de ellos se encuentra representado más de una vez en los genes predichos, debido a que muchos de estos genes son paralogos en S288c, por lo que su identificación como demuestran estudios previos puede ser difícil (Argueso et al., 2009). Sin embargo al contrastar los resultados con el ensamblaje con genoma de referencia se encuentra una mutación en el gen FLO8,

mutación sin sentido, que no afecta la funcionalidad del gen, indicando un carácter floculante de la cepa (Liu et al., 1996).

3.5 Caracterización de proteínas no identificadas

Como se mencionó anteriormente durante el análisis con el ensamblaje *de novo* y con genoma de referencia, se identificaron 4 posibles proteínas únicas para la cepa *S. cerevisiae* 202-3, dos de ellas asociadas con bajos % de identidad a genes de *S. cerevisiae* (G3708.t1 y G3706.t1) y dos genes identificados como pertenecientes a otras especies (G4.t1 y G3707.t1). Con el fin de predecir y comparar los resultados mencionados en el numeral 3.3.5 de la posible función de las proteínas que no fueron identificadas en la especie de *S. cerevisiae* G4.t1 y G3707.t1 se realiza una caracterización *in silico* mediante la predicción de la estructura de las proteínas estudiadas y alinamientos estructurales.

La predicción y alineamiento estructural (figura 3-6), concuerda con las funciones identificadas con la herramienta Blast, correspondiendo el gen G4.t1 a una subunidad de hemoglobina y el gen G3707.t1 con función metiltransferasa.

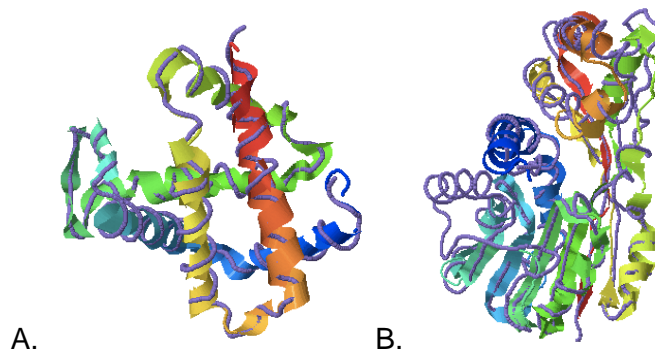


Figura 3-6. Alineamiento estructural para el gen A. G4.t1 con la subunidad de hemoglobina de *Homo sapiens*, con un TM-score de 0,992 y B. G3707.t1 con la metiltransferasa asociada a *Aspergillus fumigata*, con un TM-score de 0,910. Las proteínas modeladas para los genes desconocidos se representan con una línea más delgada en morado.

En conclusión con la función identificada para las dos proteínas que no fueron asignadas a la especie *S. cerevisiae* no logra identificarse un rol en el metabolismo de xilosa. Sin embargo los otros dos genes nuevos G3708.t1 y G3706.t1 y la presencia de la proteína G3810.t1 reportada para la levadura JAY291 sugieren una posible asignación como genes involucrados en el metabolismo de xilosa según los dominios encontrados.

La hipótesis que se plantea en esta investigación para justificar el fenotipo positivo frente al consumo de xilosa para la cepa 202-3, teniendo en cuenta los resultados obtenidos para las posibles nuevas proteínas, se menciona en la figura 3-7. Esta propuesta surge por la presencia de los dominios encontrados en los genes G3810.t1 y G3706.t1,

dominios asociados a genes como GAL4 y CEP3, que están relacionados al metabolismo de xilosa para algunas cepas de *S. cerevisiae* transformadas (Matsushika et al., 2014; Klassen et al., 2012) y que por lo tanto podrían tener mecanismos similares a los descritos en la figura 3-7. En el caso del gen G3708.t1 al tener asociado un dominio MFS y estar involucrado en fenómenos de transporte podrían asociarse a transporte de azúcares, pero dado que el dominio reportado corresponde a un dominio encontrado con gran frecuencia en varias proteínas no se menciona como gen hipotético en la figura 3-7.

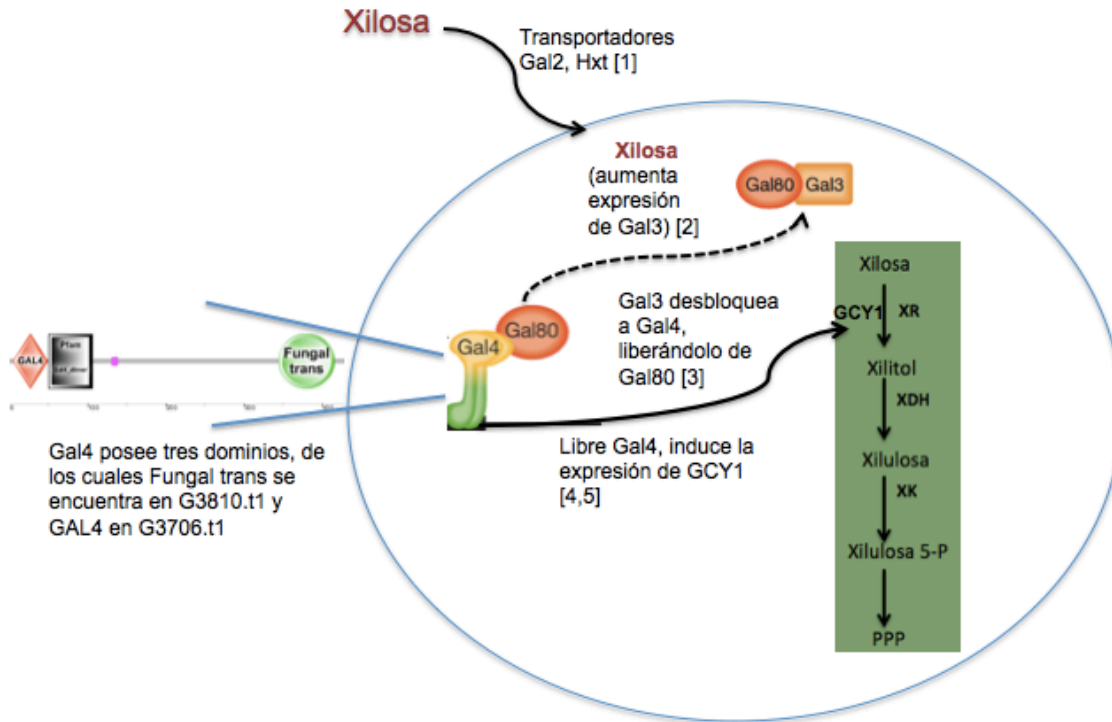


Figura 3-7. Diagrama propuesto a partir de datos en literatura del rol de GAL4 en el metabolismo de xilosa. La propuesta se hace en base a los dominios que se encuentran en los genes G3810.t1 y G3706.t1 no reportados para la cepa de referencia S288c, que podrían estar asociados al metabolismo de xilosa al igual que GAL4. A continuación se mencionan las citas empleadas para la estructuración del diagrama: [1] Nijland et al., 2014; [2] Matsushika et al., 2014; [3] Traven et al., 2006; [4] Ren et al., 2000 y [5] Angermayr et al., 2003.

4. Conclusiones y Recomendaciones

- El ensamblaje de la levadura *S. cerevisiae* 202-3 se llevó a cabo con el ensamblador ABySS con un kmer de 89, generando 146 *scaffolds* y una cobertura promedio de 100X. La anotación del genoma ensamblado con aproximación *de novo* y con la herramienta Augustus permitió identificar 5497 CDS, se identificaron 6 ARN ribosomales (5 nucleares y 1 mitocondrial), así como la presencia de 246 ARN de transferencia. De los genes identificados al menos el 97%, obtuvo un termino GO asociado.
- Con el fin de identificar y corroborar los genes nuevos y no reportados para la cepa de referencia S288c, se realizó un ensamblaje por mapeo el igual logró mapear el 97,7% del genoma de la cepa *S. cerevisiae* 202-3.
- Del porcentaje restante del genoma que no fue mapeado y junto con los resultados del ensamblaje *de novo* se identificaron 38 proteínas que no fueron identificadas en el genoma de referencia de la cepa S288c. De éstas 38 proteínas se proponen 4 proteínas exclusivas de la cepa *S. cerevisiae* 202-3.
- Como se evidenció anteriormente no se encuentra el gen XDH1, asociado al fenotipo positivo del consumo de xilosa para cepas de *S. cerevisiae*, sin embargo se propone como posibles candidatos asociados al consumo de xilosa de la cepa en estudio, la presencia de dos genes uno de ellos reportado previamente para la cepa de bioetanol JAY291. Estos genes fueron asociados a posibles factores de transcripción con dominios GAL4-like Zn2Cys6, presentes en factores de transcripción como GAL4 y CEP3, previamente relacionados al consumo de xilosa en *S. cerevisiae*.
- La identificación de mutaciones, supera lo reportado a otras cepas de *S. cerevisiae*, lo cual se justifica dado que la cepa estudiada no es haploide pues presenta los genes MATA y MAT α , ocasionando casos de genes heterocigotos en la cepa, como reportan estudios previos. Se recomienda verificar las mutaciones encontradas en el genoma.
- Como recomendaciones se sugiere implementar otras tecnologías de secuenciación que permitan obtener lecturas más largas, para poder obtener contigs mucho más largos y cercanos al genoma real de *S. cerevisiae*.
- Con el fin de validar los modelos génicos propuestos se recomiendan realizar ensayos de expresión diferencial, así como la caracterización específica y corroboración de la presencia de los posibles genes nuevos asociados a la cepa *S. cerevisiae* 202-3.
- Se recomienda un análisis exhaustivo del efecto de los genes que presentan los dominios GAL4 y Fungal trans, para el mejoramiento del consumo de xilosa y posterior fermentación de materiales lignocelulósicos, pues podrían estar asociados a la regulación del gen GCY1 (con actividad xilosa reductasa) como lo hace el factor de transcripción Gal4.

- Dados los esfuerzos que actualmente se encuentran centrados en el mejoramiento de los transportadores de azúcares en *S. cerevisiae*, se sugiere analizar con detenimiento y confirmar las mutaciones reportadas para dichos transportadores, pues estas mutaciones pueden incrementar o no la afinidad por el sustrato de xilosa.

Anexo 1: Consumo de xilosa en medios sintéticos, extracción de ADN y amplificación del gen XDH1

Consumo de Xilosa:

La caracterización del consumo de xilosa para las cepas de *S. cerevisiae* evaluadas se llevó a cabo en medios líquidos y sólidos. En el caso de los medios líquidos la cuantificación del consumo de xilosa se realizó en cromatografía líquida de alta resolución (HPLC), El programa para la cuantificación de azúcares por HPLC fue el siguiente: Columna SC1011, fase móvil agua desionizada, flujo 0,8mL/min, temperatura de detector 40°C, temperatura de horno 80°C, en el equipo Dionex Thermoscientific Ultimate® 3000 LC.

Las cepas empleadas para este estudio, fueron las cepas EC1118 y 202-3. La cepa EC1118 se empleó como control positivo para el consumo de xilosa, para descartar contaminación la cuantificación se realizó con seguimiento de los blancos (medios YNBx sin levadura).

En la figura 1, se presenta el crecimiento de la cepa 202-3 en medio sólido.

Tabla 1. Composición de los medios empleados para caracterizar el fenotipo de la cepa *S. cerevisiae* 202-3.

*Solo para los medios sólidos.

Medio Líquido (YNBx)	
Componente	g/L
YNB (Yeast Nitrogen Base)	1,67
Sulfato de amonio	5
Glucosa	30
Xilosa	30
Agar*	15

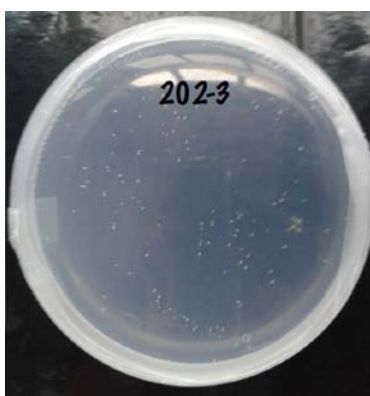


Figura 1. Crecimiento de la levadura 202-3 en medio sólido.

Extracción de ADN:

La extracción de ADN se realizó enzimáticamente, según recomendaciones del fabricante del kit YeaStar columns (Zymo Research). Los pasos llevados a cabo fueron los siguientes:

1. Con 1mL de células a una concentración de $1-5 \times 10^7$ células, se centrifugaron a 500 g por 2 minutos.
2. Una vez removido el sobrenadante se adicionó 120 μ L de la solución YD Digestion Buffer y 5 μ L la enzima R-Zymolyase TM (RNAsa A + ZymolyaseTM). El pellet se resuspendió por vórtex y se incubó a 37°C por 60 minutos.
3. A la mezcla del punto anterior se adicionó 120 μ L de la solución YD Lysis Buffer. Una vez centrifugado el sobrenadante se pasó por la columna Zymo-spin III y se centrifugó a 10000 rpm por 1 minuto. Con el fin de eliminar contaminantes se realizaron dos lavados con la solución DNA Wash Buffer del kit.
4. Se realizó finalmente el lavado de la columna para retirar el ADN con 60 μ L de agua.

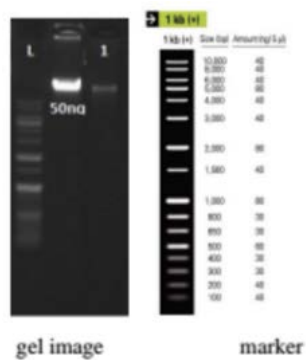
Amplificación del gen XDH1:

La amplificación del gen XDH1, se realizó con los siguientes primers:

GSP44	ATGGCTCCAATCGAAAATCC	XDH1 colony PCR forward
GSP45	TCAGGCCCAAAAATAATGGT	XDH1 colony PCR reverse

Anexo 2: Análisis de calidad del ADN sometido a secuenciación

Reporte para evaluar la calidad de ADN.



#	Library Name	Library Type	Conc. (ng/ul)	Conc. (nM)	Size (bp)	Result*
1	202-3	TruSeq Nano DNA (350)	35.45	116.05	470	Pass

Referente a las secuencias obtenidas

No	Sample	TotalBases	ReadCount	GC(%)	AT(%)	Q20(%)	Q30(%)
1	202-3	2,411,583,666	23,877,066	38.27	61.74	94.63	87.84

Anexo 3: Métricas de los ensamblajes de novo.

ABYSS

		Show heatmap										
		Worst Median Best										
Statistics without reference		31	41	51	61	71	81	91				
# contigs	415	305	252	224	214	207	186					
# contigs (>= 0 bp)	4646	2913	2128	1455	956	533	304					
# contigs (>= 1000 bp)	365	269	231	199	190	178	166					
# contigs (>= 5000 bp)	261	187	153	133	127	113	106					
# contigs (>= 10000 bp)	224	159	131	109	101	91	86					
# contigs (>= 25000 bp)	153	118	101	85	77	71	67					
# contigs (>= 50000 bp)	86	73	69	62	55	54	52					
Largest contig	212 922	364 820	478 838	546 087	531 075	547 314	548 470					
Total length	11 569 882	11 656 757	11 699 176	11 723 756	11 821 798	11 847 424	11 867 618					
Total length (>= 0 bp)	11 862 367	11 885 239	11 903 325	11 881 308	11 929 006	11 901 035	11 888 578					
Total length (>= 1000 bp)	11 531 644	11 629 539	11 682 774	11 705 457	11 804 673	11 827 113	11 852 720					
Total length (>= 5000 bp)	11 272 763	11 421 498	11 489 562	11 536 835	11 652 897	11 677 319	11 705 059					
Total length (>= 10000 bp)	11 011 770	11 230 039	11 331 321	11 370 110	11 472 063	11 524 117	11 561 466					
Total length (>= 25000 bp)	9 814 442	10 544 923	10 830 150	10 997 831	11 090 666	11 214 906	11 235 390					
Total length (>= 50000 bp)	7 413 473	8 897 132	9 680 180	10 161 479	10 291 225	10 603 652	10 698 896					
N50	68 447	112 431	158 197	214 345	222 880	234 146	253 877					
N75	36 530	52 608	67 061	82 130	101 858	120 193	135 739					
L50	59	32	24	20	19	18	17					
L75	117	70	54	42	37	34	32					
GC (%)	38.11	38.12	38.11	38.12	38.12	38.14	38.16					
Mismatches												
# N's	6615	4704	2485	2379	3350	3792	9269					
# N's per 100 kbp	57.17	40.35	21.24	20.29	28.34	32.01	78.1					
		73	75	77	79	81	83	85	87	89	91	93
212	209	199	190	207	203	194	192	184	186	186	205	
898	804	710	597	533	465	396	358	325	304	304	303	
189	185	176	172	178	180	173	173	163	166	166	185	
127	123	116	113	113	112	108	105	102	106	106	115	
102	99	98	92	91	91	88	89	87	86	86	95	
77	76	78	72	71	70	68	67	69	67	67	75	
59	58	60	54	54	53	53	51	52	52	52	60	
531 093	546 501	546 172	547 312	547 314	547 330	548 231	548 239	548 475	548 470	548 470	548 575	
11 810 228	11 835 864	11 973 337	11 837 957	11 847 424	11 878 214	11 854 910	11 851 818	11 856 708	11 867 618	11 867 618	11 876 766	
11 911 260	11 926 401	12 052 353	11 902 456	11 901 035	11 922 023	11 889 959	11 880 250	11 881 145	11 888 578	11 888 578	11 894 943	
11 793 602	11 818 638	11 956 862	11 825 177	11 827 113	11 862 069	11 839 497	11 838 385	11 841 228	11 852 720	11 852 720	11 861 970	
11 663 308	11 680 189	11 817 288	11 693 325	11 677 319	11 703 473	11 685 823	11 676 597	11 686 537	11 705 059	11 705 059	11 692 825	
11 496 534	11 520 793	11 692 834	11 553 653	11 524 117	11 555 948	11 547 057	11 563 501	11 574 796	11 561 466	11 561 466	11 553 485	
11 107 880	11 145 212	11 383 831	11 236 324	11 214 906	11 218 727	11 221 569	11 203 863	11 265 364	11 235 390	11 235 390	11 224 973	
10 445 520	10 508 803	10 745 007	10 603 814	10 603 652	10 607 206	10 674 044	10 624 137	10 667 807	10 698 896	10 698 896	10 663 957	
214 406	216 797	222 892	241 804	234 146	248 048	253 805	256 391	248 043	253 877	253 877	241 844	
110 209	110 210	110 211	114 811	120 193	120 048	120 269	132 818	132 818	135 739	135 739	110 203	
20	19	19	18	18	17	17	17	17	17	17	20	
38	37	37	34	34	33	33	32	33	33	33	38	
38.13	38.13	38.15	38.14	38.14	38.14	38.14	38.15	38.15	38.15	38.16	38.17	
3419	3264	3297	3612	3792	3678	4425	5428	6883	9269	9269	19 011	
28.95	27.58	27.54	30.51	32.01	30.96	37.33	45.8	58.05	78.1	78.1	160.07	

Anexo 4: ARN de transferencia en *S. cerevisiae* 202-3

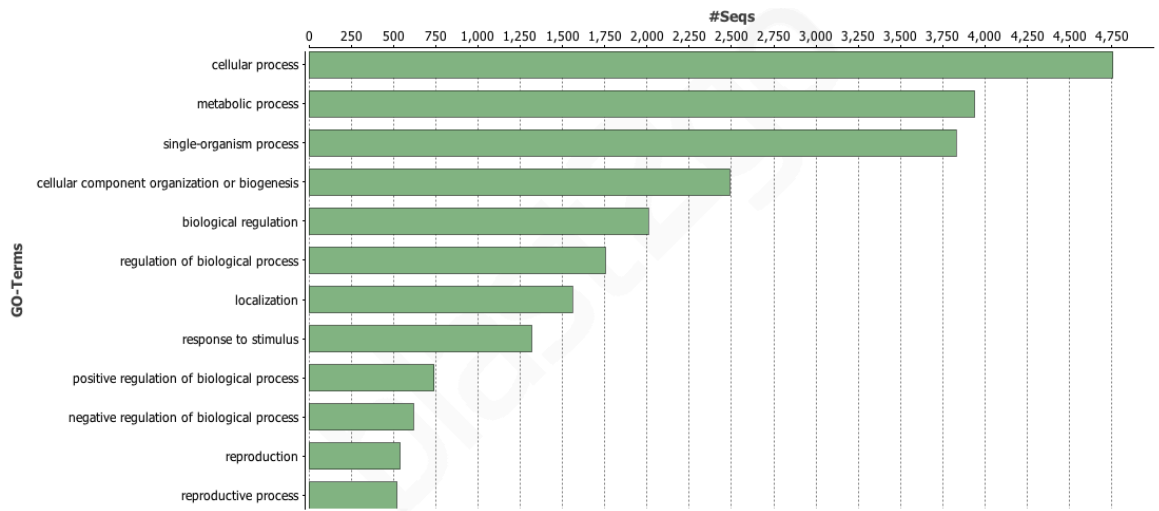
Número de ARNs de transferencia identificados en el genoma de la cepa *S. cerevisiae* 202-3.

Anticodón tRNA	Frecuencia	Anticodón tRNA	Frecuencia	Anticodón tRNA	Frecuencia	Anticodón tRNA	Frecuencia
AAA Phe	--	GAA Phe	1	CAA Leu	--	TAA Leu	7
AGA Ser	8	GGA Ser	--	CGA Ser	--	TGA Ser	3
ACA Cys	--	GCA Cys	4	CCA Trp	--	TCA seC	--
ATA Tyr	--	GTA Tyr	7	CTA Pyl	--	TTA Stop	--
AAG Leu	--	GAG Leu	1	CAG Leu	--	TAG Leu	--
AGG Pro	2	GGG Pro	--	CGG Pro	--	TGG Pro	--
ACG Arg	6	GCG Arg	--	CCG Arg	1	TCG Arg	--
ATG His	--	GTG His	6	CTG Gln	1	TTG Gln	10
AAC Val	13	GAC Val	--	CAC Val	1	TAC Val	3
AGC Ala	9	GGC Ala	--	CGC Ala	--	TGC Ala	5
ACC Gly	--	GCC Gly	13	CCC Gly	1	TCC Gly	3
ATC Asp	--	GTC Asp	16	CTC Glu	2	TTC Glu	9
AAT Ile	10	GAT Ile	1	CAT Met	10	TAT Ile	--
AGT Thr	6	GGT Thr	--	CGT Thr	1	TGT Thr	3
ACT Ser	--	GCT Ser	1	CCT Arg	1	TCT Arg	11
ATT Asn	--	GTT Asn	9	CTT Lys	12	TTT Lys	1
Anticodón tRNA	Frecuencia	Anticodón tRNA	Frecuencia	Anticodón tRNA	Frecuencia	Anticodón tRNA	Frecuencia
TTT Phe	--	TTC Phe	1	TTG Leu	--	TTA Leu	7
TCT Ser	8	TCC Ser	--	TCG Ser	--	TCA Ser	3
TGT Cys	--	TGC Cys	4	TGG Trp	--	TGA seC	--
TAT Tyr	--	TAC Tyr	7	TAG Pyl	--	TAA Stop	--
CTT Leu	--	CTC Leu	1	CTG Leu	--	CTA Leu	--
CCT Pro	1	CCC Pro	--	CCG Pro	--	CCA Pro	--
CGT Arg	7	CGC Arg	--	CGG Arg	1	CGA Arg	--
CAT His	--	CAC His	6	CAG Gln	1	CAA Gln	10
GTT Val	13	GTC Val	--	GTG Val	1	GTA Val	3
GCT Ala	9	GCC Ala	--	GCG Ala	--	GCA Ala	5
GGT Gly	--	GGC Gly	13	GGG Gly	1	GGA Gly	3
GAT Asp	--	GAC Asp	16	GAG Glu	2	GAA Glu	9
ATT Ile	10	ATC Ile	1	ATG Met	10	ATA Ile	--

ACT Thr	6	ACC Thr	--	ACG Thr	1	ACA Thr	3
AGT Ser	--	AGC Ser	1	AGG Arg	1	AGA Arg	11
AAT Asn	--	AAC Asn	9	AAG Lys	12	AAA Lys	1
Total tRNA	246						
Total tmRNA	0						

Anexo 5: Reportes obtenidos durante la anotación con términos GO

A.



B.

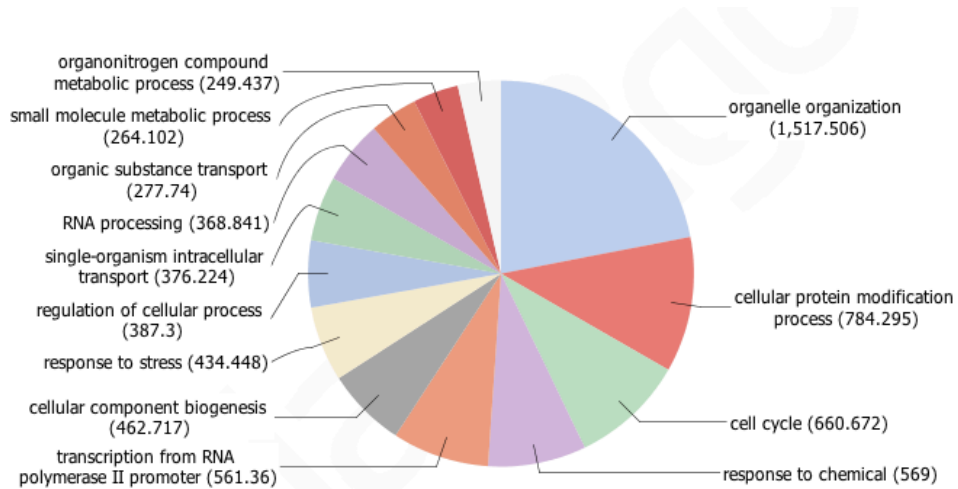
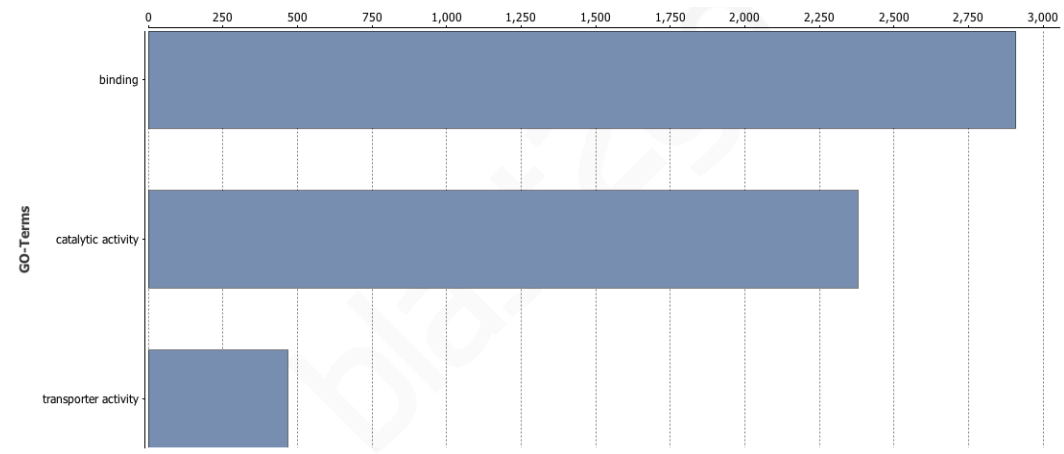


Figura 1. A. Distribución de términos GO, nivel 2, para procesos biológicos y **B.** Distribución del puntaje para las anotaciones realizadas en los procesos biológicos.

A.



B.

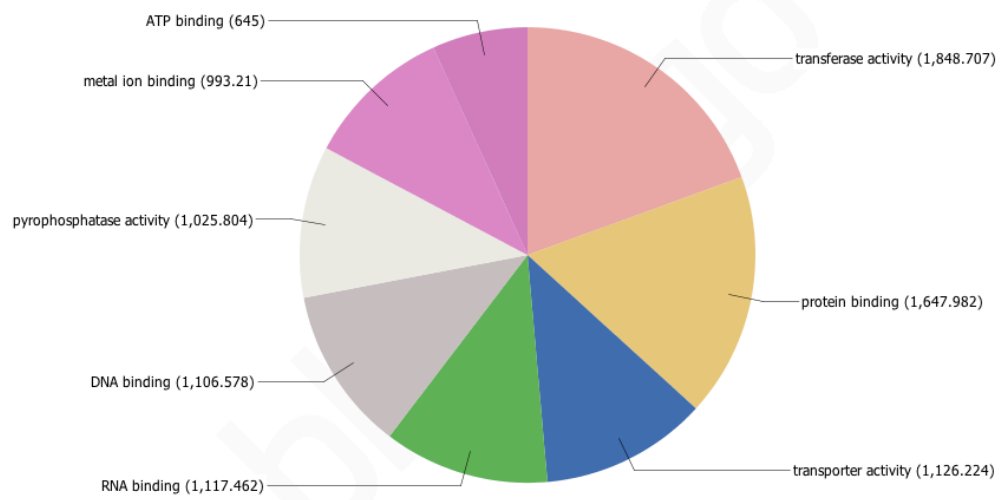


Figura 2. A. Distribución de términos GO, nivel 2, para funciones moleculares. **B.** Distribución del puntaje para las anotaciones realizadas en las funciones moleculares.

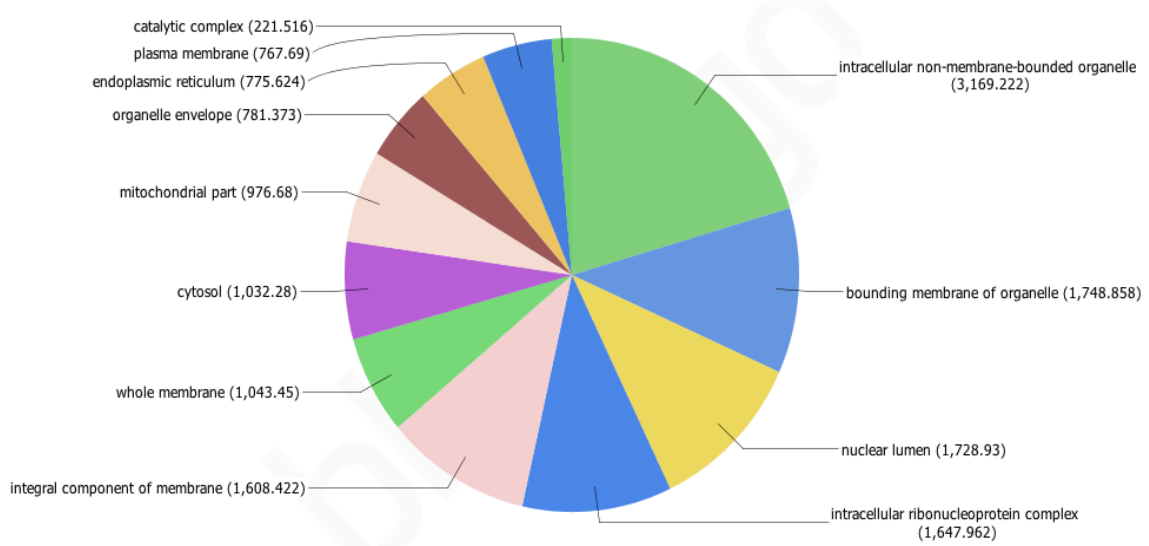
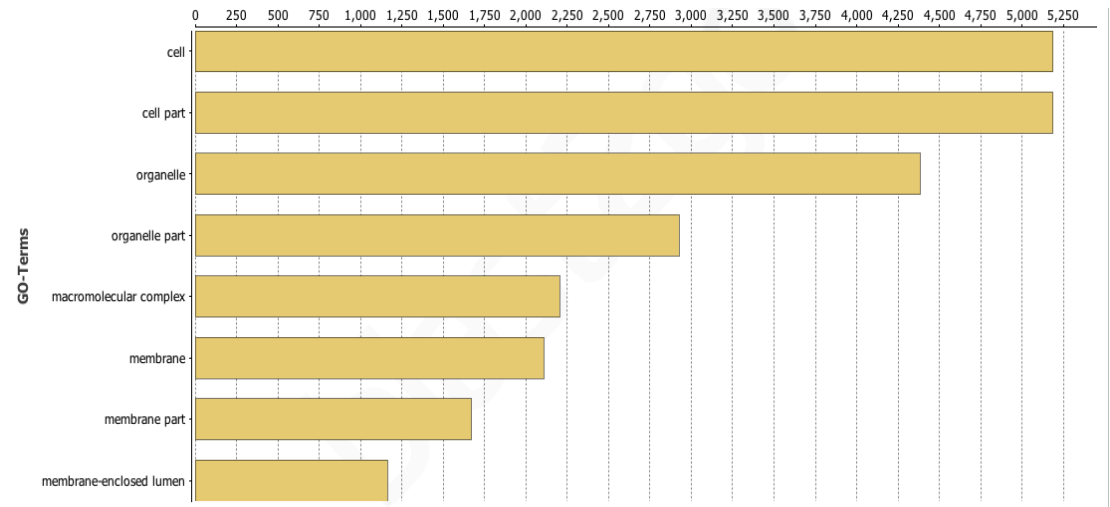


Figura 3. A. Distribución de términos GO, nivel 2, para componentes celulares. **B.** Distribución del puntaje para las anotaciones realizadas en los componentes celulares.



Figura 4. Distribución de secuencias en cada ruta.

Anexo 6: Asociación de genes identificados en ensamblaje *de novo* a rutas metabólicas.

Pathway	Seqs in Pathway	Enzyme	Ezyme ID	Seqs of Enzyme	Seqs
alpha-Linolenic	11	A2	ec:3.1.1.4	1	G5255.t1
		dehydrogenase	ec:1.1.1.1	6	G4043.t1, G152.t1, G2799.t1, G2935.t1, G3832.t1, G689.t1
		oxidase	ec:1.3.3.6	1	G196.t1
		C-acyltransferase	ec:2.3.1.16	1	G5221.t1
		hydratase	ec:4.2.1.17	2	G3670.t1, G2223.t1
Biosynthesis of	10	reductase	ec:1.1.1.100	2	G1978.t1, G2163.t1
		reductase (NADPH)	ec:1.3.1.38	2	G248.t1, G4193.t1
		oxidase	ec:1.3.3.6	1	G196.t1
		C-acyltransferase	ec:2.3.1.16	1	G5221.t1
		hydrolase	ec:3.1.2.2	1	G3573.t1
		hydratase	ec:4.2.1.17	2	G3670.t1, G2223.t1
		9-desaturase	ec:1.14.19.1	1	G4187.t1
Linoleic acid m	1	A2	ec:3.1.1.4	1	G5255.t1
Arachidonic ac	16	reductase (NADPH)	ec:1.1.1.184	6	G1902.t1, G236.t1, G1883.t1, G3904.t1, G3115.t1, G5388.t1
		A2	ec:3.1.1.4	1	G5255.t1
		peroxidase	ec:1.11.1.9	7	G1859.t1, G5044.t1, G2188.t1, G1501.t1, G5043.t1, G2728.t1, G863.t1
		epoxide hydrolase	ec:3.3.2.10	2	G3392.t1, G4449.t1
		hydrolase	ec:3.3.2.6	1	G4449.t1
Taurine and hy	5	decarboxylase	ec:4.1.1.15	1	G3882.t1
		dehydrogenase	ec:1.4.1.2	1	G793.t1
		dioxygenase	ec:1.14.11.17	1	G4870.t1
		glutamyl transpeptidase	ec:2.3.2.2	2	G1186.t1, G2692.t1
Sulfur metaboli	19	kinase	ec:2.7.1.25	1	G2214.t1

		reductase (thioredoxin)	ec:1.8.4.8	1	G1383.t1
		sulfite reductase (NADPH)	ec:1.8.1.2	2	G3732.t1, G3469.t1
		nucleotidase	ec:3.1.3.7	1	G2955.t1
		sulfurtransferase	ec:2.8.1.1	5	G1654.t1, G2504.t1, G1624.t1, G2594.t1, G1655.t1
		dioxygenase	ec:1.14.11.17	1	G4870.t1
		adenylyltransferase	ec:2.7.7.4	1	G3581.t1
		adenylyltransferase (ADP)	ec:2.7.7.5	2	G1874.t1, G876.t1
		gamma-synthase	ec:2.5.1.48	3	G3477.t1, G106.t1, G4869.t1
		synthase	ec:2.5.1.47	2	G5343.t1, G1183.t1
Steroid degradation	2	dehydrogenase	ec:1.1.1.145	2	G5354.t1, G2876.t1
Drug metabolism	39	kinase	ec:2.7.1.48	3	G2592.t1, G4599.t1, G3444.t1
		deaminase	ec:3.5.4.5	1	G1229.t1
		ali-esterase	ec:3.1.1.1	11	G3644.t1, G5325.t1, G3573.t1, G3920.t1, G3392.t1, G1650.t1, G4913.t1, G2538.t1, G4134.t1, G2112.t1, G3908.t1
		dehydrogenase (NADP+)	ec:1.3.1.2	1	G909.t1
		synthase (glutamine-hydrolysing)	ec:6.3.5.2	6	G896.t1, G4031.t1, G894.t1, G3913.t1, G26.t1, G1410.t1
		diphosphatase	ec:3.6.1.19	8	G3454.t1, G608.t1, G609.t1, G3334.t1, G2655.t1, G4344.t1, G2390.t1, G818.t1
		dehydrogenase	ec:1.1.1.205	5	G1069.t1, G133.t1, G2654.t1, G269.t1, G2653.t1
		phosphoribosyl transferase	ec:2.4.2.10	2	G3863.t1, G86.t1
		phosphoribosyl transferase	ec:2.4.2.8	2	G3474.t1, G1761.t1

		phosphoribosyl transferase	ec:2.4.2.8	2	G3474.t1, G1761.t1
Drug metabolism	22	transferase	ec:2.5.1.18	11	G4868.t1, G2637.t1, G3881.t1, G5459.t1, G1859.t1, G2596.t1, G1501.t1, G5043.t1, G1026.t1, G863.t1, G446.t1
		dehydrogenase	ec:1.1.1.1	6	G4043.t1, G152.t1, G2799.t1, G2935.t1, G3832.t1, G689.t1
		dehydrogenase [NAD(P)+]	ec:1.2.1.5	4	G4017.t1, G1733.t1, G434.t1, G3245.t1
		monooxygenase	ec:1.14.13.8	1	G266.t1
Carbapenem b	3	dehydrogenase	ec:1.2.1.41	1	G1689.t1
		5-kinase	ec:2.7.2.11	2	G4846.t1, G2095.t1
Metabolism of z	27	reductase (NADPH)	ec:1.1.1.184	6	G1902.t1, G236.t1, G1883.t1, G3904.t1, G3115.t1, G5388.t1
		transferase	ec:2.5.1.18	11	G4868.t1, G2637.t1, G3881.t1, G5459.t1, G1859.t1, G2596.t1, G1501.t1, G5043.t1, G1026.t1, G863.t1, G446.t1
		dehydrogenase	ec:1.1.1.1	6	G4043.t1, G152.t1, G2799.t1, G2935.t1, G3832.t1, G689.t1
		dehydrogenase [NAD(P)+]	ec:1.2.1.5	4	G4017.t1, G1733.t1, G434.t1, G3245.t1
Arginine and pr	34	aminotransferase	ec:2.6.1.13	1	G1064.t1
		decarboxylase	ec:4.1.1.50	1	G2967.t1
		deaminase	ec:3.5.4.1	1	G543.t1
		dehydrogenase	ec:1.2.1.41	1	G1689.t1

		synthase	ec:2.5.1.16	1	G549.t1
		synthase	ec:2.5.1.22	1	G4889.t1
		decarboxylase	ec:4.1.1.17	1	G938.t1
		transaminase	ec:2.6.1.1	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		dehydrogenase (NAD+)	ec:1.2.1.3	15	G1043.t1, G236.t1, G19.t1, G2101.t1, G2099.t1, G4017.t1, G4529.t1, G1733.t1, G3458.t1, G2019.t1, G434.t1, G3245.t1, G20.t1, G4211.t1, G5388.t1
		reductase	ec:1.5.1.2	1	G3298.t1
		acylamidase	ec:3.5.1.4	1	G4796.t1
		arginine amidinase	ec:3.5.3.1	1	G309.t1
		5-kinase	ec:2.7.2.11	2	G4846.t1, G2095.t1
		polyamine oxidase	ec:1.5.3.17	1	G4097.t1
		oxidase	ec:1.5.3.16	1	G4097.t1
		N- acetyltransfera se	ec:2.3.1.57	1	G4643.t1
Aminobenzoate	25	phosphatase	ec:3.1.3.2	10	G3454.t1, G608.t1, G609.t1, G1088.t1, G553.t1, G2655.t1, G3185.t1, G4624.t1, G4344.t1, G2390.t1
		phosphatase	ec:3.1.3.1	2	G1832.t1, G33.t1
		acylamidase	ec:3.5.1.4	1	G4796.t1
		nitrophenyl phosphatase	ec:3.1.3.41	8	G1652.t1, G4507.t1, G3105.t1, G1940.t1, G1511.t1, G4040.t1, G4197.t1, G986.t1
		acetonitrilase	ec:3.5.5.1	1	G5224.t1
		CoA- transferase	ec:2.8.3.8	1	G4230.t1
		hydratase	ec:4.2.1.17	2	G3670.t1, G2223.t1
Purine metabol	188	cyclo-ligase	ec:6.3.3.1	1	G171.t1

		kinase	ec:2.7.4.8	1	G1808.t1
		kinase	ec:2.7.4.6	1	G2152.t1
		kinase	ec:2.7.4.3	6	G3906.t1, G4782.t1, G1483.t1, G2801.t1, G2190.t1, G3152.t1
		ec:3.5.2.5 allantoinase	ec:3.5.2.5	1	G5054.t1
		carboxylase	ec:4.1.1.21	2	G3122.t1, G2333.t1
		lyase	ec:4.3.2.2	1	G1135.t1
		lyase	ec:4.3.2.3	1	G5049.t1
		deaminase	ec:3.5.4.2	1	G4535.t1
		deaminase	ec:3.5.4.3	1	G31.t1
		kinase	ec:2.7.1.40	2	G1319.t1, G1708.t1
		deaminase	ec:3.5.4.6	3	G4146.t1, G2689.t1, G3646.t1
		kinase	ec:2.7.1.25	1	G2214.t1
		kinase	ec:2.7.1.20	1	G3498.t1
		diphosphokina se	ec:2.7.6.1	5	G563.t1, G2958.t1, G941.t1, G4277.t1, G3218.t1
		cyclase	ec:4.6.1.1	1	G3594.t1
		adenylyltransfe rase	ec:2.7.7.53	2	G1874.t1, G876.t1
		synthase (glutamine- hydrolysing)	ec:6.3.5.2	6	G896.t1, G4031.t1, G894.t1, G3913.t1, G26.t1, G1410.t1
		synthase	ec:6.3.5.3	1	G5299.t1
		formyltransfera se	ec:2.1.2.3	2	G4007.t1, G4995.t1
		formyltransfera se	ec:2.1.2.2	1	G1769.t1
		synthase	ec:6.3.4.4	1	G1508.t1

		adenylpyrophosphatase	ec:3.6.1.3	74	G5042.t1, G169.t1, G2340.t1, G4038.t1, G39.t1, G358.t1, G3724.t1, G3731.t1, G1483.t1, G2129.t1, G2318.t1, G3529.t1, G803.t1, G3823.t1, G3957.t1, G3305.t1, G2981.t1, G2801.t1, G4923.t1, G1973.t1, G4482.t1, G1519.t1, G416.t1, G4236.t1, G1895.t1, G4811.t1, G2410.t1, G3753.t1, G1352.t1, G3214.t1, G1098.t1, G4934.t1, G123.t1, G349.t1, G4158.t1, G3650.t1, G2416.t1, G4590.t1, G4074.t1, G4989.t1, G1413.t1, G1937.t1, G872.t1, G2289.t1, G2779.t1, G2349.t1, G2323.t1, G1225.t1, G3613.t1, G1059.t1, G4348.t1, G3785.t1, G300.t1, G4495.t1, G2727.t1, G621.t1, G3766.t1, G3090.t1
		diphosphate phosphatase	ec:3.6.1.6	2	G3358.t1, G3318.t1
		diphosphatase	ec:3.6.1.9	2	G3334.t1, G818.t1
		ec:3.5.3.4 allantoicase	ec:3.5.3.4	1	G5052.t1
		reductase	ec:1.17.4.1	4	G3248.t1, G3608.t1, G5136.t1, G2613.t1
		synthase	ec:6.3.2.6	2	G3122.t1, G2333.t1
		metaphosphatase	ec:3.6.1.11	2	G1806.t1, G2143.t1
		diphosphatase	ec:3.6.1.13	1	G592.t1
		phosphatase	ec:3.6.1.15	9	G3454.t1, G608.t1, G609.t1, G3334.t1, G3318.t1, G2655.t1, G4344.t1, G2390.t1, G818.t1
		diphosphatase	ec:3.6.1.19	8	G3454.t1, G608.t1, G609.t1, G3334.t1, G2655.t1, G4344.t1, G2390.t1, G818.t1
		adenyllyltransferase	ec:2.7.7.4	1	G3581.t1
		RNA polymerase	ec:2.7.7.6	2	G2273.t1, G1765.t1

		DNA polymerase	ec:2.7.7.7	38	G2359.t1, G4502.t1, G148.t1, G4815.t1, G3557.t1, G2361.t1, G1695.t1, G788.t1, G4688.t1, G3370.t1, G1925.t1, G1376.t1, G1707.t1, G275.t1, G4151.t1, G828.t1, G2054.t1, G1778.t1, G4250.t1, G1472.t1, G4409.t1, G2146.t1, G2127.t1, G147.t1, G2360.t1, G3585.t1, G1520.t1, G5491.t1, G2695.t1, G5200.t1, G787.t1, G4150.t1, G2020.t1, G60.t1, G1480.t1, G3.t1, G1408.t1, G360.t1
		ligase	ec:6.3.4.13	1	G171.t1
		nucleosidase	ec:3.2.2.2	1	G1258.t1
		nucleosidase	ec:3.2.2.1	1	G1762.t1
		dehydrogenase	ec:1.1.1.205	5	G1069.t1, G133.t1, G2654.t1, G269.t1, G2653.t1
		phosphoribosyl transferase	ec:2.4.2.22	1	G3474.t1
		(alpha-D-glucose-1,6-bisphosphate-dependent)	ec:5.4.2.2	3	G3856.t1, G4023.t1, G987.t1
		phosphodeoxyribomutase	ec:5.4.2.7	1	G3856.t1
		phosphoribosyl diphosphate 5-amidotransferase	ec:2.4.2.14	1	G3835.t1
		cyclohydrolase	ec:3.5.4.10	2	G4007.t1, G4995.t1
		phosphoribosyl transferase	ec:2.4.2.7	2	G1797.t1, G4134.t1
		phosphoribosyl transferase	ec:2.4.2.8	2	G3474.t1, G1761.t1
		kinase	ec:2.7.4.10	1	G3152.t1
		phosphorylase	ec:2.4.2.1	1	G1258.t1

Naphthalene de	6	dehydrogenase	ec:1.1.1.1	6	G4043.t1, G152.t1, G2799.t1, G2935.t1, G3832.t1, G689.t1
Carbon fixation	33	carboxylase	ec:6.4.1.1	2	G4181.t1, G2751.t1
		carboxylase	ec:6.4.1.2	2	G3440.t1, G3923.t1
		dehydrogenase	ec:1.3.5.1	6	G3622.t1, G2355.t1, G4009.t1, G975.t1, G4740.t1, G969.t1
		cyclohydrolase	ec:3.5.4.9	3	G617.t1, G5463.t1, G2593.t1
		ligase (ADP- forming)	ec:6.2.1.5	2	G2558.t1, G1528.t1
		ligase	ec:6.2.1.1	2	G1305.t1, G4882.t1
		C- acetyltransfera se	ec:2.3.1.9	1	G464.t1
		reductase (NADH)	ec:1.3.1.6	2	G3549.t1, G3362.t1
		dehydrogenase (NADP+)	ec:1.5.1.5	3	G617.t1, G5463.t1, G2593.t1
		ligase	ec:6.3.4.3	2	G617.t1, G2593.t1
		reductase [NAD(P)H]	ec:1.5.1.20	2	G2045.t1, G468.t1
		hydratase	ec:4.2.1.3	2	G2674.t1, G1182.t1
		hydratase	ec:4.2.1.2	1	G2007.t1
		hydratase	ec:4.2.1.17	2	G3670.t1, G2223.t1
		dehydrogenase	ec:1.1.1.37	3	G2899.t1, G4386.t1, G1024.t1
		dehydrogenase	ec:1.1.1.35	1	G2223.t1
		dehydrogenase (NADP+)	ec:1.1.1.42	3	G4377.t1, G1941.t1, G1289.t1
Lipoic acid met	2	transferase	ec:2.3.1.181	1	G1234.t1
		synthase	ec:2.8.1.8	1	G1577.t1

Chloroalkane a	23	dehydrogenase	ec:1.1.1.1	6	G4043.t1, G152.t1, G2799.t1, G2935.t1, G3832.t1, G689.t1
		dehydrogenase (NAD+)	ec:1.2.1.3	15	G1043.t1, G236.t1, G19.t1, G2101.t1, G2099.t1, G4017.t1, G4529.t1, G1733.t1, G3458.t1, G2019.t1, G434.t1, G3245.t1, G20.t1, G4211.t1, G5388.t1
		epoxide hydrolase	ec:3.3.2.10	2	G3392.t1, G4449.t1
Toluene degrad	9	maleylacetate enol-lactonase	ec:3.1.1.45	1	G4395.t1
		dehydrogenase	ec:1.1.1.90	7	G41.t1, G2396.t1, G28.t1, G410.t1, G5489.t1, G4314.t1, G1411.t1
		dehydrogenase	ec:1.1.1.35	1	G2223.t1
Xylene degrada	7	dehydrogenase	ec:1.1.1.90	7	G41.t1, G2396.t1, G28.t1, G410.t1, G5489.t1, G4314.t1, G1411.t1
Valine, leucine	14	synthase	ec:2.2.1.6	2	G4019.t1, G843.t1
		transaminase	ec:2.6.1.42	2	G54.t1, G48.t1
		ammonia-lyase	ec:4.3.1.19	4	G907.t1, G3232.t1, G887.t1, G5226.t1
		synthase	ec:2.3.3.13	2	G3103.t1, G4504.t1
		dehydratase	ec:4.2.1.9	1	G3575.t1
		dehydratase	ec:4.2.1.33	1	G5361.t1
		reductoisomerase (NADP+)	ec:1.1.1.86	1	G1138.t1
		dehydrogenase	ec:1.1.1.85	1	G849.t1
Ubiquinone and	8	polyprenyltransferase	ec:2.5.1.39	1	G3414.t1
		dehydrogenase (quinone)	ec:1.6.5.2	1	G5011.t1
		3-O-methyltransferase	ec:2.1.1.64	1	G2924.t1
		transaminase	ec:2.6.1.5	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		methyltransferase	ec:2.1.1.114	1	G2924.t1
Biotin metabolis	11	synthase	ec:6.3.3.3	2	G3398.t1, G3399.t1
		transaminase	ec:2.6.1.62	1	G3398.t1
		reductase	ec:1.1.1.100	2	G1978.t1, G2163.t1
		dehydratase	ec:4.2.1.59	2	G940.t1, G2546.t1

		reductase (NADPH, Si-specific)	ec:1.3.1.10	3	G940.t1, G248.t1, G4193.t1
		ligase	ec:6.3.4.9	1	G2822.t1
		synthase	ec:2.8.1.6	1	G2432.t1
		(ATP-hydrolysing)] ligase	ec:6.3.4.10	1	G2822.t1
		ligase	ec:6.3.4.11	1	G2822.t1
		ligase	ec:6.3.4.15	1	G2822.t1
		synthase I	ec:2.3.1.41	2	G1978.t1, G3257.t1
Pyruvate metal	59	carboxylase	ec:6.4.1.1	2	G4181.t1, G2751.t1
		carboxylase	ec:6.4.1.2	2	G3440.t1, G3923.t1
		carboxykinase (ATP)	ec:4.1.1.49	1	G5481.t1
		kinase	ec:2.7.1.40	2	G1319.t1, G1708.t1
		ligase	ec:6.2.1.1	2	G1305.t1, G4882.t1
		dehydrogenase	ec:1.8.1.4	2	G3776.t1, G474.t1
		dehydrogenase (NAD ⁺)	ec:1.2.1.3	15	G1043.t1, G236.t1, G19.t1, G2101.t1, G2099.t1, G4017.t1, G4529.t1, G1733.t1, G3458.t1, G2019.t1, G434.t1, G3245.t1, G20.t1, G4211.t1, G5388.t1
		C- acetyltransferase	ec:2.3.1.9	1	G464.t1
		lyase	ec:4.4.1.5	1	G4117.t1
		synthase	ec:2.3.3.9	2	G5050.t1, G4514.t1
		synthase	ec:2.3.3.14	4	G4437.t1, G2387.t1, G2786.t1, G2832.t1

		synthase	ec:2.3.3.13	2	G3103.t1, G4504.t1
		hydratase	ec:4.2.1.2	1	G2007.t1
		decarboxylase	ec:4.1.1.3	2	G3312.t1, G2185.t1
		reductase	ec:1.2.99.6	1	G4017.t1
		acetyltransferase	ec:2.3.1.12	1	G4472.t1
		reductase (NADPH)	ec:1.1.1.283	4	G236.t1, G1883.t1, G2876.t1, G5388.t1
		hydrolase	ec:3.1.2.1	1	G4230.t1
		hydrolase	ec:3.1.2.6	2	G3048.t1, G4824.t1
		dehydrogenase (cytochrome)	ec:1.1.2.3	1	G135.t1
		dehydrogenase (cytochrome)	ec:1.1.2.4	3	G3130.t1, G2794.t1, G2790.t1
		reductase (NADP+)	ec:1.1.1.79	2	G1462.t1, G311.t1
		reductase (NADH)	ec:1.1.1.78	3	G2935.t1, G3832.t1, G689.t1
		dehydrogenase	ec:1.1.1.37	3	G2899.t1, G4386.t1, G1024.t1
		dehydrogenase (decarboxylating)	ec:1.1.1.39	1	G2185.t1
		dehydrogenase (oxaloacetate-decarboxylating)	ec:1.1.1.38	2	G3312.t1, G2185.t1
		dehydrogenase (oxaloacetate-decarboxylating) (NADP+)	ec:1.1.1.40	2	G3312.t1, G2185.t1
		dehydrogenase (acetyl-transferring)	ec:1.2.4.1	2	G2749.t1, G3144.t1
Butirosin and n	5	glucokinase (phosphorylating)	ec:2.7.1.2	5	G1056.t1, G155.t1, G3711.t1, G1862.t1, G868.t1
Oxidative phos	36	kinase	ec:2.7.4.1	1	G3599.t1

		dehydrogenase	ec:1.3.5.1	6	G3622.t1, G2355.t1, G4009.t1, G975.t1, G4740.t1, G969.t1
		reductase (H ⁺ -translocating)	ec:1.6.5.3	5	G2555.t1, G4393.t1, G72.t1, G3983.t1, G13.t1
		diphosphatase	ec:3.6.1.1	2	G3867.t1, G4206.t1
		ATPase	ec:3.6.3.6	2	G5360.t1, G457.t1
		oxidase	ec:1.9.3.1	11	G4987.t1, G178.t1, G1100.t1, G3885.t1, G2114.t1, G4608.t1, G207.t1, G2785.t1, G210.t1, G4456.t1, G3180.t1
		reductase	ec:1.10.2.2	7	G2062.t1, G3728.t1, G1361.t1, G4259.t1, G3343.t1, G2296.t1, G1873.t1
		dehydrogenase	ec:1.6.99.3	2	G1193.t1, G1492.t1
Biosynthesis of	240	transaminase	ec:2.6.1.11	1	G2886.t1
		aminotransferase	ec:2.6.1.13	1	G1064.t1
		cyclo-ligase	ec:6.3.3.1	1	G171.t1
		carboxylase	ec:6.4.1.2	2	G3440.t1, G3923.t1
		transaminase (isomerizing)	ec:2.6.1.16	2	G4042.t1, G1006.t1
		kinase	ec:2.7.4.6	1	G2152.t1
		kinase	ec:2.7.4.3	6	G3906.t1, G4782.t1, G1483.t1, G2801.t1, G2190.t1, G3152.t1
		kinase	ec:2.7.4.2	1	G3910.t1
		synthase	ec:4.1.1.48	1	G914.t1
		carboxykinase (ATP)	ec:4.1.1.49	1	G5481.t1
		decarboxylase	ec:4.1.1.33	1	G3413.t1
		dehydrogenase	ec:1.3.5.1	6	G3622.t1, G2355.t1, G4009.t1, G975.t1, G4740.t1, G969.t1
		kinase	ec:2.7.1.71	1	G4694.t1
		transaminase	ec:2.6.1.52	1	G1565.t1
		transaminase	ec:2.6.1.57	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		dihydroxyacetone transferase	ec:2.2.1.2	2	G5314.t1, G1139.t1
		glycolaldehyde transferase	ec:2.2.1.1	2	G554.t1, G587.t1

		synthase	ec:2.2.1.6	2	G4019.t1, G843.t1
		synthase	ec:4.1.3.27	2	G3227.t1, G914.t1
		kinase	ec:2.7.2.8	2	G3249.t1, G3647.t1
		kinase	ec:2.7.2.4	2	G1461.t1, G3268.t1
		kinase	ec:2.7.2.3	1	G829.t1
		lyase	ec:4.3.2.1	1	G2080.t1
		lyase	ec:4.3.2.2	1	G1135.t1
		reductase	ec:1.2.1.38	2	G3249.t1, G4719.t1
		dehydrogenase	ec:1.2.1.31	1	G588.t1
		kinase	ec:2.7.1.40	2	G1319.t1, G1708.t1
		deaminase	ec:3.5.4.6	3	G4146.t1, G2689.t1, G3646.t1
		dehydrogenase	ec:1.2.1.41	1	G1689.t1
		carbamoyltransferase	ec:2.1.3.3	1	G3661.t1
		transaminase	ec:2.6.1.39	3	G3169.t1, G3637.t1, G198.t1
		kinase	ec:2.7.1.36	1	G3922.t1
		3-kinase	ec:2.7.1.31	1	G2592.t1
		transaminase	ec:2.6.1.42	2	G54.t1, G48.t1
		reductase	ec:1.3.1.71	1	G5364.t1
		reductase	ec:1.3.1.70	1	G1457.t1
		gluconokinase (phosphorylating)	ec:2.7.1.12	1	G4802.t1
		phosphohexokinase	ec:2.7.1.11	2	G2561.t1, G3925.t1
		dehydratase	ec:4.2.1.51	1	G1424.t1
		diphosphate synthase	ec:2.5.1.10	2	G427.t1, G2297.t1
		ligase (ADP-forming)	ec:6.2.1.5	2	G2558.t1, G1528.t1
		ligase (GDP-forming)	ec:6.2.1.4	1	G1528.t1
		1-carboxyvinyltransferase	ec:2.5.1.19	1	G4694.t1
		ligase	ec:6.2.1.1	2	G1305.t1, G4882.t1
		synthase	ec:2.5.1.21	1	G2132.t1
		diphosphate synthase	ec:2.5.1.29	1	G427.t1
		dehydrogenase	ec:1.2.1.11	1	G4719.t1

		dehydrogenase (phosphorylating)	ec:1.2.1.12	3	G3582.t1, G3629.t1, G2604.t1
		decarboxylase	ec:4.1.1.17	1	G938.t1
		diphosphokinase	ec:2.7.6.1	5	G563.t1, G2958.t1, G941.t1, G4277.t1, G3218.t1
		3-dehydrogenase (decarboxylating)	ec:1.1.1.170	1	G5354.t1
		dehydrogenase	ec:1.8.1.4	2	G3776.t1, G474.t1
		geranyl-diphosphate synthase	ec:2.5.1.1	2	G427.t1, G2297.t1
		isomerase	ec:5.3.1.1	1	G4623.t1
		isomerase	ec:5.3.1.6	1	G3093.t1
		isomerase	ec:5.3.1.9	1	G2771.t1
		24-C-methyltransferase	ec:2.1.1.41	1	G4121.t1
		phosphatase	ec:3.1.3.3	1	G2589.t1
		dehydrogenase (NADP+)	ec:1.1.1.2	2	G3816.t1, G4312.t1
		dehydrogenase	ec:1.1.1.1	6	G4043.t1, G152.t1, G2799.t1, G2935.t1, G3832.t1, G689.t1
		dehydrogenase	ec:1.1.1.3	1	G3467.t1
		transaminase	ec:2.6.1.5	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		transaminase	ec:2.6.1.9	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		transaminase	ec:2.6.1.1	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		Delta-isomerase	ec:5.3.3.2	1	G314.t1
		diphosphorylase	ec:2.7.7.23	1	G4410.t1

		ammonia-lyase	ec:4.3.1.17	3	G907.t1, G887.t1, G5226.t1
		ammonia-lyase	ec:4.3.1.19	4	G907.t1, G3232.t1, G887.t1, G5226.t1
		dehydrogenase (NADP+)	ec:1.3.1.13	1	G710.t1
		dehydrogenase	ec:1.3.1.12	1	G710.t1
		dehydrogenase (NAD+)	ec:1.2.1.3	15	G1043.t1, G236.t1, G19.t1, G2101.t1, G2099.t1, G4017.t1, G4529.t1, G1733.t1, G3458.t1, G2019.t1, G434.t1, G3245.t1, G20.t1, G4211.t1, G5388.t1
		C-acetyltransferase	ec:2.3.1.9	1	G464.t1
		N-acetyltransferase	ec:2.3.1.1	2	G4062.t1, G3647.t1
		synthase	ec:5.5.1.4	1	G2283.t1
		gamma-lyase	ec:4.4.1.1	1	G1346.t1
		synthase	ec:6.3.5.3	1	G5299.t1
		monooxygenase	ec:1.14.13.72	1	G5300.t1
		14alpha-demethylase	ec:1.14.13.70	1	G2069.t1
		reductase	ec:1.5.1.2	1	G3298.t1
		(Si)-synthase	ec:2.3.3.1	3	G835.t1, G1950.t1, G491.t1
		dehydrogenase (NAD+, L-lysine-forming)	ec:1.5.1.7	1	G5047.t1
		1-epimerase	ec:5.1.3.15	1	G4028.t1
		synthase	ec:2.3.3.10	1	G67.t1
		formyltransferase	ec:2.1.2.3	2	G4007.t1, G4995.t1
		formyltransferase	ec:2.1.2.2	1	G1769.t1
		synthase	ec:2.3.3.14	4	G4437.t1, G2387.t1, G2786.t1, G2832.t1
		hydroxymethyltransferase	ec:2.1.2.1	2	G4969.t1, G2708.t1
		succinyltransferase	ec:2.3.1.61	1	G4712.t1
		aldolase	ec:4.1.2.13	1	G2158.t1
		synthase	ec:6.3.4.5	1	G2961.t1
		dehydrogenase (NADP+, L-glutamate-forming)	ec:1.5.1.10	1	G3406.t1

		dehydratase	ec:4.2.1.9	1	G3575.t1
		hydratase	ec:4.2.1.3	2	G2674.t1, G1182.t1
		hydratase	ec:4.2.1.2	1	G2007.t1
		arginine amidinase	ec:3.5.3.1	1	G309.t1
		equilase	ec:1.11.1.6	2	G4809.t1, G5276.t1
		5-kinase	ec:2.7.2.11	2	G4846.t1, G2095.t1
		synthase	ec:6.3.2.6	2	G3122.t1, G2333.t1
		glucokinase (phosphorylating)	ec:2.7.1.2	5	G1056.t1, G155.t1, G3711.t1, G1862.t1, G868.t1
		synthase	ec:4.2.3.5	1	G243.t1
		synthase	ec:4.2.3.4	1	G4694.t1
		decarboxylase	ec:4.1.1.1	3	G5277.t1, G4900.t1, G4982.t1
		1-epimerase	ec:5.1.3.3	3	G52.t1, G3387.t1, G4200.t1
		3-epimerase	ec:5.1.3.1	1	G3688.t1
		acetyltransferase	ec:2.3.1.12	1	G4472.t1
		hexose diphosphatase	ec:3.1.3.11	1	G1118.t1
		C-acyltransferase	ec:2.3.1.16	1	G5221.t1
		adenylyltransferase	ec:2.7.7.4	1	G3581.t1
		phosphatase	ec:3.1.3.25	2	G2109.t1, G4834.t1
		uridylyltransferase	ec:2.7.7.9	2	G2180.t1, G564.t1
		O-acetyltransferase	ec:2.3.1.31	1	G1460.t1
		N-acetyltransferase	ec:2.3.1.35	1	G4062.t1
		ligase	ec:6.3.4.13	1	G171.t1
		3-dehydrogenase	ec:1.1.1.270	1	G4928.t1
		dehydrogenase	ec:1.1.1.282	1	G4694.t1
		phosphogluconolactonase	ec:3.1.1.31	4	G5419.t1, G2554.t1, G3422.t1, G2453.t1
		synthase	ec:5.4.99.7	1	G2541.t1
		mutase	ec:5.4.99.5	1	G541.t1

		O-methyltransferase	ec:2.1.1.100	1	G1771.t1
		beta-synthase	ec:4.2.1.22	1	G2636.t1
		synthase	ec:4.2.1.20	1	G5377.t1
		dehydrogenase	ec:1.1.1.95	5	G212.t1, G3236.t1, G2400.t1, G1745.t1, G5144.t1
		gamma-synthase	ec:2.5.1.48	3	G3477.t1, G106.t1, G4869.t1
		synthase	ec:2.5.1.47	2	G5343.t1, G1183.t1
		hydratase	ec:4.2.1.11	6	G285.t1, G2443.t1, G2548.t1, G2403.t1, G268.t1, G1930.t1
		dehydratase	ec:4.2.1.10	1	G4694.t1
		synthase	ec:2.5.1.54	2	G2723.t1, G4612.t1
		hydratase	ec:4.2.1.17	2	G3670.t1, G2223.t1
		farnesyltransferase	ec:2.5.1.58	2	G2195.t1, G4399.t1
		synthase (NADH)	ec:1.4.1.14	1	G2796.t1
		reductase	ec:1.1.1.88	2	G113.t1, G1052.t1
		dehydrogenase	ec:1.1.1.87	1	G5159.t1
		reductoisomerase (NADP+)	ec:1.1.1.86	1	G1138.t1
		hydratase	ec:4.2.1.36	1	G4789.t1
		dehydrogenase	ec:1.1.1.37	3	G2899.t1, G4386.t1, G1024.t1
		dehydrogenase	ec:1.1.1.35	1	G2223.t1
		reductase (NADPH)	ec:1.1.1.34	2	G113.t1, G1052.t1
		dehydrogenase (NADP+-dependent, decarboxylating)	ec:1.1.1.44	2	G2405.t1, G259.t1
		dehydrogenase (NADP+)	ec:1.1.1.42	3	G4377.t1, G1941.t1, G1289.t1
		dehydrogenase (NAD+)	ec:1.1.1.41	2	G4442.t1, G1523.t1
		dehydrogenase (NADP+)	ec:1.1.1.49	1	G1491.t1
		(alpha-D-glucose-1,6-bisphosphate-dependent)	ec:5.4.2.2	3	G3856.t1, G4023.t1, G987.t1

		mutase	ec:5.4.2.3	1	G3373.t1
		dehydrogenase	ec:1.1.1.25	1	G4694.t1
		phosphoribosyl transferase	ec:2.4.2.18	1	G1914.t1
		phosphoribosyl diphosphate 5-amidotransferase	ec:2.4.2.14	1	G3835.t1
		dehydrogenase (succinyl-transferring)	ec:1.2.4.2	3	G3776.t1, G5186.t1, G474.t1
		dehydrogenase (acetyl-transferring)	ec:1.2.4.1	2	G2749.t1, G3144.t1
		cyclohydrolase	ec:3.5.4.10	2	G4007.t1, G4995.t1
		isomerase	ec:5.3.1.24	1	G4318.t1
		kinase	ec:2.7.4.10	1	G3152.t1
Pentose phosp	28	dihydroxyacetone transferase	ec:2.2.1.2	2	G5314.t1, G1139.t1
		glycolaldehydet ransferase	ec:2.2.1.1	2	G554.t1, G587.t1
		deoxyribokinas e	ec:2.7.1.15	1	G808.t1
		gluconokinase (phosphorylatin g)	ec:2.7.1.12	1	G4802.t1
		phosphohexoki nase	ec:2.7.1.11	2	G2561.t1, G3925.t1
		diphosphokina se	ec:2.7.6.1	5	G563.t1, G2958.t1, G941.t1, G4277.t1, G3218.t1
		isomerase	ec:5.3.1.6	1	G3093.t1
		isomerase	ec:5.3.1.9	1	G2771.t1
		aldolase	ec:4.1.2.13	1	G2158.t1
		3-epimerase	ec:5.1.3.1	1	G3688.t1
		hexose diphosphatase	ec:3.1.3.11	1	G1118.t1
		phosphoglucon olactonase	ec:3.1.1.31	4	G5419.t1, G2554.t1, G3422.t1, G2453.t1
		dehydrogenase (NADP+-dependent, decarboxylatin g)	ec:1.1.1.44	2	G2405.t1, G259.t1
		dehydrogenase (NADP+)	ec:1.1.1.49	1	G1491.t1
		(alpha-D-glucose-1,6-bisphosphate-dependent)	ec:5.4.2.2	3	G3856.t1, G4023.t1, G987.t1
		phosphodeoxyr ibomutase	ec:5.4.2.7	1	G3856.t1
Streptomycin b	11	synthase	ec:5.5.1.4	1	G2283.t1
		glucokinase (phosphorylatin g)	ec:2.7.1.2	5	G1056.t1, G155.t1, G3711.t1, G1862.t1, G868.t1

		phosphatase	ec:3.1.3.25	2	G2109.t1, G4834.t1
		(alpha-D-glucose-1,6-bisphosphate-dependent)	ec:5.4.2.2	3	G3856.t1, G4023.t1, G987.t1
Methane metab	29	transaminase	ec:2.6.1.52	1	G1565.t1
		kinase	ec:2.7.1.29	2	G118.t1, G2393.t1
		phosphohexokinase	ec:2.7.1.11	2	G2561.t1, G3925.t1
		ligase	ec:6.2.1.1	2	G1305.t1, G4882.t1
		hydrolase	ec:3.1.2.12	1	G3644.t1
		phosphatase	ec:3.1.3.3	1	G2589.t1
		dehydrogenase	ec:1.2.1.2	2	G2400.t1, G1745.t1
		hydroxymethyltransferase	ec:2.1.2.1	2	G4969.t1, G2708.t1
		aldolase	ec:4.1.2.13	1	G2158.t1
		hexose diphosphatase	ec:3.1.3.11	1	G1118.t1
		dehydrogenase	ec:1.1.1.284	2	G2799.t1, G689.t1
		dehydrogenase	ec:1.1.1.95	5	G212.t1, G3236.t1, G2400.t1, G1745.t1, G5144.t1
		hydratase	ec:4.2.1.11	6	G285.t1, G2443.t1, G2548.t1, G2403.t1, G268.t1, G1930.t1
		dehydrogenase	ec:1.1.1.37	3	G2899.t1, G4386.t1, G1024.t1
Amino sugar an	35	transaminase (isomerizing)	ec:2.6.1.16	2	G4042.t1, G1006.t1
		isomerase	ec:5.3.1.8	1	G3320.t1
		isomerase	ec:5.3.1.9	1	G2771.t1
		deacetylase	ec:3.5.1.41	2	G1178.t1, G1179.t1
		diphosphorylase	ec:2.7.7.23	1	G4410.t1
		guanylyltransferase	ec:2.7.7.13	1	G4368.t1
		uridylyltransferase	ec:2.7.7.12	1	G4201.t1

		N-acetyltransferase	ec:2.3.1.4	1	G3774.t1
		synthase	ec:2.4.1.16	3	G658.t1, G4576.t1, G4196.t1
		chitodextrinase	ec:3.2.1.14	1	G1899.t1
		glucokinase (phosphorylating)	ec:2.7.1.2	5	G1056.t1, G155.t1, G3711.t1, G1862.t1, G868.t1
		mannokinase (phosphorylating)	ec:2.7.1.7	5	G1056.t1, G155.t1, G3711.t1, G1862.t1, G868.t1
		galactokinase (phosphorylating)	ec:2.7.1.6	2	G4317.t1, G4199.t1
		fructokinase (phosphorylating)	ec:2.7.1.4	5	G1056.t1, G155.t1, G3711.t1, G1862.t1, G868.t1
		4-epimerase	ec:5.1.3.2	3	G52.t1, G3387.t1, G4200.t1
		uridylyltransferase	ec:2.7.7.9	2	G2180.t1, G564.t1
		reductase	ec:1.6.2.2	4	G967.t1, G68.t1, G102.t1, G5117.t1
		(alpha-D-glucose-1,6-bisphosphate-dependent) mutase	ec:5.4.2.2	3	G3856.t1, G4023.t1, G987.t1
		mannose phosphomutase	ec:5.4.2.3	1	G3373.t1
		mannose phosphomutase	ec:5.4.2.8	1	G3802.t1
Nitrogen metabolism	12	ligase	ec:6.3.1.2	1	G521.t1
		monooxygenase	ec:1.13.12.16	1	G47.t1
		dehydrogenase (NADP+)	ec:1.4.1.4	2	G1734.t1, G1298.t1
		dehydrogenase	ec:1.4.1.2	1	G793.t1
		dehydratase	ec:4.2.1.1	1	G4441.t1
		acetonitrilase	ec:3.5.5.1	1	G5224.t1
		synthase (ammonia)	ec:6.3.4.16	2	G3494.t1, G3697.t1
		reductase (NO-forming)	ec:1.7.2.1	1	G4456.t1
		oxide reductase (cytochrome c)	ec:1.7.2.5	1	G2567.t1
		synthase (NADH)	ec:1.4.1.14	1	G2796.t1
Glutathione metabolism	41	(ATP-hydrolysing) synthase	ec:3.5.2.9	1	G910.t1
		synthase	ec:2.5.1.16	1	G549.t1
		transferase	ec:2.5.1.18	11	G4868.t1, G2637.t1, G3881.t1, G5459.t1, G1859.t1, G2596.t1, G1501.t1, G5043.t1, G1026.t1, G863.t1, G446.t1
		synthase	ec:2.5.1.22	1	G4889.t1
		reductase (glutathione)	ec:1.8.4.2	3	G1864.t1, G5084.t1, G1657.t1
		decarboxylase	ec:4.1.1.17	1	G938.t1

		reductase	ec:1.8.1.7	4	G290.t1, G1128.t1, G4205.t1, G4333.t1
		hydrolase	ec:3.4.19.13	1	G1186.t1
		peroxidase	ec:1.11.1.9	7	G1859.t1, G5044.t1, G2188.t1, G1501.t1, G5043.t1, G2728.t1, G863.t1
		reductase	ec:1.17.4.1	4	G3248.t1, G3608.t1, G5136.t1, G2613.t1
		synthase	ec:6.3.2.3	1	G2969.t1
		ligase	ec:6.3.2.2	1	G3674.t1
		dehydrogenase (ascorbate)	ec:1.8.5.1	2	G2637.t1, G5459.t1
		dehydrogenase (NADP+-dependent, decarboxylating)	ec:1.1.1.44	2	G2405.t1, G259.t1
		dehydrogenase (NADP+)	ec:1.1.1.42	3	G4377.t1, G1941.t1, G1289.t1
		gamma-glutamyl-amino acid cyclotransferase	ec:2.3.2.4	1	G3159.t1
		glutamyl transpeptidase	ec:2.3.2.2	2	G1186.t1, G2692.t1
		dehydrogenase (NADP+)	ec:1.1.1.49	1	G1491.t1
		glutathione peroxidase	ec:1.11.1.12	3	G5044.t1, G2188.t1, G2728.t1
		thioredoxin peroxidase	ec:1.11.1.15	2	G3492.t1, G2188.t1
Aminoacyl-tRN	44	ligase	ec:6.1.1.6	2	G4614.t1, G4474.t1
		ligase	ec:6.1.1.7	2	G4445.t1, G1698.t1
		ligase	ec:6.1.1.9	2	G5270.t1, G2610.t1
		ligase	ec:6.1.1.2	2	G2923.t1, G4822.t1
		ligase	ec:6.1.1.3	2	G928.t1, G5148.t1
		ligase	ec:6.1.1.4	2	G1113.t1, G354.t1
		ligase	ec:6.1.1.5	3	G453.t1, G4283.t1, G2610.t1
		ligase	ec:6.1.1.1	2	G296.t1, G2610.t1
		synthase (glutamine-hydrolysing)	ec:6.3.5.7	3	G4286.t1, G3842.t1, G5262.t1
		formyltransferase	ec:2.1.2.9	1	G4228.t1
		ligase	ec:6.1.1.18	1	G1553.t1
		ligase	ec:6.1.1.17	2	G2982.t1, G162.t1
		ligase	ec:6.1.1.19	3	G1663.t1, G2834.t1, G2521.t1
		ligase	ec:6.1.1.14	2	G668.t1, G724.t1
		ligase	ec:6.1.1.16	1	G1486.t1

		ligase	ec:6.1.1.15	2	G3231.t1, G2082.t1
		ligase	ec:6.1.1.10	4	G2027.t1, G2412.t1, G2621.t1, G354.t1
		ligase	ec:6.1.1.12	2	G5036.t1, G302.t1
		ligase	ec:6.1.1.11	2	G2073.t1, G4602.t1
		ligase	ec:6.1.1.24	1	G2982.t1
		ligase	ec:6.1.1.21	1	G519.t1
		ligase	ec:6.1.1.20	4	G531.t1, G4967.t1, G5161.t1, G3779.t1
		ligase	ec:6.1.1.22	2	G2081.t1, G819.t1
Tryptophan me	34	decarboxylase	ec:4.1.1.43	1	G1890.t1
		3,4- dioxygenase	ec:1.13.11.6	1	G3569.t1
		transaminase	ec:2.6.1.7	2	G3637.t1, G2479.t1
		dehydrogenase (NAD ⁺)	ec:1.2.1.3	15	G1043.t1, G236.t1, G19.t1, G2101.t1, G2099.t1, G4017.t1, G4529.t1, G1733.t1, G3458.t1, G2019.t1, G434.t1, G3245.t1, G20.t1, G4211.t1, G5388.t1
		C- acetyltransfera se	ec:2.3.1.9	1	G464.t1
		acylamidase	ec:3.5.1.4	1	G4796.t1
		kynurenine formamidase	ec:3.5.1.9	2	G1786.t1, G3637.t1
		N- acetyltransfera se	ec:2.3.1.87	1	G4643.t1
		equilase	ec:1.11.1.6	2	G4809.t1, G5276.t1
		ec:3.7.1.3 kynureninase	ec:3.7.1.3	1	G1239.t1
		acetonitrilase	ec:3.5.5.1	1	G5224.t1
		2,3- dioxygenase	ec:1.13.11.52	1	G3523.t1
		3- monooxygenas e	ec:1.14.13.9	1	G4302.t1
		hydratase	ec:4.2.1.17	2	G3670.t1, G2223.t1
		dehydrogenase	ec:1.1.1.35	1	G2223.t1
		dehydrogenase (succinyl- transferring)	ec:1.2.4.2	3	G3776.t1, G5186.t1, G474.t1
Arginine biosyn	35	transaminase	ec:2.6.1.11	1	G2886.t1
		ligase	ec:6.3.1.2	1	G521.t1
		kinase	ec:2.7.2.8	2	G3249.t1, G3647.t1
		lyase	ec:4.3.2.1	1	G2080.t1
		reductase	ec:1.2.1.38	2	G3249.t1, G4719.t1

		carbamoyltransferase	ec:2.1.3.3	1	G3661.t1
		transaminase	ec:2.6.1.2	2	G4681.t1, G4940.t1
		transaminase	ec:2.6.1.1	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		N-acetyltransferase	ec:2.3.1.1	2	G4062.t1, G3647.t1
		hydrolase	ec:3.5.1.54	1	G2760.t1
		dehydrogenase (NADP+)	ec:1.4.1.4	2	G1734.t1, G1298.t1
		dehydrogenase	ec:1.4.1.2	1	G793.t1
		glutaminase I	ec:3.5.1.2	12	G1042.t1, G896.t1, G4031.t1, G2539.t1, G4032.t1, G895.t1, G894.t1, G890.t1, G2724.t1, G893.t1, G26.t1, G1410.t1
		synthase	ec:6.3.4.5	1	G2961.t1
		carboxylase	ec:6.3.4.6	1	G2760.t1
		arginine amidinase	ec:3.5.3.1	1	G309.t1
		N-acetyltransferase	ec:2.3.1.35	1	G4062.t1
		synthase (ammonia)	ec:6.3.4.16	2	G3494.t1, G3697.t1
Carbon fixation	25	carboxykinase (ATP)	ec:4.1.1.49	1	G5481.t1
		glycolaldehydetransferase	ec:2.2.1.1	2	G554.t1, G587.t1
		kinase	ec:2.7.2.3	1	G829.t1
		dehydrogenase (phosphorylating)	ec:1.2.1.12	3	G3582.t1, G3629.t1, G2604.t1
		isomerase	ec:5.3.1.1	1	G4623.t1
		isomerase	ec:5.3.1.6	1	G3093.t1
		transaminase	ec:2.6.1.2	2	G4681.t1, G4940.t1

		transaminase	ec:2.6.1.1	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		aldolase	ec:4.1.2.13	1	G2158.t1
		SBPase	ec:3.1.3.37	1	G2253.t1
		3-epimerase	ec:5.1.3.1	1	G3688.t1
		hexose diphosphatase	ec:3.1.3.11	1	G1118.t1
		dehydrogenase	ec:1.1.1.37	3	G2899.t1, G4386.t1, G1024.t1
		dehydrogenase (decarboxylatin g)	ec:1.1.1.39	1	G2185.t1
		dehydrogenase (oxaloacetate- decarboxylatin g) (NADP+)	ec:1.1.1.40	2	G3312.t1, G2185.t1
Geraniol degra	3	C- acyltransferase	ec:2.3.1.16	1	G5221.t1
		hydratase	ec:4.2.1.17	2	G3670.t1, G2223.t1
		dehydrogenase	ec:1.1.1.35	1	G2223.t1
Valine, leucine	25	transaminase	ec:2.6.1.42	2	G54.t1, G48.t1
		dehydrogenase	ec:1.8.1.4	2	G3776.t1, G474.t1
		dehydrogenase (NAD+)	ec:1.2.1.3	15	G1043.t1, G236.t1, G19.t1, G2101.t1, G2099.t1, G4017.t1, G4529.t1, G1733.t1, G3458.t1, G2019.t1, G434.t1, G3245.t1, G20.t1, G4211.t1, G5388.t1
		C- acetyltransfera se	ec:2.3.1.9	1	G464.t1
		synthase	ec:2.3.3.10	1	G67.t1
		C- acyltransferase	ec:2.3.1.16	1	G5221.t1
		hydrolase	ec:3.1.2.4	1	G4613.t1
		hydratase	ec:4.2.1.17	2	G3670.t1, G2223.t1
		dehydrogenase	ec:1.1.1.35	1	G2223.t1
Primary bile ac	1	dehydrogenase	ec:1.1.1.35	1	G2223.t1
Pantothenate a	23	decarboxylase	ec:4.1.1.36	3	G3057.t1, G5456.t1, G1021.t1
		synthase	ec:2.2.1.6	2	G4019.t1, G843.t1
		kinase	ec:2.7.1.33	1	G1875.t1
		transaminase	ec:2.6.1.42	2	G54.t1, G48.t1

		kinase	ec:2.7.1.24	2	G4757.t1, G2423.t1
		synthase	ec:2.7.8.7	3	G1978.t1, G239.t1, G342.t1
		2-reductase	ec:1.1.1.169	2	G2123.t1, G2819.t1
		dehydrogenase (NADP+)	ec:1.3.1.2	1	G909.t1
		diphosphatase	ec:3.6.1.9	2	G3334.t1, G818.t1
		dehydratase	ec:4.2.1.9	1	G3575.t1
		ligase	ec:6.3.2.5	1	G5150.t1
		ligase (AMP- forming)	ec:6.3.2.1	1	G5205.t1
		adenylyltransfe rase	ec:2.7.7.3	1	G2423.t1
		hydroxymethylt ransferase	ec:2.1.2.11	1	G718.t1
		reductoisomera se (NADP+)	ec:1.1.1.86	1	G1138.t1
Other types of	9	mannosyltransf erase	ec:2.4.1.109	9	G1974.t1, G4403.t1, G3463.t1, G4402.t1, G1335.t1, G1369.t1, G2598.t1, G4852.t1, G1687.t1
Various types of	12	alpha-1,6- mannosyltransf erase	ec:2.4.1.232	2	G4178.t1, G5387.t1
		alpha-1,3- mannosyltransf erase	ec:2.4.1.258	1	G4288.t1
		alpha-1,6- mannosyltransf erase	ec:2.4.1.260	1	G3427.t1
		alpha-1,2- mannosyltransf erase	ec:2.4.1.261	1	G1509.t1
		1,2-alpha- mannosidase	ec:3.2.1.113	3	G3476.t1, G4970.t1, G58.t1
		alpha-1,3- mannosyltransf erase	ec:2.4.1.132	1	G4178.t1
		alpha-1,2- mannosyltransf erase	ec:2.4.1.131	1	G4452.t1
		N- acetylglucosam inyltransferase	ec:2.4.1.141	2	G630.t1, G5394.t1
		beta- mannosyltransf erase	ec:2.4.1.142	1	G593.t1
Sesquiterpenoi	1	synthase	ec:2.5.1.21	1	G2132.t1
		dimethylallyltra nsferase	ec:2.5.1.75	1	G1645.t1

		alpha-D-mannosidase	ec:3.2.1.24	1	G237.t1
Citrate cycle (T	32	carboxylase	ec:6.4.1.1	2	G4181.t1, G2751.t1
		carboxykinase (ATP)	ec:4.1.1.49	1	G5481.t1
		dehydrogenase	ec:1.3.5.1	6	G3622.t1, G2355.t1, G4009.t1, G975.t1, G4740.t1, G969.t1
		ligase (ADP-forming)	ec:6.2.1.5	2	G2558.t1, G1528.t1
		ligase (GDP-forming)	ec:6.2.1.4	1	G1528.t1
		dehydrogenase	ec:1.8.1.4	2	G3776.t1, G474.t1
		(Si)-synthase	ec:2.3.3.1	3	G835.t1, G1950.t1, G491.t1
		succinyltransferase	ec:2.3.1.61	1	G4712.t1
		hydratase	ec:4.2.1.3	2	G2674.t1, G1182.t1
		hydratase	ec:4.2.1.2	1	G2007.t1
		acetyltransferase	ec:2.3.1.12	1	G4472.t1
		dehydrogenase	ec:1.1.1.37	3	G2899.t1, G4386.t1, G1024.t1
		dehydrogenase (NADP+)	ec:1.1.1.42	3	G4377.t1, G1941.t1, G1289.t1
		dehydrogenase (NAD+)	ec:1.1.1.41	2	G4442.t1, G1523.t1
		dehydrogenase (succinyl-transferring)	ec:1.2.4.2	3	G3776.t1, G5186.t1, G474.t1
		dehydrogenase (acetyl-transferring)	ec:1.2.4.1	2	G2749.t1, G3144.t1
One carbon po	18	cyclo-ligase	ec:6.3.3.2	3	G4436.t1, G2787.t1, G3139.t1
		cyclohydrolase	ec:3.5.4.9	3	G617.t1, G5463.t1, G2593.t1
		synthase	ec:2.1.1.45	1	G3074.t1
		synthase	ec:2.1.1.13	1	G3226.t1
		reductase	ec:1.5.1.3	1	G1611.t1
		dehydrogenase (NADP+)	ec:1.5.1.5	3	G617.t1, G5463.t1, G2593.t1
		formyltransferase	ec:2.1.2.3	2	G4007.t1, G4995.t1

		formyltransferase	ec:2.1.2.2	1	G1769.t1
		hydroxymethyltransferase	ec:2.1.2.1	2	G4969.t1, G2708.t1
		formyltransferase	ec:2.1.2.9	1	G4228.t1
		ligase	ec:6.3.4.3	2	G617.t1, G2593.t1
		reductase [NAD(P)H]	ec:1.5.1.20	2	G2045.t1, G468.t1
		dehydrogenase (NAD+)	ec:1.5.1.15	1	G5463.t1
		S-aminomethyl dihydrolipoylprotein:(6S)-tetrahydrofolate aminomethyltransferase (ammonia-forming)	ec:2.1.2.10	1	G4598.t1
N-Glycan biosy	19	kinase	ec:2.7.1.108	1	G4104.t1
		alpha-1,2-mannosyltransferase	ec:2.4.1.259	1	G1509.t1
		alpha-1,3-mannosyltransferase	ec:2.4.1.258	1	G4288.t1
		alpha-1,6-mannosyltransferase	ec:2.4.1.260	1	G3427.t1
		alpha-1,2-mannosyltransferase	ec:2.4.1.261	1	G1509.t1
		1,3-alpha-glucosidase	ec:3.2.1.84	1	G2743.t1
		1,2-alpha-mannosidase	ec:3.2.1.113	3	G3476.t1, G4970.t1, G58.t1
		beta-D-mannosyltransferase	ec:2.4.1.83	2	G1974.t1, G1369.t1
		glucosidase	ec:3.2.1.106	2	G5378.t1, G3934.t1
		beta-glucosyltransferase	ec:2.4.1.117	1	G1974.t1
		alpha-1,3-mannosyltransferase	ec:2.4.1.132	1	G4178.t1
		alpha-1,2-mannosyltransferase	ec:2.4.1.131	1	G4452.t1
		N-acetylglucosaminyltransferase	ec:2.4.1.141	2	G630.t1, G5394.t1
		beta-mannosyltransferase	ec:2.4.1.142	1	G593.t1
		dolichol diphosphatase	ec:3.6.1.43	1	G5320.t1
		N-acetylglucosaminephosphotransferase	ec:2.7.8.15	1	G2729.t1

Limonene and	17	dehydrogenase (NAD+)	ec:1.2.1.3	15	G1043.t1, G236.t1, G19.t1, G2101.t1, G2099.t1, G4017.t1, G4529.t1, G1733.t1, G3458.t1, G2019.t1, G434.t1, G3245.t1, G20.t1, G4211.t1, G5388.t1
		hydratase	ec:4.2.1.17	2	G3670.t1, G2223.t1
beta-Alanine m	25	transaminase	ec:2.6.1.19	1	G5337.t1
		synthase	ec:2.5.1.16	1	G549.t1
		synthase	ec:2.5.1.22	1	G4889.t1
		decarboxylase	ec:4.1.1.15	1	G3882.t1
		dehydrogenase [NAD(P)+]	ec:1.2.1.5	4	G4017.t1, G1733.t1, G434.t1, G3245.t1
		dehydrogenase (NAD+)	ec:1.2.1.3	15	G1043.t1, G236.t1, G19.t1, G2101.t1, G2099.t1, G4017.t1, G4529.t1, G1733.t1, G3458.t1, G2019.t1, G434.t1, G3245.t1, G20.t1, G4211.t1, G5388.t1
		dehydrogenase (NADP+)	ec:1.3.1.2	1	G909.t1
		ligase (AMP-forming)	ec:6.3.2.1	1	G5205.t1
		polyamine oxidase	ec:1.5.3.17	1	G4097.t1
		oxidase	ec:1.5.3.16	1	G4097.t1
		hydrolase	ec:3.1.2.4	1	G4613.t1
		hydratase	ec:4.2.1.17	2	G3670.t1, G2223.t1
Glucosinolate b	2	transaminase	ec:2.6.1.42	2	G54.t1, G48.t1
Terpenoid back	14	kinase	ec:2.7.4.2	1	G3910.t1
		decarboxylase	ec:4.1.1.33	1	G3413.t1
		kinase	ec:2.7.1.36	1	G3922.t1
		diphosphate synthase	ec:2.5.1.10	2	G427.t1, G2297.t1
		diphosphate synthase	ec:2.5.1.29	1	G427.t1
		synthase [(2E,6E)-farnesyl-diphosphate specific]	ec:2.5.1.31	1	G4214.t1
		geranyl-diphosphate synthase	ec:2.5.1.1	2	G427.t1, G2297.t1
		Delta-isomerase	ec:5.3.3.2	1	G314.t1

		C-acetyltransferase	ec:2.3.1.9	1	G464.t1
		synthase	ec:2.3.3.10	1	G67.t1
		O-methyltransferase	ec:2.1.1.100	1	G1771.t1
		farnesyltransferase	ec:2.5.1.58	2	G2195.t1, G4399.t1
		reductase	ec:1.1.1.88	2	G113.t1, G1052.t1
		reductase (NADPH)	ec:1.1.1.34	2	G113.t1, G1052.t1
D-Glutamine ar	12	glutaminase I	ec:3.5.1.2	12	G1042.t1, G896.t1, G4031.t1, G2539.t1, G4032.t1, G895.t1, G894.t1, G890.t1, G2724.t1, G893.t1, G26.t1, G1410.t1
Lysine degrada	29	dehydrogenase	ec:1.2.1.31	1	G588.t1
		transaminase	ec:2.6.1.39	3	G3169.t1, G3637.t1, G198.t1
		c]-lysine N-methyltransferase	ec:2.1.1.59	1	G2506.t1
		dehydrogenase (NAD+)	ec:1.2.1.3	15	G1043.t1, G236.t1, G19.t1, G2101.t1, G2099.t1, G4017.t1, G4529.t1, G1733.t1, G3458.t1, G2019.t1, G434.t1, G3245.t1, G20.t1, G4211.t1, G5388.t1
		C-acetyltransferase	ec:2.3.1.9	1	G464.t1
		dehydrogenase (NAD+, L-lysine-forming)	ec:1.5.1.7	1	G5047.t1
		succinyltransferase	ec:2.3.1.61	1	G4712.t1
		dehydrogenase (NADP+, L-glutamate-forming)	ec:1.5.1.10	1	G3406.t1
		hydratase	ec:4.2.1.17	2	G3670.t1, G2223.t1
		dehydrogenase	ec:1.1.1.35	1	G2223.t1

		dehydrogenase (succinyl-transferring)	ec:1.2.4.2	3	G3776.t1, G5186.t1, G474.t1
Tropane, piperi	5	transaminase	ec:2.6.1.57	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		transaminase	ec:2.6.1.5	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		transaminase	ec:2.6.1.9	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		transaminase	ec:2.6.1.1	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
Porphyrin and c	17	decarboxylase	ec:4.1.1.37	1	G4621.t1
		dehydrogenase	ec:1.3.1.76	2	G2756.t1, G5453.t1
		synthase	ec:4.2.1.75	1	G1648.t1
		ferrochelata	ec:4.99.1.4	1	G2756.t1
		ferro-protoporphyrin chelatase	ec:4.99.1.1	1	G1559.t1
		ceruloplasmin	ec:1.16.3.1	3	G4426.t1, G4065.t1, G3798.t1
		oxidase	ec:1.3.3.4	1	G3308.t1
		oxidase	ec:1.3.3.3	1	G4618.t1
		synthase	ec:4.4.1.17	2	G1318.t1, G1022.t1
		synthase	ec:2.3.1.37	1	G4787.t1
		C-methyltransferase	ec:2.1.1.107	1	G5453.t1
		synthase	ec:4.2.1.24	1	G5389.t1
		synthase	ec:2.5.1.61	1	G2367.t1
		ligase	ec:6.1.1.17	2	G2982.t1, G162.t1
Cysteine and m	47	decarboxylase	ec:4.1.1.50	1	G2967.t1
		transaminase	ec:2.6.1.57	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		kinase	ec:2.7.2.4	2	G1461.t1, G3268.t1
		transaminase	ec:2.6.1.42	2	G54.t1, G48.t1
		transaminase	ec:2.6.1.44	1	G3787.t1
		synthase	ec:2.5.1.16	1	G549.t1
		synthase	ec:2.5.1.22	1	G4889.t1

		dehydrogenase	ec:1.2.1.11	1	G4719.t1
		adenosyltransferase	ec:2.5.1.6	2	G1283.t1, G1852.t1
		(R)-S-oxide reductase	ec:1.8.4.14	1	G2149.t1
		dehydrogenase	ec:1.1.1.3	1	G3467.t1
		synthase	ec:2.1.1.13	1	G3226.t1
		S-methyltransferase	ec:2.1.1.14	1	G3226.t1
		transaminase	ec:2.6.1.5	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		S-adenosylhomocysteine synthase	ec:3.3.1.1	1	G3278.t1
		transaminase	ec:2.6.1.1	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		ammonia-lyase	ec:4.3.1.17	3	G907.t1, G887.t1, G5226.t1
		beta-lyase	ec:4.4.1.8	3	G213.t1, G2503.t1, G3709.t1
		gamma-lyase	ec:4.4.1.1	1	G1346.t1
		synthase	ec:3.1.3.77	1	G3354.t1
		synthase	ec:6.3.2.3	1	G2969.t1
		ligase	ec:6.3.2.2	1	G3674.t1
		dioxygenase [iron(II)-requiring]	ec:1.13.11.54	1	G4108.t1
		dioxygenase (Ni ²⁺ -requiring)	ec:1.13.11.53	1	G4108.t1
		O-acetyltransferase	ec:2.3.1.31	1	G1460.t1
		beta-synthase	ec:4.2.1.22	1	G2636.t1
		aminocarboxypopyltransferase	ec:2.5.1.49	1	G1183.t1
		gamma-synthase	ec:2.5.1.48	3	G3477.t1, G106.t1, G4869.t1
		synthase	ec:2.5.1.47	2	G5343.t1, G1183.t1
		1-phosphate dehydratase	ec:4.2.1.109	1	G3570.t1

		dehydrogenase	ec:1.1.1.37	3	G2899.t1, G4386.t1, G1024.t1
		phosphorylase	ec:2.4.2.28	1	G5006.t1
		isomerase	ec:5.3.1.23	4	G2239.t1, G754.t1, G1192.t1, G5281.t1
Nicotinate and	16	kinase	ec:2.7.1.23	3	G3357.t1, G379.t1, G3551.t1
		kinase	ec:2.7.1.22	1	G4524.t1
		adenylyltransferase	ec:2.7.7.18	2	G5344.t1, G1161.t1
		nicotinamide deaminase	ec:3.5.1.19	1	G5386.t1
		synthase (glutamine-hydrolysing)	ec:6.3.5.1	1	G2539.t1
		diphosphatase	ec:3.6.1.9	2	G3334.t1, G818.t1
		adenylyltransferase	ec:2.7.7.1	3	G872.t1, G5344.t1, G1161.t1
		diphosphatase	ec:3.6.1.22	1	G4176.t1
		nucleosidase	ec:3.2.2.3	1	G1762.t1
		nucleosidase	ec:3.2.2.1	1	G1762.t1
		diphosphorylase (carboxylating)	ec:2.4.2.19	2	G3716.t1, G1587.t1
		phosphorylase	ec:2.4.2.1	1	G1258.t1
Sphingolipid m	17	kinase	ec:2.7.1.91	2	G1216.t1, G1554.t1
		reductase	ec:1.1.1.102	1	G2706.t1
		phosphatase	ec:3.1.3.4	4	G1853.t1, G4832.t1, G3965.t1, G4494.t1
		aldolase	ec:4.1.2.27	2	G4840.t1, G3782.t1
		N-acyltransferase	ec:2.3.1.24	5	G2782.t1, G3837.t1, G411.t1, G2058.t1, G2206.t1
		C-palmitoyltransferase	ec:2.3.1.50	2	G3839.t1, G4634.t1
		sulfatase	ec:3.1.6.1	1	G5490.t1
C5-Branched d	6	synthase	ec:2.2.1.6	2	G4019.t1, G843.t1
		aldolase	ec:4.1.3.17	1	G3312.t1
		ligase (ADP-forming)	ec:6.2.1.5	2	G2558.t1, G1528.t1
		dehydrogenase	ec:1.1.1.85	1	G849.t1
Glycolysis / Glu	64	carboxykinase (ATP)	ec:4.1.1.49	1	G5481.t1

		kinase	ec:2.7.2.3	1	G829.t1
		kinase	ec:2.7.1.40	2	G1319.t1, G1708.t1
		phosphohexokinase	ec:2.7.1.11	2	G2561.t1, G3925.t1
		ligase	ec:6.2.1.1	2	G1305.t1, G4882.t1
		dehydrogenase (phosphorylating)	ec:1.2.1.12	3	G3582.t1, G3629.t1, G2604.t1
		dehydrogenase	ec:1.8.1.4	2	G3776.t1, G474.t1
		isomerase	ec:5.3.1.1	1	G4623.t1
		isomerase	ec:5.3.1.9	1	G2771.t1
		dehydrogenase (NADP+)	ec:1.1.1.2	2	G3816.t1, G4312.t1
		dehydrogenase	ec:1.1.1.1	6	G4043.t1, G152.t1, G2799.t1, G2935.t1, G3832.t1, G689.t1
		dehydrogenase [NAD(P)+]	ec:1.2.1.5	4	G4017.t1, G1733.t1, G434.t1, G3245.t1
		dehydrogenase (NAD+)	ec:1.2.1.3	15	G1043.t1, G236.t1, G19.t1, G2101.t1, G2099.t1, G4017.t1, G4529.t1, G1733.t1, G3458.t1, G2019.t1, G434.t1, G3245.t1, G20.t1, G4211.t1, G5388.t1
		1-epimerase	ec:5.1.3.15	1	G4028.t1
		aldolase	ec:4.1.2.13	1	G2158.t1
		glucokinase (phosphorylating)	ec:2.7.1.2	5	G1056.t1, G155.t1, G3711.t1, G1862.t1, G868.t1
		decarboxylase	ec:4.1.1.1	3	G5277.t1, G4900.t1, G4982.t1
		1-epimerase	ec:5.1.3.3	3	G52.t1, G3387.t1, G4200.t1
		acetyltransferase	ec:2.3.1.12	1	G4472.t1
		hexose diphosphatase	ec:3.1.3.11	1	G1118.t1
		hydratase	ec:4.2.1.11	6	G285.t1, G2443.t1, G2548.t1, G2403.t1, G268.t1, G1930.t1
		(alpha-D-glucose-1,6-bisphosphate-dependent)	ec:5.4.2.2	3	G3856.t1, G4023.t1, G987.t1
		dehydrogenase (acetyl-transferring)	ec:1.2.4.1	2	G2749.t1, G3144.t1
Starch and suc	55	endo-1,4-beta-D-glucanase	ec:3.2.1.4	1	G1415.t1

		1,4-alpha-glucosidase	ec:3.2.1.3	1	G5164.t1
		isomerase	ec:5.3.1.9	1	G2771.t1
		alpha-glucosidase	ec:3.2.1.48	16	G2662.t1, G2660.t1, G283.t1, G2442.t1, G5223.t1, G2436.t1, G5229.t1, G2433.t1, G18.t1, G898.t1, G3456.t1, G270.t1, G902.t1, G2677.t1, G2870.t1, G2391.t1
		1,3-beta-glucosidase	ec:3.2.1.58	4	G1185.t1, G4814.t1, G2428.t1, G1571.t1
		maltase	ec:3.2.1.20	15	G2662.t1, G2660.t1, G283.t1, G2442.t1, G2436.t1, G5229.t1, G2433.t1, G18.t1, G898.t1, G3456.t1, G902.t1, G2677.t1, G2870.t1, G2391.t1, G4777.t1
		synthase	ec:2.4.1.34	3	G1150.t1, G5324.t1, G3829.t1
		trehalase	ec:3.2.1.28	3	G512.t1, G4216.t1, G4324.t1
		invertase	ec:3.2.1.26	16	G2662.t1, G2660.t1, G283.t1, G2442.t1, G5223.t1, G2436.t1, G5229.t1, G2433.t1, G18.t1, G898.t1, G3456.t1, G270.t1, G902.t1, G2677.t1, G2870.t1, G2391.t1
		endo-1,3-beta-D-glucosidase	ec:3.2.1.39	5	G2679.t1, G5379.t1, G3390.t1, G2428.t1, G4891.t1
		amylase-1,6-glucosidase	ec:3.2.1.33	1	G1368.t1
		branching enzyme	ec:2.4.1.18	1	G3330.t1
		synthase (UDP-forming)	ec:2.4.1.15	3	G3873.t1, G92.t1, G671.t1
		synthase	ec:2.4.1.11	2	G1218.t1, G3746.t1
		limit dextrinase (erroneous)	ec:3.2.1.10	15	G2662.t1, G2660.t1, G283.t1, G2442.t1, G2436.t1, G5229.t1, G2433.t1, G18.t1, G898.t1, G3456.t1, G270.t1, G902.t1, G2677.t1, G2870.t1, G2391.t1

		disproportionating enzyme	ec:2.4.1.25	1	G1368.t1
		pectin depolymerase	ec:3.2.1.15	1	G43.t1
		diphosphatase	ec:3.6.1.9	2	G3334.t1, G818.t1
		glucokinase (phosphorylating)	ec:2.7.1.2	5	G1056.t1, G155.t1, G3711.t1, G1862.t1, G868.t1
		fructokinase (phosphorylating)	ec:2.7.1.4	5	G1056.t1, G155.t1, G3711.t1, G1862.t1, G868.t1
		trehalose 6-phosphatase	ec:3.1.3.12	3	G3873.t1, G92.t1, G4646.t1
		uridylyltransferase	ec:2.7.7.9	2	G2180.t1, G564.t1
		(alpha-D-glucose-1,6-bisphosphate-dependent)	ec:5.4.2.2	3	G3856.t1, G4023.t1, G987.t1
Ether lipid metabolism	10	reductase	ec:1.1.1.101	1	G5185.t1
		EPT	ec:2.7.8.1	1	G2492.t1
		cholinephosphotransferase	ec:2.7.8.2	2	G4525.t1, G2492.t1
		A2	ec:3.1.1.4	1	G5255.t1
		phosphatase	ec:3.1.3.4	4	G1853.t1, G4832.t1, G3965.t1, G4494.t1
		D	ec:3.1.4.4	2	G301.t1, G2244.t1
Cutin, suberine	1	O-acyltransferase	ec:2.3.1.20	1	G1619.t1

Glycerophosph	45	phosphodiesterase	ec:3.1.4.46	2	G308.t1, G393.t1
		kinase (ATP)	ec:2.7.1.107	1	G1679.t1
		kinase	ec:2.7.1.82	2	G4711.t1, G4901.t1
		kinase	ec:2.7.1.32	2	G4711.t1, G4901.t1
		decarboxylase	ec:4.1.1.65	2	G4559.t1, G2622.t1
		reductase	ec:1.1.1.101	1	G5185.t1
		EPT	ec:2.7.8.1	1	G2492.t1
		cholinephosphotransferase	ec:2.7.8.2	2	G4525.t1, G2492.t1
		3-phosphatidyltransferase	ec:2.7.8.5	1	G840.t1
		O-phosphatidyltransferase	ec:2.7.8.8	1	G3295.t1
		A2	ec:3.1.1.4	1	G5255.t1
		lecithinase B	ec:3.1.1.5	5	G129.t1, G3061.t1, G4109.t1, G4110.t1, G3003.t1
		phosphatase	ec:3.1.3.4	4	G1853.t1, G4832.t1, G3965.t1, G4494.t1
		N-methyltransferase	ec:2.1.1.71	1	G3528.t1
		dehydrogenase (NAD+)	ec:1.1.1.8	2	G2960.t1, G4343.t1
		cytidyltransferase	ec:2.7.7.41	2	G5312.t1, G4191.t1
		N-methyltransferase	ec:2.1.1.17	2	G2634.t1, G3528.t1
		cytidyltransferase	ec:2.7.7.14	1	G5347.t1
		cytidyltransferase	ec:2.7.7.15	1	G2595.t1
		dehydrogenase	ec:1.1.5.3	1	G5215.t1
		diphosphate phosphatase	ec:3.1.3.81	1	G4832.t1
		deacylase	ec:3.1.1.52	1	G3782.t1
		1-O-acyltransferase	ec:2.3.1.15	4	G3991.t1, G5451.t1, G4227.t1, G3778.t1
		phosphatidylglycerol phosphate phosphatase	ec:3.1.3.27	1	G2514.t1

		O-acyltransferase	ec:2.3.1.23	2	G772.t1, G1558.t1
		O-acyltransferase	ec:2.3.1.51	4	G4365.t1, G4930.t1, G771.t1, G1558.t1
		O-acyltransferase	ec:2.3.1.42	4	G3991.t1, G5451.t1, G4227.t1, G3778.t1
		3-phosphatidyltransferase	ec:2.7.8.11	1	G749.t1
		dehydrogenase [NAD(P)+]	ec:1.1.1.94	2	G2960.t1, G4343.t1
		phosphodiesterase	ec:3.1.4.2	1	G308.t1
		D	ec:3.1.4.4	2	G301.t1, G2244.t1
Synthesis and	2	C-acyltransferase	ec:2.3.1.9	1	G464.t1
		synthase	ec:2.3.3.10	1	G67.t1
Glycosylphosph	7	phospholipase D	ec:3.1.4.50	1	G3699.t1
		N-acetylglucosaminyltransferase	ec:2.4.1.198	5	G368.t1, G1793.t1, G2581.t1, G4443.t1, G421.t1
		deacetylase	ec:3.5.1.89	1	G3853.t1
Fatty acid degr	33	ligase	ec:6.2.1.3	5	G3884.t1, G3307.t1, G5088.t1, G655.t1, G1684.t1
		dehydrogenase	ec:1.1.1.1	6	G4043.t1, G152.t1, G2799.t1, G2935.t1, G3832.t1, G689.t1
		isomerase	ec:5.3.3.8	2	G1199.t1, G1563.t1
		dehydrogenase (NAD+)	ec:1.2.1.3	15	G1043.t1, G236.t1, G19.t1, G2101.t1, G2099.t1, G4017.t1, G4529.t1, G1733.t1, G3458.t1, G2019.t1, G434.t1, G3245.t1, G20.t1, G4211.t1, G5388.t1
		C-acyltransferase	ec:2.3.1.9	1	G464.t1
		oxidase	ec:1.3.3.6	1	G196.t1
		C-acyltransferase	ec:2.3.1.16	1	G5221.t1
		hydratase	ec:4.2.1.17	2	G3670.t1, G2223.t1
		dehydrogenase	ec:1.1.1.35	1	G2223.t1

Inositol phosph	23	3-kinase	ec:2.7.1.127	2	G4735.t1, G4596.t1
		4-kinase	ec:2.7.1.67	4	G1468.t1, G922.t1, G3673.t1, G1181.t1
		5-kinase	ec:2.7.1.68	1	G4767.t1
		phospholipase C	ec:3.1.4.11	1	G2013.t1
		3-kinase	ec:2.7.1.137	1	G1233.t1
		2-kinase	ec:2.7.1.158	1	G2858.t1
		5-kinase	ec:2.7.1.150	1	G3742.t1
		multikinase	ec:2.7.1.151	1	G4735.t1
		isomerase	ec:5.3.1.1	1	G4623.t1
		synthase	ec:5.5.1.4	1	G2283.t1
		5-phosphatase	ec:3.1.3.36	4	G3104.t1, G4505.t1, G2954.t1, G5080.t1
		3-phosphatase	ec:3.1.3.67	1	G4523.t1
		inositol-1,3-bisphosphate 3-phosphatase	ec:3.1.3.64	4	G3104.t1, G4505.t1, G913.t1, G3493.t1
		phosphatase	ec:3.1.3.25	2	G2109.t1, G4834.t1
		3-phosphatidyltransferase	ec:2.7.8.11	1	G749.t1
Glycerolipid me	55	kinase (ATP)	ec:2.7.1.107	1	G1679.t1
		kinase	ec:2.7.1.30	1	G584.t1
		3-kinase	ec:2.7.1.31	1	G2592.t1
		kinase	ec:2.7.1.29	2	G118.t1, G2393.t1
		acyltransferase	ec:2.3.1.158	1	G3448.t1
		2-dehydrogenase (NADP+)	ec:1.1.1.156	2	G1902.t1, G3115.t1
		lipase	ec:3.1.1.3	8	G3083.t1, G3822.t1, G5185.t1, G3081.t1, G5472.t1, G5424.t1, G4630.t1, G2763.t1
		phosphatase	ec:3.1.3.4	4	G1853.t1, G4832.t1, G3965.t1, G4494.t1
		dehydrogenase (NADP+)	ec:1.1.1.2	2	G3816.t1, G4312.t1
		dehydrogenase (NAD+)	ec:1.2.1.3	15	G1043.t1, G236.t1, G19.t1, G2101.t1, G2099.t1, G4017.t1, G4529.t1, G1733.t1, G3458.t1, G2019.t1, G434.t1, G3245.t1, G20.t1, G4211.t1, G5388.t1
		diphosphate phosphatase	ec:3.1.3.81	1	G4832.t1
		1-O-acyltransferase	ec:2.3.1.15	4	G3991.t1, G5451.t1, G4227.t1, G3778.t1
		alpha-glycerophosphatase	ec:3.1.3.21	2	G3256.t1, G5126.t1

		lipase	ec:3.1.1.23	5	G247.t1, G4416.t1, G1799.t1, G4370.t1, G1015.t1
		O-acyltransferase	ec:2.3.1.20	1	G1619.t1
		O-acyltransferase	ec:2.3.1.22	1	G1619.t1
		O-acyltransferase	ec:2.3.1.51	4	G4365.t1, G4930.t1, G771.t1, G1558.t1
		reductase	ec:1.1.1.21	4	G1902.t1, G4429.t1, G3115.t1, G3507.t1
Novobiocin bios	6	transaminase	ec:2.6.1.57	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		transaminase	ec:2.6.1.5	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		transaminase	ec:2.6.1.9	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		transaminase	ec:2.6.1.1	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		dehydrogenase	ec:1.3.1.12	1	G710.t1
Phenylalanine,	17	synthase	ec:4.1.1.48	1	G914.t1
		kinase	ec:2.7.1.71	1	G4694.t1
		transaminase	ec:2.6.1.57	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		synthase	ec:4.1.3.27	2	G3227.t1, G914.t1
		dehydratase	ec:4.2.1.51	1	G1424.t1
		1-carboxyvinyltransferase	ec:2.5.1.19	1	G4694.t1
		transaminase	ec:2.6.1.5	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		transaminase	ec:2.6.1.9	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		transaminase	ec:2.6.1.1	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		dehydrogenase (NADP+)	ec:1.3.1.13	1	G710.t1

		dehydrogenase	ec:1.3.1.12	1	G710.t1
		synthase	ec:4.2.3.5	1	G243.t1
		synthase	ec:4.2.3.4	1	G4694.t1
		dehydrogenase	ec:1.1.1.282	1	G4694.t1
		mutase	ec:5.4.99.5	1	G541.t1
		synthase	ec:4.2.1.20	1	G5377.t1
		dehydratase	ec:4.2.1.10	1	G4694.t1
		synthase	ec:2.5.1.54	2	G2723.t1, G4612.t1
		dehydrogenase	ec:1.1.1.25	1	G4694.t1
		phosphoribosyl transferase	ec:2.4.2.18	1	G1914.t1
		isomerase	ec:5.3.1.24	1	G4318.t1
Cyanoamino ac	6	asparaginase II	ec:3.5.1.1	1	G2853.t1
		hydroxymethyl transferase	ec:2.1.2.1	2	G4969.t1, G2708.t1
		acetonitrilase	ec:3.5.5.1	1	G5224.t1
		glutamyl transpeptidase	ec:2.3.2.2	2	G1186.t1, G2692.t1
Isoquinoline alk	5	transaminase	ec:2.6.1.57	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		transaminase	ec:2.6.1.5	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		transaminase	ec:2.6.1.1	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
Lysine biosynth	20	transaminase	ec:2.6.1.57	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		kinase	ec:2.7.2.4	2	G1461.t1, G3268.t1
		dehydrogenase	ec:1.2.1.31	1	G588.t1

		transaminase	ec:2.6.1.39	3	G3169.t1, G3637.t1, G198.t1
		dehydrogenase	ec:1.2.1.11	1	G4719.t1
		dehydrogenase	ec:1.1.1.3	1	G3467.t1
		dehydrogenase (NAD+, L-lysine-forming)	ec:1.5.1.7	1	G5047.t1
		synthase	ec:2.3.3.14	4	G4437.t1, G2387.t1, G2786.t1, G2832.t1
		dehydrogenase (NADP+, L-glutamate-forming)	ec:1.5.1.10	1	G3406.t1
		dehydrogenase	ec:1.1.1.87	1	G5159.t1
		hydratase	ec:4.2.1.36	1	G4789.t1
Fluorobenzoate	1	maleylacetate enol-lactonase	ec:3.1.1.45	1	G4395.t1
Benzoate degra	5	aldolase	ec:4.1.3.17	1	G3312.t1
		C-acetyltransferase	ec:2.3.1.9	1	G464.t1
		C-acyltransferase	ec:2.3.1.16	1	G5221.t1
		hydratase	ec:4.2.1.17	2	G3670.t1, G2223.t1
		dehydrogenase	ec:1.1.1.35	1	G2223.t1
Chlorocyclohex	1	maleylacetate enol-lactonase	ec:3.1.1.45	1	G4395.t1
Phenylalanine t	22	decarboxylase	ec:4.1.1.43	1	G1890.t1
		transaminase	ec:2.6.1.57	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		transaminase	ec:2.6.1.5	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		transaminase	ec:2.6.1.9	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1

		transaminase	ec:2.6.1.1	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		dehydrogenase [NAD(P)+]	ec:1.2.1.5	4	G4017.t1, G1733.t1, G434.t1, G3245.t1
		acylamidase	ec:3.5.1.4	1	G4796.t1
		N-acetyltransferase	ec:2.3.1.36	2	G3380.t1, G1359.t1
		dehydrogenase	ec:1.1.1.90	7	G41.t1, G2396.t1, G28.t1, G410.t1, G5489.t1, G4314.t1, G1411.t1
		hydratase	ec:4.2.1.17	2	G3670.t1, G2223.t1
T cell receptor	8	phosphatase	ec:3.1.3.16	8	G1349.t1, G4638.t1, G433.t1, G1133.t1, G2065.t1, G33.t1, G5004.t1, G5030.t1
Monobactam b	4	kinase	ec:2.7.2.4	2	G1461.t1, G3268.t1
		dehydrogenase	ec:1.2.1.11	1	G4719.t1
		adenylyltransferase	ec:2.7.7.4	1	G3581.t1
Glycine, serine	42	transaminase	ec:2.6.1.51	1	G3787.t1
		transaminase	ec:2.6.1.52	1	G1565.t1
		kinase	ec:2.7.2.4	2	G1461.t1, G3268.t1
		kinase	ec:2.7.1.39	1	G2087.t1
		3-kinase	ec:2.7.1.31	1	G2592.t1
		transaminase	ec:2.6.1.44	1	G3787.t1
		O-phosphatidyltransferase	ec:2.7.8.8	1	G3295.t1
		dehydrogenase	ec:1.2.1.11	1	G4719.t1
		dehydrogenase	ec:1.8.1.4	2	G3776.t1, G474.t1
		phosphatase	ec:3.1.3.3	1	G2589.t1
		dehydrogenase	ec:1.1.1.1	6	G4043.t1, G152.t1, G2799.t1, G2935.t1, G3832.t1, G689.t1
		dehydrogenase	ec:1.1.1.3	1	G3467.t1
		ammonia-lyase	ec:4.3.1.17	3	G907.t1, G887.t1, G5226.t1
		ammonia-lyase	ec:4.3.1.18	2	G907.t1, G203.t1
		ammonia-lyase	ec:4.3.1.19	4	G907.t1, G3232.t1, G887.t1, G5226.t1

		dehydrogenase	ec:1.2.1.8	2	G3961.t1, G3960.t1
		gamma-lyase	ec:4.4.1.1	1	G1346.t1
		hydroxymethyltransferase	ec:2.1.2.1	2	G4969.t1, G2708.t1
		dehydrogenase (aminomethyltransferring)	ec:1.4.4.2	5	G3776.t1, G1313.t1, G474.t1, G4598.t1, G3941.t1
		synthase	ec:4.2.3.1	1	G5434.t1
		synthase	ec:2.3.1.37	1	G4787.t1
		S-aminomethyl dihydrolipoylprotein:(6S)-tetrahydrofolate aminomethyltransferase (ammonia-forming)	ec:2.1.2.10	1	G4598.t1
		beta-synthase	ec:4.2.1.22	1	G2636.t1
		synthase	ec:4.2.1.20	1	G5377.t1
		dehydrogenase	ec:1.1.1.95	5	G212.t1, G3236.t1, G2400.t1, G1745.t1, G5144.t1
		reductase	ec:1.1.1.81	2	G1462.t1, G311.t1
Steroid biosynt	14	reductase	ec:1.3.1.71	1	G5364.t1
		reductase	ec:1.3.1.70	1	G1457.t1
		synthase	ec:2.5.1.21	1	G2132.t1
		3-dehydrogenase (decarboxylating)	ec:1.1.1.170	1	G5354.t1
		24-C-methyltransferase	ec:2.1.1.41	1	G4121.t1
		monooxygenase	ec:1.14.13.72	1	G5300.t1
		14alpha-demethylase	ec:1.14.13.70	1	G2069.t1
		esterase	ec:3.1.1.13	5	G3081.t1, G5472.t1, G5032.t1, G976.t1, G5003.t1
		3-dehydrogenase	ec:1.1.1.270	1	G4928.t1
		synthase	ec:5.4.99.7	1	G2541.t1
Vitamin B6 met	18	5'-phosphate synthase	ec:1.4.3.5	3	G1379.t1, G661.t1, G1046.t1

		transaminase	ec:2.6.1.52	1	G1565.t1
		kinase	ec:2.7.1.35	2	G3347.t1, G3430.t1
		5'-phosphate synthase (glutamine hydrolysing)	ec:4.3.3.6	10	G1042.t1, G896.t1, G4031.t1, G4032.t1, G895.t1, G894.t1, G890.t1, G893.t1, G26.t1, G1410.t1
		synthase	ec:4.2.3.1	1	G5434.t1
		4-dehydrogenase	ec:1.1.1.65	1	G761.t1
Phosphatidylinositol	26	kinase (ATP)	ec:2.7.1.107	1	G1679.t1
		3-kinase	ec:2.7.1.127	2	G4735.t1, G4596.t1
		4-kinase	ec:2.7.1.67	4	G1468.t1, G922.t1, G3673.t1, G1181.t1
		5-kinase	ec:2.7.1.68	1	G4767.t1
		phospholipase C	ec:3.1.4.11	1	G2013.t1
		3-kinase	ec:2.7.1.137	1	G1233.t1
		2-kinase	ec:2.7.1.158	1	G2858.t1
		5-kinase	ec:2.7.1.150	1	G3742.t1
		multikinase	ec:2.7.1.151	1	G4735.t1
		kinase C	ec:2.7.11.13	1	G4308.t1
		cytidyltransferase	ec:2.7.7.41	2	G5312.t1, G4191.t1
		5-phosphatase	ec:3.1.3.36	4	G3104.t1, G4505.t1, G2954.t1, G5080.t1
		3-phosphatase	ec:3.1.3.67	1	G4523.t1
		inositol-1,3-bisphosphate 3-phosphatase	ec:3.1.3.64	4	G3104.t1, G4505.t1, G913.t1, G3493.t1
		phosphatase	ec:3.1.3.25	2	G2109.t1, G4834.t1
		3-phosphatidyltransferase	ec:2.7.8.11	1	G749.t1
		kinase	ec:2.7.4.24	1	G1088.t1
		kinase	ec:2.7.4.21	2	G1088.t1, G4596.t1
Butanoate metabolism	18	transaminase	ec:2.6.1.19	1	G5337.t1
		dehydrogenase	ec:1.3.5.1	6	G3622.t1, G2355.t1, G4009.t1, G975.t1, G4740.t1, G969.t1
		dehydrogenase (NAD ⁺)	ec:1.2.1.24	1	G4211.t1
		synthase	ec:2.2.1.6	2	G4019.t1, G843.t1
		dehydrogenase [NAD(P) ⁺]	ec:1.2.1.16	1	G4211.t1
		decarboxylase	ec:4.1.1.15	1	G3882.t1

		dehydrogenase	ec:1.1.1.4	1	G1300.t1
		reductase [(R)-acetoin forming]	ec:1.1.1.303	1	G1299.t1
		C-acetyltransferase	ec:2.3.1.9	1	G464.t1
		synthase	ec:2.3.3.10	1	G67.t1
		CoA-transferase	ec:2.8.3.8	1	G4230.t1
		hydratase	ec:4.2.1.17	2	G3670.t1, G2223.t1
		dehydrogenase	ec:1.1.1.35	1	G2223.t1
Fatty acid elong	8	hydrolase	ec:3.1.2.22	2	G4913.t1, G4134.t1
		reductase (NADPH)	ec:1.3.1.38	2	G248.t1, G4193.t1
		C-acyltransferase	ec:2.3.1.16	1	G5221.t1
		hydrolase	ec:3.1.2.2	1	G3573.t1
		hydratase	ec:4.2.1.17	2	G3670.t1, G2223.t1
		dehydrogenase	ec:1.1.1.35	1	G2223.t1
Fatty acid biosy	15	carboxylase	ec:6.4.1.2	2	G3440.t1, G3923.t1
		reductase	ec:1.1.1.100	2	G1978.t1, G2163.t1
		dehydratase	ec:4.2.1.59	2	G940.t1, G2546.t1
		ligase	ec:6.2.1.3	5	G3884.t1, G3307.t1, G5088.t1, G655.t1, G1684.t1
		hydrolase	ec:3.1.2.14	1	G940.t1
		reductase (NADPH, Si-specific)	ec:1.3.1.10	3	G940.t1, G248.t1, G4193.t1
		reductase (NADH)	ec:1.3.1.9	1	G940.t1
		synthase	ec:2.3.1.86	2	G1978.t1, G940.t1
		S-malonyltransferase	ec:2.3.1.39	2	G940.t1, G1598.t1
		synthase I	ec:2.3.1.41	2	G1978.t1, G3257.t1
Peptidoglycan t	1	MraY transferase	ec:2.7.8.13	1	G2729.t1
Selenocompou	15	reductase	ec:1.8.1.9	2	G1915.t1, G2509.t1
		synthase	ec:2.1.1.13	1	G3226.t1
		S-methyltransferase	ec:2.1.1.14	1	G3226.t1
		beta-lyase	ec:4.4.1.8	3	G213.t1, G2503.t1, G3709.t1
		gamma-lyase	ec:4.4.1.1	1	G1346.t1
		beta-lyase	ec:4.4.1.13	1	G3709.t1
		adenylyltransferase	ec:2.7.7.4	1	G3581.t1
		gamma-synthase	ec:2.5.1.48	3	G3477.t1, G106.t1, G4869.t1

		ligase	ec:6.1.1.10	4	G2027.t1, G2412.t1, G2621.t1, G354.t1
Phenylpropano	1	lactoperoxidase	ec:1.11.1.7	1	G4.t1
Tyrosine metab	23	transaminase	ec:2.6.1.57	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		dehydrogenase [NAD(P)+]	ec:1.2.1.16	1	G4211.t1
		dehydrogenase	ec:1.1.1.1	6	G4043.t1, G152.t1, G2799.t1, G2935.t1, G3832.t1, G689.t1
		transaminase	ec:2.6.1.5	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		transaminase	ec:2.6.1.9	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		transaminase	ec:2.6.1.1	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		dehydrogenase [NAD(P)+]	ec:1.2.1.5	4	G4017.t1, G1733.t1, G434.t1, G3245.t1
		dehydrogenase	ec:1.1.1.90	7	G41.t1, G2396.t1, G28.t1, G410.t1, G5489.t1, G4314.t1, G1411.t1
Aflatoxin biosyn	2	carboxylase	ec:6.4.1.2	2	G3440.t1, G3923.t1
Tetracycline bic	2	carboxylase	ec:6.4.1.2	2	G3440.t1, G3923.t1
Alanine, aspart	42	transaminase (isomerizing)	ec:2.6.1.16	2	G4042.t1, G1006.t1
		transaminase	ec:2.6.1.19	1	G5337.t1
		ligase	ec:6.3.1.2	1	G521.t1
		dehydrogenase (NAD+)	ec:1.2.1.24	1	G4211.t1
		lyase	ec:4.3.2.1	1	G2080.t1
		lyase	ec:4.3.2.2	1	G1135.t1
		carbamoyltransferase	ec:2.1.3.2	2	G3494.t1, G3697.t1
		transaminase	ec:2.6.1.44	1	G3787.t1
		dehydrogenase [NAD(P)+]	ec:1.2.1.16	1	G4211.t1
		decarboxylase	ec:4.1.1.15	1	G3882.t1
		transaminase	ec:2.6.1.2	2	G4681.t1, G4940.t1
		transaminase	ec:2.6.1.1	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1

		synthase (glutamine- hydrolysing)	ec:6.3.5.4	3	G776.t1, G5243.t1, G96.t1
		synthase (glutamine- hydrolysing)	ec:6.3.5.5	3	G3494.t1, G3697.t1, G1671.t1
		dehydrogenase (NADP+)	ec:1.4.1.4	2	G1734.t1, G1298.t1
		dehydrogenase	ec:1.4.1.2	1	G793.t1
		glutaminase I	ec:3.5.1.2	12	G1042.t1, G896.t1, G4031.t1, G2539.t1, G4032.t1, G895.t1, G894.t1, G890.t1, G2724.t1, G893.t1, G26.t1, G1410.t1
		asparaginase II	ec:3.5.1.1	1	G2853.t1
		synthase	ec:6.3.4.4	1	G1508.t1
		synthase	ec:6.3.4.5	1	G2961.t1
		synthase (ammonia)	ec:6.3.4.16	2	G3494.t1, G3697.t1
		synthase (NADH)	ec:1.4.1.14	1	G2796.t1
		phosphoribosyl diphosphate 5- amidotransferase	ec:2.4.2.14	1	G3835.t1
Riboflavin meta	10	reductase	ec:1.1.1.193	1	G697.t1
		synthase	ec:4.1.99.12	1	G1838.t1
		kinase	ec:2.7.1.26	1	G4791.t1
		synthase	ec:2.5.1.9	1	G2715.t1
		synthetase	ec:2.7.7.2	2	G4791.t1, G4359.t1
		reductase [NAD(P)H]	ec:1.5.1.41	1	G243.t1
		reductase (NADPH)	ec:1.5.1.30	1	G243.t1
		reductase (NADPH)	ec:1.5.1.38	1	G5011.t1
		reductase [NAD(P)H]	ec:1.5.1.39	1	G5011.t1
		deaminase	ec:3.5.4.26	2	G4353.t1, G2953.t1
		cyclohydrolase II	ec:3.5.4.25	1	G4248.t1
Styrene degrad	2	acylamidase	ec:3.5.1.4	1	G4796.t1
		acetonitrilase	ec:3.5.5.1	1	G5224.t1
Ethylbenzene d	1	C- acyltransferase	ec:2.3.1.16	1	G5221.t1
Propanoate me	22	carboxylase	ec:6.4.1.2	2	G3440.t1, G3923.t1
		transaminase	ec:2.6.1.19	1	G5337.t1
		lyase	ec:4.1.3.30	2	G496.t1, G3253.t1

		ligase (ADP-forming)	ec:6.2.1.5	2	G2558.t1, G1528.t1
		ligase (GDP-forming)	ec:6.2.1.4	1	G1528.t1
		ligase	ec:6.2.1.1	2	G1305.t1, G4882.t1
		dehydratase	ec:4.2.1.79	1	G492.t1
		dehydrogenase	ec:1.8.1.4	2	G3776.t1, G474.t1
		C-acetyltransferase	ec:2.3.1.9	1	G464.t1
		synthase	ec:2.3.3.5	1	G491.t1
		reductase (NADPH)	ec:1.1.1.283	4	G236.t1, G1883.t1, G2876.t1, G5388.t1
		CoA-transferase	ec:2.8.3.8	1	G4230.t1
		hydrolase	ec:3.1.2.4	1	G4613.t1
		hydratase	ec:4.2.1.17	2	G3670.t1, G2223.t1
Ascorbate and	17	dehydrogenase (NAD ⁺)	ec:1.2.1.3	15	G1043.t1, G236.t1, G19.t1, G2101.t1, G2099.t1, G4017.t1, G4529.t1, G1733.t1, G3458.t1, G2019.t1, G434.t1, G3245.t1, G20.t1, G4211.t1, G5388.t1
		dehydrogenase (ascorbate)	ec:1.8.5.1	2	G2637.t1, G5459.t1
Galactose meta	39	phosphohexokinase	ec:2.7.1.11	2	G2561.t1, G3925.t1
		uridylyltransferase	ec:2.7.7.12	1	G4201.t1
		maltase	ec:3.2.1.20	15	G2662.t1, G2660.t1, G283.t1, G2442.t1, G2436.t1, G5229.t1, G2433.t1, G18.t1, G898.t1, G3456.t1, G902.t1, G2677.t1, G2870.t1, G2391.t1, G4777.t1
		invertase	ec:3.2.1.26	16	G2662.t1, G2660.t1, G283.t1, G2442.t1, G5223.t1, G2436.t1, G5229.t1, G2433.t1, G18.t1, G898.t1, G3456.t1, G270.t1, G902.t1, G2677.t1, G2870.t1, G2391.t1
		limit dextrinase (erroneous)	ec:3.2.1.10	15	G2662.t1, G2660.t1, G283.t1, G2442.t1, G2436.t1, G5229.t1, G2433.t1, G18.t1, G898.t1, G3456.t1, G270.t1, G902.t1, G2677.t1, G2870.t1, G2391.t1

		glucokinase (phosphorylating)	ec:2.7.1.2	5	G1056.t1, G155.t1, G3711.t1, G1862.t1, G868.t1
		galactokinase (phosphorylating)	ec:2.7.1.6	2	G4317.t1, G4199.t1
		1-epimerase	ec:5.1.3.3	3	G52.t1, G3387.t1, G4200.t1
		4-epimerase	ec:5.1.3.2	3	G52.t1, G3387.t1, G4200.t1
		uridylyltransferase	ec:2.7.7.9	2	G2180.t1, G564.t1
		(alpha-D-glucose-1,6-bisphosphate-dependent)	ec:5.4.2.2	3	G3856.t1, G4023.t1, G987.t1
		reductase	ec:1.1.1.21	4	G1902.t1, G4429.t1, G3115.t1, G3507.t1
Fructose and m	29	phosphofructokinase 2	ec:2.7.1.105	5	G5170.t1, G2285.t1, G2890.t1, G1147.t1, G3925.t1
		triose kinase	ec:2.7.1.28	2	G118.t1, G2393.t1
		phosphohexokinase	ec:2.7.1.11	2	G2561.t1, G3925.t1
		isomerase	ec:5.3.1.1	1	G4623.t1
		isomerase	ec:5.3.1.8	1	G3320.t1
		isomerase	ec:5.3.1.6	1	G3093.t1
		guanylyltransferase	ec:2.7.7.13	1	G4368.t1
		aldolase	ec:4.1.2.13	1	G2158.t1
		2-phosphatase	ec:3.1.3.46	2	G2285.t1, G1147.t1
		6-phosphatase	ec:3.1.3.54	1	G1832.t1
		mannokinase (phosphorylating)	ec:2.7.1.7	5	G1056.t1, G155.t1, G3711.t1, G1862.t1, G868.t1
		fructokinase (phosphorylating)	ec:2.7.1.4	5	G1056.t1, G155.t1, G3711.t1, G1862.t1, G868.t1
		hexose diphosphatase	ec:3.1.3.11	1	G1118.t1
		mannose phosphomutase	ec:5.4.2.8	1	G3802.t1
		2-dehydrogenase	ec:1.1.1.14	4	G3622.t1, G4958.t1, G1037.t1, G969.t1
		reductase	ec:1.1.1.21	4	G1902.t1, G4429.t1, G3115.t1, G3507.t1

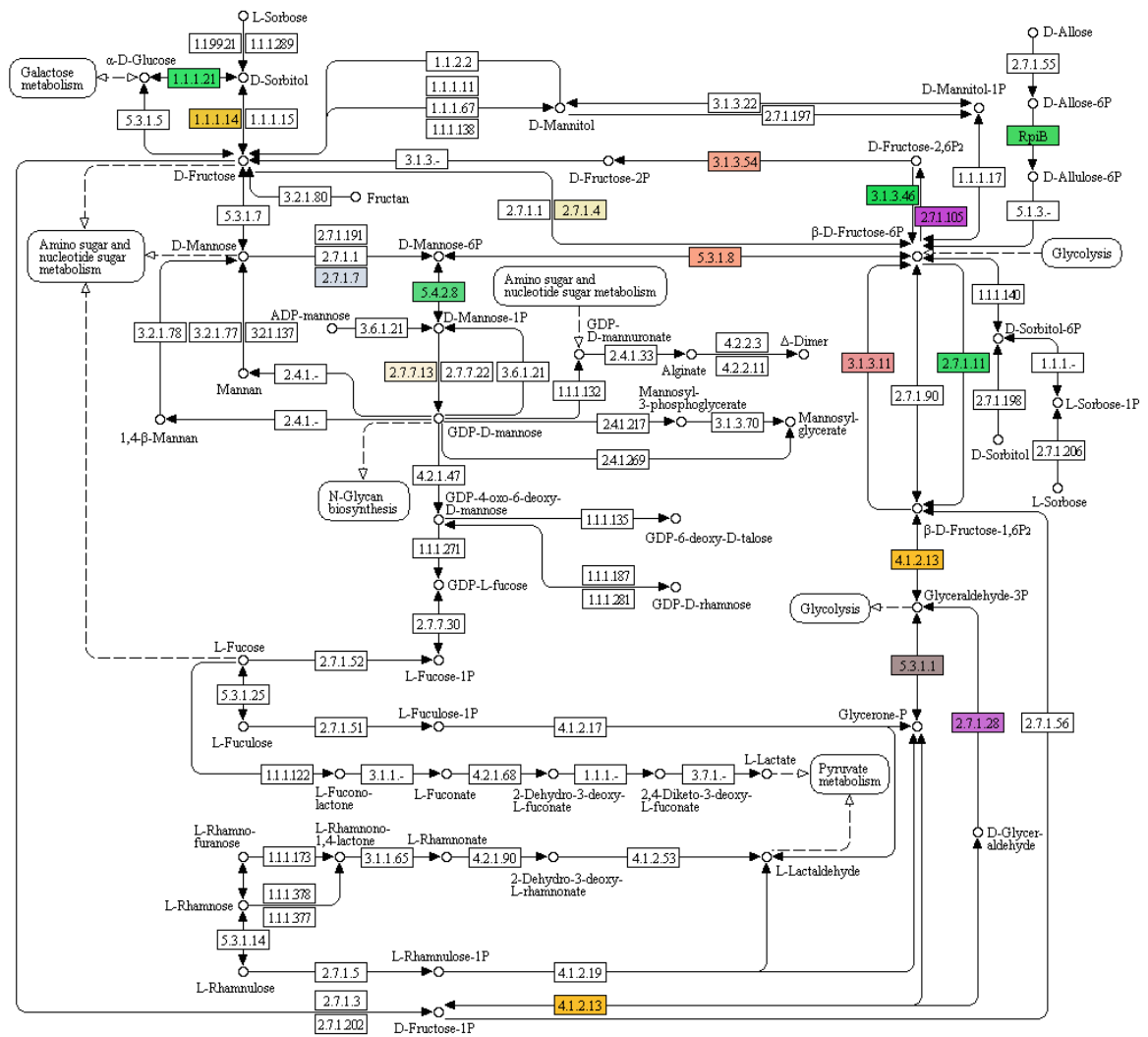
Biosynthesis of	2	glycolaldehydet ransferase	ec:2.2.1.1	2	G554.t1, G587.t1
Phosphonate a	4	EPT	ec:2.7.8.1	1	G2492.t1
		cholinesphosph otransferase	ec:2.7.8.2	2	G4525.t1, G2492.t1
		cytidyllyltransfer ase	ec:2.7.7.14	1	G5347.t1
		cytidyllyltransfer ase	ec:2.7.7.15	1	G2595.t1
Caprolactam de	7	dehydrogenase (NADP+)	ec:1.1.1.2	2	G3816.t1, G4312.t1
		dehydrogenase (NADP+)	ec:1.2.1.4	3	G1883.t1, G434.t1, G2876.t1
		hydratase	ec:4.2.1.17	2	G3670.t1, G2223.t1
		dehydrogenase	ec:1.1.1.35	1	G2223.t1
Histidine metab	26	transaminase	ec:2.6.1.9	5	G5177.t1, G1004.t1, G198.t1, G2479.t1, G4996.t1
		N- methyltransfera se	ec:2.1.1.22	1	G4492.t1
		dehydrogenase [NAD(P)+]	ec:1.2.1.5	4	G4017.t1, G1733.t1, G434.t1, G3245.t1
		dehydrogenase (NAD+)	ec:1.2.1.3	15	G1043.t1, G236.t1, G19.t1, G2101.t1, G2099.t1, G4017.t1, G4529.t1, G1733.t1, G3458.t1, G2019.t1, G434.t1, G3245.t1, G20.t1, G4211.t1, G5388.t1
		histidinol phosphate phosphatase	ec:3.1.3.15	1	G3736.t1
		diphosphatase	ec:3.6.1.31	1	G858.t1
		dehydratase	ec:4.2.1.19	1	G1581.t1
		phosphoribosyl transferase	ec:2.4.2.17	1	G3265.t1
		dehydrogenase	ec:1.1.1.23	1	G858.t1

		isomerase	ec:5.3.1.16	1	G5099.t1
		cyclohydrolase	ec:3.5.4.19	1	G858.t1
mTOR signaling	9	protein kinase	ec:2.7.11.11	3	G954.t1, G391.t1, G2294.t1
		protein kinase	ec:2.7.11.24	6	G536.t1, G5317.t1, G4918.t1, G4231.t1, G2092.t1, G959.t1
Retinol metabo	6	dehydrogenase	ec:1.1.1.1	6	G4043.t1, G152.t1, G2799.t1, G2935.t1, G3832.t1, G689.t1
Pyrimidine met	82	kinase	ec:2.7.4.6	1	G2152.t1
		carbamoylaspartic dehydrase	ec:3.5.2.3	3	G1080.t1, G3494.t1, G3697.t1
		kinase	ec:2.7.4.9	1	G3543.t1
		decarboxylase	ec:4.1.1.23	1	G3340.t1
		kinase	ec:2.7.1.48	3	G2592.t1, G4599.t1, G3444.t1
		deaminase	ec:3.5.4.1	1	G543.t1
		deaminase	ec:3.5.4.5	1	G1229.t1
		carbamoyltransferase	ec:2.1.3.2	2	G3494.t1, G3697.t1
		synthase	ec:4.2.1.70	1	G1295.t1
		reductase	ec:1.8.1.9	2	G1915.t1, G2509.t1
		synthase	ec:2.1.1.45	1	G3074.t1
		dehydrogenase (fumarate)	ec:1.3.98.1	1	G909.t1
		dehydrogenase (NADP+)	ec:1.3.1.2	1	G909.t1
		synthase (glutamine-hydrolysing)	ec:6.3.5.5	3	G3494.t1, G3697.t1, G1671.t1
		synthase (glutamine hydrolysing)	ec:6.3.4.2	2	G3500.t1, G4254.t1

		diphosphate phosphatase	ec:3.6.1.6	2	G3358.t1, G3318.t1
		reductase	ec:1.17.4.1	4	G3248.t1, G3608.t1, G5136.t1, G2613.t1
		diphosphatase	ec:3.6.1.12	1	G3531.t1
		diphosphatase	ec:3.6.1.19	8	G3454.t1, G608.t1, G609.t1, G3334.t1, G2655.t1, G4344.t1, G2390.t1, G818.t1
		diphosphatase	ec:3.6.1.23	2	G2720.t1, G3531.t1
		RNA polymerase	ec:2.7.7.6	2	G2273.t1, G1765.t1
		DNA polymerase	ec:2.7.7.7	38	G2359.t1, G4502.t1, G148.t1, G4815.t1, G3557.t1, G2361.t1, G1695.t1, G788.t1, G4688.t1, G3370.t1, G1925.t1, G1376.t1, G1707.t1, G275.t1, G4151.t1, G828.t1, G2054.t1, G1778.t1, G4250.t1, G1472.t1, G4409.t1, G2146.t1, G2127.t1, G147.t1, G2360.t1, G3585.t1, G1520.t1, G5491.t1, G2695.t1, G5200.t1, G787.t1, G4150.t1, G2020.t1, G60.t1, G1480.t1, G3.t1, G1408.t1, G360.t1

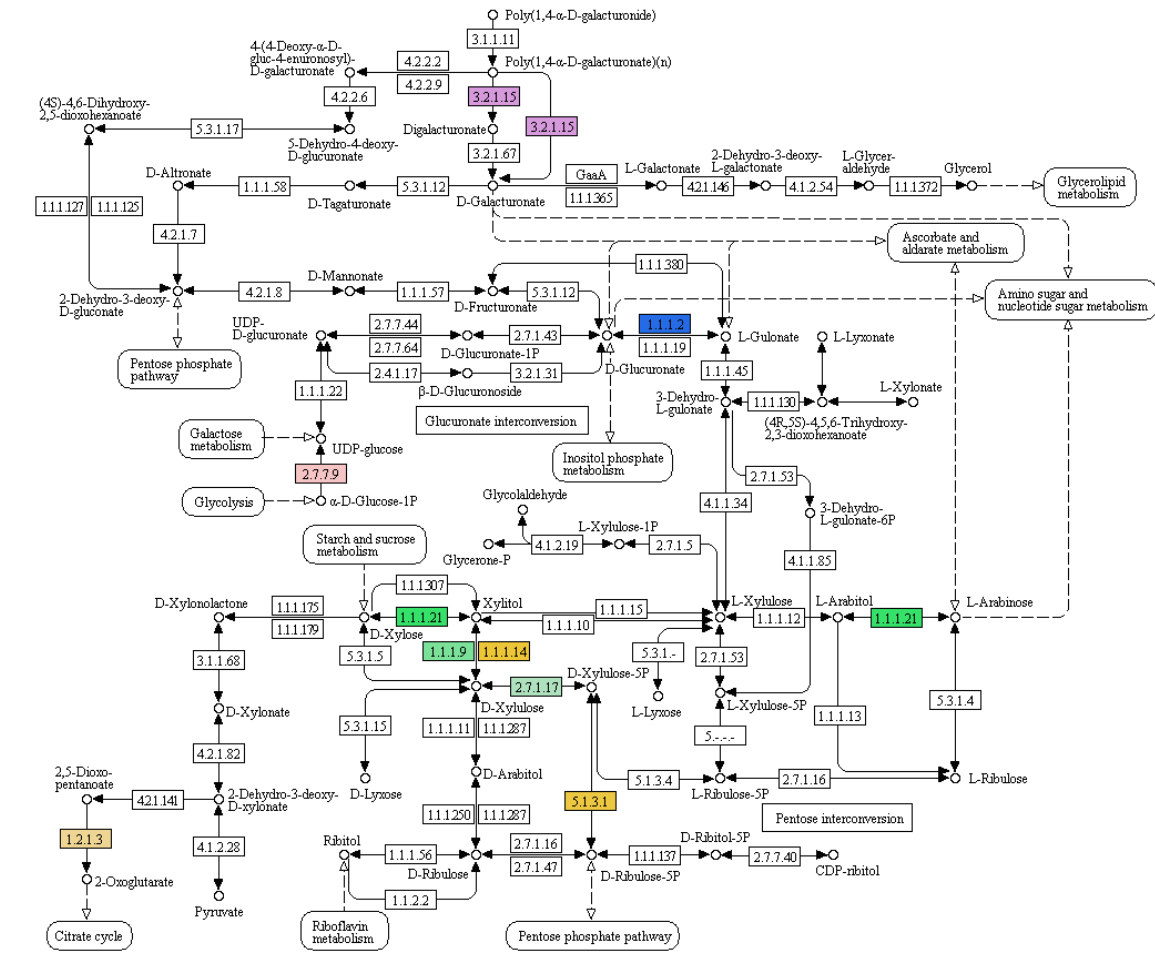
		nucleosidase	ec:3.2.2.3	1	G1762.t1
		phosphoribosyl transferase	ec:2.4.2.10	2	G3863.t1, G86.t1
		deaminase	ec:3.5.4.12	1	G2471.t1
		phosphoribosyl transferase	ec:2.4.2.9	1	G2488.t1
		kinase	ec:2.7.4.14	1	G2190.t1
		phosphorylase	ec:2.4.2.1	1	G1258.t1
Thiamine meta	16	kinase	ec:2.7.4.7	3	G2004.t1, G757.t1, G2964.t1
		kinase	ec:2.7.1.50	1	G401.t1
		kinase	ec:2.7.1.49	3	G2004.t1, G757.t1, G2964.t1
		diphosphokina se	ec:2.7.6.2	1	G1529.t1
		phosphate synthase	ec:2.5.1.3	1	G401.t1
		desulfurase	ec:2.8.1.7	2	G3272.t1, G848.t1
		phosphatase	ec:3.6.1.15	9	G3454.t1, G608.t1, G609.t1, G3334.t1, G3318.t1, G2655.t1, G4344.t1, G2390.t1, G818.t1
		aminohydrolas e	ec:3.5.99.2	3	G2004.t1, G757.t1, G2964.t1
Atrazine degrad	2	hydratase	ec:4.2.1.69	1	G891.t1
		hydrolase	ec:3.5.1.54	1	G2760.t1
		carboxylase	ec:6.3.4.6	1	G2760.t1
Folate biosynth	10	lyase	ec:4.1.3.38	1	G3845.t1
		synthase	ec:2.5.1.15	1	G1477.t1
		synthase	ec:2.6.1.85	1	G3423.t1
		diphosphokina se	ec:2.7.6.3	1	G1477.t1
		phosphatase	ec:3.1.3.1	2	G1832.t1, G33.t1
		reductase	ec:1.5.1.3	1	G1611.t1
		aldolase	ec:4.1.2.25	1	G1477.t1
		synthase	ec:6.3.2.12	2	G4014.t1, G983.t1
		synthase	ec:6.3.2.17	3	G1615.t1, G4014.t1, G983.t1
		cyclohydrolase I	ec:3.5.4.16	1	G2414.t1
Steroid hormon	3	dehydrogenase	ec:1.1.1.145	2	G5354.t1, G2876.t1
		sulfatase	ec:3.1.6.1	1	G5490.t1
Glyoxylate and	26	ligase	ec:6.3.1.2	1	G521.t1
		3-kinase	ec:2.7.1.31	1	G2592.t1
		ligase	ec:6.2.1.8	1	G2748.t1
		dehydrogenase	ec:1.2.1.2	2	G2400.t1, G1745.t1

		C-acetyltransferase	ec:2.3.1.9	1	G464.t1
		synthase	ec:2.3.3.9	2	G5050.t1, G4514.t1
		(Si)-synthase	ec:2.3.3.1	3	G835.t1, G1950.t1, G491.t1
		kynurenine formamidase	ec:3.5.1.9	2	G1786.t1, G3637.t1
		hydroxymethyltransferase	ec:2.1.2.1	2	G4969.t1, G2708.t1
		hydratase	ec:4.2.1.3	2	G2674.t1, G1182.t1
		lyase	ec:4.1.3.1	2	G496.t1, G3253.t1
		equilase	ec:1.11.1.6	2	G4809.t1, G5276.t1
		reductase (NADP+)	ec:1.1.1.79	2	G1462.t1, G311.t1
		reductase	ec:1.1.1.81	2	G1462.t1, G311.t1
		dehydrogenase	ec:1.1.1.37	3	G2899.t1, G4386.t1, G1024.t1
		reductase	ec:1.1.1.26	1	G1462.t1
Pentose and gl	30	xylulokinase (phosphorylating)	ec:2.7.1.17	1	G2602.t1
		dehydrogenase (NADP+)	ec:1.1.1.2	2	G3816.t1, G4312.t1
		reductase	ec:1.1.1.9	1	G4958.t1
		dehydrogenase (NAD+)	ec:1.2.1.3	15	G1043.t1, G236.t1, G19.t1, G2101.t1, G2099.t1, G4017.t1, G4529.t1, G1733.t1, G3458.t1, G2019.t1, G434.t1, G3245.t1, G20.t1, G4211.t1, G5388.t1
		pectin depolymerase	ec:3.2.1.15	1	G43.t1
		3-epimerase	ec:5.1.3.1	1	G3688.t1
		uridylyltransferase	ec:2.7.7.9	2	G2180.t1, G564.t1
		2-dehydrogenase	ec:1.1.1.14	4	G3622.t1, G4958.t1, G1037.t1, G969.t1
		reductase	ec:1.1.1.21	4	G1902.t1, G4429.t1, G3115.t1, G3507.t1
				2906	



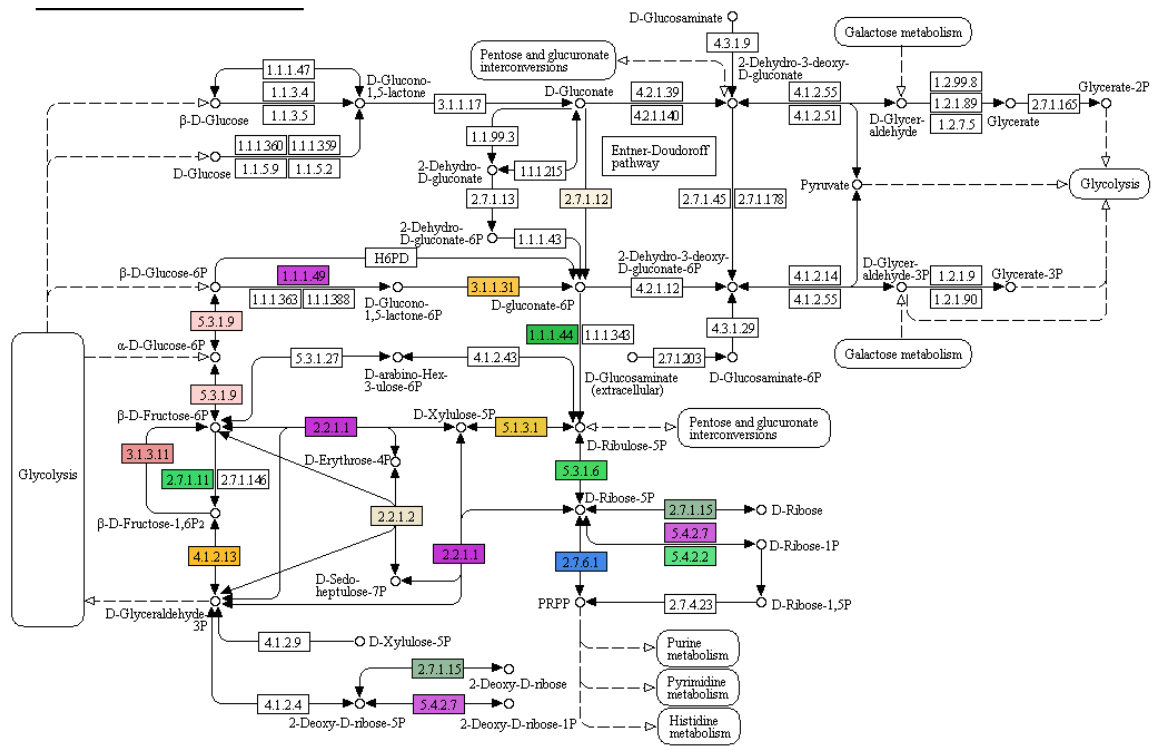
Color	Enzima	Secuencias
ec:2.7.1.105	- phosphofruktokinase 2	G5170.t1, G2890.t1, G2285.t1, G1147.t1, G3925.t1
ec:2.7.1.28	- triose kinase	G118.t1, G2393.t1
ec:2.7.1.11	- phosphohexokinase	G2561.t1, G3925.t1
ec:5.3.1.1	- isomerase	G4623.t1
ec:5.3.1.8	- isomerase	G3320.t1
ec:5.3.1.6	- isomerase	G3093.t1
ec:2.7.7.13	- guanilyltransferase	G4368.t1
ec:4.1.2.13	- aldolase	G2158.t1
ec:3.1.3.46	- 2-phosphatase	G2285.t1, G1147.t1
ec:3.1.3.54	- 6-phosphatase	G1832.t1
ec:2.7.1.7	- mannosinase (phosphorylating)	G3711.t1, G155.t1, G1056.t1, G1862.t1, G868.t1
ec:2.7.1.4	- fruktokinase (phosphorylating)	G3711.t1, G155.t1, G1056.t1, G1862.t1, G868.t1
ec:3.1.3.11	- hexose diphosphatase	G1118.t1
ec:5.4.2.8	- mannosine phosphomutase	G3802.t1
ec:1.1.1.14	- 2-dehydrogenase	G1037.t1, G4958.t1, G3622.t1, G969.t1,
ec:1.1.1.21	- reductase	G1902.t1, G2510.t1, G4429.t1, G3115.t1, G3507.t1

Figura 1. Enzimas mapeadas en el metabolismo de manosa y fructosa



Color	Enzima	Secuencias
Green	ec:2.7.1.17 - xylulokinase (phosphorylating)	G2602.t1
Blue	ec:1.1.1.2 - dehydrogenase (NADP+)	G3816.t1, G4312.t1
Green	ec:1.1.1.9 - reductase	G4958.t1
Orange	ec:1.2.1.3 - dehydrogenase (NAD+)	G1043.t1, G236.t1, G19.t1, G2101.t1, G2099.t1, G4017.t1, G4529.t1, G1733.t1, G3458.t1, G2019.t1
Purple	ec:3.2.1.15 - pectin depolymerase	G43.t1
Yellow	ec:5.1.3.1 - 3-epimerase	G3688.t1
Pink	ec:2.7.7.9 - uridylyltransferase	G2180.t1, G564.t1
Orange	ec:1.1.1.14 - 2-dehydrogenase	G1037.t1, G4958.t1, G3622.t1, G969.t1
Red	ec:1.1.1.21 - reductase	G1902.t1, G2510.t1, G4429.t1, G3115.t1, G3507.t1

Figura 2. Enzimas mapeadas en la interconversión de pentosa y gluconato.



Color	Enzima	Secuencias
	ec:2.2.1.2 - dihydroxyacetonttransferase	G5314.t1, G1139.t1
	ec:2.2.1.1 - glycolaldehydtransferase	G554.t1, G587.t1
	ec:2.7.1.15 - deoxyribokinase	G808.t1
	ec:2.7.1.12 - gluconokinase (phosphorylating)	G4802.t1
	ec:2.7.1.11 - phosphohexokinase	G2561.t1, G3925.t1
	ec:2.7.6.1 - diphosphokinase	G2958.t1, G563.t1, G3218.1, G941.t1, G4277.t1
	ec:5.3.1.6 - isomerase	G3093.t1
	ec:5.3.1.9 - isomerase	G2771.t1
	ec:4.1.2.13 - aldolase	G2158.t1
	ec:5.1.3.1 - 3-epimerase	G3688.t1
	ec:3.1.3.11 - hexose diphosphatase	G1118.t1
	ec:3.1.1.31 - phosphogluconolactonase	G5419.t1, G3422.t1, G2453.t1, G2554.t1
	ec:1.1.1.44 - dehydrogenase (NADP+-dependent, decarboxylating)	G2405.t1, G259.t1
	ec:1.1.1.49 - dehydrogenase (NADP+)	G149.t1
	ec:5.4.2.2 - (alpha-D-glucose-1,6-bisphosphate-dependent)	G3856.t1, G4023.t1, G987.t1
	ec:5.4.2.7 - phosphodeoxyribomutase	G3856.t1

Figura 3. Enzimas mapeadas en el metabolismo de pentosas fosfato.

Anexo 7: Mutaciones encontradas con MUDI para la cepa 202-3 respecto a la cepa S288c.

Gen	Cromosoma	Posición	Alelo de referencia	Alelo mutado	Descripción de la mutación
YHR104W (GRE3)	chr08	324023	T	G	Mutación sinónima Silenciosa
		324254	G	A	Mutación sinónima Silenciosa
		323398	ATTTTTT	ATTTTT	UPSTREAM
YOR120W (GKY1)	chr15	551437	A	G	Mutación no sinónima MISSENSE
		551234	A	G	Mutación sinónima Silenciosa
YDR368W (YPR1)	chr04	1213821	A	T	UPSTREAM
		1213825	C	A	UPSTREAM
		1214004	A	G	Mutación no sinónima MISSENSE
		1214210	T	C	DOWNSTREAM
		1214860	T	A	DOWNSTREAM
		1214842	AGCGCGC	AGCGC	Mutación sinónima Silenciosa
YJR096W	chr10	611484	G	A	Mutación no sinónima MISSENSE
		612120	T	C	DOWNSTREAM
		612069	C	T	DOWNSTREAM
		612134	G	A	DOWNSTREAM
YDL124W	chr04	240804	C	T	UPSTREAM
		240868	G	C	UPSTREAM
		240480	T	C	UPSTREAM
		241166	G	A	Mutación sinónima Silenciosa
		240408	G	A	Mutación sinónima Silenciosa

		240336	T	C	Mutación no sinónima [MISSENSE]
		240400	T	A	Mutación sinónima [Silenciosa]
		241206	T	C	Mutación no sinónima [MISSENSE]
		240506	C	A	Mutación no sinónima [MISSENSE]
		240647	T	C	Mutación no sinónima [MISSENSE]
		240657	A	C	Mutación no sinónima [MISSENSE]
		240375	G	C	Mutación sinónima [Silenciosa]
		240232	G	A	Mutación sinónima [Silenciosa]
		240532	C	A	Mutación no sinónima [MISSENSE]
		240185	C	T	Mutación no sinónima [MISSENSE]
		240564	T	G	Mutación sinónima [Silenciosa]
		240544	A	G	Mutación no sinónima [MISSENSE]
		240612	A	G	Mutación sinónima [Silenciosa]
		241065	A	G	Mutación no sinónima [MISSENSE]
		240258	G	A	DOWNSTREAM

		240610	C	T	Mutación sinónima [Silenciosa]
YJR159W (SOR1)	chr10	735992	G	A	UPSTREAM
		736655	C	T	Mutación sinónima [Silenciosa]
		736160	C	T	Mutación sinónima [Silenciosa]
		736916	A	C	Mutación sinónima [Silenciosa]
YLR070C (XYL2)	chr12	275287	C	T	Mutación no sinónima [MISSENSE]
		274580	G	C	DOWNSTREAM, UPSTREAM
		274651	G	C	Mutación sinónima [Silenciosa]
		275047	T	C	Mutación sinónima [Silenciosa]
		274098	T	C	DOWNSTREAM
		274115	A	G	DOWNSTREAM
		274243	G	C	DOWNSTREAM
		274041	A	G	Mutación no sinónima [MISSENSE]
		274059	AATATATATATA TATA	AATATATATAT ATATATA	DOWNSTREAM
YGR194C (XKS1)	chr07	887091	T	C	Mutación sinónima [Silenciosa]
		886772	C	T	Mutación no sinónima [MISSENSE]

		886758	C	T	Mutación no sinónima [MISSENSE]
		887135	T	C	Mutación sinónima [Silenciosa]
		886023	T	C	DOWNSTREAM
		886742	C	T	Mutación sinónima [Silenciosa]
		887698	G	T	Mutación no sinónima [MISSENSE]
		887966	C	A	Mutación no sinónima [MISSENSE]
		887484	T	C	Mutación sinónima [Silenciosa]
		886106	G	A	UPSTREAM
		887887	T	TA	UPSTREAM
		887882	A	G	UPSTREAM
		886003	AT	ATGT	DOWNSTREAM
YLR354C (TAL1)	chr12	836339	G	A	DOWNSTREAM
YPR074C (TKL1)	chr16	693600	C	T	Mutación sinónima [Silenciosa]
		693603	A	G	Mutación sinónima [Silenciosa]
		693567	T	G	Mutación sinónima [Silenciosa]
		694413	C	T	Mutación sinónima [Silenciosa]
		694716	T	C	Mutación sinónima [Silenciosa]

		693141	C	T	Mutación sinónima [Silenciosa]
		694434	C	A	Mutación sinónima [Silenciosa]
		693774	G	A	Mutación sinónima [Silenciosa]
		693795	A	G	Mutación sinónima [Silenciosa]
TRANSPORTADORES					
YHR094C (HXT1)	chr08	291084	T	A	Mutación sinónima [Silenciosa]
		292176	T	G	Mutación sinónima [Silenciosa]
		292435	G	A	Mutación sinónima [Silenciosa]
		292444	A	G	Mutación no sinónima [MISSENSE]
		292143	T	G	Mutación no sinónima [MISSENSE]
		291726	G	A	Mutación no sinónima [MISSENSE]
		291731	T	C	Mutación sinónima [Silenciosa]
		291987	A	C	Mutación sinónima [Silenciosa]
		291315	A	G	Mutación sinónima [Silenciosa]
		291312	G	A	Mutación sinónima [Silenciosa]

		291318	G	A	Mutación sinónima [Silenciosa]
		291471	G	A	Mutación sinónima [Silenciosa]
		291411	G	A	Mutación sinónima [Silenciosa]
		290985	G	A	Mutación sinónima [Silenciosa]
		291399	A	G	Mutación sinónima [Silenciosa]
		291396	G	A	Mutación sinónima [Silenciosa]
		291327	T	G	Mutación sinónima [Silenciosa]
		291381	G	A	Mutación sinónima [Silenciosa]
YMR011 W (HXT2)	chr13	289726	CT	CTTT	Mutación sinónima [Silenciosa]
		288558	A	G	Mutación sinónima [Silenciosa]
		289227	T	C	Mutación sinónima [Silenciosa]
		288433	T	A	Mutación no sinónima [MISSENSE]
		289416	C	T	UPSTREAM
		288031	T	C	UPSTREAM
		288024	T	C	DOWNSTREAM

		289183	T	C	Mutación sinónima [Silenciosa]
YDR345C (HXT3)	chr04	1164250	G	A	Mutación sinónima [Silenciosa]
		1164181	G	A	Mutación sinónima [Silenciosa]
		1163713	C	T	Mutación sinónima [Silenciosa]
		1163806	G	A	Mutación sinónima [Silenciosa]
		1163436	G	A	Mutación no sinónima [MISSENSE]
		1163794	C	T	Mutación sinónima [Silenciosa]
		1163209	A	G	Mutación sinónima [Silenciosa]
		1163271	T	C	Mutación sinónima [Silenciosa]
		1163487	T	C	Mutación no sinónima [MISSENSE]
		1163502	A	G	UPSTREAM
		1164726	A	C	Mutación sinónima [Silenciosa]
		1164487	T	G	Mutación sinónima [Silenciosa]
		1163371	G	T	Mutación sinónima [Silenciosa]
		1164502	A	G	Mutación sinónima [Silenciosa]

		1163374	A	G	Mutación sinónima [Silenciosa]
		1163378	AC	A	FRAME_SHIFT
YHR092C (HXT4)	chr08	287702	A	G	Mutación sinónima [Silenciosa]
		287699	G	A	Mutación sinónima [Silenciosa]
		288892	C	T	UPSTREAM
		288884	T	G	UPSTREAM
		287468	A	G	Mutación sinónima [Silenciosa]
		288389	T	C	Mutación sinónima [Silenciosa]
		288392	T	G	Mutación sinónima [Silenciosa]
		288388	A	G	Mutación sinónima [Silenciosa]
		287678	A	G	Mutación sinónima [Silenciosa]
		287675	G	C	Mutación sinónima [Silenciosa]
		287666	A	G	Mutación sinónima [Silenciosa]
		287669	A	C	Mutación no sinónima [MISSENSE]
		287671	A	T	Mutación no sinónima [MISSENSE]
		YHR096C (HXT5)	chr08	296046	G

		294915	G	A	Mutación sinónima [Silenciosa]
		295200	T	G	Mutación sinónima [Silenciosa]
		294900	G	A	Mutación sinónima [Silenciosa]
		295445	A	G	Mutación sinónima [Silenciosa]
		295470	C	T	Mutación sinónima [Silenciosa]
		295376	G	A	Mutación sinónima [Silenciosa]
		294768	T	C	Mutación sinónima [Silenciosa]
		296436	A	G	Mutación sinónima [Silenciosa]
		295620	G	A	Mutación sinónima [Silenciosa]
		296229	T	G	Mutación no sinónima [MISSENSE]
YDR343C (HXT6)	chr04	1161404	GTTTTTTT	GTTTTTTT	UPSTREAM
		1159689	A	G	Mutación sinónima [Silenciosa]
		1159797	A	G	Mutación sinónima [Silenciosa]
		1159860	A	G	Mutación sinónima [Silenciosa]
		1161147	T	A	Mutación sinónima [Silenciosa]
		1161206	G	T	Mutación no sinónima [MISSENSE]

		1161300	C	T	Mutación sinónima [Silenciosa]
		1161302	T	C	Mutación no sinónima [MISSENSE]
		1161213	C	T	Mutación sinónima [Silenciosa]
		1161214	A	G	Mutación sinónima [Silenciosa]
		1161289	T	C	UPSTREAM
		1159713	T	C	UPSTREAM
		1161282	T	A	UPSTREAM
		1161283	G	A	UPSTREAM
		1161228	T	G	UPSTREAM
		1161225	G	T	UPSTREAM
		1161230	A	G	UPSTREAM
		1161053	G	A	Mutación sinónima [Silenciosa]
YDR342C (HXT7)	chr04	1154321	C	T	Mutación sinónima [Silenciosa]
		1154583	T	A	Mutación no sinónima MISSENSE
		1154468	A	G	Mutación sinónima Silenciosa
		1154539	C	A	Mutación no sinónima [MISSENSE]

1154538	G	C	Mutación no sinónima [MISSENSE]
1154576	A	G	Mutación sinónima [Silenciosa]
1154574	A	T	Mutación no sinónima [MISSENSE]
1154573	A	G	Mutación sinónima [Silenciosa]
1154575	C	T	Mutación no sinónima [MISSENSE]
1154405	A	G	Mutación no sinónima [MISSENSE]
1154411	G	A	Mutación sinónima [Silenciosa]
1154412	A	T	Mutación sinónima [Silenciosa]
1154507	C	G	Mutación no sinónima [MISSENSE]
1154509	A	T	Mutación sinónima [Silenciosa]
1154523	G	T	Mutación no sinónima [MISSENSE]
1154525	A	G	Mutación no sinónima [MISSENSE]
1154528	A	G	Mutación sinónima [Silenciosa]

		1154513	A	G	Mutación no sinónima [MISSENSE]
		1154518	C	A	Mutación no sinónima [MISSENSE]
		1154524	T	G	Mutación sinónima [Silenciosa]
		1154517	G	C	Mutación no sinónima [MISSENSE]
		1155755	T	A	Mutación sinónima [Silenciosa]
		1155785	G	T	Mutación no sinónima [MISSENSE]
		1155661	T	C	Mutación no sinónima [MISSENSE]
		1155623	C	T	Mutación sinónima [Silenciosa]
		1155625	T	G	Mutación no sinónima [MISSENSE]
		1155620	A	G	Mutación sinónima [Silenciosa]
		1154783	C	A	Mutación no sinónima [MISSENSE]
		1155607	T	A	Mutación no sinónima [MISSENSE]

		1155571	A	C	Mutación no sinónima [MISSENSE]
		1155578	C	T	Mutación sinónima [Silenciosa]
		1155569	G	A	Mutación sinónima [Silenciosa]
YJL214W (HXT8)	chr10	27942	C	T	Mutación sinónima [Silenciosa]
		27906	C	A	Mutación sinónima [Silenciosa]
		27894	G	C	Mutación sinónima [Silenciosa]
		28178	C	T	Mutación sinónima [Silenciosa]
		28167	T	C	Mutación no sinónima [MISSENSE]
		28147	G	C	Mutación no sinónima [MISSENSE]
		27376	G	A	Mutación no sinónima [MISSENSE]
		27780	T	C	Mutación sinónima [Silenciosa]
		28086	G	A	Mutación sinónima [Silenciosa]
		27405	T	C	Mutación sinónima [Silenciosa]
		27318	C	T	Mutación sinónima [Silenciosa]

		27330	G	A	Mutación sinónima [Silenciosa]
		27417	T	C	Mutación sinónima [Silenciosa]
		27429	A	T	Mutación sinónima [Silenciosa]
		27420	C	T	Mutación sinónima [Silenciosa]
		27435	T	C	Mutación sinónima [Silenciosa]
		28422	T	C	Mutación sinónima [Silenciosa]
		26958	G	A	Mutación sinónima [Silenciosa]
YJL219W (HXT9)	chr10	20360	A	G	Mutación sinónima [Silenciosa]
		20300	G	C	Mutación sinónima [Silenciosa]
		19679	C	T	Mutación sinónima [Silenciosa]
		19685	C	T	Mutación sinónima [Silenciosa]
		20771	C	T	Mutación sinónima [Silenciosa]
		19970	T	C	Mutación sinónima [Silenciosa]
		20708	T	A	Mutación sinónima [Silenciosa]
		20718	T	C	Mutación sinónima [Silenciosa]

19946	T	C	Mutación sinónima [Silenciosa]
20021	A	C	Mutación sinónima [Silenciosa]
20279	G	A	Mutación sinónima [Silenciosa]
20045	T	C	Mutación sinónima [Silenciosa]
20048	C	T	Mutación sinónima [Silenciosa]
20276	T	C	Mutación sinónima [Silenciosa]
19709	C	A	Mutación sinónima [Silenciosa]
19871	C	A	Mutación no sinónima [MISSENSE]
20081	A	T	Mutación sinónima [Silenciosa]
20118	T	G	Mutación no sinónima [MISSENSE]
20525	C	T	Mutación sinónima [Silenciosa]
20240	A	G	Mutación sinónima [Silenciosa]
19766	T	C	Mutación sinónima [Silenciosa]
19769	C	T	Mutación sinónima [Silenciosa]
20198	C	T	Mutación sinónima [Silenciosa]

		20504	T	C	Mutación no sinónima [MISSENSE]
		20216	C	T	Mutación sinónima [Silenciosa]
		20220	A	G	Mutación sinónima [Silenciosa]
YFL011W (HXT10)	chr06	113247	T	C	Mutación sinónima [Silenciosa]
		113328	C	T	Mutación sinónima [Silenciosa]
		113181	T	A	Mutación sinónima [Silenciosa]
		113778	T	C	Mutación sinónima [Silenciosa]
		112624	C	T	Mutación no sinónima [MISSENSE]
		113586	C	A	Mutación sinónima [Silenciosa]
		113799	C	T	UPSTREAM
		112607	G	A	Mutación no sinónima [MISSENSE]
		112326	CAAAAAAA	CAAAAAAAA,CAA AAAAAAA	Mutación sinónima [Silenciosa]
		112984	T	C	Mutación sinónima [Silenciosa]

		113910	A	G	Mutación sinónima [Silenciosa]
		112349	T	C	UPSTREAM
		112269	C	A	Mutación sinónima [Silenciosa]
		113124	G	A	Mutación no sinónima [MISSENSE]
		113076	C	A	Mutación sinónima [Silenciosa]
		112425	A	G	Mutación sinónima [Silenciosa]
		112435	A	G	Mutación no sinónima [MISSENSE]
		113011	T	C	Mutación sinónima [Silenciosa]
YOL156W	chr15	26289	C	T	Mutación sinónima [Silenciosa]
		25974	C	T	Mutación sinónima [Silenciosa]
		26211	T	A	Mutación sinónima [Silenciosa]
		25318	A	G	Mutación sinónima [Silenciosa]
		25319	T	A	Mutación no sinónima [MISSENSE]
		25917	A	T	Mutación no sinónima [MISSENSE]

		25350	C	T	Mutación sinónima [Silenciosa]
		25746	T	C	Mutación sinónima [Silenciosa]
		25365	A	G	Mutación sinónima [Silenciosa]
		25821	T	C	Mutación sinónima [Silenciosa]
		25797	A	C	Mutación sinónima [Silenciosa]
		25795	C	T	Mutación sinónima [Silenciosa]
		25824	C	T	Mutación sinónima [Silenciosa]
		26547	C	T	Mutación sinónima [Silenciosa]
		25809	C	A	Mutación sinónima [Silenciosa]
		25869	T	C	Mutación sinónima [Silenciosa]
		26112	T	C	Mutación sinónima [Silenciosa]
		26106	C	T	Mutación sinónima [Silenciosa]
		26586	C	T	Mutación sinónima [Silenciosa]

		26595	T	C	Mutación sinónima [Silenciosa]
		26487	G	A	Mutación sinónima [Silenciosa]
		26484	T	A	Mutación sinónima [Silenciosa]
		26494	T	C	Mutación sinónima [Silenciosa]
		26088	A	C	Mutación no sinónima [MISSENSE]
YIL170W (HXT12)	chr09	20110	A	C	EXON
		20299	A	G	EXON
		20296	C	T	EXON
		20236	C	T	EXON
		20320	G	C	EXON
		20325	A	G	EXON
		20260	G	A	EXON
		20356	T	C	EXON
		20332	A	C	EXON
		19803	CAAAAAAAAA	CAAAAAA	EXON
		20494	A	G	EXON,UPSTREAM
		20889	C	T	EXON
		21185	A	G	EXON
		21151	C	T	EXON
		21234	G	A	EXON
21133	G	A	DOWNSTREAM		

		21121	G	A	EXON
		21239	A	G	DOWNSTREAM
		21273	C	T	DOWNSTREAM
		21252	C	T	DOWNSTREAM
		21145	A	G	DOWNSTREAM
		21279	C	T	EXON
		21142	G	A	DOWNSTREAM
		21139	G	A	DOWNSTREAM
		21251	G	C	EXON
		21289	C	T	EXON
		21264	T	A	DOWNSTREAM
		21250	A	G	DOWNSTREAM
YEL069C (HXT13)	chr05	21849	T	C	Mutación sinónima [Silenciosa]
		21876	C	T	Mutación sinónima [Silenciosa]
		22493	A	G	Mutación sinónima [Silenciosa]
		22143	G	A	Mutación sinónima [Silenciosa]a
		21532	T	C	DOWNSTREAM
		21762	C	A	Mutación sinónima [Silenciosa]
		22542	G	C	Mutación sinónima [Silenciosa]
		22221	G	A	Mutación sinónima [Silenciosa]

		21741	A	T	Mutación sinónima [Silenciosa]
		22158	G	A	Mutación sinónima [Silenciosa]
		22191	G	A	Mutación sinónima [Silenciosa]
		22188	A	C	Mutación sinónima [Silenciosa]
		22200	A	G	Mutación sinónima [Silenciosa]
		21951	A	G	Mutación sinónima [Silenciosa]
		21954	A	T	Mutación sinónima [Silenciosa]
		22614	C	G	DOWNSTREAM
		21440	T	G	Mutación sinónima [Silenciosa]
		22239	A	G	Mutación sinónima [Silenciosa]
		22434	A	G	Mutación sinónima [Silenciosa]
		22302	T	G	Mutación sinónima [Silenciosa]
		22305	A	G	Mutación sinónima [Silenciosa]
		22863	G	A	Mutación sinónima [Silenciosa]
		23322	C	G	UPSTREAM

		22764	T	C	Mutación sinónima Silenciosa
		23309	A	G	UPSTREAM
		22722	G	A	Mutación sinónima Silenciosa
		22746	C	T	Mutación sinónima Silenciosa
		22728	G	A	Mutación sinónima Silenciosa
		23283	GCC	GC	UPSTREAM
		23230	ATTTTTTTTTTTTT	ATTTTTTTTTTTTT,A TTTTTTTTTTTT	UPSTREAM,UPSTREAM
YNL318C (HXT14)	chr14	40397	A	G	Mutación sinónima Silenciosa
		39421	A	G	UPSTREAM
		39865	C	T	Mutación sinónima Silenciosa
		39122	T	C	DOWNSTREAM 40 YNL3 19W YNL319W,Mutación no sinónima MISSENSE
		38800	G	A	Mutación sinónima Silenciosa ggA/ggT G41 Y NL319W YNL319W,Mutaci ón sinónima Silenciosa

		38764	A	T	Mutación sinónima [Silenciosa]agG/agA R53 Y NL319W YNL319W,Mutaci ón sinónima [Silenciosa]
YDL245C	chr04	11807	C	T	Mutación sinónima [Silenciosa]
		12146	T	C	Mutación sinónima [Silenciosa]
		12110	G	A	Mutación sinónima [Silenciosa]
		12113	T	G	Mutación sinónima [Silenciosa]
		12158	T	C	Mutación sinónima [Silenciosa]
		12074	T	A	Mutación sinónima [Silenciosa]
		12068	G	A	Mutación sinónima [Silenciosa]
		12050	T	C	Mutación sinónima [Silenciosa]
		11587	C	T	DOWNSTREAM
		12613	A	G	Mutación sinónima [Silenciosa]
		12884	C	T	Mutación sinónima [Silenciosa]
		11608	T	C	DOWNSTREAM
11927	G	A	Mutación sinónima [Silenciosa]		

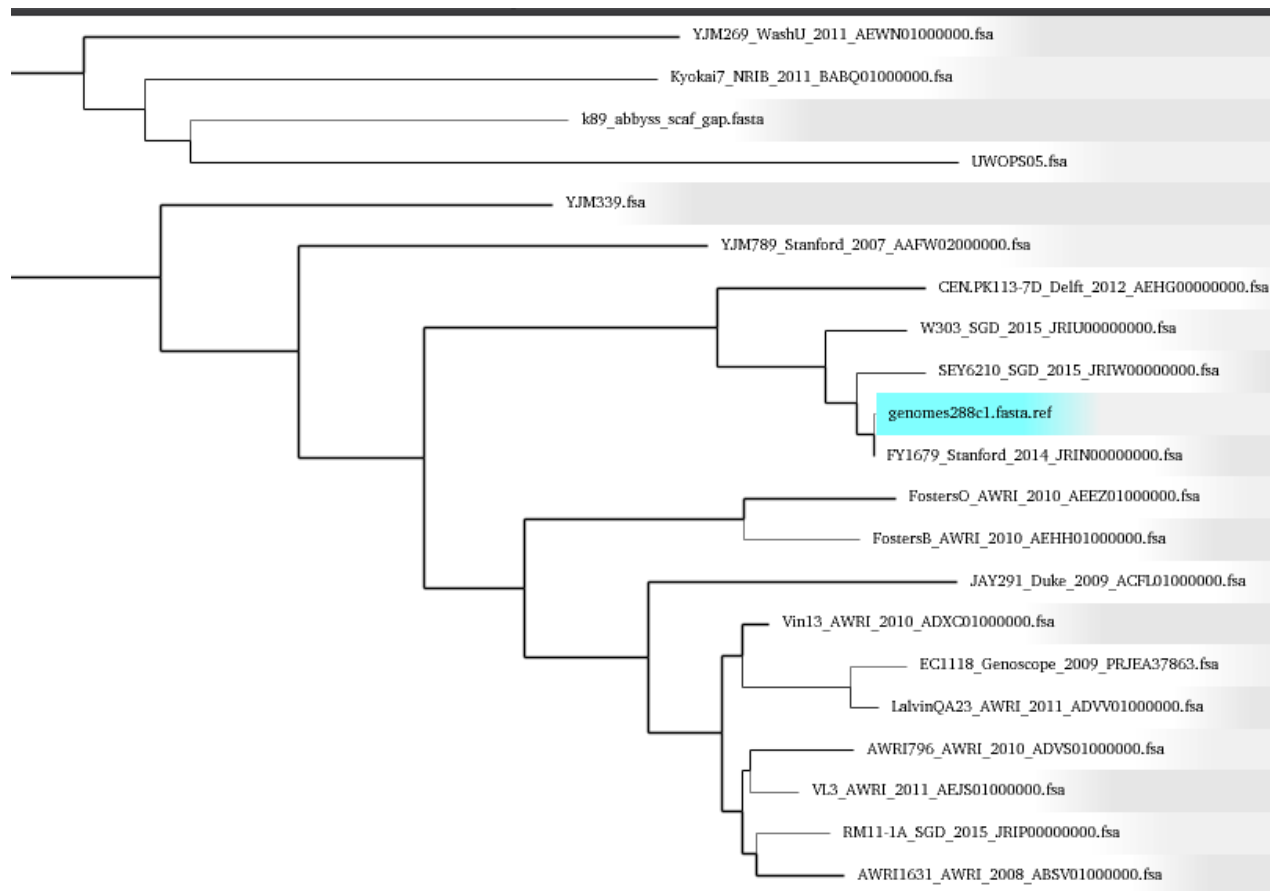
		12689	C	T	Mutación sinónima [Silenciosa]
		11696	G	A	Mutación sinónima [Silenciosa]
		11640	T	C	DOWNSTREAM
		13052	C	T	Mutación sinónima [Silenciosa]
		11664	A	G	Mutación no sinónima [MISSENSE]
		11653	T	A	DOWNSTREAM
		11654	T	C	DOWNSTREAM
		11657	T	C	STOP_LOST[MISSENSE]
		11658	C	T	Sinónima_STOP[Silenciosa]
YLR081W (GAL2)	chr12	291267	C	T	Mutación sinónima[Silenciosa]
		291216	G	A	Mutación sinónima[Silenciosa]
		290159	T	C	UPSTREAM
		290922	C	T	Mutación sinónima[Silenciosa]
		291459	T	C	Mutación sinónima[Silenciosa]
		290328	T	C	Mutación sinónima[Silenciosa]

		291386	A	G	Mutación no sinónima MISSENSE
		291384	A	G	Mutación sinónima Silenciosa
		290359	T	C	Mutación no sinónima MISSENSE
		290421	C	T	Mutación sinónima Silenciosa
YHR182C -A (GND1)	chr08	471466	C	T	Mutación sinónima Silenciosa,UPST REAM
YNL241C (ZWF1)	chr14	196819	T	C	Mutación sinónima Silenciosa
		197191	T	C	Mutación sinónima Silenciosa
		197710	G	A	Mutación sinónima Silenciosa
		196345	G	A	DOWNSTREAM
		197080	G	A	Mutación sinónima Silenciosa
		197788	C	A	Mutación sinónima Silenciosa
		196326	T	G	DOWNSTREAM
		197764	GTCCTCCTCC	GTCCTCC	CODON_DELETION
		196374	GAAAAAAAAA	GAAAAAAAAA,GA AAAAAAA	DOWNSTREAM,DOWNST REAM

YJL121C	chr10	190302	T	A	DOWNSTREAM,DOWNST REAM
		190443	C	T	Mutación sinónima Silenciosa
		190451	C	T	Mutación no sinónima MISSENSE
		190584	T	C	Mutación sinónima Silenciosa

Anexo 8: Filogenía para algunas cepas de *S. cerevisiae*

Mediante la plataforma de Harvest (Treangen, 2014), con el uso de la herramienta Parsnp y Gingr. En este caso el alineamiento combina alineamiento de todo el genoma y mapeo de lecturas, mediante el uso de MUMs, para finalmente contruir un árbol basado en los SNPs encontrados en las cepas con FastTree2 (aproximación de máxima verosimilitud) (Treangen, 2014). La cepa de estudio *S. cerevisiae* 202-3, corresponde a la asignada como k89_abyss_scaf_gap.fasta.



Anexo 9: Publicaciones y participaciones en eventos

Trabajos asociados al estudio fenotípico de las cepas *S. cerevisiae* aisladas:
1.1. European Biotechnology Congress 2014, Italia, Lecce: **Effect of inorganic nutrients supplementation in sugar cane molasses on fusel alcohols production by a *Saccharomyces cerevisiae* native strain.** *Journal of Biotechnology*, Volume 185, Supplement, September 2014, Pages S123 European Biotechnology Congress 2014.

<http://dx.doi.org/10.1016/j.jbiotec.2014.07.419>

Effect of inorganic nutrients supplementation in sugar cane molasses on fusel alcohols production by a *Saccharomyces cerevisiae* native strain



Juan Pablo Ortiz Rosas*, Yina Alejandra Cifuentes Triana, Mario Enrique Velasquez Lozano

Department of Chemical and Environmental Engineering, National University of Colombia, Bogotá, Colombia

E-mail address: jportizr@unal.edu.co (J.P.O. Rosas).

Fusel alcohols include mainly 1-propanol, isobutanol and amyl alcohols. They are usually produced in small amounts during alcoholic fermentation by *Saccharomyces cerevisiae*. Recently, isobutanol has obtained great attention because it could be an alternative drop in fuel and it could be use as building block molecule. In order to increase isobutanol production, inorganic nutrients supplementation effects on fermentation were evaluated. One 2⁵ complete factorial design with 2 central points per block and four blocks was performed. Experimental variables were (NH₄)₂SO₄, KH₂PO₄, ZnSO₄, MgSO₄ and MnSO₄ concentrations. Sugar cane molasses were diluted to 16 °Bx (150 g/L sucrose). The experiments were carried out in 50 mL bottles, with an airlock to guarantee anaerobiosis and incubated at 30 °C for 48 h and magnetically agitated. Supplementation with potassium phosphate stimulated isobutanol and isoamyl alcohol production. Ammonium sulfate had negative effect. Ammonium sulfate and potassium phosphate increased propanol production. Other nutrients showed little effect on fermentation products (*P* > 0.05).

- 1.2. European Biotechnology Congress 2014, Italia, Lecce: **Production of ethanol through very high gravity fermentation by *Saccharomyces cerevisiae* wild strains**, *Journal of Biotechnology*, Volume **185**, Supplement, September 2014, Pages S123 European Biotechnology Congress 2014.

<http://dx.doi.org/10.1016/j.jbiotec.2014.07.421>

Production of ethanol through very high gravity fermentation by *Saccharomyces cerevisiae* wild strains



Carlos Rafael Castillo Saldarriaga, Mario Enrique Velasquez Lozano*, Yina Cifuentes Triana

Department of Chemical and Environmental Engineering, Universidad Nacional de Colombia, Bogota, Colombia

E-mail address: mevelasquezl@gmail.com (M.E.V. Lozano).

The objective of this study was to select a wild yeast strain with a good performance in very high gravity (VHG) fermentation. A number of yeast strains were isolated from a fermented sugar cane juice and sugar cane cutting mill in an ethanol production company. Interdelta analysis with primer pair (delta12-delta21) was done to confirm the differences between the isolated strains previously identified as *Saccharomyces cerevisiae* with ITS method. VHG fermentation performance was evaluated using YPD modified medium with four different initial glucose concentration: 180 g/L, 250 g/L, 300 g/L and 350 g/L. Ethanol Red Fermentis®, a VHG commercial yeast was used as a control. Specific growth rate, ethanol concentration, yield and productivity were determined. Wild yeast strain, 119-3, was the most promising with $0.305 \pm 0.009 \text{ h}^{-1}$ specific growth rate and $31.53 \pm 0.07 \text{ g/L}$ ethanol concentration with a corresponding yield and productivity of 0.09 g ethanol/g glucose and 4.50 g/L/h, respectively.

- 1.3. Tokyo International Conference on Life Science and Biological Engineering, 2014, Tokyo, Japón: **Fermentation of Enzymatic Hydrolysates of Sugar Cane Bagasse by a Colombian Native Strain of *Saccharomyces cerevisiae* for the Production of Cellulosic Ethanol**. ISBN: 978-986-5654-04-7, 2014. p. 287.

ILSBE-46

Fermentation of Enzymatic Hydrolysates of Sugar Cane Bagasse by a Colombian Native Strain of *Saccharomyces Cerevisiae* for the Production of Cellulosic Ethanol

**Diana Morales-Fonseca^a, Yina Cifuentes-Triana^b Richard Ruiz-Merchan^c,
Jose Montoya-Baca^d and Mario Velásquez-Lozano^{e*}**

^{a,c,d,e}Department of Environmental and Chemical Engineering, Universidad Nacional de Colombia, Bogotá, Colombia.

^bDepartment of Chemistry, Universidad Nacional de Colombia, Bogotá, Colombia.

*Corresponding Author: mevelasquezl@unal.edu.co

ABSTRACT

In this study, ethanol production was evaluated from enzymatic hydrolysates of sugarcane bagasse by a group of native *Saccharomyces cerevisiae* strains isolated from sugar cane juice. The biomass was pretreated with dilute sulfuric acid (2% w/w) at 170°C and subjected to fed-batch enzymatic hydrolysis with a solid loading 17 % w/w. The hydrolysate fermentation performance by the native yeasts was carried out with an initial total reducing sugar concentration: 100 g/l and low concentration inhibitors (0.1g/l hidroxymethylfurfural, 1.0 g/l furfural and 6 g/l acetic acid).

The native group of yeasts was characterized by the ITS regions and Interdelta profiles to discriminate at a species and subspecies level. The commercial yeast Ethanol Red fermentis[®] was used as a process control. The maximum ethanol yield obtained was $0.48 \pm 0.01 \text{ g g}^{-1}$ and an ethanol volumetric productivity of $1.40 \text{ g l}^{-1} \text{ h}^{-1}$ (for data up to 24h) by the native strain 202-3. These results indicated that the native capacity of the strain for cellulosic ethanol production was 93.48% of the theoretical ethanol yield.

Keyword: Sugar cane bagasse, Enzymatic hydrolysis, Dilute sulfuric acid pretreatment, *Saccharomyces cerevisiae*, native strain.

2. Trabajos asociados al estudio del genoma de *S. cerevisiae* 202-3:

2.1. Bogotá Microbial Meeting 2015, Colombia, Bogotá, póster: **Molecular characterization of the metabolic pathway of xylose for a native *Saccharomyces cerevisiae* strain from Colombia with potential for the production of second-generation ethanol.**

Abstract: Bioethanol is an ecological, clean, and renewable biofuel that merge as an opHon to dismiss the environmental impact of fossil fuels. The predominant microorganisms responsible for the efficient produchOn of ethanol are selected strains of the yeast *Saccharomyces cerevisiae*. In this work we characterize the genome of a natural isolated strain of *S. cerevisiae* that was selected from a group of yeast because its potential to produce ethanol of first and second generaHon, comparable with other natural isolated and commercial strains (like Red Fermentis®).

This strain has the innate capability to consume, in a slow rate, xylose without modification. For these reasons we generated a de novo assembly and annotation for *Saccharomyces cerevisiae* isolate 202-3, and provide ideas for metabolic engineering for the production of bioethanol from biomass. With the draft of the genome sequence, we wanted to find out if our strain has the ability to consume xylose because the presence of some genes discovered in a few strains of *S. cerevisiae*, mainly as wine yeast and contribute with information of genetic variability of the *S. cerevisiae* yeast as other authors suggested with their jobs.

El comité organizador del Bogotá Microbial Meeting (Edición I) y la Asociación Colombiana de Microbiología

certifican que

Yina Alejandra Cifuentes Triana

con C.C. 1012364645, participó en calidad de PRESENTACIÓN DE POSTER en el



que se realizó en la Pontificia Universidad Javeriana el día 21 de Agosto de 2015



Con el apoyo de ASM (American Society of Microbiology) e ISME (International Society of Microbial Ecology)



2.1. ICCABS congress 2015, USA, Miami. Presentación oral: Draft Genome Sequence of a natural isolated *Saccharomyces cerevisiae* from Colombia (<http://doi.ieeecomputersociety.org/10.1109/ICCABS.2015.7344727>).



October 12, 2015

To

Yina Alejandra Cifuentes Triana
B.Sc. Química
Estudiante de M. Sc.-Microbiología
Universidad Nacional de Colombia

Dear Yina,

On behalf of the organizers of the International Conference on Computational Advances in Bio and medical Sciences (ICCABS) 2015, I am delighted to invite you to attend this great conference. As you know, the conference will be held in the Sofitel Luxury Hotels, Miami, FL, USA from October 15 to October 17, 2015. Your attendance is essential given that one of your papers has been accepted for presentation in this conference. Each accepted paper must be presented by one of the authors – otherwise the paper will not be included in the proceedings.

Your participation will greatly add to the success of the conference. Looking forward to seeing you soon in Miami!

Sincerely,

Sanguthevar Rajasekaran
Co-General Chair, ICCABS 2015
UTC Chair Professor of CSE and
Director of Booth Engineering Center for Advanced Technology
University of Connecticut, Storrs, CT, USA

Draft Genome Sequence of a natural isolated *Saccharomyces cerevisiae* from Colombia

Yina Cifuentes
Department of Environmental and
Chemical Engineering, Universidad
Nacional de Colombia, Bogotá.
Avenida Carrera 30 N° 45-03
+57 3004416557
yacifuentest@unal.edu.co

Sergio Latorre
Chemistry Department, Universidad
Nacional de Colombia, Bogotá.
Avenida Carrera 30 N° 45-03
+57 3005362051
smlatorreo@unal.edu.co

Andrés Pinzón
Genetic institute, Universidad Nacional
de Colombia, Bogotá. Avenida Carrera
30 N° 45-03
+57(1)3165000 Ext.11618
ampinzonv@unal.edu.co

Mario Velásquez
Department of Environmental and
Chemical Engineering, Universidad
Nacional de Colombia, Bogotá.
Avenida Carrera 30 N° 45-03
+57 3133197570
mevelasquezl@unal.edu.co

Abstract—Here, we report the draft genome sequence of *Saccharomyces cerevisiae* strain 202-3, natural isolated from Colombia, used in bioethanol production from sugarcane bagasse hydrolysates with the ability of innate xylose consumption. The draft genome information provides useful insights into metabolic engineering for the production of bioethanol from biomass.

Keywords—*Saccharomyces cerevisiae*, bioethanol, genome sequencing, *de novo* assembly.

I. INTRODUCTION

Bioethanol is an ecological, clean, and renewable biofuel that merge as an option to dismiss the environmental impact of fossil fuels [1]. The predominant microorganisms responsible for the efficient production of ethanol are selected strains of the yeast *Saccharomyces cerevisiae* [1]. In this work we characterize the genome of a natural isolated strain of *S. cerevisiae* that was selected from a group of yeast because its potential to produce ethanol of first and second generation, comparable with other natural isolated and commercial strains (like *Red Fermentis*®) [2]. This strain has the innate capability to consume, in a slow rate, xylose without modification. For these reasons we generated a *de novo* assembly and annotation for *Saccharomyces cerevisiae* isolate 202-3, and provide ideas for metabolic engineering for the production of bioethanol from biomass.

With the draft of the genome sequence, we wanted to find out if our strain has the ability to consume xylose because the presence of some genes discovered in a few strains of *S. cerevisiae*, mainly as wine yeast [3] and contribute with information of genetic variability of the *S. cerevisiae* yeast as other authors suggested with their jobs [3,4,5,6].

II. MATERIALS AND METHODS

A. Yeast strain and growth media

The natural isolated *S. cerevisiae* 202-3, was chosen from a group of well characterized yeasts [2] by its bioethanol

production from sugarcane bagasse hydrolysate. The corroboration of the use of xylose as sole carbon source was made in agar YNBX medium (xylose 5 %, agar 2 %, Yeast Nitrogen Base 0,67%) under standard conditions. Yeast cells were routinely grown on YPD medium (1 % yeast extract, 2 % peptone, and 2 % glucose) under standard conditions.

B. Genome sequence analysis

Genomic DNA was isolated from strain 202-3 cells with YeaStar columns (Zymo Research) according to the manufacturer's recommendations.

Shotgun genome sequencing was performed using *Illumina HiSeq 2000* paired end. Quality assessment was performed using *FastQC*, and quality score based cleaning was performed with *Prinseq* software [7].

A *de novo* assembly was done using software *SOAPdenovo* version 2.04 [8]. Kmer size of 61 was used to perform the process.

Gene predictions were made by hidden Markov model (HMM) approach, using *AUGUSTUS* [9], with the reference strain S288c of *S. cerevisiae*. An automatic annotation was performed using *BLASTp* algorithm [10] by comparing extracted predicted proteins against *S. cerevisiae* proteins from the *Saccharomyces* Genome Database (SGD) [11].

Finally, the identifiers of annotated proteins were used as inputs for Database for Annotation, Visualization and Integrated Discovery (DAVID) [12] to get retrieve general classification categories.

III. RESULTS

The sequencing process generated paired-end read files R1 and R2, consisting of a total of 21'730.524 reads.

The resulting assembly has 120 scaffolds with a total length of 12'855.005 bp, an N50 of 330 kb and GC content of 38%.

Genome annotation performed by *AUGUSTUS* revealed 5.366 putative open reading frames (ORFs), which were potential predicted proteins. An automatic annotation using *BLASTp* algorithm revealed 5.230 potential proteins with significant similarity.

DAVID analysis revealed the presence of common genes reported in *S. cerevisiae* S288c in the xylose metabolism. To confirm the presence of the gene that confers the ability to growth on xylose, a *BLASTx* was made with the reported genes of the metabolism of xylose in some *S. cerevisiae* strains, with special attention in the gene *XDH1* [3], which confers the capability to growth on xylose to some strains [3]. These data reveal that 202-3 strain has a potential genetic background beneficial for effective bioethanol production from C6 and C5 carbon sources, without the presence of the *XDH1* gene. Genomic analyses of 202-3 are therefore essential for further refinement of the xylose metabolic pathways.

IV. CONCLUSIONS

The *S. cerevisiae* 202-3 strain has the capability to consume xylose without the presence of the *XDH1* gene. With the *BLASTp* annotation we found new ORF that does not have the reference yeast S288c.

ACKNOWLEDGEMENT

We thank to Colciencias and ECOPEPETROL for the financial support of this work.

REFERENCES

- [1] Babrzadeh, F., Jalili, R., Wang, C., Shokralla, S., Pierce, S., Mosher, A., Nyren, P., Shafer, R., Basso, L., de Amorim, H., Oliveira, A., Davis, R., Ronaghi, M., Gharizadeh, B., Stambuk, B. 2012. Whole-genome sequencing of the efficient industrial fuel- ethanol fermentative *Saccharomyces cerevisiae* strain CAT-1, Mol Genet Genomics, 287:485- 494.
- [2] Fonseca, D., Cifuentes, Y ., Ruiz, R., Montoya, J., Velásquez, M. 2014. Fermentation of Enzymatic Hydrolysates of Sugar Cane Bagasse by a Colombian Native Strain of *Saccharomyces Cerevisiae* for the Production of Cellulosic Ethanol. Conference Proceedings, Tokyo International Conference on Life Science and Biological Engineering, 287.
- [3] Wenger, J., Schwartz, K., Sherlock, G. 2010. Bulk Segregant Analysis by High-Throughput Sequencing Reveals a Novel Xylose Utilization Gene from *Saccharomyces cerevisiae*, PLoS Genetics, 6:1-17.
- [4] Novo, M., Bigey, F., Beyne, E., Galeote, V., Gavory, F., Mallet, S., Cambon, B., Luc-Légras, J., Wicker, P., Casaregola, S., Dequin, S. 2009. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118, PNAS, 106:16333-38.
- [5] Borneman, A., Desany, B., Riches, D., Affourtit, J., Forgan, A., Pretorius, I., Egholm, M., Chambers, P. 2011. Whole- Genome Comparison Reveals Novel Genetic Elements That Characterize the Genome of Industrial Strains of *Saccharomyces cerevisiae*, PLoS Genetics, 7:1-10.
- [6] Sahara, T., Fujimori, K., Nezu, M., Tsukahara, M., Tochigi, Y ., Ohgiya, S., Kamagata, Y . 2014. Draft Genome Sequence of *Saccharomyces cerevisiae* IR-2, a Useful Industrial Strain for Highly Efficient Production of Bioethanol, Genome Announc, 2(1):e01160-13.
- [7] Schmieder, R., and Edwards, R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27:863-864.
- [8] Luo et al. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*, 1:18.
- [9] Stanke, M., Diekhans, M., Baertsch, R., Haussler, D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, 24:637- 644.
- [10] Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, L. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*, 10:421.
- [11] Cherry, J., Hong, E., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E., Christie, K., Costanzo, M., Dwight, S., Engel, S., Fisk, D., Hirschman, J., Hitz, B., Karra, K., Krieger, C., Miyasato, S., Nash, R., Park, J., Skrzypek, M., Simison, M., Weng, S., Wong, E. 2012. *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* , 40:D700 -D705.
- [12] Huang, D., Sherman, B., Lempicki, R. 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4:44-57.

Bibliografía

1. Akao T, Yashiro I, Hosoyama A, Kitagaki H, Horikawa H, Watanabe D, Akada R, Ando Y, Harashima S, Inoue T, Inoue Y, Kajiwara S, Kitamoto K, Kitamoto N, Kobayashi O, Kuhara S, Masubuchi T, Mizoguchi H, Nakao Y, Nakazato A, Namise M, Oba T, Ogata T, Ohta A, Sato M, Shibasaki S, Takatsume Y, Tanimoto S, Tsuboi H, Nishimura A, Yoda K, Ishikawa T, Iwashita K, Fujita N, Shimoi H: **Whole-Genome Sequencing of Sake Yeast *Saccharomyces cerevisiae* Kyokai no. 7.** *DNA Research* 2011, **18**:423–434.
2. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403–410.
3. Andrews S: **FastQC: a quality control tool for high throughput sequence data.** Disponible en: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. 2015.
4. Angermayr M, Bandlow W: **Permanent Nucleosome Exclusion from the Gal4p-inducible Yeast *GCY1* Promoter.** *THE JOURNAL OF BIOLOGICAL CHEMISTRY* 2006, **278**: 11026-11031.
5. Aoki-Kinoshita KF, Kanehisa M: **Gene annotation and pathway mapping in KEGG.** *Methods Mol Biol* 2007, **396**:71–91.
6. Apel A, Ouellet M, Szmidi H, Keasling J, Mukhopadhyay A: **Evolved hexose transporter enhances xylose uptake and glucose/xylose co-utilization in *Saccharomyces cerevisiae*.** *Scientific Reports* 2016, **6**:1-10.
7. Ashburner M, Ball C, Blake J, Botstein D, Butler H, Cherry M, Davis A, Dolinski K, Dwight S, Harris E, Hill D, Tarver L, Kasarskis A, Lewis S, Matese J, Richardson J, Ringwald M, Rubin G, Sherlock G: **Gene Ontology: tool for the unification of biology.** *Nature genetics* 2000, **25**: 25-29.
8. Atiyeh H, Duvnjak Z: **Production of Fructose and Ethanol from Sugar Beet Molasses Using *Saccharomyces cerevisiae* ATCC 36858.** *Biotechnology Progress* 2002, **18**: 234-239.
9. Atfield P, Bell P: **Use of population genetics to derive nonrecombinant *Saccharomyces cerevisiae* strains that grow using xylose as a sole carbon source.** *FEMS Yeast Res* 2006, **6**: 862–868.
10. Babrzadeh F, Jalili R, Wang C, Shokralla S, Pierce S, Robinson A, Nyren P, Shafer R, Basso L, Amorim H, Oliveira A, Davis R, Ronaghi M, Gharizadeh B, Stambuk B: **Whole-genome sequencing of the efficient industrial fuel-ethanol fermentative *Saccharomyces cerevisiae* strain CAT-1.** *Mol Genet Genomics* 2012, **287**:485–494.
11. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin A V, Sirotkin A V, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA: **SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing.** *J Comput Biol* 2012, **19**:455–477.
12. Bauer M, Derbyshire M, Gonzales N, Lu S, Chitsaz F, Geer R, He J, Gwadz M, Hurwitz D, Lanczycki C, Lu F, Marchler G, Song J, Thanki N, Wang Z, Yamashita

- R, Zhang D, Zheng C, Bryant S: **CDD: NCBI's conserved domain database.** *Nucleic Acids Res.* 2015, **43**: D222-6.
13. Bengtsson-Palme J, Ryberg M, Hartmann M, Branco S, Wang Z, Godhe A, De Wit P, Sánchez-García M, Ebersberger I, de Sousa F, Amend A, Jumpponen A, Unterseher M, Kristiansson E, Abarenkov K, Bertrand Y. J. K, Sanli K, Eriksson K. M, Vik U, Veldre V, Nilsson R. H: **Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data.** *Methods in Ecology and Evolution* 2013, **4**: 914–919.
 14. Blais EM, Chavali AK, Papin JA: **Linking genome-scale metabolic modeling and genome annotation.** *Methods Mol Biol* 2013, **985**:61–83.
 15. Blake J: **Ten Quick Tips for Using the Gene Ontology.** *PLOS Computational Biology* 2013, **9**:1-2.
 16. Borneman A, Desany B, Riches D, Affourtit J, Forgan A, Pretorius I, Egholm M, Chambers P: **Whole-Genome Comparison Reveals Novel Genetic Elements That Characterize the Genome of Industrial Strains of *Saccharomyces cerevisiae*.** *PLoS Genetics*, 2011, **7**:1-10.
 17. Borneman A, Forgan A, Pretorius I, Chambers P: **Comparative genome analysis of a *Saccharomyces cerevisiae* wine strain.** *FEMS Yeast Res* 2008, **8**:1185–1195.
 18. Botes A: **Affinity purification and characterization of a yeast epoxide hydrolase.** *Biotechnology Letters* 1999, **21**: 511-517.
 19. Chan P, Lowe T: **GtRNAdb: a database of transfer RNA genes detected in genomic sequence.** *Nucleic Acids Research* 2009, **37**: D93-D97.
 20. Chen H, Fu X: **Industrial technologies for bioethanol production from lignocellulosic biomass.** *Renewable and Sustainable Energy Reviews* 2016, **57**:468–478.
 21. Chen R, Dou J: **Biofuels and bio-based chemicals from lignocellulose: metabolic engineering strategies in strain development.** *Biotechnol Lett* 2016, **38**:213–221.
 22. Conesa et al. A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinforma* 2005, **21**:3674–3676.
 23. Conesa et al. A, Götz S: **Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics.** *Int J Plant Genomics* 2008, **2008**:1-12.
 24. Denton J, Lugo J, Tucker A, Schrider D, Warren W, Hahn M: **Extensive Error in the Number of Genes Inferred from Draft Genome Assemblies.** *PLoS Computational Biology* 2014, **10**: e1003998- e1003998.
 25. Dozdova P, Tarasov O, Matveenko A, Radchenko A, Sopova J, Polev D, Inge-Vechtomov S, Dobrynin P: **Genome Sequencing and Comparative Analysis of *Saccharomyces cerevisiae* Strains of the Peterhof Genetic Collection.** *PLoS ONE* 2016, **11**: e0154722.

26. Eliasson A, Boles E, Johansson B: **Xylulose fermentation by mutant and wild-type strains of *Zygosaccharomyces* and *Saccharomyces cerevisiae*.** *Appl Microbiol Biotechnol* 2000, **53**: 376-382.
27. Engel S, Dietrich F, Fisk D, Binkley G, Balakrishna R, Costanzo M, Dwight S, Hitz B, Karra K, Nash R, Weng S, Wong E, Lloyd P, Skrzypek M, Miyasato S, Simison M, Cherry M: **The Reference Genome Sequence of *Saccharomyces cerevisiae*: Then and Now.** *G3 (Bethesda)*. 2014, **4**: 389–398.
28. Fedebiocombustibles, Preguntas Frecuentes de los Biocombustibles, 2007. Disponible En: <http://www.fedebiocombustibles.com/v3/nota-web-id-923.htm#sthash.ffazLSfm.dpuf>
29. Fisk D, Ball C, Dolinski K, Engel S, Hong E, Tarver L, Schwartz K, Sethuraman A, Botstein D, Cherry M, The *Saccharomyces* Genome Database Project: ***Saccharomyces cerevisiae* S288C genome annotation: a working hypothesis.** *Yeast* 2006, **23**: 857-865.
30. Gasch A, Payseur B, Pool J: **The Power of Natural Variation for Model Organism Biology.** *Trends in Genetics* 2016, **32**: 147-154.
31. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A: **Protein Identification and Analysis Tools on the ExPASy Server** 2005, *John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press*, pp. 571-607.
32. Gaudet P, Chisholm R, Berardini T, Dimmer E, Engel SR, Fey P, Hill DP, Howe D, Hu JC, Huntley R, Khodiyar VK, Kishore R, Li D, Lovering RC, McCarthy F, Ni L, Petri V, Siegele DA, Tweedie S, Van Auken K, Wood V, Basu S, Carbon S, Dolan M, Mungall CJ, Dolinski K, Thomas P, Ashburner M, Blake JA, Cherry JM, et al.: **The gene ontology's reference genome project: A unified framework for functional annotation across species.** *PLoS Comput Biol* 2009, **5**:1-8.
33. Gorsich S, Dien B, Nichols N, Slinger P, Liu Z, Skory C: **Tolerance to furfural-induced stress is associated with pentose phosphate pathway genes *ZWF1*, *GND1*, *RPE1*, and *TKL1* in *Saccharomyces cerevisiae*.** *Appl Microbiol Biotechnol* 2006, **71**: 339–349.
34. Graovac M, Chen N: **Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences.** *Current Protocols in Bioinformatics* 2009, **25**:4.10:4.10.1–4.10.14.
35. Green M, Karp P: **Genome annotation errors in pathway databases due to semantic ambiguity in partial EC numbers.** *Nucleic Acids Research* 2005, **33**: 4035–4039.
36. Gurevich A, Saveliev V, Vyahhi N, Tesler G: **QUAST: quality assessment tool for genome assemblies.** *Bioinforma* 2013, **29**:1072–1075.
37. Hani J, Feldmann H: **tRNA genes and retroelements in the yeast genome.** *Nucleic Acids Research* 1998, **26**: 689-696.
38. Marcus D, Dignard D, Lépine G, Askew C, Raymond M, Whiteway M, Wu C: **Comparative Xylose Metabolism among the Ascomycetes *C. albicans*, *S. stipitis* and *S. cerevisiae*.** *PLOS ONE* 2013, **8**: 1-10.

39. Hasper A, Visser J, Graaff L: **The *Aspergillus niger* transcriptional activator XlnR, which is involved in the degradation of the polysaccharides xylan and cellulose, also regulates D-xylose reductase gene expression.** *Molecular Microbiology* 2000, **36**: 193-200.
40. Huaiyu Mi, Qing Dong, Muruganujan A, Gaudet P, Lewis S, Thomas P: **PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium.** *Nucl. Acids Res.* 2010, **38**: D204-D210.
41. Huxley C, Green E, Dunham I: **Rapid assessment of *S. cerevisiae* mating type by PCR.** *Trends Genet.* 1990, **6**:236.
42. Iben J, Maraia R: **tRNAomics: tRNA gene copy number variation and codon use provide bioinformatic evidence of a new anticodon:codon wobble pair in a eukaryote.** *RNA* 2012, **18**: 1358-1372.
43. Ibraheem O, Ndimba B: **Molecular Adaptation Mechanisms Employed by Ethanologenic Bacteria in Response to Lignocellulose derived Inhibitory Compounds.** *Int. J. Biol. Sci.* 2013, **9**:598-612.
44. Iida N, Yamao F, Nakamura Y, Iida T: **Mudi, a web tool for identifying mutations by bioinformatics analysis of whole-genome sequence.** *Genes to Cells* 2014, **19**:517–527.
45. Imelfort M: **gff2fasta.** 2013.
46. Jakočiūnas T, Jensen M, Keasling J: **CRISPR/Cas9 advances engineering of microbial cell factories.** *Metabolic Engineering* 2016, **34**:44–59.
47. Jayakody L, Hayashi N, Kitagaki H. **Molecular mechanisms for detoxification of major aldehyde inhibitors for production of bioethanol by *Saccharomyces cerevisiae* from hot-compressed water-treated lignocellulose.** En: Materials and processes for energy communicating current research and technological developments (A. Méndez-Vilas, Ed) 2013, pp: 302-311.
48. Jongsik Chun Lab: **EzFungi.** Disponible en: <http://www.ezbiocloud.net/ezfungi/identify>, 2015. Consultado el 2016.
49. Kahar P, Taku K, Tanaka S: **Enhancement of xylose uptake in 2-deoxyglucose tolerant mutant of *Saccharomyces cerevisiae*,** *Journal of Bioscience and Bioengineering* 2011, **5**: 557-563.
50. Kavšček M, Stražar M, Curk T, Natter K, Petrovič U: **Yeast as a cell factory: current state and perspectives.** *Microb Cell Fact* 2015, **14**:1-10.
51. Klassen P, Gielesen B, Gijssberdina S, Van P, Sarantinopoulos P, Herman W, Heijne M, Greeve A: **Yeast cell capable of converting sugars including arabinose and xylose.** Patente: **WO 2012143513 A2**, 2012.
52. Langmead B, Salzberg S: **Fast gapped-read alignment with Bowtie 2.** *Nature Methods.* 2012, **9**:357-359.
53. Laslett D, Canback B: **ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences.** *Nucl. Acids Res.* 2004, **32**:11-16.
54. Legras J, Karst F: **Optimisation of interdelta analysis for *Saccharomyces cerevisiae* strain characterisation.** *FEMS Microbiol Lett.* 2003, **221**: 249-255.

55. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler Transform.** *Bioinformatics* 2009 (a), **25**:1754-60.
56. Li H, Handsaker B, Wysoker A, Fennell T, Rujan J, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics* 2009 (b), **25**: 2078-9.
57. Li X, Park A, Estrela R, Kim S, Jin Y, Cate J: **Comparison of xylose fermentation by two high-performance engineered strains of *Saccharomyces cerevisiae*.** *Biotechnology Reports* 2016, **9**: 53–56.
58. Lin D, Yin X, Wang X, Zhou P, Guo F: **Re-Annotation of Protein-Coding Genes in the Genome of *Saccharomyces cerevisiae* Based on Support Vector Machines.** *PLOS ONE* 2013, **8**:1-6.
59. Liu H, Styles C, Fink G: ***Saccharomyces cerevisiae* S288C Has a Mutation in FL08, a Gene Required for Filamentous Growth.** *Genetics* 1996, **144**:967-978.
60. Lomsadze A, Ter-Hovhannisyan V, Chernoff Y, Borodovsky M: **Gene identification in novel eukaryotic genomes by self-training algorithm.** *Nucleic Acids Research* 2005, **33**: 6494–6506.
61. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung D, Ming S, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Wah T, Wang J: **SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler.** *GigaScience* 2012, **1**:1-6.
62. Maqueda M, Zamora E, Rodríguez N, Ramírez M: **Wine yeast molecular typing using a simplified method for simultaneously extracting mtDNA, nuclear DNA and virus dsRNA.** *Food Microbiology* 2010, **27**: 205-209.
63. Martin M: **Cutadapt removes adapter sequences from high-throughput sequencing reads.** *Technical Notes EMBnet.journal* 2011, **17**:10-12.
64. Matsushika A, Goshima T, Hoshino T: **Transcription analysis of recombinant industrial and laboratory *Saccharomyces cerevisiae* strains reveals the molecular basis for fermentation of glucose and xylose.** *Microbial Cell Factories* 2014, **13**:1-18.
65. Merchant S, Wood D, Salzberg S: **Unexpected cross-species contamination in genome sequencing projects.** *PeerJ* 2014, **2**:e675.
66. Metwally S, Hamza T, Zakaria M, Helmy M: **Next-Generation Sequence Assembly: Four Stages of Data Processing and Computational Challenges.** *PLOS Computational Biology* 2013, **9**:1-19.
67. Morales D, Cifuentes Y, Ruiz R, Baca J, Velásquez M: **Fermentation of Enzymatic Hydrolysates of Sugar Cane Bagasse by a Colombian Native Strain of *Saccharomyces Cerevisiae* for the Production of Cellulosic Ethanol.** *Conference Proceedings, Tokyo International Conference on Life Science and Biological Engineering* 2014, 287.
68. Nijland J, Shin H, de Jong R, de Waal P, Klaassen P, Driessen A: **Engineering of an endogenous hexose transporter into a specific D-xylose transporter facilitates glucose-xylose co-consumption in *Saccharomyces cerevisiae*.** *Biotech Biofuels* 2014, **7**: 1-11.

69. Nogueira D, Branco V, Moreira J, Pepe M, Gonçalves F: **Xylose Fermentation by *Saccharomyces cerevisiae*: Challenges and Prospects.** *Int. J. Mol. Sci.* 2016, **17**: 1-18.
70. Oreb M, Dietz H, Farwick A, Boles E: **Novel strategies to improve co-fermentation of pentoses with d-glucose by recombinant yeast strains in lignocellulosic hydrolysates.** *Bioengineered* 2012, **3**:347-351.
71. Otero J, Vongsangnak W, Asadollahi M, Olivares R, Maury J, Farinelli L, Barlocher L, Osteras M, Schalk M, Clark A: **Whole genome sequencing of *Saccharomyces cerevisiae*: from genotype to phenotype for improved metabolic engineering applications.** *BMC Genomics* 2010,**11**: 1-17.
72. Passoth V: **Chapter 9: Molecular Mechanisms in Yeast Carbon Metabolism: Bioethanol and Other Biofuels.** En: *Molecular Mechanisms in Yeast Carbon Metabolism* (Springer) 2014, pp: 217-259.
73. Paszkiewicz K, Studholme D: **De novo assembly of short sequence reads.** *Brief. Bioinform* 2010, **11**: 457-472.
74. Pruitt KD, Tatusova T, Maglott DR: **NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins.** *Nucleic Acids Res* 2005, **33** (suppl 1):D501–D504.
75. Ren B, Robert F, Wyrick J, Aparicio O, Jennings E, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert T, Wilson C, Bell S, Young R: **Genome-Wide Location and Function of DNA Binding Proteins.** *Science* 2000, **290**: 2306-2309.
76. Sahara T, Fujimori K, Nezu M, Tsukahara M, Tochigi Y, Ohgiya S, Kamagata Y: **Draft Genome Sequence of *Saccharomyces cerevisiae* IR-2, a Useful Industrial Strain for Highly Efficient Production of Bioethanol.** *Genome Announcements* 2014 (a), **2**:1-2.
77. Sahara T, Fujimori K, Nezu M, Tsukahara M, Tochigi Y, Ohgiya S, Kida Y, Taguchi H, Akamatsu T, Kamagata Y: **Draft Genome Sequence of *Saccharomyces cerevisiae* NAM34-4C, a Lactic Acid-Assimilating Industrial Yeast Strain.** *Genome Announcements* 2014 (b), **2**:1-2.
78. Sánchez B, Nielsen J: **Genome scale models of yeast: towards standardized evaluation and consistent omic integration.** *Integr. Biol.* 2015, **7**:846-858.
79. Sánchez V, Karhumaa K: **Xylose fermentation as a challenge for commercialization of lignocellulosic fuels and chemicals.** *Biotechnol Lett* 2015, **37**:761–772.
80. Sánchez V, Bettiga M: **Isolation and characterization of a resident tolerant *Saccharomyces cerevisiae* strain from a spent sulfite liquor fermentation plant.** *AMB Express* 2012, **2**:1-11.
81. Schmieder R, Edwards R: **Quality control and preprocessing of metagenomic datasets.** *Bioinformatics* 2011, **27**:863-864.
82. SGD: *Saccharomyces* GENOME DATABASE, *Saccharomyces cerevisiae* Genome Snapshot, 2016. Disponible en: <http://www.yeastgenome.org/genomesnapshot>.

83. SGD: Saccharomyces GENOME DATABASE, *Saccharomyces cerevisiae* Genome Snapshot, 2016-b. Disponible en: <http://www.yeastgenome.org/locus/S000029683/overview>
84. SGD: Saccharomyces GENOME DATABASE, *Saccharomyces cerevisiae* Genome Snapshot, 2016-c. Disponible en: <http://www.yeastgenome.org/locus/S000005698/overview>.
85. Simao F, Waterhouse R, Ioannidis P, Kriventseva E, Zdobnov: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics* 2015, **31**:3210-3212.
86. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I: **ABYSS: A parallel assembler for short read sequence data.** *Genome Res* 2009, **19**:1117-1123.
87. Sravanthi B, Ulaganathan K: **Draft Genome Sequence of Saccharomyces cerevisiae Strain NCIM3186 Used in the Production of Bioethanol from Sweet Sorghum,** *Genome Announcements* 2015, **3**:1-2.
88. Stanke M, Morgenstern B: **Augustus: a web server for gene prediction in eukaryotes that allows user-defined constraints.** *Nucleic Acids Research*, 2005, **33**:W465-W467.
89. Stovicek V, Borodina I, Foster J: **CRISPR-Cas system enables fast and simple genome editing of industrial Saccharomyces cerevisiae strains.** *Metabolic Engineering Communications* 2015, **2**:13-22.
90. Sung C, Ionk I, Lesmana A, Million G, Chang G, Rin S, Su Y: **Rapid and marker-free Cas9/CRISPR refactoring yields equivalent xylose-utilization performance in yeast.** *Biotechnol Bioeng.* 2015, **112**: 2406-2411.
91. Tamari Z, Rosin D, Voichek Y, Barkai N: **Coordination of Gene Expression and Growth-Rate in Natural Populations of Budding Yeast.** *PLoS ONE* 2014, **9**:1-7.
92. Toivari M, Salusjarvi L, Ruohonen L, Penttila M: **Endogenous xylose pathway in Saccharomyces cerevisiae.** *Appl Environ Microbiol* 2004, **70**:3681-3686.
93. Traff K, Jonsson L, Hahn-Haerdal B: **Putative xylose and arabinose reductases in Saccharomyces cerevisiae.** *Yeast* 2002, **19**: 1233-1241.
94. Traven A, Jelcic B, Sopta M: **Yeast Gal4: a transcriptional paradigm revisited.** *EMBO* 2006, **7**: 496-499.
95. Treangen T, Ondov B, Koren S, Phillippy A: **The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes.** *Genome Biology* 2014, **15**: 1-15.
96. Treu L, Toniolo C, Nadai C, Sardu A, Giacomini A, Corich V, Campanaro S: **The impact of genomic variability on gene expression in environmental Saccharomyces cerevisiae strains.** *Environmental Microbiology* 2014, **15**:1378-1397.
97. VBC, Barnap: <http://www.vicbioinformatics.com/software.barnap.shtml>, consultado el 2016.
98. Venema J, Tollervey D: **Ribosome synthesis in Saccharomyces cerevisiae.** *Annu. Rev. Genet.* 1999, **33**: 261-311.

99. Wang M, Yu C, Zhao H: **Directed Evolution of Xylose Specific Transporters to Facilitate Glucose-Xylose Co-Utilization.** *Biotechnology and Bioengineering* 2016, **113**: 484-491.
100. Wenger J, Schwartz K, Sherlock G: **Bulk Segregant Analysis by High-Throughput Sequencing Reveals a Novel Xylose Utilization Gene from *Saccharomyces cerevisiae*.** *PLoS Genetics* 2010, **6**: 1-17.
101. Wicker R, Fujimura T, Esteban R: **Chapter One – Viruses and Prions of *Saccharomyces cerevisiae*, en Advances in Virus Research.** *Mycoviruses* 2013, **86**: 1-36.
102. Wolters J, Chiu K, Fiumera H: **Population structure of mitochondrial genomes in *Saccharomyces cerevisiae*.** *BMC genomics* 2015, **16**: 1-13.
103. Won C, Tamaki H, Nakayama R, Yamamoto K, Kumagai H: **G-Protein Coupled Receptor from Yeast *Saccharomyces cerevisiae*.** *Biochemical and Biophysical Research Communications* 1997, **240**: 287-292.
104. Yandell M, Ence D: **A beginner's guide to eukaryotic genome annotation.** *Nature Reviews* 2012,**13**:329-342.
105. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y: **The I-TASSER Suite: Protein structure and function prediction.** *Nature Methods* 2015, **12**: 7-8.
106. Xu Z, Wang H: **LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons.** *Nucleic Acids Res.* 2007, **35**: W265-W268.
107. Zerbino DR, Birney E: **Velvet: Algorithms for de novo short read assembly using de Bruijn graphs.** *Genome Res* 2008, **18**:821–829.
108. Zhang J, Zhu F, Vongsangnak W, Tang Y, Zhang W, Shen B: **Evaluation and Comparison of Multiple Aligners for Next-Generation Sequencing Data Analysis.** *BioMed Research International* 2014, **2014**: 1-16.
109. Zheng D, Wang P, Chen J, Zhang K, Liu T, Wu X, Li Y, Zhao Y: **Genome sequencing and genetic breeding of a bioethanol *Saccharomyces cerevisiae* strain YJS329.** *BMC Genomics* 2012, **13**:1-13.