

UCH-FC
Biotecnología
M32.1
E.1



FACULTAD DE CIENCIAS
UNIVERSIDAD DE CHILE

**IDENTIFICACIÓN Y CARACTERIZACIÓN DE GENES QUE SE EXPRESAN
EN EL METABOLISMO FERMENTATIVO DE *Xanthophyllomyces*
*dendrorhous***

Seminario de Título

Entregado a la

Universidad de Chile

en cumplimiento parcial de los requisitos

para optar al Título de

Ingeniero en Biotecnología Molecular

por

Andrés Esteban Marcoleta Caldera

Enero, 2007

Santiago – Chile



Director de Memoria de Título: Dr. Víctor Hugo Cifuentes Guzmán

A mis padres, a mis otros padres, a mis hermanos, a mis otros hermanos y a Vero.

Con los ojos tapados desde la punta del abismo sigo viendo como nuestra maravillosa vida no termina de empezar.





**“IDENTIFICACIÓN Y CARACTERIZACION DE GENES QUE SE
EXPRESAN EN EL METABOLISMO FERMENTATIVO DE
XANTHOPHYLLOMYCES DENDRORHOUS”**

Seminario de Título entregado a la Facultad de Ciencias de la Universidad de Chile, en cumplimiento parcial de los requisitos para optar al título profesional de Ingeniero en Biotecnología Molecular.

ANDRÉS ESTEBAN MARCOLETA CALDERA

Dr. Víctor Cifuentes Guzmán
Tutor Seminario de Título

Firma manuscrita de Víctor Cifuentes Guzmán sobre una línea horizontal.

Comisión Revisora y Evaluadora

Dra. Margarita Carú Marambio
Presidente

Firma manuscrita de Margarita Carú Marambio sobre una línea horizontal.

Dra. Rosa Alba Lagos Mónaco

Firma manuscrita de Rosa Alba Lagos Mónaco sobre una línea horizontal.



SANTIAGO, Mayo de 2007.

II. Agradecimientos

Con amor quiero agradecer a todos aquellos que hacen posible mis logros, sin duda una lista no sería suficiente para incluirlos a todos.

Agradezco a quienes estuvieron directamente involucrados en este trabajo, en primer lugar a todos quienes forman parte del laboratorio de Genética de la Facultad de Ciencias de la Universidad de Chile. Gracias por lo que me han enseñado, por la crítica constructiva a mi trabajo y sobretodo por los buenos momentos que hemos pasado. Agradezco especialmente al Dr. Víctor Cifuentes por haberme acogido en su grupo de trabajo y por brindarme su apoyo en los buenos y no tan buenos momentos. Agradezco también al Dr. Marcelo Baeza por la revisión crítica de esta memoria y de nuestro trabajo en general. Gracias al Dr. Antonio Jiménez y a quienes formaban parte del laboratorio 121 del centro de Biología Molecular Severo Ochoa mientras estuve por allí. Su buena acogida y ayuda fue indispensable para la realización de este trabajo.

Quiero agradecer de corazón a mis padres por ser un ejemplo intachable de dedicación y esfuerzo, por todo el amor que me han entregado y por sobretodo por haberme dado esta vida tan maravillosa. Gracias Ricardo por enseñarme tantas cosas, sin duda fuiste mi segundo padre. Agradezco también a mis hermanos por ser como son y por todo lo que hemos vivido juntos... Marco, Isis, Verónica, Lorena, Ricardo, todos han sido muy importantes para mí.

Gracias a tí Vero por tu amor, por aceptarme tal como soy y por ser siempre mi apoyo, sobretodo en aquellos momentos en que faltaban ganas de continuar. Gracias además por ser un ejemplo de esfuerzo y por recordarme que soy muy afortunado al vivir como vivo.

No puedo dejar de agradecer muy especialmente a mis otros hermanos, aquellos que fui conociendo en el camino. Gracias por enseñarme a ver con mis propios ojos, por lo bien que lo hemos pasado desde hace tanto y sobretodo por compartir tantos sueños, los que tarde o temprano se han ido haciendo realidad.





III. Índice

I Dedicatoria.....	1
II Agradecimientos.....	2
III Índice.....	3
IV Resumen.....	5
1 Introducción.....	6
1.1 Antecedentes generales.....	6
1.2 Metabolismo fermentativo y biosíntesis de carotenoides.....	8
1.3 Planteamiento del problema.....	13
1.4 Propósito del trabajo.....	16
1.5 Objetivo general.....	16
1.6 Objetivos específicos.....	16
2 Materiales y Métodos.....	17
2.1 Cepas y condiciones de cultivo.....	17
2.2 Construcción de genotecas de cDNA en <i>X. dendrorhous</i> : estrategia general.....	17
2.3 Extracción de RNA total.....	20
2.4 Purificación de mRNA poliA ⁺	21
2.5 Síntesis de cDNA de doble hebra.....	21
2.6 Fraccionamiento del cDNA por tamaño.....	21
2.7 Determinación del rango de tamaño de cada fracción obtenida.....	22
2.8 Reacción de ligado del cDNA.....	22
2.9 Transformación de <i>E.coli</i>	22
2.10 Secuenciación de DNA.....	23
2.11 Cálculo del índice de adaptación de codones (CAI).....	23
3 Resultados.....	25
3.1 Construcción de genotecas de cDNA de <i>X. dendrorhous</i>	25
3.1.1 Extracción de RNA total de <i>X. dendrorhous</i> y purificación de mRNA.....	25
3.1.2 Síntesis de cDNA para la construcción de las genotecas.....	27

3.1.3 Fraccionamiento del cDNA por tamaño.....	28
3.1.4 Ligado del cDNA al vector plasmidial y transformación de <i>E. coli</i>	29
3.1.5 Determinación del tamaño medio de inserto de la genoteca de cDNA....	31
3.1.6 Almacenamiento y conservación de la genoteca.....	34
3.2 Aislamiento e Identificación de genes expresados en metabolismo fermentativo de <i>X. dendrorhous</i>	35
3.3 Uso de codones en <i>X. dendrorhous</i>	41
3.4 Determinación de la estructura exón-intrón de genes expresados en Metabolismo fermentativo de <i>X. dendrorhous</i>	49
4 Discusión.....	53
4.1 Construcción de las genotecas de cDNA de <i>X. dendrorhous</i>	53
4.2 Aislamiento e identificación de genes expresados en metabolismo fermentativo de <i>X. dendrorhous</i>	56
4.3 Uso de codones en <i>X. dendrorhous</i>	59
4.4 Estructura exón-intrón de genes expresados en metabolismo fermentativo de <i>X. dendrorhous</i>	64
5 Conclusiones.....	67
6 Bibliografía.....	68



IV. RESUMEN

X. dendrorhous es una levadura con la particular capacidad de sintetizar astaxantina a partir de la fermentación de distintos azúcares. La astaxantina es un pigmento rojo-anaranjado indispensable para la coloración de salmónidos y constituye uno de los aditivos más costosos incluidos en el alimento de peces cultivados en salmoneculturas. El elevado interés comercial del pigmento junto con su costosa síntesis química han motivado el desarrollo de fuentes naturales de astaxantina, entre las cuales *X. dendrorhous* constituye una de las alternativas más promisorias. Sin embargo, los escasos conocimientos que se tienen sobre aspectos genéticos y biológicos de esta levadura han dificultado en gran medida los intentos por aumentar su rendimiento carotenogénico. Una de las grandes dificultades que enfrentan las investigaciones actuales en esta levadura es el reducido número de genes que se conocen, lo que ocasiona entre otras cosas que se cuente con muy pocos marcadores de selección y se desconozcan muchas de las características genómicas relevantes tales como los patrones de uso de codones y de la estructura exón-intrón de sus genes. En este trabajo se construyó una genoteca de cDNA a partir de la cual se identificaron y secuenciaron 35 nuevos genes de *X. dendrorhous*, aumentando en cerca de un 100% los genes de este organismo publicados en la base de datos GENBANK. Los nuevos genes identificados más los disponibles a la fecha en la base de datos fueron utilizados para construir una tabla de uso de codones y predecir sus niveles de expresión en base al cálculo del índice de adaptación de codones. De entre los nuevos genes aislados, 13 resultaron ser de interés para futuras investigaciones en este organismo y su estructura exón-intrón fue determinada clonando las respectivas versiones genómicas. Finalmente estos 13 genes más aquellos cuya estructura exón-intrón se encontraba publicada en GENBANK, fueron utilizados para determinar las secuencias consenso de los sitios dadores y aceptores de *splicing*.

1. INTRODUCCIÓN

1.1 Antecedentes generales

Xanthophyllomyces dendrorhous (ex *Phaffia rhodozyma*) es una levadura basidiomicete inicialmente aislada a partir de exudados de árboles caducifolios en regiones montañosas de Japón y Alaska, por Herman Jan Phaff y colaboradores a fines de la década del 60 (Phaff y cols., 1972). Esta levadura posee la particular capacidad de producir compuestos carotenoides a partir de la fermentación de diversos azúcares, siendo el principal el pigmento de color rojo-anaranjado denominado astaxantina (Andrewes y Starr, 1976). La astaxantina (3,3'-dihidroxi- β,β' -caroteno-4,4'-diona) es un carotenoide natural producido por un limitado número de especies de microalgas, bacterias marinas y hongos (Johnson y Schroeder, 1996). Al través de la cadena alimentaria, este pigmento es acumulado en animales como crustáceos, salmónidos y algunas aves otorgándoles un intenso color rojo-anaranjado a sus caparazones, musculatura, o plumas.

La producción de salmónidos en pisciculturas se ha vuelto una industria muy importante en muchos países tales como Noruega, Chile y Reino Unido, entre otros. Sin embargo, una de las principales dificultades que enfrenta esta industria es la necesidad absoluta del suministro de astaxantina, o determinados precursores de carotenoides, en el alimento de los peces para que alcancen su color característico, y con ello su aceptación entre los consumidores. La astaxantina constituye el ingrediente alimentario de mayor costo en la industria de la acuicultura, y junto con su desarrollo se ha creado un mercado del pigmento estimado en \$200 millones de dólares al año (Lorenz y Cysewski, 2000). Tradicionalmente la producción de astaxantina se ha efectuado por síntesis química. Sin embargo, además de la desventaja de su alto

costo, tanto los consumidores como las autoridades sanitarias coinciden en evitar aditivos químicos en alimentos, por lo que se ha incentivado la investigación en el desarrollo de alternativas biológicas para la producción del pigmento. Los organismos que han mostrado un mayor atractivo para la producción industrial de astaxantina son *X. dendrorhous* (Andrewes, 1976) y la microalga *Haematococcus pluviialis* (Lorenz, 2000), los cuales se encuentran en diferentes etapas de desarrollo comercial.

Las principales limitaciones para la producción industrial de pigmentos a partir de sistemas biológicos suelen ser el bajo rendimiento de cepas silvestres y de la mayoría de sus cepas mutantes, además de costosos métodos de extracción y purificación del metabolito de interés (An y Johnson, 1991; Cañizares-Villanueva y cols., 1998). Se han utilizado variadas aproximaciones para aumentar el rendimiento carotenogénico de *X. dendrorhous*, incluyendo la optimización de metodologías de fermentación (Vázquez y cols., 1998; Flores-Cotera y cols., 2001), mutagénesis (An y cols., 1989; Bielich y cols., 1991) y elementos de ingeniería metabólica-genética (Misawa y Shimada, 1997). Sin embargo, estos esfuerzos se han visto dificultados por el escaso conocimiento que se tenía hasta hace unos pocos años sobre aspectos biológicos y genéticos del microorganismo. Investigaciones relativamente recientes han permitido establecer, entre otras cosas, la composición cromosómica de cepas silvestres y mutantes (Cifuentes y col, 1997), la identificación de los genes responsables de la biosíntesis de astaxantina (Verdoes y cols., 1999 a 1999 b; Hoshino y cols., 2000), el hallazgo de procesamiento alternativo de los mRNAs de al menos dos de estos genes (Lodato y cols., 2003), y el desarrollo de sistemas relativamente eficientes de transformación genética para esta levadura (Wery y cols., 1997; Martínez y cols., 1998).

A pesar de los avances en el estudio de *X. dendrorhous*, aún siguen siendo insuficientes los conocimientos que se tienen sobre los mecanismos de regulación de

la ruta de biosíntesis de astaxantina. Además, se desconoce qué posibles genes expresados en metabolismo fermentativo pudieran estar involucrados indirectamente en la carotenogénesis, actuando en etapas anteriores o paralelas a las clásicamente consideradas dentro de la vía. Es por esto que resulta de gran relevancia realizar estudios que permitan incrementar el conocimiento y finalmente entender los posibles mecanismos de regulación de la biosíntesis de carotenoides, vista como un proceso global enmarcado dentro del metabolismo fermentativo de la levadura. Este entendimiento constituye un elemento clave a la hora de pretender desarrollar programas de optimización con fines industriales de la producción de pigmentos.

1.2 Metabolismo Fermentativo y Biosíntesis de carotenoides.

X. dendrorhous es capaz de crecer en diversas fuentes de carbono, entre las cuales se encuentran glucosa, sacarosa, maltosa, xilosa, almidón, succinato, glicerol y etanol (Vazquez y cols., 1997; Johnson, 2003; Wozniak y cols., 2005; Marin y cols., 2006). Diversos estudios han mostrado que existe una relación directa entre la fuente de carbono utilizada por la levadura y la biosíntesis de carotenoides. Este efecto se observa tanto en la cantidad de pigmentos totales como en su composición, es decir, la cantidad relativa de astaxantina y los distintos intermediarios que se producen en la ruta de carotenogénesis (Vázquez y cols., 1997; Flores-Cotera y cols., 2001; Johnson, 2003)

Dependiendo de la fuente de carbono disponible en el medio, *X. dendrorhous* es capaz de llevar a cabo básicamente dos tipos de metabolismos. Una de las alternativas consiste en la oxidación de azúcares como glucosa u otras que son finalmente convertidas en esta última a través de la vía glicolítica. Como producto de la glicólisis se obtiene principalmente piruvato, el que luego de una reacción de decarboxilación y

reducción mediada por NADH, es convertido en etanol. La producción de etanol permite la síntesis de ATP por medio de fosforilación a nivel de sustrato (de manera independiente de oxígeno) y la regeneración del NAD⁺ convertido en NADH durante la glicólisis. A este conjunto de procesos que comprenden la oxidación glicolítica seguida por fermentación alcohólica se les denomina de manera genérica *metabolismo fermentativo* (Figura 1).

Por otra parte, fuentes de carbono no fermentables tales como el succinato o el glicerol no son sustrato de la vía glicolítica y son incorporadas al metabolismo a nivel del ciclo de Krebs. A pesar que luego de múltiples reacciones el producto final obtenido a partir de este tipo de fuentes de carbono también es piruvato; por mecanismos regula-

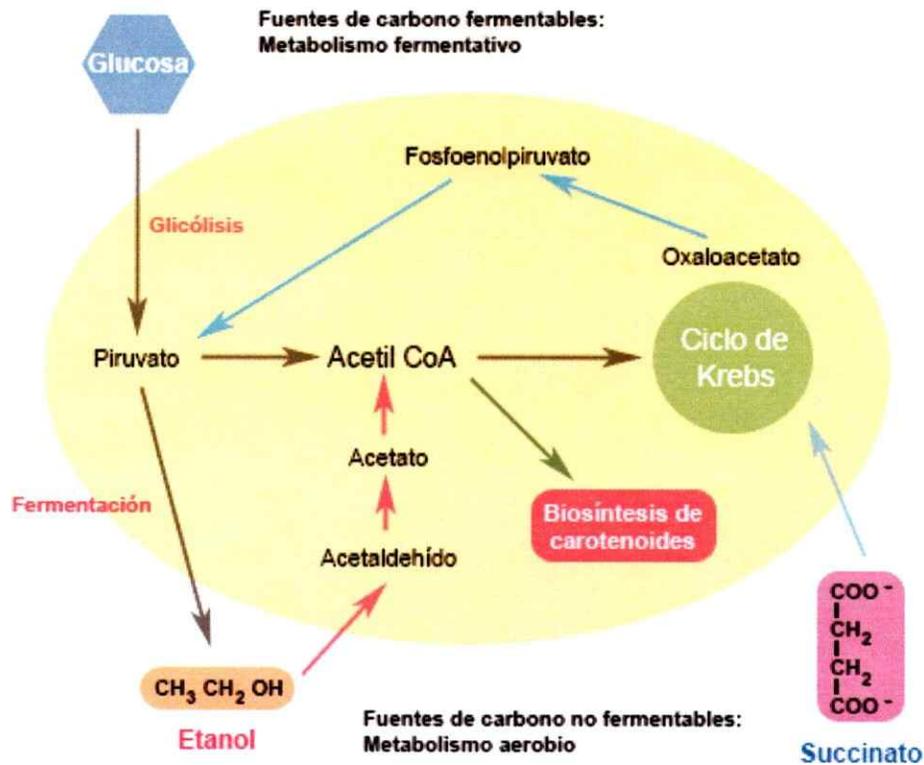


Figura 1. Metabolismo del carbono en *X. dendrorhous*. Representación esquemática de la asimilación de distintas fuentes de carbono (fermentables y no fermentables) en la levadura.

torios aún desconocidos, este último es destinado mayoritariamente a la formación de acetil CoA y no a la producción de etanol. La acetil CoA producida es luego distribuida entre el ciclo de Krebs (que permite la generación del poder reductor requerido para la síntesis de ATP por fosforilación oxidativa), y las enzimas que sintetizan los precursores necesarios para la carotenogénesis. La asimilación del carbono mediada por los procesos anteriormente descritos recibe el nombre de *metabolismo aerobio o no fermentativo* (Figura 1).

Un caso particular ocurre cuando *X. dendrorhous* es crecida en glucosa como única fuente de carbono. En estas condiciones, la levadura comienza a consumir el azúcar por medio de los procesos del metabolismo fermentativo, que culminan con la producción de etanol y su paulatina acumulación en el medio. Una vez que la glucosa se ha agotado, ocurre un cambio metabólico y comienza el consumo del etanol anteriormente producido, dando paso al metabolismo no fermentativo (Lodato y cols, 2004). Estudios realizados en nuestro laboratorio indican que el agotamiento de la glucosa y el comienzo del consumo de etanol coinciden con un aumento abrupto en la síntesis de carotenoides (Figura 2) (Lodato y cols., 2003). Este momento del crecimiento de *X. dendrorhous* en glucosa recibe el nombre de *fase de inducción de la carotenogénesis*, la que coincide con la entrada del cultivo en fase estacionaria.

Bastante distinta es la situación observada cuando se utiliza succinato como única fuente de carbono. Bajo dichas condiciones, la biosíntesis de pigmentos es mucho más constante a lo largo del tiempo y comienza en etapas tempranas del cultivo, no observándose la fase de inducción (Wozniak y cols., 2005). Las evidencias anteriormente mencionadas sugieren la existencia de mecanismos regulatorios de la biosíntesis de carotenoides, de alguna manera relacionados con la transición entre el metabolismo fermentativo y el aerobio.

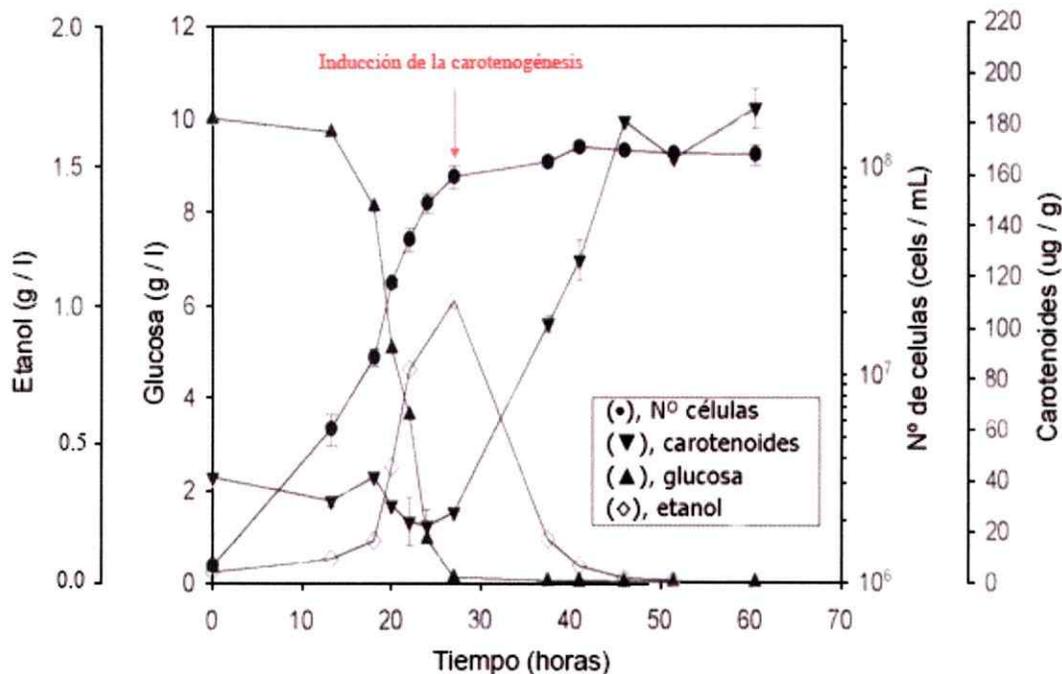


Figura 2. Crecimiento de *X. dendrorhous* utilizando glucosa como única fuente de carbono (Lodato y cols., 2003). A partir de un cultivo de *X. dendrorhous* en medio mínimo glucosa, se determinó en función del tiempo el número de células, la cantidad de carotenoides totales sintetizados, y la concentración de glucosa y etanol en el medio extracelular.

La biosíntesis de carotenoides requiere, como sustrato principal, isopentenil pirofosfato (IPP). El IPP es sintetizado a partir de acetil-CoA en una serie de pasos que involucran, en primera instancia, la condensación de dos moléculas de acetil-CoA para formar acetoacetil-CoA. Esta reacción es catalizada por la enzima acetil-CoA C-acetiltransferasa (EC 2.3.1.9). En el siguiente paso la Hidroximetilglutaril-CoA sintasa (EC 2.3.3.10) cataliza la unión de una molécula de acetoacetil-CoA con otra de acetil-CoA, formando 3-hidroxi-3-metil-glutaril-CoA, el que es posteriormente reducido a mevalonato por la actividad hidroximetilglutaril-CoA reductasa (EC 1.1.1.34). El mevalonato es fosforilado dos veces, primero por la mevalonato kinasa (EC 2.7.1.36) y luego por la fosfomevalonato kinasa (EC 2.7.4.2) para obtener difosfomevalonato. Por

último, el difosfomevalonato es decarboxilado por la difosfomevalonato decarboxilasa (EC 4.1.1.33) formando isopentenil pirofosfato (Figura 3).

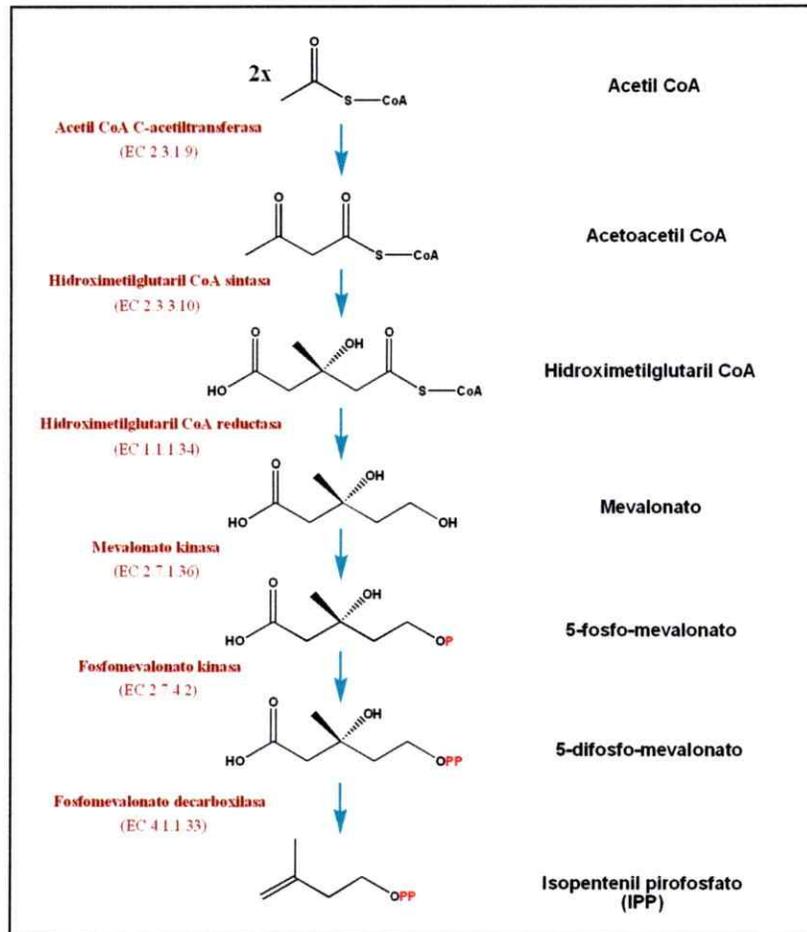


Figura 3. Ruta de Biosíntesis de IPP. En hongos y específicamente en basidiomicetes se lleva a cabo la biosíntesis de IPP a partir de acetil CoA, a través de la vía del mevalonato.

En una etapa intermedia, previa a la ruta de carotenogénesis, el IPP es susceptible de ser convertido en su isómero activo dimetilalil pirofosfato (DMAPP), por la enzima Isopentenil pirofosfato isomerasa (producto del gen *idi*). Luego, se lleva a cabo la condensación de una molécula de IPP con otra de DMAPP originando geranil pirofosfato, el que luego de la sucesiva adición de dos moléculas de IPP forma geranilgeranil pirofosfato (GGPP). La producción de GGPP a partir de IPP y DMAPP es

catalizada por la geranylgeranyl pirofosfato sintasa (producto del gen *crtE*) (Niklitscheck, 2007).

En *X. dendrorhous*, la biosíntesis de carotenoides propiamente tal comienza con la condensación de dos moléculas de GGPP para dar origen a una molécula de *cis*-fitoeno; en una reacción catalizada por la actividad fitoeno sintasa de la enzima bifuncional Fitoeno/ β -caroteno sintasa, producto del gen *crtYB* (Verdoes y cols., 1999b). El siguiente paso involucra 4 reacciones de desaturación (formación de dobles enlaces C-C) catalizadas por la enzima fitoeno desaturasa (producto del gen *crtI*), las que conducen a la formación de licopeno (Verdoes y cols., 1999a). El licopeno es posteriormente convertido en β -caroteno por la actividad licopeno ciclasa del producto del gen *crtYB* (Verdoes y cols., 1999b). Finalmente, luego de dos hidroxilaciones y dos adiciones de grupos ceto, el β -caroteno es transformado en astaxantina por la astaxantina sintasa (producto del gen *crtS*) (Alvarez y cols., 2006; Ojima y cols., 2006) (Figura 4). Resulta interesante notar que con tan solo 3 genes la levadura produce sobre 15 carotenoides diferentes, cuya composición en la célula puede depender de las condiciones de crecimiento.

1.3 Planteamiento del problema.

Actualmente se han publicado en la base de datos internacional "GenBank" tan solo 28 genes de *X. dendrorhous*, entre los cuales se encuentran los que codifican para las 5 proteínas involucradas en la biosíntesis de astaxantina a partir de isopentenil pirofosfato (IPP); 15 proteínas ribosómicas, además de los genes *act* y *gpdh* que codifican para actina y para gliceraldehído-3-fosfato deshidrogenasa, respectivamente, constituyendo este último el único gen identificado hasta ahora que está relacionado con el metabolismo de carbono.

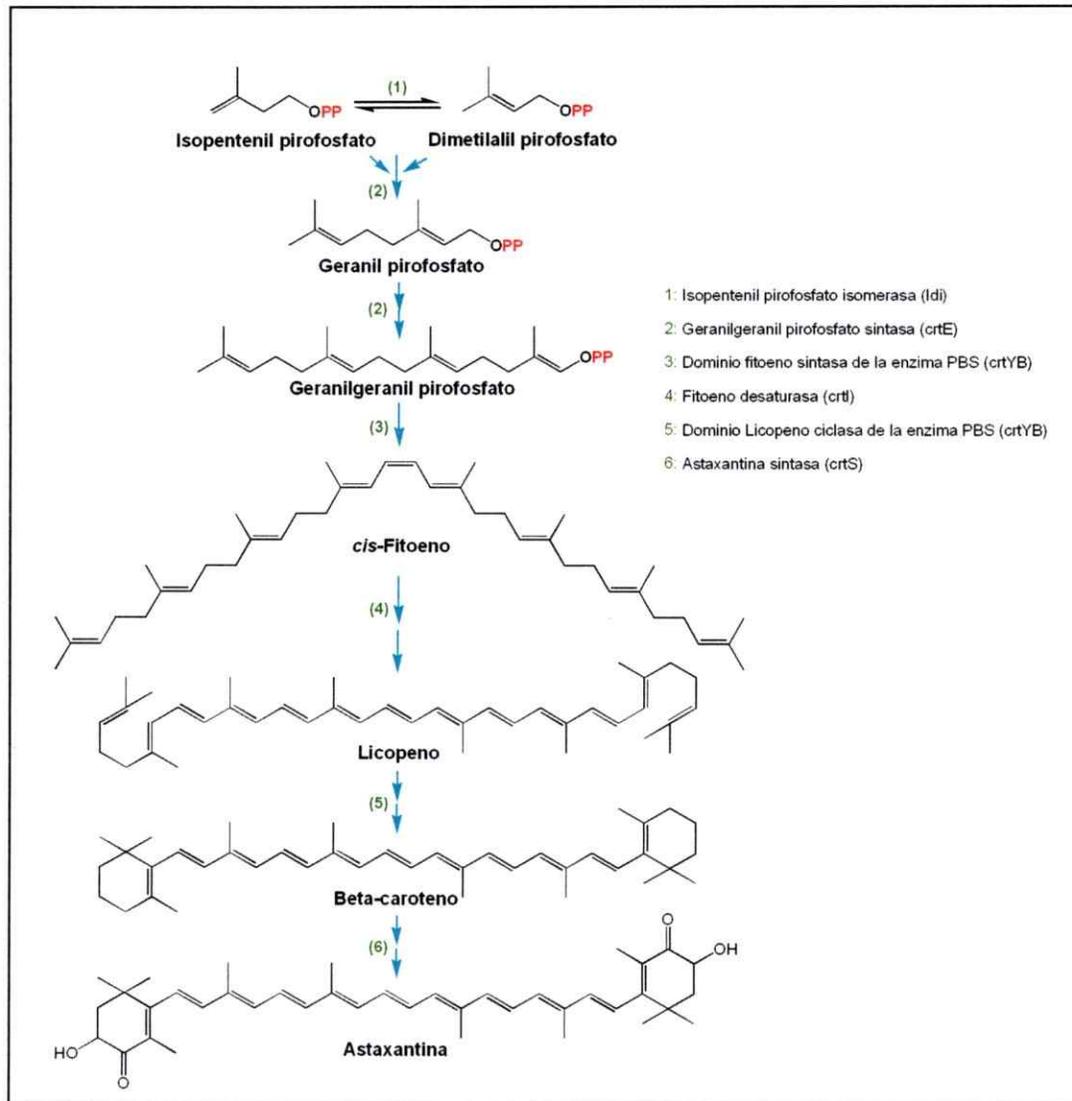


Figura 4. Ruta de Biosíntesis de astaxantina a partir de IPP en *X. dendrorhous*. La biosíntesis de astaxantina a partir de IPP en *X. dendrorhous* involucra 6 actividades enzimáticas codificadas por 5 genes (mostrados entre paréntesis en la leyenda).

El evidentemente reducido número de genes identificados en la levadura ha dificultado el desarrollo de estudios sobre aspectos relevantes de la fisiología del gen tales como el uso preferencial de codones, estructura exón-intrón, procesamiento alternativo de ácidos ribonucleicos mensajeros (mRNAs) y tamaño medio de estos últimos. Similarmente, se desconoce la existencia de elementos de secuencia comunes en grupos de genes cuyos productos proteicos están funcionalmente relacionados, y si su

presencia cumple un rol regulador del metabolismo. Igualmente hay escasa información con respecto a los genes que están involucrados directamente en los procesos del metabolismo del carbono (glucólisis, Ciclo de Krebs, etc.), o bien de proteínas reguladoras de los mismos que bien podrían cumplir un papel clave dentro de la biosíntesis de carotenoides, ya sea directamente sobre reacciones de la vía, o bien controlando la disponibilidad de sustratos precursores provenientes de los procesos metabólicos mencionados anteriormente.

El presente trabajo propone la identificación y caracterización estructural de nuevos genes aislados a partir de *X. dendrorhous* expresados en condiciones de metabolismo fermentativo; además de la realización de un análisis de uso de codones en la levadura y el estudio genético-estructural de aquellos genes que puedan tener alguna relación con la biosíntesis de astaxantina o un potencial interés biotecnológico.

1.4 Propósito del trabajo

La construcción de genotecas de cDNA en *X. dendrorhous* permitirá aislar un número significativo de genes hasta ahora desconocidos de este organismo. La caracterización de dichos genes hará posible determinar los patrones de uso preferencial de codones y de la estructura exón-intrón ocurrentes en la levadura, aportando nuevos antecedentes sobre sus características genómicas.

1.5 Objetivo general

Identificar y caracterizar genes expresados durante el metabolismo fermentativo de *X. dendrorhous*, especialmente aquellos que podrían tener relación con la biosíntesis de astaxantina o un potencial interés biotecnológico.

1.6 Objetivos específicos

1. Construir una genoteca de cDNA de *X. dendrorhous* crecida en una fuente de carbono fermentable.
2. Seleccionar clones al azar a partir de la genoteca de cDNA, secuenciar e identificar genes que podrían tener relación con el proceso de carotenogénesis, o que posean un potencial interés biotecnológico.
3. Construir una tabla de uso de codones de *X. dendrorhous* a partir de los genes identificados anteriormente y los que están publicados en la base de datos internacional.
4. Clonar y secuenciar los segmentos de DNA genómico correspondientes a distintos genes identificados a partir de la genoteca de cDNA.
5. Determinar la estructura exón-intrón de genes de interés aislados desde la genoteca.

2. MATERIALES Y MÉTODOS

2.1 Cepas y condiciones de cultivo.

Para la realización del presente estudio, y a menos que se diga expresamente lo contrario se utilizó la cepa silvestre UCD 67-385 de *X. dendrorhous*. Esta cepa fue crecida a 22 °C en medio YM (1% glucosa; 0,3% extracto de levadura; 0,3% extracto de malta; 0,5% peptona) o en medio mínimo Vogel (modificado según Retamales y col., 2002) suplementado con una fuente de carbono adecuada, según corresponda.

La genoteca de cDNA fue construida y propagada en la cepa XL10-GOLD de *E. coli*, la cual fue crecida a 37 °C en medio LB (1% triptona, 0,5% extracto de levadura, 1% NaCl) suplementado con ampicilina (100 µg/ml).

El medio NZY⁺ fue preparado colocando en un recipiente 10 g de NZ amina, 5 g de extracto de levadura y 5 g de NaCl, agregando agua destilada hasta completar un volumen final de 1 litro. Posteriormente se ajustó el pH hasta 7,5 usando NaOH, luego de lo cual la mezcla fue autoclavada 15 min a 1,5 atm. Inmediatamente antes de usar se le añadió al medio 12,5 ml de MgCl₂ 1M; 12,5 ml de MgSO₄ 1M y 20 ml de una solución 20% (p/v) glucosa.

2.2 Construcción de genotecas de cDNA en *X. dendrorhous*: estrategia general.

Para la construcción de las genotecas se empleó el "pBluescript II XR cDNA library construction kit" (Stratagene), el cual usa como sustrato inicial mRNA poli A⁺. El proceso de obtención de estas librerías contempló a grandes rasgos los siguientes pasos: obtención de RNA total, purificación de RNA mensajero, síntesis de cDNA de doble hebra, ligado de adaptadores, fraccionamiento por tamaño, ligación con vector plasmidial y transformación (Figura 5A). Una vez obtenido el mRNA, los pasos ulteriores fueron realizados de acuerdo a los protocolos designados por el fabricante

del kit. Luego de contar con el mRNA poliA⁺ purificado se llevo a cabo la síntesis de la primera hebra de cDNA, proceso que requirió un molde y un partidor. El molde usado fue el mRNA y como partidor se empleó un oligonucleótido de 50 bases compuesto por un sitio de restricción *Xho*I, una secuencia repetida GAGA para proteger dicho sitio de restricción, y una secuencia poli-dT de 18 bases (Figura 5B).

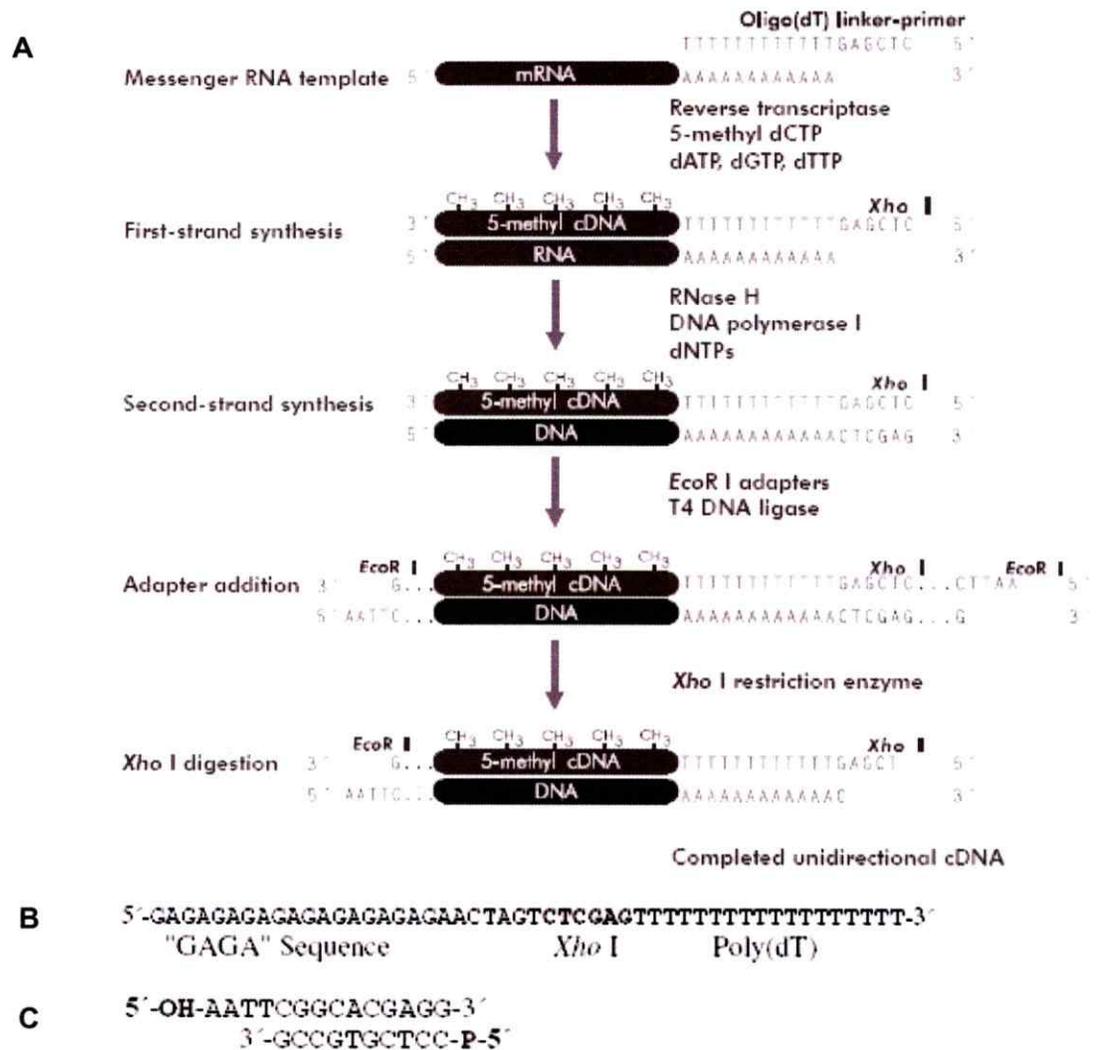


Figura 5. Estrategia general de síntesis de cDNA a partir de mRNA para la construcción de las genotecas. (A) Resumen esquemático de la estrategia general (adaptado del manual de usuario del kit de construcción de la genoteca). (B) Oligonucleótido de 50 bases usado en la síntesis de la primera hebra de cDNA. (C) Adaptador *Eco*RI.

El sitio de restricción *Xho*I permitió que el cDNA obtenido al final del proceso pudiese ser insertado en el vector pBluescript II en una determinada orientación (*Eco*RI-*Xho*I). La mezcla de nucleótidos utilizada para la síntesis de la primera hebra contenía dATP, ³²P-γ-dATP, dGTP y dTTP, además del análogo 5-metil-dCTP. De esta manera, la primera hebra sintetizada poseía un grupo metilo en cada citosina, lo que permitió proteger el cDNA de la digestión por enzimas de restricción utilizadas en pasos posteriores.

La síntesis de la segunda hebra de cDNA se hizo según el método de "nick translation", donde la enzima RNasa H corta el RNA que permanece unido a la primera hebra de cDNA generando una multitud de fragmentos que sirven como partidores para la DNA polimerasa I, que sintetiza la segunda hebra de cDNA. La mezcla de nucleótidos utilizados en esta reacción fue suplementada con dCTP sin modificaciones de modo de disminuir la probabilidad de que la 5-metil-dCTP fuese incorporada en la segunda hebra. Esto permitió que el sitio de restricción presente en el partidor de 50 bases fuese susceptible a la digestión con enzimas de restricción. Posteriormente los extremos de hebra simple del cDNA fueron rellenados usando *Pfu* DNA polimerasa, de modo de generar extremos romos, a los cuales se les ligó los adaptadores *Eco*RI (figura 5C). Estos adaptadores están compuestos de oligonucleótidos de 10 y 14 bases que son complementarios entre sí, los cuales al aparearse generan un extremo cohesivo *Eco*RI. El oligonucleótido de 10 bases se encuentra fosforilado, lo que le permite ligarse a otros extremos romos disponibles en el cDNA u otros adaptadores. El oligonucleótido de 14 bases se mantuvo desfosforilado para prevenir que se ligase a otros extremos cohesivos. Una vez inactivada la ligasa y finalizada la ligación de los adaptadores, el extremo del oligonucleótido de 14 bases fue fosforilado, de manera de posibilitar su ligación al vector desfosforilado.

Posteriormente se hizo una digestión con *Xho*I, lo que permitió liberar el adaptador *Eco*RI y el fragmento residual del partidor de 50 bases del extremo 3' del cDNA. Luego la mezcla fue fraccionada por tamaño en una columna de filtración en gel, donde los fragmentos provenientes del extremo 3' fueron separados del cDNA de doble hebra. El cDNA fraccionado por tamaño fue posteriormente precipitado y ligado al vector pBluescript II SK(+) predigerido con *Eco*RI y *Xho*I (Stratagene).

Finalmente, con el DNA ligado se transformó células competentes XL10-Gold. Debido a que la mayoría de las cepas de *E. coli* digieren DNA hemimetilado, fue de vital importancia utilizar una cepa *McrA*⁻ *McrB*⁻ (mutaciones en el sistema de restricción-modificación) para obtener un buen rendimiento en la transformación.

2.3 Extracción de RNA total.

Para la extracción de RNA total se utilizó células provenientes de cultivos en fase exponencial tardía, crecidos en MMv glucosa o MMv succinato. Una vez alcanzada la densidad óptica deseada, las células fueron colectadas centrifugando durante 5 min a 10000 rpm. El sedimento resultante fue transferido a un mortero tratado con dietilpirocarbonato (DEPC), y se agregó 1 volumen de nitrógeno líquido. Una vez evaporado el nitrógeno y congeladas las células, se adicionó 10 ml de TRIAGENT, y se homogenizó durante 5 min. Posteriormente se agregó a la mezcla 2 ml de cloroformo, se homogenizó, se alicuotó en tubos eppendorff y se agitó en vortex durante 5 min. Se centrifugó las muestras durante 15 min a 4 °C y 10000 rpm, se recuperó la fase acuosa y se le añadió 1 volumen de isopropanol, incubando posteriormente esta mezcla a -20 °C durante 2 h. Posteriormente se centrifugó las muestras durante 10 min a 10000 rpm (4 °C). El RNA sedimentado se lavó con etanol 75% y finalmente se resuspendió en H₂O tratada con DEPC.

2.4 Purificación de mRNA poli A⁺.

40 mg de Oligo-dT celulosa fueron lavados 3 veces con "Elution Buffer" (10 mM Tris pH 7,5; 1 mM EDTA; 0,05% SDS), y posteriormente 2 veces con "1X Binding Buffer" (0,5 M NaCl; 10 mM Tris pH 7,5; 1 mM EDTA; 0,05% SDS). De forma paralela, el RNA total fue llevado a un volumen final de 600 µl con H₂O libre de RNasa y se incubó a 65° C durante 4 min. Luego se agregó 600 µl de 2X Binding Buffer y los 40 mg de oligo-dT celulosa equilibrada en 1X Binding Buffer. La mezcla resultante se incubó durante 15 min a temperatura ambiente con agitación suave permanente. Al cabo de este periodo la mezcla fue centrifugada para posteriormente quitar el sobrenadante. El pellet fue lavado 2 veces con 1X Binding Buffer y otras 2 veces con "Wash Buffer" (0,2 M NaCl; 10 mM Tris pH 7,5; 1 mM EDTA; 0,05% SDS), después de lo cual fue resuspendido en H₂O libre de RNasa e incubado a 37° C durante 5 min Finalmente la mezcla fue centrifugada y el sobrenadante se traspasó a un tubo eppendorff estéril. Este último paso fue repetido una segunda vez, juntando los volúmenes obtenidos en cada caso.

2.5 Síntesis de cDNA de doble hebra.

La síntesis de cDNA de doble hebra fue realizada de acuerdo a los protocolos indicados en el manual de usuario del kit "pBluescript® II XR cDNA Library Construction" (Stratagene).

2.6 Fraccionamiento del cDNA por tamaño.

El fraccionamiento por tamaños del cDNA obtenido se realizó por cromatografía de filtración en gel utilizando una columna de sefarosa CL-2B. El medio filtrante fue empacado usando como columna una pipeta de 1 ml (graduada hasta -0,3 ml en su extremo superior) y una jeringa desechable de 10 ml como reservorio de buffer, la cual fue unida al extremo superior de la pipeta. Inicialmente la columna fue lavada con 10 ml de buffer STE (0,1 M NaCl; 20 mM Tris-HCl pH 7,5; 10 mM EDTA) y luego se

procedió a agregar el medio filtrante, llenando hasta 5 mm por debajo de la unión entre la jeringa y la pipeta. Posteriormente se lavó la columna haciendo pasar 5 ml de buffer STE. Luego se cargó la muestra de cDNA previamente mezclada con colorante de carga (50% v/v glicerol; 10% v/v 10X STE; 40% p/v azul de bromofenol saturado). Apenas el frente delantero de la banda de colorante alcanzó la marca de -0,3 ml, se comenzó a coleccionar fracciones de tres gotas cada una ($\approx 100 \mu\text{l}$). La última fracción se coleccionó cuando el extremo trasero de la banda de colorante alcanzó la marca de 0,3 ml.

2.7 Determinación del rango de tamaño de cada fracción obtenida.

Alícuotas de 8 μl de cada fracción coleccionada luego de la cromatografía fueron sometidas a electroforesis en un minigel TAE-agarosa 1,1% teñido con bromuro de etidio (0,5 $\mu\text{g/ml}$). Posteriormente el gel fue colocado sobre papel 3MM y se dejó secar en un secageles de vacío a 40 °C durante 1 h. Una vez transcurrido este periodo, el gel fue puesto sobre una película para rayos X, la que fue expuesta durante 48 horas a -70 °C siendo posteriormente revelada y analizada.

2.8 Reacción de Ligado del cDNA.

El ligado del cDNA al vector plasmidial (pBluescript II) fue realizado en una mezcla de reacción compuesta por 20 ng de vector predigerido pBluescript II XR (Stratagene); 2U de T4 DNA ligasa (Stratagene); 1X buffer de ligasa (Stratagene); 1 mM rATP y cantidades variables de cDNA en un volumen final de 5 μl . La mezcla de reacción fue incubada durante 24 h a 10 °C.

2.9 Transformación de *E. coli*.

La transformación con la mezcla de ligado del cDNA se realizó utilizando células competentes de *E. coli* XL10-Gold (Stratagene). Para ello 150 μl de células competentes fueron alicuotadas en tubos de polipropileno Falcon 2059 previamente enfriados en hielo. Posteriormente se agregó 4 μl de mezcla de β -mercaptoetanol

(Stratagene), incubando en hielo durante 10 min agitando suavemente cada 2 min. Al cabo de este período se agregó la mezcla completa de ligación, incubando en hielo durante 30 minutos. Luego se dió un pulso de temperatura a 42 °C durante 30 s, se incubó en hielo 2 min, se agregó 0,9 ml de medio NZY⁺ precalentado a 42 °C, y se incubó la mezcla durante 1 h a 37 °C. Finalmente, la mezcla de transformación se distribuyó en 5 placas de medio LB-ampicilina, las que fueron incubadas a 37 °C entre 14 y 18 h.

2.10 Secuenciación de DNA.

La secuencia nucleotídica de muestras de DNA plasmidial, o bien de fragmentos amplificados por PCR, se realizó utilizando un secuenciador automático ABI 3100 Avant (Applied Biosystems). En el primer caso, el DNA plasmidial fue previamente purificado empleando el kit Wizard[®] Plus SV Minipreps DNA Purification System (Promega). En el caso de fragmentos amplificados por PCR, estos fueron purificados directamente a partir de un gel de agarosa, luego de ser sometidos a electroforesis. Para la purificación del DNA a partir de geles de agarosa se utilizó el kit GeneClean (QBiogen).

El análisis de las secuencias nucleotídicas obtenidas se hizo empleando el paquete de software Vector NTI Suite 9.0 (Informax), junto con otras aplicaciones disponibles en la WEB. Para la búsqueda de secuencias homólogas en bases de datos internacionales se usó el conjunto de aplicaciones BLAST, disponible en el sitio WEB <http://www.ncbi.nlm.nih.gov/blast/>.

2.11 Cálculo del Índice de Adaptación de Codones (CAI)

El cálculo del CAI se realizó de acuerdo a las ecuaciones formuladas por Sharp y Li (Sharp y Li, 1987). Brevemente, en primer lugar se construyó a partir de un set de genes de referencia (designado arbitrariamente) una tabla de uso relativo de codones

sinónimos (RSCU). Los valores de RSCU corresponden simplemente al cociente entre la frecuencia observada de cada codón y la esperada asumiendo igual uso de codones sinónimos para cada aminoácido. Luego, a partir de los valores de RSCU, se calculó el grado de adaptación relativa de cada codón (W_{ij}), que representa la frecuencia de uso de un determinado codón comparada con la frecuencia del codón más usado (óptimo) para cada aminoácido. Una vez construida la tabla de referencia, se calculó el valor de CAI para cada gen. El CAI se obtiene como la media geométrica de los valores de RSCU (tabla de referencia) correspondientes a cada uno de los codones usados en el gen, dividido por el CAI máximo posible (usando solo los codones preferidos) para un gen de la misma composición aminoacídica.

3. RESULTADOS

3.1 Construcción de genotecas de cDNA de *X. dendrorhous*

3.1.1 Extracción de RNA total de *X. dendrorhous* y purificación de mRNA poli A⁺.

Dado que se ha demostrado que la fuente de carbono para el crecimiento de *X. dendrorhous* influye en la producción de pigmentos y en la expresión génica de dicho organismo, se construyeron dos genotecas de cDNA; una usando glucosa como única fuente de carbono y la otra utilizando succinato.

A partir de 400 ml de cultivo en medio mínimo glucosa y otros 400 ml de cultivo en medio mínimo succinato se obtuvo, luego del proceso de extracción, una solución de aproximadamente 500 µl de RNA total para cada caso. Cada una de estas muestras fue posteriormente cuantificada empleando un espectrofotómetro NANODROP. A partir de ese mismo ensayo espectrofotométrico se obtuvo los parámetros de interés que dan cuenta de la pureza de la muestra de RNA, los cuales corresponden a las relaciones de absorbancia $A^{260\text{nm}}/A^{230\text{nm}}$ y $A^{260\text{nm}}/A^{280\text{nm}}$ (Tabla 1). Una razón cercana a 2 indica la ausencia de cantidades detectables de contaminantes, aunque valores superiores a 1,7 se aceptan como suficientes para diversas aplicaciones (Sambrook y cols., 2001).

Tabla 1. Cuantificación y determinación de parámetros de interés del RNA total extraído.

Muestra	Concentración (µg/µl)	A^{260}	A^{260}/A^{230}	A^{260}/A^{280}	Masa total (µg)
MMv-glucosa	1,3069	32,672	1,69	1,91	650
MMv-succinato	0,9984	28,661	1,81	2,11	500

Con el objeto de analizar la integridad del RNA total obtenido se realizó un ensayo de electroforesis capilar en condiciones desnaturantes, empleando un equipo Bioanalyzer 2000 (Figura 6). Este ensayo permitió establecer si el RNA total obtenido estaba suficientemente libre de degradación como para ser utilizado en la posterior construcción de las genotecas de cDNA.

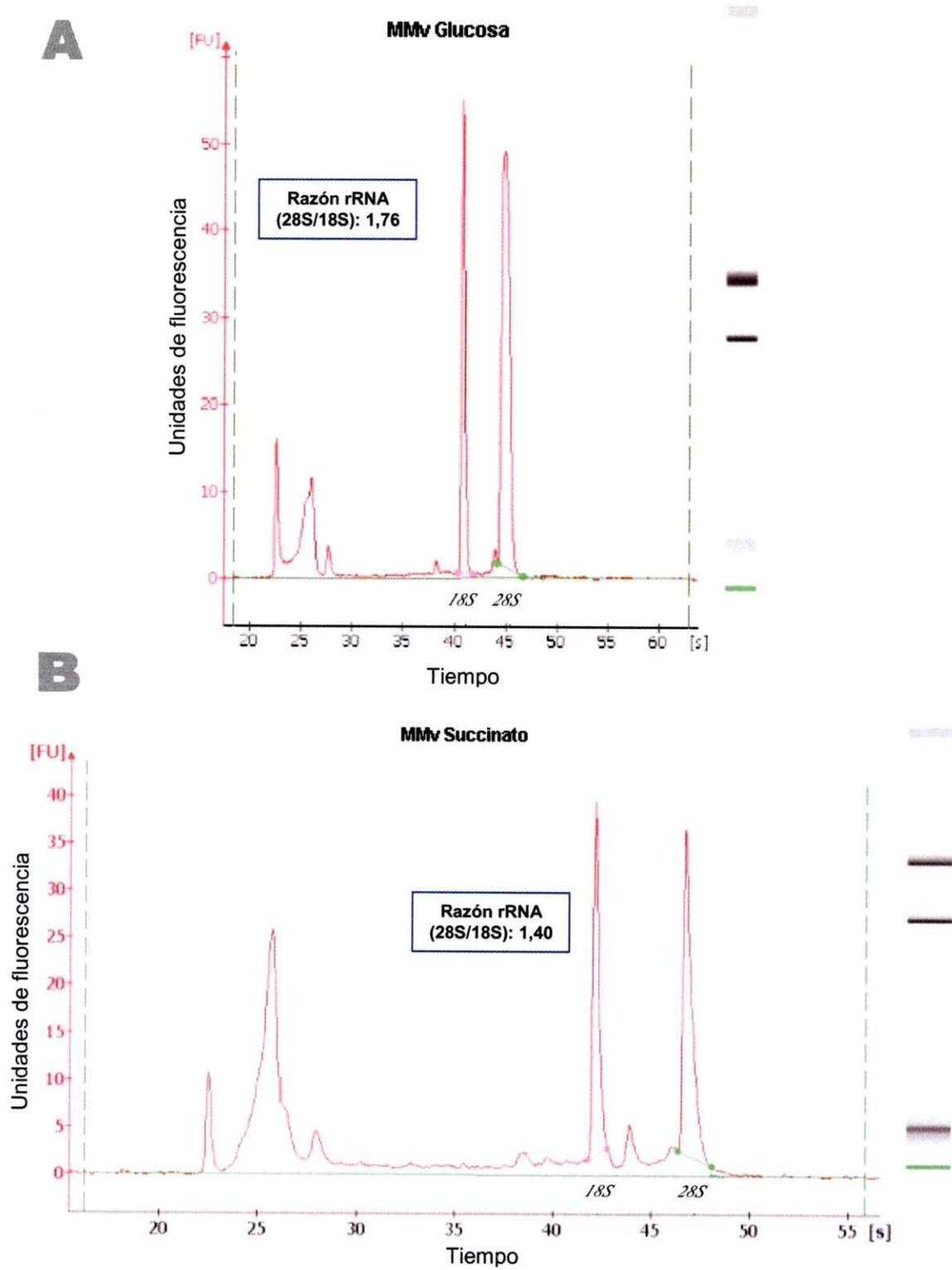


Figura 6. Electroforesis capilar en condiciones denaturantes de las muestras de RNA total extraído. El RNA fue extraído a partir de células crecidas en medio mínimo glucosa (A) y medio mínimo succinato (B). Para ambos casos se muestra en un recuadro la relación de áreas de los peaks correspondientes a los rRNAs 28S y 18S.

A partir del ensayo de electroforesis capilar se obtuvieron dos electroferogramas (uno con cada fuente de carbono), donde se muestran las unidades arbitrarias de fluorescencia versus el tiempo en que las moléculas de RNA tardan en llegar al detector. En ambos casos se puede ver claramente los picos correspondientes a los RNAs ribosómicos 18S y 28S marcados en la figura con letras rosadas y verdes, respectivamente. Un parámetro comúnmente utilizado para determinar integridad del RNA en este tipo de ensayos consiste en calcular la relación de áreas bajo los picos 28S y 18S. El valor ideal esperado es 2,0, aunque normalmente valores superiores a 1,2 caen en un rango de integridad aceptable. En el caso de las células crecidas en glucosa, el RNA extraído presenta una relación 28S/18S de 1,76, mientras que para el RNA de las células crecidas en succinato, la relación 28S/18S es de 1,40.

Los resultados de la cuantificación, análisis de pureza y análisis de integridad indican para ambos casos un RNA suficientemente íntegro y puro como para ser usado en la construcción de las genotecas de cDNA.

Luego de obtener un RNA total íntegro y puro se procedió a purificar el RNA mensajero poliadenilado en el extremo 3' (mRNA poli A⁺), por medio de cromatografía en lotes usando como resina oligo-dT celulosa. Una vez finalizado el procedimiento de purificación se obtuvo, a partir de los 500-600 µg de RNA total extraído para cada fuente de carbono, una masa total cercana a 5 µg de una fracción enriquecida en mRNA poli A⁺. Debido a la pequeña cantidad de mRNA obtenido luego de la purificación no se realizó un análisis de la integridad de este último.

3.1.2 Síntesis de cDNA para la construcción de las genotecas.

Para realizar la síntesis de cDNA de doble hebra se utilizaron los 5 µg de la fracción enriquecida en mRNA poli A⁺ obtenidos para cada fuente de carbono como se

describió en la sección anterior. La estrategia general de síntesis utilizada corresponde a la descrita en el apartado 2.2 de Materiales y Métodos, mientras que el detalle de los protocolos se encuentra en el manual de usuario del kit de construcción de la genoteca (pBlueScript II XR Instruction manual). Los procedimientos de síntesis llevados a cabo en esta etapa permitieron obtener un cDNA de doble hebra hemimetilado y marcado radiactivamente, con extremos cohesivos correspondientes a sitios *EcoRI* y *XhoI* en los extremos 5' y 3' respectivamente.

3.1.3 Fraccionamiento del cDNA por tamaño

El siguiente paso en la construcción de la genoteca fue fraccionar el cDNA en base a su tamaño, de modo de remover los adaptadores de bajo peso molecular generados luego de la digestión con *XhoI*, y además seleccionar el rango de tamaño de cDNA deseado para ligar con el vector plasmidial. Para esto se montó una columna de sefarosa CL-2B, por la cual se hizo pasar el cDNA de doble hebra previamente sintetizado a partir del RNA obtenido en cada fuente de carbono. Durante el proceso, en cada muestra (glucosa y succinato) se colectaron 11 fracciones de aproximadamente 100 µl cada una, los que fueron analizados por electroforesis en un gel TAE-agarosa teñido con bromuro de etidio, usando como estándar de peso molecular DNA de fago lambda digerido con *HindIII*. Luego de registrar la migración de las bandas del estándar en un transiluminador U.V., el gel fue secado y utilizado para exponer una película de rayos X. La figura 7 muestra los resultados de la autoradiografía de los geles en que se analizaron las fracciones colectadas para cada caso.

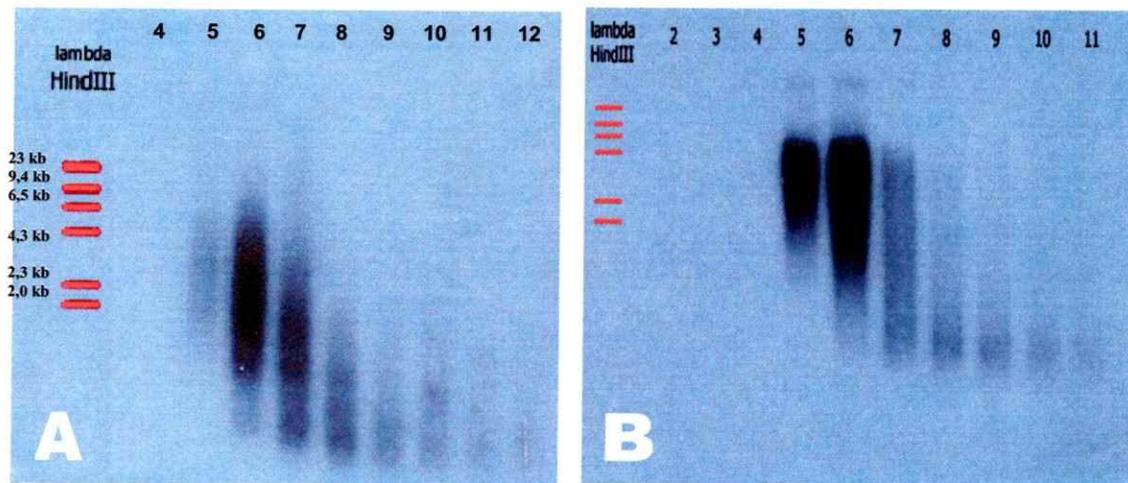


Figura 7. Autoradiografía de las fracciones obtenidas luego de la separación por tamaño del cDNA (marcado con ^{32}P), posteriormente sometidas a electroforesis en geles de agarosa. (A) Fracciones 4 a 12 del cDNA obtenido en glucosa. (B) Fracciones 2 a 11 del cDNA obtenido en succinato. Las bandas rojas corresponden al estándar de peso molecular λ HindIII.

En base a los resultados obtenidos para cada condición, se seleccionaron las fracciones que mostraron un rango de tamaño adecuado para efectuar la posterior ligación con el vector plasmidial. Tanto para MMv-glucosa como para MMv-succinato las fracciones escogidas fueron la número 5, 6 y 7 puesto a que contenían un cDNA de tamaño diverso, pero que en promedio se acercaba a las 1,5 kb. El cDNA de las fracciones escogidas para cada caso fue reunido en un solo tubo y concentrado mediante precipitación con etanol absoluto. Posteriormente el precipitado resultante fue resuspendido en H_2O bidestilada, constituyendo un volumen final de 10 μl .

3.1.4 Ligado del cDNA al vector plasmidial y transformación en *E. coli*

El cDNA sintetizado para cada fuente de carbono fue ligado al vector plasmidial pBluescript II SK (+) digerido con las enzimas de restricción *EcoRI* y *XhoI*. Luego de ligar se procedió a transformar células competentes de *E. coli*. Para de obtener el número deseado de clones fue necesario realizar múltiples intentos de reacciones de

ligación y transformación, variando la cantidad de cDNA y/o las condiciones de reacción utilizadas. Pese a los múltiples intentos de ligado y transformación, sólo fue posible obtener un número aceptable de clones a partir del cDNA de la levadura crecida en glucosa. Inexplicablemente en el caso del cDNA obtenido en succinato nunca fue posible lograr un número aceptable de clones, por lo que al final del proceso solo se logró construir una genoteca de cDNA de células crecidas en glucosa. En adelante los resultados mostrados corresponden sólo al caso de la genoteca de cDNA en glucosa.

La tabla 2 muestra los resultados de las pruebas de transformación con reacciones de ligado realizadas en distintas condiciones. Se muestra además el experimento definitivo (nº6), en el cual se logró sobre 25.000 clones recombinantes. Cabe destacar que para los primeros cuatro intentos se ligó con 1 µl de una dilución preparada tomando 1 µl del volumen total de cDNA de cada fuente de carbono, diluido 50 veces. Los tres primeros intentos corresponden básicamente a réplicas del mismo procedimiento, efectuados para descartar errores experimentales. En el cuarto intento se quiso probar otro método de transformación posiblemente más eficiente, y que además permitiese averiguar si las células competentes incluidas en el kit estaban estropeadas, y si eran de este modo las responsables de la baja eficiencia (aunque al transformar con pUC18 como control se obtuvo siempre un alto número de clones). Con este objetivo se preparó células electrocompetentes a partir de una alícuota de células XL10-Gold de acuerdo al protocolo indicado en el manual "Molecular Cloning" (Sambrook y cols., 2001). Al transformar por electroporación tampoco se logró obtener un número suficiente de clones (intento 4), aún aumentando en 10 veces la cantidad de cDNA utilizada para ligar (intento 5a), o empleando una T4 DNA ligasa diferente (intento 5b).

Tabla 2. Pruebas de ligación y transformación con el cDNA MMv-glucosa (incluido en el vector pBluescript II) realizadas hasta conseguir el número de clones deseado.

Intento Nº	Ligado		Transformación			
	Cantidad cDNA	Tiempo y Tº incubación	Ligasa y Buffer utilizado	Método Empleado	Células Competentes	Número total de clones obtenidos
1	1 µl dilución 1:50	48 horas, 4 °C	Reactivos kit Stratagene	Shock térmico a 42 °C	XL10-Gold Stratagene	≈900
2	1 µl dilución 1:50	48 horas, 4 °C	Reactivos kit Stratagene	Shock térmico a 42 °C	XL10-Gold Stratagene	≈30
3*	2 µl dilución 1:50	48 horas, 10°C	Reactivos kit Stratagene	Shock térmico a 42 °C	XL10-Gold Stratagene	3
4*	2 µl dilución 1:50	24 horas, 10°C	Reactivos kit Stratagene	Electroporación	XL10-Gold Electrocomp.	0
5a*	1 µl dilución 1:5	48 horas, 4 °C	Reactivos kit Stratagene	Electroporación	XL10-Gold Electrocomp.	0
5b*	1 µl dilución 1:5	48 horas, 4 °C	T4 Ligasa y buffer Roche	Electroporación	XL10-Gold Electrocomp.	0
6	Total cDNA restante	24 horas, 10°C	Reactivos kit Stratagene	Shock térmico a 42 °C	XL10-Gold Stratagene	≈25000

(*) Paralelamente se transformó con pBS o pUC18 como control de eficiencia de las células competentes, obteniendo un número de transformantes dentro de lo esperado en todos los casos. Los resultados corresponden sólo al cDNA obtenido en glucosa.

En la prueba definitiva a los aproximadamente 8 µl de cDNA restantes (sin diluir) se les agregó 400 µl de isopropanol y 50 µl de acetato de amonio 7,5M. Se incubó durante 2 h a -20 °C de modo que precipitara el cDNA, y posteriormente se centrifugó 15 min a 4 °C. Luego de eliminar el sobrenadante se lavó el pellet con etanol 70%, se secó, y finalmente se resuspendió en 3 µl de buffer TE (10 mM Tris-HCl pH 7,5; 1mM EDTA). Luego el total del volumen obtenido fue usado para ligar y finalmente transformar células XL10-Gold por tratamiento térmico a 42 °C. En esta ocasión se obtuvo un número suficiente de transformantes, estimado en aproximadamente 25000. Una vez obtenido este número de transformantes se procedió a almacenar la genoteca para su posterior utilización.

3.1.5 Determinación del tamaño medio de inserto de la genoteca de cDNA

Una vez obtenida la genoteca en MMv-glucosa se procedió a determinar el tamaño medio de inserto que poseen los clones que conforman dicha librería. Para esto se realizó 50 reacciones de PCR usando como molde el DNA de 50 clones seleccionados

al azar, con partidores que flanquean la zona del vector pBluescript II donde se introdujo el inserto. Luego, las reacciones de PCR fueron sometidas a electroforesis en geles de agarosa, los cuales fueron posteriormente fotografiados y digitalizados para calcular el tamaño del amplicón obtenido en cada caso (Figura 8).

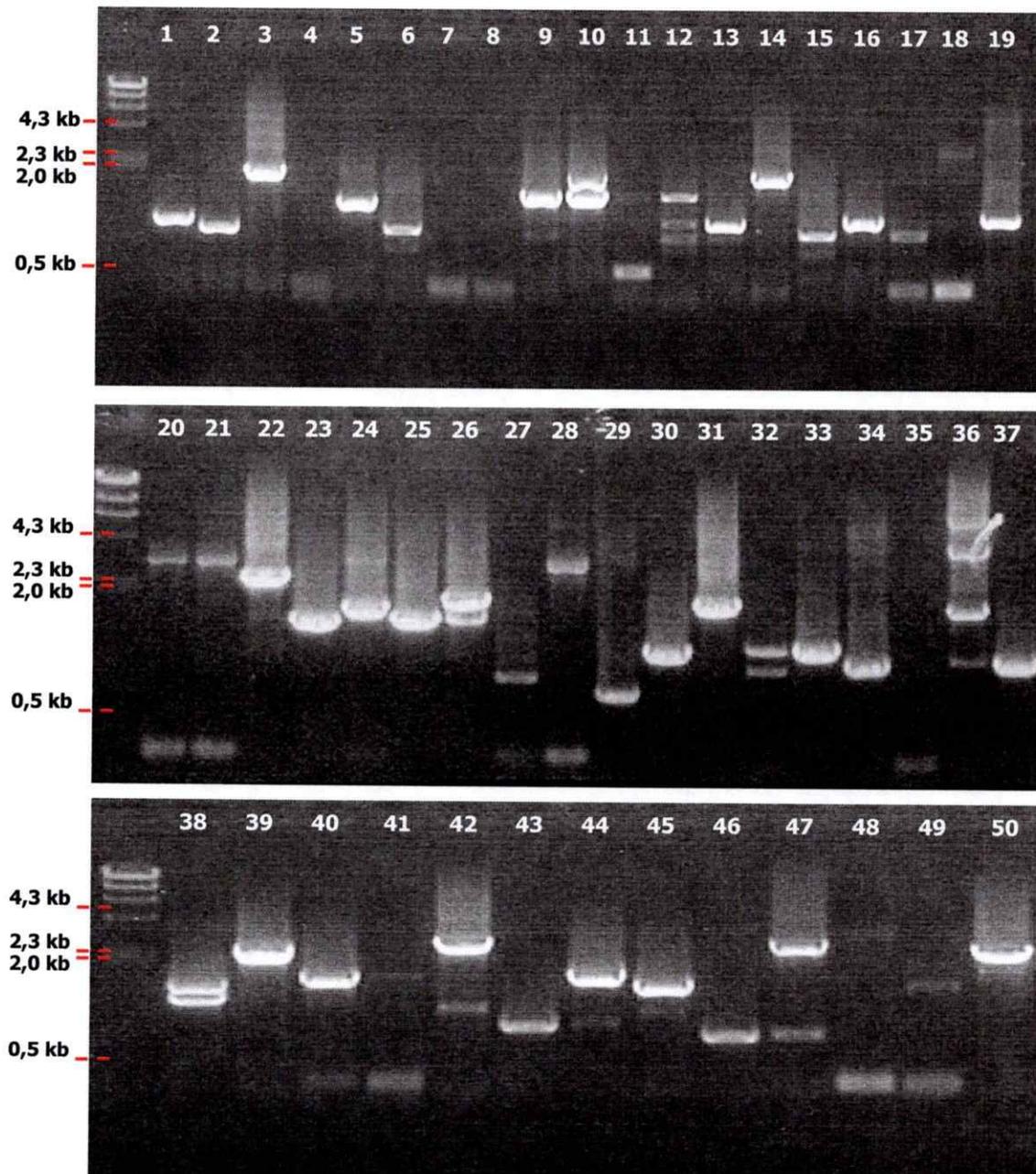


Figura 8. Electroforesis en gel de agarosa de los amplicones obtenidos por PCR correspondientes al inserto de cDNA de 50 clones de la genoteca seleccionados al azar. El primer carril de cada gel corresponde al estándar de peso molecular λ /HindIII.

Una vez determinado el tamaño del inserto que portaba cada clon, se calculó la media aritmética de dichos tamaños junto con su desviación estándar. Los tamaños calculados para cada amplicón junto con la media calculada se presentan en la tabla 3.

Tabla 3. Tamaño del inserto portado por 50 clones tomados al azar desde la genoteca de cDNA.

Nº clon	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Tamaño (kb)	1,0	0,9	1,8	R	1,3	0,9	R	R	1,4	1,6*	0,4	1,5*	1,0	1,7	0,9	1,1	0,9
Nº clon	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
Tamaño (kb)	3,1	1,3	3,0	3,0	2,3	1,5	1,7	1,6	1,7*	0,7	2,9	0,6	1,0	1,8	1,2*	1,1	0,8
Nº clon	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	
Tamaño (kb)	R	1,6*	0,8	1,1*	2,2	1,6	R	2,5	0,8	1,7	1,5	0,7	2,5	R	1,8	2,4	

Se muestra además la media calculada a partir de 45 clones (no se consideran los religados). (*) En los casos en que aparece más de una banda se consideró el tamaño promedio entre ellas (clones 10, 26, 32, 38) o bien la banda más intensa (clones 12 y 36). Los clones marcados con "R" son aquellos que no generan amplicones pues portaban vector sin inserto. El tamaño medio determinado es de $1,5 \pm 0,7$ kb.

A partir de los resultados obtenidos fue posible estimar el rango de tamaños de inserto que presenta la genoteca recién construida, comprobando que estaba de acuerdo con lo esperado al considerar los tamaños de las fracciones de cDNA elegidas para ligar. Fue posible constatar además la baja proporción de clones que portaban vector sin inserto (12%) y una gran diversidad de tamaños, denotada por una desviación estándar considerable (0,7 kb). De acuerdo a lo anterior se puede inferir que la genoteca posee insertos diversos (y por lo tanto mensajeros diversos), probablemente en su mayoría completos (dado el tamaño medio observado) y con una baja proporción de vectores sin inserto, lo que favorece la probabilidad de tener representados la mayoría de los genes expresados en glucosa de *X. dendrorhous*.

3.1.6 Almacenamiento y conservación de la genoteca.

La genoteca en MMv-glucosa fue fraccionada en 5 partes (correspondientes a las 5 placas obtenidas con aproximadamente 4.800 transformantes cada una). De cada placa se obtuvo 5 copias, cada una de las cuales fueron guardadas a -80 °C en criotubos con un volumen final de 2 ml LB-ampicilina-50% glicerol.

Para facilitar la posterior búsqueda de genes específicos en la genoteca, ésta fue fraccionada en mezclas con un menor número de clones, de acuerdo a la metodología que se describe a continuación. A partir de cada una de las cinco fracciones originales almacenadas a -80 °C (conteniendo ~4.800 transformantes cada una), se extrajo una alícuota que se usó para inocular 50 ml de medio LB-ampicilina. Luego de incubar cada cultivo a 37° C hasta saturación, un determinado volumen de la suspensión de células se usó para extraer DNA plasmidial, el cual fue guardado a -20 °C constituyendo una colección de los cDNAs contenidos en los 4.800 clones de la mezcla original. Una fracción del volumen de células restante fue diluido y esparcido en placas LB-ampicilina de modo de obtener colonias aisladas. Estas colonias fueron replicadas de modo de obtener 50 placas con 200 clones cada una (mezclas de 200 clones, por duplicado), y cada una de las placas de 200 clones fue luego replicada de modo de obtener 4 placas de 50 clones (mezclas de 50 clones, por duplicado). A una copia de cada mezcla (tanto de 200 como de 50 clones), se le extrajo DNA plasmidial y fue almacenado a -20 °C; la otra copia fue resuspendida en 1 ml de LB-ampicilina-50% glicerol y almacenada a -80 °C. De este modo, los cerca de 25000 clones que componen la genoteca original fueron repartidos en 5 mezclas con aproximadamente 5.000 clones cada una, las que a su vez se dividieron en 50 mezclas de 200 clones; siendo cada mezcla de 200 clones posteriormente dividida en 4 mezclas de 50 clones. Como a cada mezcla se le extrajo DNA plasmidial, la búsqueda de un determinado gen

en la genoteca puede hacerse directamente por PCR, acotando cada vez más el universo de clones distintos entre los cuales se encuentra el gen de interés.

3.2 Aislamiento e Identificación de genes expresados en metabolismo fermentativo de *X. dendrorhous*.

La genoteca de cDNA construida constituye una gran herramienta para la identificación de genes hasta ahora desconocidos de *X. dendrorhous*, ya sea realizando una selección al azar o bien una búsqueda dirigida de determinados genes de interés. En una primera instancia y con el objeto de realizar un aporte significativo a la base de datos que recopila los genes identificados de *X. dendrorhous*, además de tener una aproximación respecto a qué genes se están expresando en esa etapa del metabolismo, se seleccionó 113 clones al azar a partir de la genoteca. Cada uno de estos clones fue crecido en medio líquido y posteriormente se le extrajo DNA plasmidial, el que fue usado para determinar la secuencia del extremo 5' del cDNA incluido en cada plasmidio (originando lo que se denomina *expressed sequence tag* o EST). La secuencia obtenida en cada caso fue luego corregida y comparada con la base de datos internacional no redundante (nr) usando la herramienta *tBlastX* (Altschul y cols., 1997; <http://www.ncbi.nlm.nih.gov/BLAST>), la cual traduce virtualmente la secuencia problema en los seis marcos de lectura posibles y los compara con la base de datos de secuencias nucleotídicas también traducida, de modo de buscar homología a nivel de secuencia aminoacídica con todos los genes publicados en dicha base de datos y así determinar la posible función del supuesto gen representado por cada EST. De esta manera se generó una base de datos propia que actualmente contiene 113 EST's, de los cuales a 55 se les asignó una posible función en base a la

homología que presentan con otros genes identificados a partir de organismos filogenéticamente relacionados. Los 58 EST's restantes no presentaron homología significativa con ningún gen descrito hasta ahora.

Aquellos ESTs a los cuales se les asignó una posible función fueron posteriormente secuenciados por ambas hebras. Una vez que finalizó el proceso de secuenciación, los electroferogramas obtenidos para cada clon fueron corregidos y empalmados utilizando el software *contig express* (Informax). Luego, la secuencia reconstruida a partir de los distintos electroferogramas fue analizada con la aplicación ORF Finder (Open Reading Frame Finder, <http://www.ncbi.nlm.nih.gov/gorf/gorf.html>), la cual permite detectar posibles marcos de lectura abiertos (ORFs) al interior de una secuencia nucleotídica problema. De este modo se logró determinar en la mayoría de los clones un ORF designado arbitrariamente como principal, en base al tamaño de dicho marco, a la coherencia entre el codón de inicio y el codón de término y el tamaño de las regiones 5' y 3' no traducidas. Las secuencias nucleotídicas que no cumplían con los criterios anteriormente descritos probablemente constituían cDNAs generados a partir de mensajeros incompletos, producto de degradación parcial del mRNA usado para construir la genoteca. Esto último fue posteriormente comprobado a través de la comparación de las secuencias con la base de datos usando nuevamente la herramienta *tBlastX*, donde se pudo constatar que esos cDNAs eran más pequeños y carecían de algún segmento en relación a sus genes ortólogos.

Luego de la detección de un marco de lectura abierto principal (en los casos en que fue posible), la secuencia aminoacídica deducida a partir de este último fue comparada con la base de datos de proteínas (no redundante) utilizando la herramienta *Blastp* (Altschul y cols., 1997; <http://www.ncbi.nlm.nih.gov/BLAST>).

En la tabla 4 se encuentran resumidos los 55 genes identificados a partir de los clones seleccionados al azar desde la genoteca de cDNA, además del resultado del análisis con Blastp y el tamaño del ORF predicho (cuando fue posible) o bien el resultado del análisis con tBlastX.

Tabla 4. Resumen de los genes identificados a partir de la selección de clones al azar desde la genoteca de cDNA de *X. dendrorhous*.

Clon N°	Descripción	Tamaño ORF (pb)	Score (BlastP) ó (tBlastX)	E-value	I (%)	+ (%)	Organismo
1	Subunidad 20S proteosoma	-	189	6×10^{-42}	73	91	<i>Cryptococcus neoformans</i>
2	Factor de elongación de la traducción 1A (EFT1-A)	1305	816	0	92	95	<i>Schizophyllum commune</i>
3	Factor de elongación de la traducción 2 (EFT-2)	2493	1334	0	78	88	<i>Cryptococcus neoformans</i>
4	Factor de elongación de la traducción 1-gamma (EFT1-γ)	1248	424	7×10^{-117}	53	68	<i>Cryptococcus neoformans</i>
5	Antiportador ATP:ADP (ATP)	933	553	7×10^{-156}	88	93	<i>Cryptococcus neoformans</i>
6	Gen de stress sti35. Enzima de biosíntesis de tiazol THI2p	1080	389	1×10^{-106}	68	80	<i>Uromyces viciae-fabae</i>
7	Subunidad de 16 kDa complejo arp2/3	-	88	8×10^{-25}	42	59	<i>Cryptococcus neoformans</i>
8	RNA polimerasa II (45kDa) subunidad 3 (Rpb3)	1041	288	6×10^{-76}	50	67	<i>Cryptococcus neoformans</i>
9	Orotato fosforibosil transferasa	-	102	6×10^{-49}	60	79	<i>Saccharomyces cerevisiae</i>
10	Antiportador K ⁺ /H ⁺	-	183	6×10^{-44}	61	81	<i>Cryptococcus neoformans</i>
11	Disulfuro isomerasa	-	107	1×10^{-40}	75	84	<i>Aspergillus Fumigatus</i>
12	Factor citosólico SEC14	-	188	6×10^{-60}	58	76	<i>Saccharomyces cerevisiae</i>
13	Inositol-3-fosfato sintasa (MIPS)	1722	747	0	67	81	<i>Cryptococcus neoformans</i>
14	D-aminoácido oxidasa	-	99	1×10^{-11}	71	77	<i>Cryptococcus neoformans</i>
15	2,3 Triptófano dioxigenasa	-	81	2×10^{-34}	54	67	<i>Cryptococcus neoformans</i>
16	Zuotina	-	108	2×10^{-64}	65	74	<i>Cryptococcus neoformans</i>
17	Proteína de transporte Reticulo endoplásmico – Aparato de Golgi (Erv14)	411	207	2×10^{-52}	83	87	<i>Cryptococcus neoformans</i>

18	Gen de beta tubulina	1347	806	0	86	93	<i>Melampsora lini</i>
19	Transportador de aminoácidos neutros	-	136	8×10^{-40}	47	64	<i>Cryptococcus neoformans</i>
20	Partícula de reconocimiento de señales (SRP54)	-	113	3×10^{-46}	75	82	<i>Ustilago maydis</i>
21	Gen de poliubiquitina	-	385	$\frac{2 \times 10^7}{104}$	93	95	<i>Gracilaria Verrucosa</i>
22	Complejo adaptador de clatrina subunidad sigma (AP-1)	483	231	1×10^{-59}	72	88	<i>Aspergillus nidulans</i>
23	GDP-manosa pirofosforilasa (Psa2p)	1230	454	5×10^{-126}	56	72	<i>Cryptococcus neoformans</i>
24	NMT-1(THI5)	1041	462	1×10^{-128}	64	79	<i>Uromyces fabae</i>
25	CipC	-	101	3×10^{-19}	50	66	<i>Paxillus involutus</i>
26	STE7	-	81,3	2×10^{-36}	66	84	<i>Cryptococcus neoformans</i>
27	MOB2	759	289	1×10^{-76}	55	73	<i>Cryptococcus neoformans</i>
28	Proteasa aspartica (PR1)	1218	801	0	100	100	<i>Xanthophyllomyces dendrorhous</i>
29	Folilpoliglutamato sintasa	-	238	3×10^{-98}	68	81	<i>Cryptococcus neoformans</i>
30	Proteína ribosomal S3	330	146	6×10^{-34}	69	84	<i>Schizosaccharomyces pombe</i>
31	Proteína ribosomal S4	786	389	1×10^{-106}	71	83	<i>Schizosaccharomyces pombe</i>
32	Proteína ribosomal S5	-	402	2×10^{-109}	81	89	<i>Cryptococcus neoformans</i>
33	Proteína ribosomal S8	-	208	2×10^{-77}	76	88	<i>Schizophyllum commune</i>
34	Proteína ribosomal S17	453	214	2×10^{-54}	76	84	<i>Ustilago maydis</i>
35	Proteína ribosomal S18	378	217	1×10^{-55}	82	92	<i>Cryptococcus neoformans</i>
36	Proteína ribosomal S20	357	177	3×10^{-43}	74	85	<i>Cryptococcus neoformans</i>
37	Proteína ribosomal L3	1167	639	0	77	88	<i>Cryptococcus neoformans</i>
38	Proteína ribosomal L7	-	222	1×10^{-80}	70	82	<i>Cryptococcus neoformans</i>
39	Proteína ribosomal L2	759	403	6×10^{-111}	76	88	<i>Ashbya gossypii</i>
40	Proteína ribosomal L19	594	299	7×10^{-80}	77	86	<i>Cryptococcus neoformans</i>
41	Proteína ribosomal L25	564	176	1×10^{-42}	59	73	<i>Cryptococcus neoformans</i>
42	Proteína ribosomal L31	366	169	8×10^{-41}	67	82	<i>Cryptococcus neoformans</i>
43	Fructosa bifosfato aldolasa (FBA)	1083	572	$\times 10^{-161}$	78	88	<i>Cryptococcus neoformans</i>
44	Enolasa (ENO)	1338	642	0	74	83	<i>Cryptococcus neoformans</i>
45	Transportador de monosacáridos	-	324	2×10^{-87}	60	73	<i>Cryptococcus neoformans</i>
46	3-Fosfoglicerato kinasa (PGK)	1260	617	5×10^{-175}	74	85	<i>Cryptococcus neoformans</i>

47	Transportador de ácidos grasos de cadena larga (ACBP)	318	114	2×10^{-24}	58	70	<i>Cryptococcus neoformans</i>
48	NADH ubiquinona oxidoreductasa (NUO)	2256	1026	0	68	81	<i>Cryptococcus neoformans</i>
49	Proteína ACN9	489	150	2×10^{-35}	53	68	<i>Cryptococcus neoformans</i>
50	3-Fosfoglicerato deshidrogenasa (PDH)	1553	638	0	65	76	<i>Cryptococcus neoformans</i>
51	Piruvato decarboxilasa (PDC)	1818	572	5×10^{-161}	50	65	<i>Cryptococcus neoformans</i>
52	Glicerol-3-fosfato deshidrogenasa	-	342	1×10^{-91}	70	80	<i>Cryptococcus neoformans</i>
53	Gliceraldehído-3-fosfato deshidrogenasa (GPD)	1017	675	0	100	100	<i>Xanthophyllomyces dendrorhous</i>
54	Oxidoreductasa de cadena corta (OXR)	966	258	4×10^{-67}	52	68	<i>Kineococcus radiotolerans</i>
55	Aldo-keto reductasa, Aril-alcohol deshidrogenasa (AKR)	990	329	2×10^{-88}	52	71	<i>Deinococcus geothermalis</i>

Se presenta para cada clon el homólogo más cercano (de acuerdo a Score y E-value), el organismo de origen, además de los valores de identidad (I) y sustituciones aminoacídicas positivas (+). Para los genes secuenciados completamente (verde) además se muestra el tamaño del ORF predicho.

De entre los 55 genes identificados, 35 fueron secuenciados completamente y resultaron estar completos (en base a la comparación con sus genes ortólogos). Los 20 genes restantes carecían de algún segmento en sus extremos, o bien simplemente se decidió no continuar con la determinación de su secuencia debido a que no serían relevantes para nuestro estudio.

La población de genes identificados resultó ser bastante diversa, incluyendo aquellos que codifican para proteínas ribosómicas, β -tubulina, enzimas proteolíticas, enzimas del metabolismo de carbohidratos, de ácidos grasos y de aminoácidos; proteínas relacionadas con señalización intracelular, nivel de ploidía, tipo de apareamiento, transcripción y traducción. En la mayoría de los casos, los ortólogos que presentaban mayor identidad de secuencia con los genes identificados provenían del basidiomicete *Cryptococcus neoformans*, lo cual es consistente con la estrecha relación filogenética

existente entre *X. dendrorhous* y este último microorganismo. Un factor importante a tener en consideración es el alto nivel de identidad que presentan las proteínas deducidas a partir de los distintos ORFs con sus contrapartes de otras especies (Identidad \geq 50%, Identidad + sustituciones positivas \geq 65%), hecho especialmente importante en el caso de los factores de elongación de la traducción y las proteínas ribosómicas. Lo anterior sustenta en gran medida la identificación y asignación inequívoca de funcionalidad a los genes obtenidos a partir de la genoteca de cDNA.

3.3 Uso de codones en *Xanthophyllomyces dendrorhous*.

Sin duda uno de los mayores inconvenientes que han debido superar las distintas investigaciones en *X. dendrorhous* es el escaso número de genes conocidos de este microorganismo, lo que se traduce en diversas dificultades tanto para su manipulación genética y mejora del rendimiento carotenogénico, como para el estudio de sus características biológicas básicas. Al respecto, actualmente existen tan solo 17 genes publicados en la base de datos internacional (que reúne los datos de GenBank, EMBL, DDBJ y PDB), además de 19 secuencias incluidas dentro de distintas patentes. De entre estas últimas existen 11 secuencias anotadas como proteínas ribosómicas, una anotada como escualeno sintasa y una anotada como acetilCoA carboxilasa. Las 8 secuencias restantes no están anotadas y sólo la búsqueda de posibles ORFs y posterior comparación con la base de datos nos ha permitido su identificación (datos no mostrados).

De las 11 proteínas ribosómicas patentadas, 10 fueron utilizadas por Verdoes para construir la única tabla de uso de codones de *X. dendrorhous* existente hasta el momento (Verdoes y cols., 2000). Sin embargo, el bajo número de genes usados para la realización de dicho estudio junto con el pequeño tamaño de los mismos, constituye una seria limitación para la generalización de sus conclusiones dado que el número total de codones analizados resultó ser inconvenientemente bajo. Debido a lo anterior es necesario validar los datos obtenidos por Verdoes incluyendo en el análisis un número mayor de tripletes, de modo de construir una tabla de uso de codones robusta y confiable sobre la cual se puedan hacer diversas inferencias.

Con este objetivo, se empalmaron las secuencias nucleotídicas de 51 ORFs (34 provenientes de la genoteca de cDNA construida en este estudio, 9 publicados en

bases de datos y 8 patentados), generando un concatémero de 42945 pares de bases (14315 codones). Luego, la secuencia resultante fue analizada con la aplicación *Countcodon* disponible en el sitio web de la base de datos *codon usage database* (<http://www.kazusa.or.jp/codon.html>), que alberga todas las tablas de uso de codones publicadas a la fecha. La aplicación *Countcodon* determina la frecuencia total y absoluta de cada uno de los 64 tripletes posibles presentes en una secuencia problema.

A partir de esta información se construyó una nueva tabla de uso de codones de *X. dendrorhous* (Figura 9a), la cual incluye un total de 14315 codones. Una vez obtenida la tabla se calcularon las frecuencias relativas de todos los tripletes que codifican para un mismo aminoácido, de modo de determinar si existe uso preferencial de codones (Figura 9b).

A simple vista se puede apreciar que para cada aminoácido e incluso para la señal de término de la traducción existe un triplete preferido, mientras que el resto de los codones aparece con mucho menor frecuencia. Si no existiese uso preferencial de codones, la frecuencia esperada para cada triplete sería sencillamente $1/N$, donde N representa el número de tripletes distintos que codifican para un mismo aminoácido. Dicho de otra manera, si la elección del triplete que codifica para un determinado aminoácido (de entre N posibles) fuese aleatoria, todos los codones que codifican para este último debiesen aparecer con la misma frecuencia $1/N$. A modo de ejemplo, el aminoácido serina es especificado por seis tripletes distintos (es decir, $N=6$). Si no existiese uso preferencial de codones, la frecuencia relativa esperada para cada uno de los seis tripletes sería $1/6 \approx 0,17$ (17%). Sin embargo al analizar los datos se puede ver que cada uno de los seis tripletes presenta una frecuencia relativa distinta, siendo TCC usado en un 40% de las veces (codón preferido), mientras que AGT presenta una

A

UUU 6.0 (86)	UCU 18.5 (265)	UAU 5.2 (75)	UGU 4.3 (61)
UUC 31.3 (448)	UCC 27.2 (389)	UAC 23.8 (341)	UGC 7.1 (101)
UUA 1.9 (27)	UCA 4.2 (60)	UAA 2.5 (36)	UGA 0.6 (8)
UUG 15.5 (222)	UCG 10.4 (149)	UAG 0.4 (6)	UGG 11.2 (161)
CUU 23.8 (340)	CCU 16.8 (241)	CAU 4.4 (63)	CGU 1.7 (24)
CUC 33.1 (474)	CCC 22.0 (315)	CAC 18.1 (259)	CGC 0.6 (8)
CUA 2.7 (38)	CCA 4.4 (63)	CAA 5.7 (81)	CGA 34.6 (495)
CUG 4.3 (62)	CCG 4.5 (65)	CAG 26.1 (373)	CGG 2.5 (36)
AUU 11.8 (169)	ACU 14.3 (205)	AAU 4.0 (57)	AGU 3.3 (47)
AUC 43.9 (628)	ACC 36.5 (522)	AAC 34.2 (489)	AGC 4.8 (69)
AUA 0.5 (7)	ACA 3.9 (56)	AAA 7.5 (108)	AGA 4.8 (68)
AUG 21.7 (310)	ACG 4.8 (69)	AAG 65.3 (935)	AGG 4.6 (66)
GUU 17.9 (256)	GCU 31.1 (445)	GAU 24.8 (355)	GGU 26.5 (380)
GUC 44.8 (641)	GCC 50.9 (729)	GAC 29.1 (416)	GGC 8.2 (117)
GUA 2.6 (37)	GCA 4.5 (65)	GAA 8.5 (121)	GGA 42.5 (609)
GUG 8.5 (121)	GCG 5.9 (84)	GAG 50.2 (719)	GGG 3.0 (43)

B

ACN9	NUO
ACN9	β-tubulina
AKR	OXR
ATP	PDC
NUO	PDH
AP-1	PGK
EF1A	PRL3
EF1γ	PRL19
EFT2	PRL25
ENO	PRL31
THI5	PRS3
Erv14	PRS4
ACBP	PRS17
FBA	PRS18
Psa2p	PRS20
GPD	MOB2
MIPS	Rpb3
PR1	Sti35
HIS3	Act1
EPH1	Bg1
GRG1	GRG2
PRS22	PRL13A
PRL41	PRS27
PRS16	PRL37A
PRS10	PRS25
PRP2	PRL28
PRA1	

Total ORFs: 51

- ORFs genoteca cDNA
- ORFs publicados
- ORFs patentados

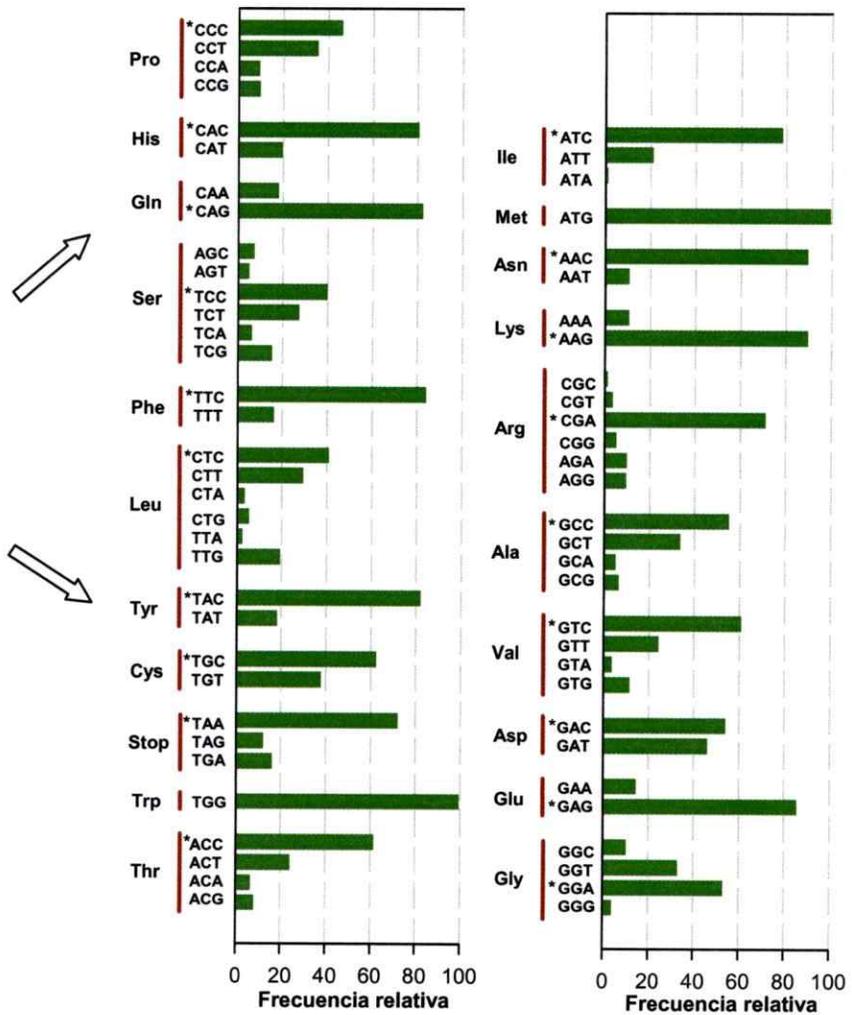


Figura 9. Uso de codones en *X. dendrorhous*. Las secuencias nucleotídicas de 51 ORFs (14315 codones) fueron analizadas para generar una tabla de uso de codones (A), donde se muestra la frecuencia relativa (x1000) y la frecuencia absoluta (entre paréntesis) de cada triplete. A partir de estos datos se construyó un histograma (B) que muestra la frecuencia relativa de todos los posibles codones para cada aminoácido y la señal de término de la traducción. (*) Triplete preferido para cada aminoácido.

frecuencia relativa de sólo un 5% (codón menos usado). Al hacer este análisis para el resto de los aminoácidos se puede notar que en todos los casos existe un triplete preferido, y que las frecuencias relativas se alejan de las esperadas para el caso de uso aleatorio de codones. Lo anteriormente mencionado es especialmente cierto para aminoácidos como histidina, glutamina, fenilalanina, tirosina, isoleucina, asparragina, lisina y glutamato, donde el codón preferido presenta una frecuencia relativa cercana al 80%. En el otro extremo, los dos tripletes que codifican para aspartato son utilizados con una frecuencia prácticamente igual, lo que muestra que el uso preferencial de codones ocurre con distinta intensidad para cada aminoácido.

Luego de construir la tabla, ésta fue comparada con la obtenida anteriormente por Verdoes (Verdoes y cols., 2000). En la Figura 10 se muestran las frecuencias relativas de los distintos tripletes que codifican para cada aminoácido, de acuerdo a los cálculos realizados en este trabajo (a partir de 14.315 codones) y los obtenidos previamente por Verdoes (1295 codones). Si bien las frecuencias obtenidas con cada set de datos son ligeramente diferentes, las tendencias observadas en ambos casos son similares tanto en la existencia de un marcado uso preferencial de codones, como en cuál de los tripletes es el mas frecuentemente usado para cada aminoácido. De esta manera los resultados deducidos a partir de este trabajo concuerdan con los obtenidos previamente, y dado el significativo aumento del número de codones analizados, permiten una estimación más representativa de las frecuencias de uso de los distintos tripletes.

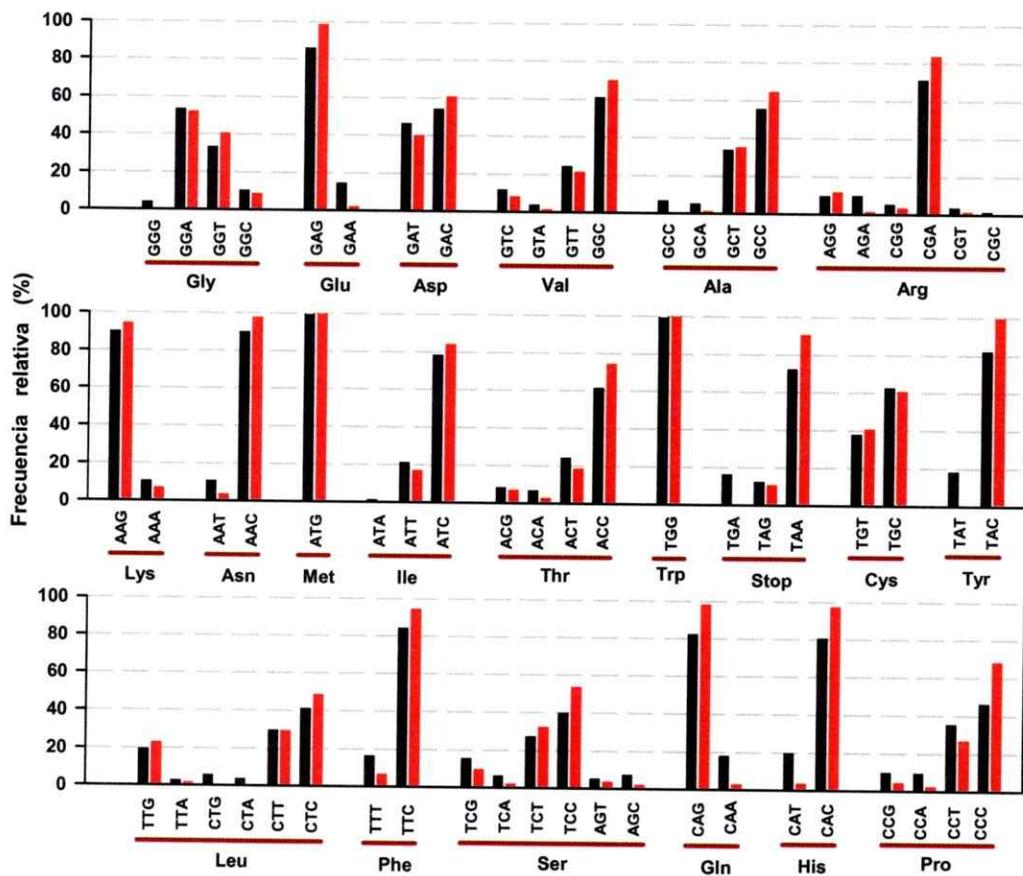


Figura 10. Comparación entre la tabla de uso de codones obtenida en este estudio y la publicada previamente por Verdoes (Verdoes y cols., 2000). La gráfica muestra las frecuencias relativas de los distintos tripletes que codifican para cada aminoácido, calculadas a partir de los datos obtenidos en este trabajo (14.315 codones, barras negras), y los datos publicados previamente por Verdoes (1.295 codones, barras rojas).

Las tablas de uso de codones entregan información mas bien cualitativa respecto a cuáles son los tripletes más frecuentemente usados. Sin embargo para su aplicación, se hace necesario poder cuantificar el grado de sesgo que presenta un determinado gen y cuán bien se adapta al uso preferencial de codones de un organismo determinado. Para ello se han desarrollado diversos algoritmos matemáticos, entre los cuales destaca el índice de adaptación de codones CAI. En esencia el CAI basa su análisis en el patrón de uso de codones determinado a partir de un set de genes altamente expresados. Este patrón revelaría cuál de los codones sinónimos para un aminoácido dado es más eficientemente traducido, además de cuan desventajosos

resultan ser el resto de los codones (Sharp y Li, 1987). De esta manera, el análisis consiste en la comparación de una determinada secuencia codificante (o un grupo de secuencias) con una tabla de uso de codones construida usando una serie de genes de referencia altamente expresados, los cuales normalmente incluyen a aquellos que codifican proteínas ribosómicas, enzimas del metabolismo de carbono, factores de elongación de la traducción, etc. El resultado obtenido al utilizar este algoritmo es un número real entre 0 y 1 el cual refleja, en primer lugar, cuan bien se ajusta la secuencia analizada al uso de codones observado en el set de referencia, y en segundo lugar, constituye una predicción del nivel de expresión que tendría dicha secuencia (gen) en el organismo a partir del cual se construyó la tabla de uso de codones. Valores de CAI cercanos a 1 indican que la secuencia analizada se ajusta al set de referencia, y por lo tanto tendría un alto nivel de expresión en ese organismo. Valores cercanos a 0 indican un bajo nivel de adaptación al sistema traduccional, y por ello, se espera que dichos genes sean expresados en bajos niveles.

Los genes aislados a partir de la genoteca más los que se encontraron publicados a la fecha, fueron utilizados para realizar un estudio de CAI. Lamentablemente en *X. dendrorhous* no se cuenta con datos de niveles de expresión, lo que constituye una complicación para elegir los genes que componen el set de referencia necesario para el análisis. Sin embargo, como se mencionó previamente, aquellos genes que codifican para proteínas ribosómicas, factores de elongación y enzimas de la vía glicolítica, entre otros; son altamente expresados en la mayoría de los organismos estudiados, lo que sustenta su utilización como set de referencia en organismos donde sus niveles de expresión son desconocidos (Sharp y cols., 1986; Sharp y Li, 1987; Sharp y Cowe, 1991; Liu y cols., 2004). De esta manera, se seleccionó 28 genes de *X. dendrorhous* para construir el set de referencia (Tabla 5), el cual fue luego usado para

calcular los valores de CAI de cada uno de los genes de la levadura conocidos hasta ahora (Figura 11). Las ecuaciones usadas para los cálculos corresponden a las publicadas por Sharp y Li (Sharp y Li, 1987; ver apartado 2.11 de Materiales y Métodos).

Los valores de CAI obtenidos resultaron ser bastante diversos, y van desde 0,2 hasta 0,88. Al analizar la distribución de dichas cifras (datos no mostrados) se observa una concentración en torno a dos valores promedio, los cuales permiten distinguir al menos dos poblaciones de genes. La primera corresponde al grupo de genes que presentan un bajo CAI, el cual está en torno a 0,3, mientras que la segunda población contiene aquellos genes que poseen un CAI alto, con un valor cercano a 0,8. De todos modos los valores intermedios se presentan con una frecuencia no despreciable, lo que ilustra la diversidad de los índices calculados. De acuerdo a lo esperado, los genes que codifican para proteínas ribosómicas, factores de elongación y algunas enzimas del metabolismo de carbono (principalmente glicólisis), presentan valores de CAI bastante altos en comparación al resto de los genes analizados. La única excepción a lo señalado anteriormente es la proteína ribosómica L25, la cual presenta un índice relativamente bajo en relación al resto de las proteínas ribosómicas. Otros genes que presentan un alto CAI son ACT1 (actina), ATP1 (antiportador ATP:ADP) y MNSOD (Superóxido dismutasa). Notablemente, los genes que participan en la biosíntesis de carotenoides presentan un CAI bastante bajo, con valores entre 0,2 (crfYB) y 0,37 (idi).

Tabla 5. Genes de *X. dendrorhous* utilizados para construir el set de referencia necesario para el cálculo del índice de adaptación de codones (CAI).

ENO	EF1A	PRL25	PRS18	PRS22	PRP2
FBA	EF1g	PRL31	PRS20	PRS16	PRL28
GPD	EFT2	PRS3	PRL13A	PRL37A	PRS27
PDC	PRL3	PRS4	PRS2	PRS10	
PGK	PRL19	PRS17	PRL41	PRS25	

Metabolismo del carbono ■ Factores de elongación de la traducción ■ Proteínas Ribosómicas ■
 Biosíntesis de carotenoides ■ Otras funciones ■ Patentes ■

GEN	CAI
ENO	0,88
FBA	0,83
GPDH	0,87
PDC	0,72
PGK	0,77
AKR	0,34
PDH	0,41
ACBP	0,51
ACN9	0,21
NUO	0,35
OXR	0,51
Psa2p	0,38

EF1g	0,72
EF1A	0,87
EFT2	0,81

PRA1	0,81
PRL13A	0,78
PRL19	0,74
PRL25	0,51
PRL28	0,66
PRL3	0,77
PRL31	0,72
PRL37A	0,85
PRL41	0,88
PRP2	0,76
PRS10	0,79
PRS16	0,82
PRS17	0,71
PRS18	0,83
PRS2	0,81

GEN	CAI
PRS20	0,81
PRS22	0,72
PRS25	0,74
PRS27	0,73
PRS4	0,79

Idi	0,37
CrtE	0,21
CrtYB	0,20
CrtI	0,31
AST	0,25

AP-1	0,23
ACT1	0,67
ATP	0,78
Bgl	0,30
β-tubulina	0,51
EPH1	0,23
Erv14	0,50
HIS3	0,37
MIPS	0,44
MOB2	0,24
PR1	0,38
Rpb3	0,22
STI35	0,37
THI5	0,42

ACC	0,29
MK	0,22
MNSOD	0,70
SQS	0,29

Figura 11. Índice de adaptación de codones (CAI) calculado para diversos genes de *X. dendrorhous*. El índice de adaptación de codones fue calculado de acuerdo con la formulación original de Sharp (Sharp P. y Li, 1987). Para el cálculo se consideró el uso de codones de un set de genes de referencia supuestos como altamente expresados.

3.4 Determinación de la estructura exón-intrón de genes expresados en metabolismo fermentativo de *X. dendrorhous*.

Una vez completada la secuenciación de los distintos cDNAs obtenidos a partir de la genoteca, se procedió a clonar y secuenciar las versiones genómicas de los genes relevantes para investigaciones en curso de nuestro laboratorio. Gracias a esto fue posible determinar la estructura exón-intrón de dichos genes, además de contar con el sustrato necesario para la obtención a futuro de cepas mutantes para los mismos (generados por recombinación homóloga entre el gen situado en el cromosoma de la levadura y una versión mutada introducida por transformación).

La estrategia general para el clonado de la versión genómica de cada gen consistió en el diseño de partidores en base a la secuencia del cDNA. Estos partidores fueron usados luego para realizar una reacción de PCR usando como molde DNA genómico de *X. dendrorhous* (cepa UCD-67 385). Las mezclas de PCR fueron luego separadas en un gel TAE agarosa al 0,8 %, y las bandas obtenidas en cada caso fueron purificadas a partir del gel, ligadas al vector plasmidial pBluescript y secuenciadas por ambas hebras. Luego de completadas las secuencias se realizó un alineamiento entre el cDNA y el DNA genómico, de modo de identificar las regiones intrónicas y exónicas. Al analizar los alineamientos se puede constatar que el número de intrones y la extensión de los mismos varía ampliamente en cada caso, al igual que el tamaño de cada gen (Figura 12). Así, el gen EF1A posee solo dos intrones, los cuales tienen un tamaño de alrededor de 120 pares de bases; mientras que el gen AKR, de un tamaño similar a EF1A, posee 9 intrones los cuales tienen una extensión cercana a las 70 pares de bases (exceptuando el primer intrón cuyo largo es considerablemente mayor).

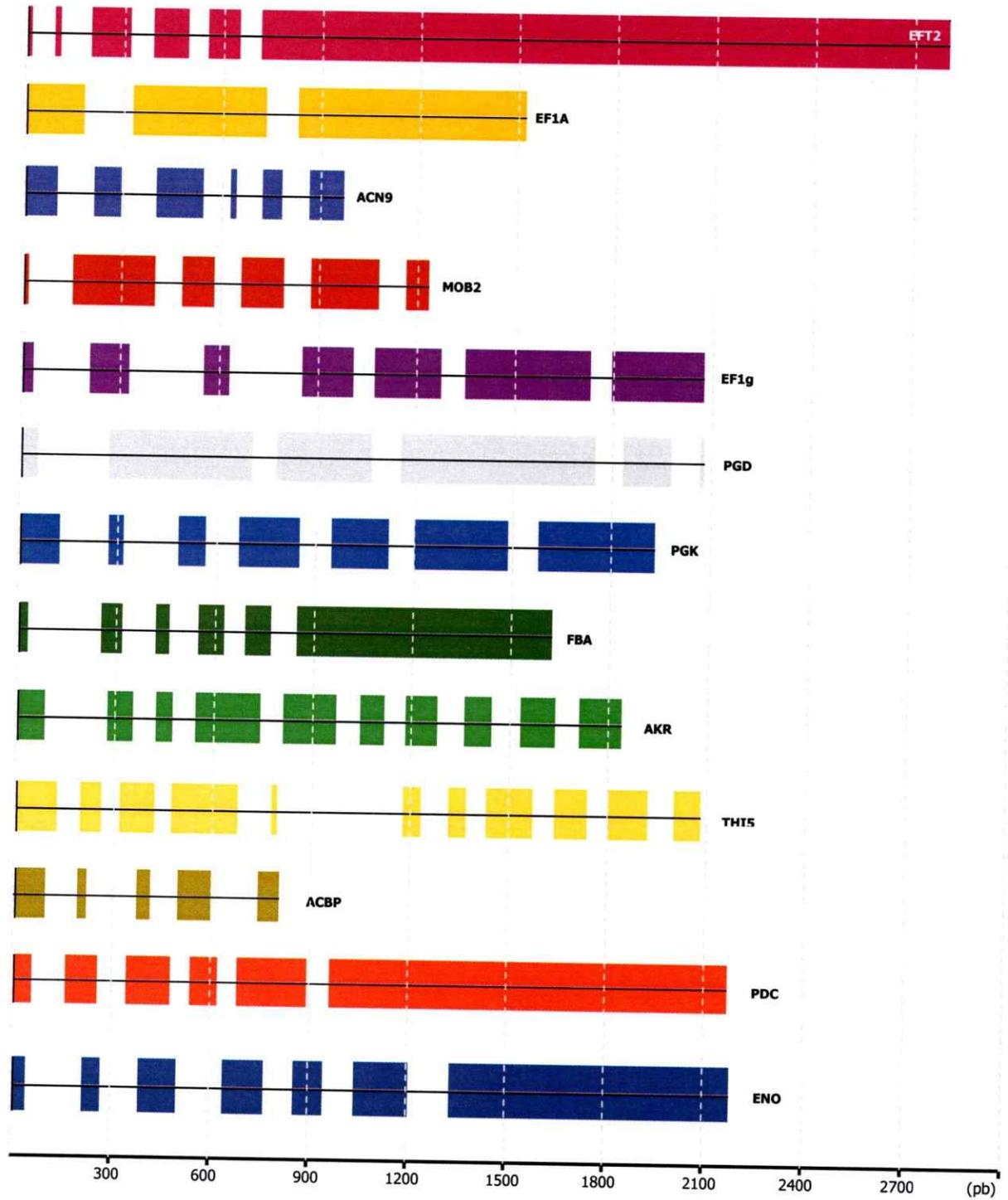


Figura 12. Estructura exón-intrón de genes expresados en metabolismo fermentativo de *X. dendrorhous*. Para cada gen se presenta la extensión del DNA genómico desde el codón de inicio hasta el codón de término. Los exones se muestran como zonas coloreadas, mientras que las regiones intrónicas corresponden a los espacios en blanco.

Aparentemente el tamaño de cada gen no se correlaciona con el número de intrones presentes. Lo anterior queda ilustrado con el gen ACBP el cual, al igual que el gen PGD, contiene 4 intrones; sin embargo el tamaño del segundo gen es casi tres veces superior. A pesar de la gran variedad de tamaños y cantidad de intrones observada en los genes analizados, es posible establecer que en promedio el número de intrones presentes en cada gen es cercano a 6 (obtenido al dividir el número total de intrones por el número de genes cuya estructura exón-intron se conoce). La ubicación de los intrones al interior de cada gen es también bastante variable, pudiendo estar concentrados en la mitad 5' como en el caso de EFT2, FBA y PDC; o bien distribuidos homogéneamente como en ACBP, AKR y THI5. Al igual que las regiones intrónicas, los exones presentan gran diversidad tanto en tamaño, número y ubicación al interior del gen.

Luego de determinadas las regiones intrónicas y exónicas se procedió a analizar los sitios dadores y aceptores de *splicing*, los cuales se ha visto que poseen elementos de secuencia conservados entre los distintos organismos (Jurica y Moore, 2003; Patel y Steitz, 2003). Para ello se realizó un alineamiento entre los sitios dadores y aceptores de *splicing* de todos los genes de *X. dendrorhous* cuya estructura exón-intrón ha sido determinada (incluyendo los genes identificados en este estudio y los publicados hasta la fecha en la base de datos Genbank). Una vez realizado el alineamiento se calcularon las frecuencias relativas de ocurrencia de una determinada base en cada posición del sitio dador y el sitio aceptor. Las posiciones de las bases que conforman los sitios dadores y aceptores fueron definidas considerando como centro del sitio las bases universalmente conservadas GT y AG, respectivamente. El número total de sitios dadores y aceptores considerados para el análisis fue de 144, de los cuales el 100% poseían las bases centrales GT y AG (Figura 13).

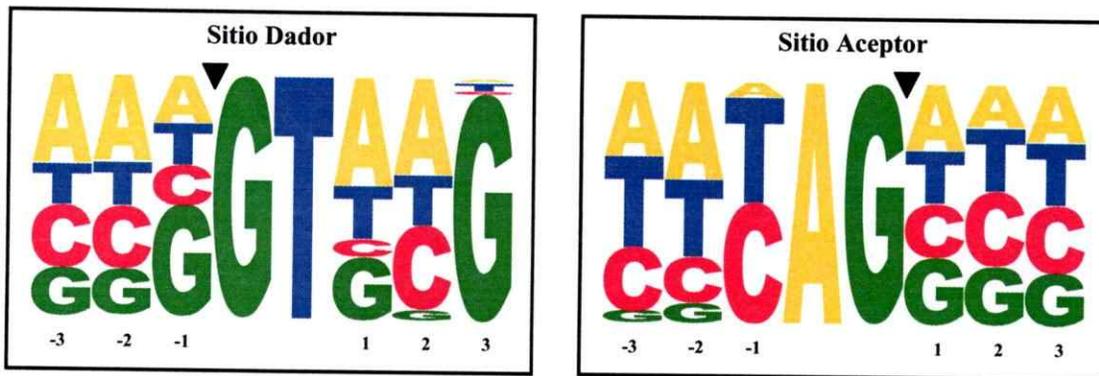


Figura 13. Secuencia consenso de sitios dadores y aceptores de *splicing* de *X. dendrorhous*. El tamaño de cada letra es proporcional a la frecuencia con que cada base aparece en una determinada posición. El sitio específico de inicio y término de la región intrónica en el sitio dador y aceptor, respectivamente, se encuentra señalado con una flecha negra.

La secuencia consenso del sitio dador posee posiciones altamente conservadas, entre las cuales destaca la posición 3. Cerca del 95% de los sitios dadores poseen una guanina en la posición 3, mientras que el 5% restante tiene adenina, timina o citosina en casi igual proporción. De manera similar aunque con menor conservación, en la posición -1 aparece guanina en el 50% de los casos. Otra característica del sitio dador es la baja frecuencia con que aparece guanina en la posición 2 y citosina en la posición 1, representando sólo el 7% y el 4% de los casos, respectivamente. De esta manera la secuencia consenso del sitio dador podría resumirse como NNGGTDHG (utilizando la nomenclatura internacional para las posiciones degeneradas). La secuencia consenso del sitio aceptor, por otra parte, se caracteriza por la muy baja frecuencia de aparición de guanina en la posición -3 y -2, llegando a ser cero en la posición -1. En esta última posición destaca además la baja frecuencia con que aparece adenina. El resto de las posiciones del sitio aceptor son poco conservadas, siendo las frecuencias para cada base cercanas al 25%. De esta manera la secuencia consenso del sitio aceptor podría resumirse como HHYAGNNN.

4. DISCUSIÓN

4.1 Construcción de las genotecas de cDNA de *X. dendrorhous*

El escaso número de genes de *X. dendrorhous* clonados, identificados y secuenciados hasta ahora constituye una seria limitación para la realización de estudios genético-moleculares en dicha levadura. Entre los principales inconvenientes que ocasiona esta falta de información se encuentran la casi nula disponibilidad de marcadores de selección, secuencias promotoras y secuencias terminadoras; el total desconocimiento de la existencia de vectores episómicos; y la imposibilidad de realizar estudios de uso preferencial de codones, que faciliten el clonamiento por genética reversa y optimización de la expresión de diversos genes.

La construcción de genotecas de cDNA de *X. dendrorhous* constituye una herramienta poderosa para la identificación y caracterización de genes anteriormente desconocidos, además de representar una visión general de los genes que se están expresando en las condiciones de cultivo utilizadas para la construcción de dicha librería. Dado que la fuente de carbono disponible en el medio de crecimiento influye directamente en la producción de pigmentos y en la expresión y procesamiento de genes relacionados con la síntesis de los mismos (Flores-Cotera, 2001; Gu, 1997; Vázquez, 2000); se construyó dos versiones de la genoteca, variando la fuente de carbono en que fueron crecidas las células. De esta manera, se utilizó glucosa como fuente de carbono fermentable y succinato como sustrato no fermentable. Si bien los procedimientos empleados para la construcción de ambas genotecas fueron los mismos, sólo fue posible obtener la genoteca en glucosa. Como se discutirá más abajo, entre los múltiples procedimientos realizados probablemente los más críticos para el éxito de la

construcción son la obtención de un mRNA de buena calidad y el ligado y transformación del cDNA sintetizado.

El primer paso para la construcción de las genotecas fue la obtención de RNA total de buena calidad y la posterior purificación de mRNA poliA+. La calidad del RNA total extraído fue evaluada considerando su pureza e integridad, es decir, se determinó la presencia de contaminantes como proteínas y oligosacáridos y la existencia de degradación. La integridad del RNA total fue evaluada mediante electroforesis capilar en un equipo Agilent BioAnalyzer, el cual permite una determinación rápida y precisa de muestras de ácidos nucleicos y proteínas (Mueller y cols., 2000; Mueller y cols, 2006). Luego de los procedimientos de extracción se obtuvo un RNA íntegro y prácticamente con ausencia de contaminación, aunque en el caso de MMV-succinato la presencia de una degradación leve podría haber sido uno de los factores que imposibilitaron la construcción de dicha genoteca.

Luego de la purificación de mRNA poliA+ a partir del RNA total y la posterior síntesis de cDNA, se realizó el fraccionamiento de este último, de modo de enriquecer la mezcla en moléculas de tamaño superior a 500 pares de bases. Esto permitió, por una parte, eliminar los oligonucleótidos adaptadores que no se ligaron al cDNA y los extremos generados luego de la digestión con *Xho*I; además de disminuir la proporción de mensajeros parcialmente degradados (incompletos) y favorecer los de mayor tamaño. Sin duda uno de los pasos críticos en la construcción de la genoteca fue el ligado del cDNA al vector plasmidial y la posterior transformación de *E. coli*. Inexplicablemente fueron necesarios 6 intentos de ligado y transformación para conseguir un número suficiente de transformantes. Antes de lograrlo se probaron distintas concentraciones de DNA, diferentes DNA-ligasas, se varió el tiempo y condiciones de incubación en la reacción de ligado, e incluso se utilizó distintos protocolos de transformación y células

competentes. Aparentemente el único factor que logró un aumento en la eficiencia de transformación fue la precipitación previa del cDNA, lo que podría haber eliminado algún compuesto que de alguna manera actuara inhibiendo la reacción de ligado o la transformación misma. Sin embargo, la transformación de una reacción de ligado control (pBluescript SKII + DNA control), usando las mismas condiciones y reactivos empleados en el intento definitivo pero sin precipitar, genera un alto número de transformantes. Por otra parte, en el caso de la genoteca en succinato se probaron las mismas condiciones usadas en los distintos intentos realizados para la genoteca en glucosa, no obteniendo resultados satisfactorios en ninguno de ellos. Lo anterior sugiere que podrían haber otros factores que estarían determinando la eficiencia de transformación.

Luego de obtenido el número suficiente de clones se determinó el tamaño medio de inserto de la genoteca. De los resultados obtenidos se observa una gran diversidad de tamaños en los clones analizados, y probablemente en el total de la genoteca. El tamaño medio de inserto se condice con los resultados del fraccionamiento del cDNA, y está de acuerdo con las fracciones que se eligieron para ligar. Finalmente la posterior secuenciación de distintos clones permitió confirmar la diversidad de mensajeros clonados, los cuales en su mayoría se encontraban completos incluso aquellos de tamaño superior a 2 kb.

El procesamiento y partición de la genoteca en grupos de menor número de clones, incluidos en grupos de mayor tamaño, hace de la genoteca una herramienta poderosa y rápida para la búsqueda dirigida de genes a través de reacciones de PCR. De esta manera se han podido aislar hasta la fecha diversos genes de *X. dendrorhous* relevantes para los estudios realizados en nuestro laboratorio.

4.2 Aislamiento e Identificación de genes expresados en metabolismo fermentativo de *X. dendrorhous*.

La genoteca de cDNA construida a partir de células crecidas en glucosa como única fuente de carbono, representa un instante en el crecimiento de *X. dendrorhous* donde la producción de biomasa decae, redistribuyendo el flujo de carbono hacia la producción de metabolitos secundarios y más específicamente hacia la producción de pigmentos carotenoides (Lodato y cols., 2003). Dicha fase, denominada de inducción de la carotenogénesis, coincide con el máximo de producción de etanol y el agotamiento de la glucosa disponible en el medio, además de presentar un aumento sustancial en los niveles de expresión de gran parte de los genes que codifican para las enzimas de biosíntesis de carotenoides (Lodato y cols., 2003; Wozniak y cols., 2005). No obstante, los patrones temporales de expresión génica y de producción de pigmentos difieren considerablemente si se utiliza una fuente de carbono no fermentable como el succinato (Vazquez y cols., 1997; Wozniak y cols., 2005). Lo anterior sugiere fuertemente que es en éste instante del crecimiento donde probablemente mecanismos regulatorios dependientes de factores como disponibilidad y tipo de nutrientes y/o metabolitos estarían actuando para controlar la producción de carotenoides. Bajo esta perspectiva, y teniendo en mente el reducido número de genes conocidos de este microorganismo, resulta de suma importancia identificar al menos una parte de los genes que se están expresando bajo estas condiciones. De este modo la identificación de nuevos genes expresados en metabolismo fermentativo tiene 2 implicancias importantes; la primera es aumentar significativamente el número de genes conocidos de *X. dendrorhous*. Esto a su vez permite la realización de estudios respecto a la estructura y organización genómica de dichos genes, la identificación de genes de *housekeeping* y potenciales marcadores de selección hasta ahora

desconocidos, y la realización de un estudio de uso preferencial de codones, entre otros. La segunda implicancia relevante tiene que ver con el momento del crecimiento de *X. dendrorhous* donde la genoteca fue construida, pues es en esta fase donde probablemente se estén expresando genes que tengan que ver con la regulación de la vía de carotenogénesis, o bien con el control de la disponibilidad de sustratos tempranos de la vía, tales como piruvato o acetil coenzima A. Por este motivo, la genoteca de cDNA ofrece una oportunidad bastante atractiva para el aislamiento de genes reguladores de la carotenogénesis.

La identificación de nuevos genes de *X. dendrorhous* comenzó con la selección de clones al azar a partir de la genoteca y la posterior secuenciación del extremo 5' del inserto que portaba cada uno de estos, dando lugar a un conjunto de ESTs. De todos los ESTs analizados, aproximadamente el 50% no presentó identidad significativa ($E\text{-value} < 1 \times 10^{-30}$) con ninguna secuencia de función conocida presente en la base de datos, razón por la cual en estos casos no fue posible la identificación.

Entre los genes identificados se encuentran aquellos que codifican para proteínas ribosómicas, β -tubulina, enzimas proteolíticas, enzimas del metabolismo de carbohidratos, de ácidos grasos y de aminoácidos; proteínas relacionadas con señalización intracelular, nivel de ploidía, tipo de apareamiento, transcripción y traducción. Sin duda un factor clave que permitió la identificación de una gran cantidad de genes es la disponibilidad del genoma anotado del basidiomicete *Cryptococcus neoformans* (Loftus y cols., 2005), ya que posee una estrecha relación filogenética con *X. dendrorhous*. Las proteínas deducidas a partir de los distintos ORFs presentaron un alto nivel de identidad de secuencia con sus contrapartes de otras especies (Identidad $\geq 50\%$, Identidad + sustituciones positivas $\geq 65\%$), hecho particularmente claro en el caso de los factores de elongación de la traducción y las proteínas ribosómicas.

Tal como se mencionó en apartados anteriores, la identificación de nuevos genes tiene un valor *per sé*, pero además permite obtener nuevas herramientas que ayuden al desarrollo de las investigaciones del organismo en cuestión. Bajo esta perspectiva y a modo de ejemplo, la secuencia promotora del factor de elongación 1 alfa aislado a partir de la genoteca ha sido usada en nuestro laboratorio como parte integral de un cassette de resistencia a higromicina, el cual permite delecionar genes específicos y seleccionar los transformantes en medio de crecimiento suplementado con dicho antibiótico (Niklitschek y cols., 2006). El gen NMT-1, por otra parte codifica para una enzima esencial para la biosíntesis de tiamina, el cual al ser eliminado genera mutantes auxótrofos para este compuesto en distintos hongos (Schweiggruber y cols., 1991; McColl y cols., 2003; Wightman y Meacock, 2003). De esta manera, se podrían construir cepas mutantes para el gen NMT-1 (auxótrofas para tiamina), y al transformar estas cepas con algún gen de interés, NMT-1 podría ser usado como marcador de selección.

Entre los genes aislados e identificados a partir de la genoteca, otro que posee gran relevancia para futuros estudios es el homólogo a ACN9. ACN9 es un gen identificado originalmente en *S. cerevisiae*, que codifica para una proteína mitocondrial requerida para el crecimiento en acetato y otras fuentes de carbono no fermentables (McCammon, 1996; Dennis y McCammon, 1999). Se ha descrito que mutaciones en este gen causan elevados niveles de expresión de enzimas del ciclo del glioxilato, gluconeogénesis y del metabolismo de acetil-CoA (McCammon, 1996). Debido a que la acetilCoA es uno de los precursores de la biosíntesis de IPP, y éste último es el sustrato fundamental de la vía de carotenogénesis, resulta de gran relevancia generar mutantes de ACN9 en *X. dendrorhous* y estudiar su fenotipo, determinando el efecto que produce dicha mutación sobre la producción de pigmentos.

Finalmente, la construcción de la genoteca permitió aislar una importante cantidad de genes cuyos productos proteicos tienen que ver con el metabolismo del carbono, especialmente con la glicólisis y procesos fermentativos. Entre estos se encuentran los que codifican para las enzimas enolasa (ENO), fructosa bifosfato aldolasa (FBA), gliceraldehído-3-fosfato deshidrogenasa (GPD), piruvato decarboxilasa (PDC), fosfoglicerato kinasa (PGK) y fosfoglicerato deshidrogenasa (PDH). Los genes anteriormente mencionados poseen especial relevancia para nuestros estudios dado que tienen que ver directamente con la utilización de distintas fuentes de carbono y la generación de acetilCoA, la cual como ya se discutió, es fundamental para la biosíntesis de carotenoides.

4.3 Uso de codones en *X. dendrorhous*.

El uso no uniforme de codones sinónimos ha sido descrito en la mayoría de los organismos uni y multicelulares estudiados (Grantham y cols., 1981). Sin embargo, cuan marcado es el sesgo en el uso de codones varía ampliamente entre genes de distintas especies. En organismos unicelulares tales como *E. coli* y *S. cerevisiae*, el grado de sesgo ha sido correlacionado tanto con el contenido de tRNAs isoaceptores como con los niveles de expresión génica, como resultado de un proceso de selección para aumentar la eficiencia traduccional (Sharp y Li, 1987; Comeron y Aguadé, 1998). En este sentido, los genes altamente expresados se encuentran adaptados al aparato traduccional y por tanto tienden a usar principalmente aquellos codones para los cuales el tRNA respectivo es más abundante. Visto de otra manera, aquellos genes que utilicen en mayor proporción los codones preferenciales, se espera que presenten superiores niveles de expresión. Por este motivo, y por otros que se discutirán más

adelante, resulta necesario conocer cuáles son los codones más frecuentemente utilizados en un organismo en particular.

En el único estudio de uso de codones realizado hasta ahora en *X. dendrorhous*, Verdoes y cols. construyeron una tabla de frecuencias de uso de tripletes, determinando cuáles eran los más usados (Verdoes y cols., 2000). Sin embargo el principal inconveniente de este estudio es el bajo número de codones analizados, lo que le resta sustento a sus conclusiones. El alto número de nuevos genes identificados en este estudio junto con los que están publicados en la base de datos Genbank, fueron utilizados para construir una nueva tabla de uso de codones, la cual consideró diez veces más tripletes que los analizados originalmente por Verdoes. Interesantemente, salvo por pequeñas diferencias numéricas, los resultados obtenidos en ambos casos son similares. Lo anterior podría explicarse debido a que en el primer estudio se utilizaron solo proteínas ribosómicas, las cuales se sabe que son altamente expresadas en la mayoría de los organismos (Sharp y Li, 1987), y por lo tanto, presentan un alto grado de sesgo en el uso de codones. De esta manera, el estudio aquí presentado permite ratificar los resultados obtenidos por Verdoes, ampliando significativamente el universo de codones analizados. Los aminoácidos que presentaron un mayor uso preferencial de codones son Glu, Arg, Lys, Asn, Ile, Tyr, Phe, Gln e His.

Tal como se pudo ver, los distintos codones sinónimos que codifican para cada aminoácido no son usados de manera aleatoria. Si bien en el sesgo en el uso de tripletes sinónimos es una característica bastante común en organismos de distintas especies, hay grandes diferencias en cuanto a cuales aminoácidos presentan un mayor grado de sesgo y cuál es el triplete preferido. Dichas diferencias han sido utilizadas para predecir y comparar la expresión esperada de un gen determinado en distintos

organismos e incluso hacer inferencias filogenéticas. Para estos y otros fines es necesario cuantificar el grado de sesgo en el uso de codones de manera tal que se puedan efectuar comparaciones tanto intra como inter especies. Muchos investigadores de esta área han formulado distintos métodos y algoritmos que permitan realizar análisis cuantitativos de uso de codones, los cuales no solo difieren en su formulación matemática, sino además en su precisión y variación artefactual producto de factores como el tamaño de las secuencias analizadas y el grado de uso preferencial de codones. Dado lo anterior, trabajos previos (Comeron y Aguadé, 1998; Novembre, 2002; Supek y Vlahovicek, 2005) han focalizado su atención en evaluar y comparar las distintas medidas frecuentemente utilizadas, entre las cuales se encuentran el número efectivo de codones (N_c) (Wright, 1990), el índice de sesgo de codones (CBI) (Morton, 1993), el índice intrínseco de sesgo de codones (ICDI) (Freire-Picos y cols., 1994) y el índice de adaptación de codones (CAI) (Sharp y Li, 1987), entre otros. De las mencionadas anteriormente, una de las medidas que produce menos artefactos y tiene un comportamiento más regular frente a variables como el largo de la secuencia analizada es el CAI (Sharp y Li, 1987; Comeron y Aguadé, 1998). Si bien este es uno de los índices más usados, posee el inconveniente de que basa sus cálculos en la construcción de un set de genes de referencia altamente expresados. Los datos de expresión génica muchas veces son difíciles de obtener y en general se encuentran disponibles para un limitado número de organismos. Por este motivo, normalmente se utiliza un set de genes que se espera que sean altamente expresados (por extrapolación de datos de otros organismos). Esta fue la estrategia utilizada en este trabajo.

El CAI posee múltiples aplicaciones. En primer lugar, puede ser usado para predecir los niveles de expresión de un gen determinado en su organismo de origen, dado que

los genes altamente expresados debiesen estar mucho más adaptados a su sistema traduccional y por lo tanto presentar un alto grado de uso preferencial de codones (Sharp y Li, 1987). En segundo lugar, el CAI permite predecir niveles de expresión heteróloga de un gen. Existen evidencias de que ciertos codones pueden afectar los niveles de expresión. Por ejemplo, el codón AGG incide marcadamente sobre la tasa de traducción de genes en *E. coli* (Bonekamp y cols., 1985). Esto sugiere que para que un gen heterólogo tenga un nivel de expresión máximo su uso de codones debe corresponder al del huésped.

El cálculo del CAI para los distintos genes de *X. dendrorhous* muestra distintas subpoblaciones de genes, donde los valores más altos son alcanzados por proteínas ribosomales y factores de elongación de la traducción, entre otros. Según lo anterior, y de acuerdo a lo esperado, los genes altamente expresados poseen valores de CAI cercanos a 1. Interesantemente, los genes que codifican para las enzimas de biosíntesis de carotenoides poseen valores de CAI bastante bajos, lo que indicaría que son poco expresados en comparación al resto de los genes. El bajo valor de CAI se podría explicar en parte por el hecho de que se trata de genes que codifican para enzimas del metabolismo secundario. Una situación similar se ha observado en *S. cerevisiae*, donde en general genes del metabolismo secundario y especialmente algunas proteínas regulatorias como Gal4 y PPR1 presentan valores de CAI cercanos a 0,2 (Sharp y Li, 1987). Si bien el valor de CAI obtenido para los genes de carotenogénesis es en general bajo, hay diferencias entre ellos, abarcando un rango desde 0,2 para crtYB a 0,37 para el gen idi. Dado que este método permite una predicción cualitativa de los niveles de expresión de una proteína determinada, diferencias pequeñas en el valor del CAI pueden no ser significativas (Sharp y Li, 1987). Sin embargo, si se considera sólo la tendencia entregada por los datos, es

esperable que el gen *idi* tenga los mayores niveles de expresión entre todos los genes que participan en la carotenogénesis, dado que participa en muchos otros procesos además del mencionado anteriormente.

Al interpretar los resultados expuestos en el párrafo anterior, se debe tener en cuenta que el CAI nos entrega información respecto sólo a la eficiencia con la que se traduciría un determinado gen, pero no dice nada respecto a la actividad de la proteína sintetizada. Así, un gen poco expresado (con un CAI bajo) podría dar lugar a una proteína con alta actividad y una baja tasa de recambio, factores que finalmente compensarían sus bajos niveles de expresión, permitiendo suplir las necesidades de la célula por esa actividad. Por lo tanto, el hecho de que un gen tenga un CAI bajo no necesariamente implica que su producto proteico sea limitante.

A pesar de la problemática planteada anteriormente, si la secuencia nucleotídica de un gen que posee un bajo CAI es modificada de modo que se adapte de mejor manera al uso preferencial de codones de los genes altamente expresados; se consigue un aumento significativo en la expresión de dicho gen, tal como se ha visto para la expresión de proteínas virales en células de mamíferos (Demi y cols., 2001; Ko y cols, 2005), y para la expresión de IG kappa humana en *S. cerevisiae* (Kotula y Curtis, 1991). Teniendo en cuenta lo anterior, la modificación del uso de codones en los genes de carotenogénesis surge como una nueva y atractiva alternativa de optimización de la producción de pigmentos en *X. dendrorhous*.

Los bajos índices de adaptación de codones de los genes de carotenogénesis en *X. dendrorhous* podrían mejorar si estos genes son expresados en otro sistema. Para estudiar esta posibilidad se utilizó un set de genes referencia de *S. cerevisiae* descrito originalmente por Sharp (Sharp y Li, 1987) para calcular los valores de CAI de los genes de carotenogénesis en *S. cerevisiae*. Desafortunadamente los valores obtenidos

son aún más bajos que en *X. dendrorhous*, lo que sugiere que *S. cerevisiae* no es un huésped óptimo para la expresión de la vía de carotenogénesis, al menos desde el punto de vista del uso de codones.

4.4 Estructura exón-intrón de genes expresados en metabolismo fermentativo de *X. dendrorhous*.

El clonado y secuenciación de la versión genómica de algunos de los genes identificados en este trabajo permitió la determinación de la estructura exón-intrón de los mismos. El número de intrones presentes en los genes de *X. dendrorhous* analizados resultó ser variable, aunque en promedio es cercano a 6 por gen. En algunos genes tales como crtS y AKR el número de intrones llega a ser superior a 10, por lo que se podría considerar a *X. dendrorhous* como una levadura rica en intrones. La abundancia de intrones es una característica compartida por la mayoría de los hongos basidiomicetes, a diferencia de los hemiascomicetes los cuales presentan la menor cantidad de intrones entre todos los hongos (Loftus y cols., 2005; Stajich y Dietrich, 2005; Stajich y Dietrich, 2006). No sólo la densidad de intrones, sino además el tamaño de los mismos varían dramáticamente entre los distintos linajes. Así, se encuentran 0,05 intrones por gen con un largo promedio de 367 pares de bases en *S. cerevisiae*; 1,7 intrones por gen con un tamaño promedio de 135 pb en *N. crassa* y 5,3 intrones por gen con un largo cercano a 67 pares de bases en *C. neoformans* (Stajich y Dietrich, 2005).

Algunos autores han especulado que el contenido de intrones estaría correlacionado con el tamaño del genoma. Por ejemplo, se ha propuesto que los genes procarióticos carentes de intrones serían el resultado de un proceso de selección contra la presencia de intrones, de modo de aumentar la eficiencia de los procesos de replicación,

transcripción y procesamiento de mRNAs (Deutsch y Long, 1999). Sin embargo, esta hipótesis es insuficiente para explicar las enormes diferencias en el contenido de intrones existentes entre organismos de distintos linajes de hongos, que poseen un genoma de tamaño similar.

La remoción de intrones ocurre durante el proceso de corte y empalme de las moléculas de pre-mRNA, y este proceso es llevado a cabo por una compleja maquinaria compuesta por partículas ribonucleoproteicas (snRNPs) denominada *spliceosoma* (Maniatis y Reed, 1987). El procesamiento, entre otras cosas, requiere el reconocimiento y unión de la snRNP U1 al sitio dador, además del reconocimiento del sitio aceptor por la snRNP U5. La importancia de estos sitios ha permitido la conservación de su estructura entre los distintos organismos, lo que desde un punto de vista práctico puede ser utilizado para la predicción de la estructura exón-intrón de distintos genes. Aunque los sitios dadores y aceptores poseen elementos conservados, al menos en parte difieren entre los distintos organismos. Por este motivo es relevante determinar la estructura exón-intrón de un organismo determinado, en este caso, *X. dendrorhous*, de modo de contar con herramientas más precisas para la predicción de la estructura de genes cuyas propiedades no han podido ser experimentalmente determinadas. El cálculo de las frecuencias de aparición de las distintas bases en cada posición del sitio dador y/o aceptor a partir de genes de estructura exón-intrón conocidas, permite el desarrollo de algoritmos que identifiquen posibles sitios de procesamiento (Reese y cols., 1997; Thanaraj y Robinson, 2000). Los algoritmos se basan en la búsqueda elementos de secuencia similares a un patrón determinado, asignando mayor puntaje a las zonas donde la secuencia se asemeje más al patrón utilizado. Normalmente no todas las posiciones dentro del patrón poseen igual importancia, por lo que se le asigna mayor puntaje a los posibles sitios que coinciden

en las posiciones más relevantes. Por esta razón es importante conocer específicamente cuales son las posiciones más conservadas, además de la frecuencia exacta en que aparecen las distintas bases en esos sitios.

5. CONCLUSIONES

- A partir de la genoteca de cDNA construida se aislaron un significativo número de genes de *X. dendrorhous* hasta ahora desconocidos, incluyendo los que codifican para proteínas ribosómicas, enzimas del metabolismo de carbono y aminoácidos, señalización intracelular, transcripción y traducción. Muchos de los genes identificados en este estudio poseen especial relevancia para futuros estudios en *X. dendrorhous*, ya sea por su potencial uso como marcadores de selección o bien por su relación con el metabolismo del carbono y la biosíntesis de carotenoides.
- El aislamiento de la versión genómica de genes de interés identificados a partir de la genoteca permitió determinar la estructura exón-intrón de los mismos y establecer la secuencia consenso de los sitios dadores (NNGGTDHG) y aceptores (HHYAGNNN) de *splicing*. Esta información resulta de gran relevancia para la predicción de regiones intrónicas y exónicas de genes de estructura desconocida.
- El análisis de uso de codones en *X. dendrorhous* permitió establecer que éstos no son empleados al azar, llegando a determinar cuáles son los tripletes usados con mayor frecuencia y aquellos que son raramente empleados. La tabla de uso de codones construida con esta información constituye una herramienta poderosa para la búsqueda y clonamiento de nuevos genes a través del uso de la genética reversa.
- La cuantificación del sesgo en el uso de tripletes a través del cálculo del índice de adaptación de codones permitió realizar una predicción respecto a los niveles de expresión de los distintos genes identificados hasta ahora en *X. dendrorhous*. De esta manera se determinó que los genes de carotenogénesis poseen una baja adaptación al sistema traduccional de *X. dendrorhous*, identificando un nuevo blanco para la optimización de la producción de pigmentos en este organismo.

6. BIBLIOGRAFÍA

- Altschul S. F., Madden T.L., Schäffer A.A., Zhang J, Zhang Z., Miller W. y Lipman D.J.** (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs". *Nucleic Acids Res.* 25:3389-3402.
- Alvarez V., Rodriguez-Saiz M., de la Fuente J.L., Gudina E.J., Godio R.P., Martin J.F. y Barredo J.L.** (2006) "The crtS gene of *Xanthophyllomyces dendrorhous* encodes a novel cytochrome-P450 hydroxylase involved in the conversion of beta-carotene into astaxanthin and other xanthophylls". *Fungal Genet. Biol.* 43:261-72.
- Andrewes A.G. y Starr M.P.** (1976) "(3R,3'R)-Astaxanthin from the yeast *Phaffia Rhodozyma*". *Phytochemistry* 15:1009-1011.
- An G.H., Schuman D.B. y Johnson, E.A.** (1989) "Isolation of *Phaffia rhodozyma* mutants with increased astaxanthin content". *Appl. Environ. Microbiol.* 55:116-124.
- An G.H. y Johnson E.A.** (1991) "Astaxanthin from microbial sources". *Crit. Rev. Biotechnol.* 11:297-326.
- Bielich J., An G.H., Auerbach R. y Johnson, E.A.** (1991) "Isolation and characterization of carotenoid hyperproducing mutants of yeast by flow cytometry and cell sorting". *Biotechnology* 9:70-73.
- Bonekamp F., Andersen H.D., Christensen T. y Jensen K.F.** (1985) "The AGG codon is translated slowly in *E. coli* even at very low expression levels". *Nucleic Acids Res.* 13:4113-4123.
- Cañizares-Villanueva R.O., Rios-Leal E., Olvera Noyola T. y Marquez Rocha, F.** (1998) "Microbial sources of pigments". *Rev Latinoam Microbiol.* 40:87-107.
- Cifuentes V., Hermosilla G., Martínez C., Leon R., Pincheira G. y Jimenez A.** (1997) "Genetics and electrophoretic karyotyping of wild-type and astaxanthin mutant strains of *Phaffia rhodozyma*". *Antonie Van Leeuwenhoek* 72:111-117.
- Cameron J. y Agudé M.** (1998) "An Evaluation of Measures of Synonymous Codon Usage Bias". *J. Mol. Evol.* 47:268-274.
- Deml L., Bojak A., Steck S., Graf M., Wild J., Schirmbeck R., Wolf H. y Wagner R.** (2001) "Multiple Effects of Codon Usage Optimization on Expression and Immunogenicity of DNA Candidate Vaccines Encoding the Human Immunodeficiency Virus Type 1 Gag Protein" *J. Virol.* 75:10991:11001.
- Dennis R.A. y McCammon M.T.** (1999) "Acn9 is a novel protein of gluconeogenesis that is located in the mitochondrial intermembrane space". *Eur. J. Biochem.* 261:236-43.
- Deutsch M. y Long M.** (1999) "Intron-exon structures of eukaryotic model organisms". *Nucleic Acids Res.* 27:3219-3228.

- Flores-Cotera L.B., Martín R. y Sánchez S.** (2001) "Citrate, a possible precursor of astaxanthin in *Phaffia rhodozyma*: influence of varying levels of ammonium, phosphate and citrate in a chemically defined medium". *Appl. Microbiol. Biotechnol.* 55:341-347.
- Freire-Picos M., González-Siso M., Rodríguez-Belmonte E., Rodríguez-Torres A., Ramil E. y Cerdán M.** (1994) "Codon usage in *Kluyveromyces lactis* and in yeast cytochrome c-encoding genes". *Gene* 139:43-49.
- Gouy M. y Gautier C.** (1982) "Codon usage in bacteria: correlation with gene expressivity". *Nucleic acids Res.* 10:7055-7074.
- Grantham R., Gautier C., Gouy M., Jacobzone M. y Mercier R.** (1981) "Codon catalog usage is a genome strategy modulated for gene expressivity". *Nucleic Acids Res.* 9:43-74
- Ikemura T.** (1985) "Codon usage and tRNA content in unicellular and multicellular organisms". *Mol. Biol. Evol.* 2:13-34.
- Johnson E.A. y Schroeder W.A.** (1996) "Microbial carotenoids". *Adv. Biochem. Eng. Biotechnol.* 53:119-78.
- Johnson E. A.** (2003) "*Phaffia rhodozyma*: colourful odyssey". *Int. Microbiol.* 6:169-174.
- Jurica M.S. y Moore M.J.** (2003) "Pre-mRNA splicing: awash in a sea of proteins". *Molec. Cell* 12:5-14.
- Karlin S. y Mrazek J.** (2000) "Predicted highly expressed genes of diverse prokaryotic genomes". *J. Bacteriol.* 182:5238-5250.
- Ko H.J., Ko S.Y., Kim Y.J., Lee E.G., Cho S.N. y Kang C.Y.** (2005) "Optimization of codon usage enhances the immunogenicity of a DNA vaccine encoding mycobacterial antigen Ag85B". *Infect. Immun.* 73:5666-74.
- Kotula L. y Curtis P.J.** (1991) "Evaluation of foreign gene codon optimization in yeast: expression of a mouse IG kappa chain". *Biotechnology (N Y)* 9:1386-1389.
- Liu Q., Feng Y., Zhao x., Dong H. y Xue Q.** (2004) "Synonymous codon usage bias in *Oryza sativa*". *Plant sci.* 167:101-105.
- Lodato P., Alcaino J., Barahona S., Retamales P. y Cifuentes V.** (2003) "Alternative splicing of transcripts from crtI and crtYB genes of *Xanthophyllomyces dendrorhous*". *Appl. Environ. Microbiol.* 69:4676-4682.
- Lodato P., Alcaino J., Barahona S., Retamales P., Jimenez A. y Cifuentes V.** (2004) "Study of the expression of carotenoid biosynthesis genes in wild-type and deregulated strains of *Xanthophyllomyces dendrorhous* (Ex.: *Phaffia rhodozyma*)". *Biol. Res.* 37:83-93.
- Loftus B.J., Fung E., Roncaglia P., Rowley D., Amedeo P., Bruno D., Vamathevan J., Miranda M., Anderson IJ., Fraser JA., Allen JE., Bosdet IE., Brent M.R., Chiu R., Doering T.L., Donlin M.J., D'Souza C.A., Fox D.S., Grinberg V., Fu J., Fukushima M., Haas B.J., Huang J.C., Janbon G., Jones S.J., Koo H.L., Krzywinski M.I., Kwon-Chung J.K., Lengeler K.B., Maiti R., Marra M.A., Marra R.E., Mathewson C.A., Mitchell T.G., Perteau M., Riggs F.R., Salzberg S.L., Schein J.E., Shvartsbeyn A., Shin H., Shumway M., Specht C.A., Suh B.B., Tenney A., Utterback T.R., Wickes B.L., Wortman J.R., Wye N.H., Kronstad J.W., Lodge J.K., Heitman J., Davis R.W., Fraser C.M. y Hyman R.W.** (2005) "The genome of the

- basidiomycetous yeast and human pathogen *Cryptococcus neoformans*". *Science* 307:1321-1324.
- Lorenz R. y Cysewski G.** (2000) "Commercial potential for *Haematococcus* microalgae as a natural source of astaxanthin". *Trends Biotechnol.* 18:160-167.
- Maniatis T. y Reed R.** (1987) "The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing," *Nature* 325:673-678.
- Marchler-Bauer A, y Bryant SH** (2004). "CD-Search: protein domain annotations on the fly." *Nucleic Acids Res.* 32:327-331.
- Marin D., Linde D. y Lobato M.F.** (2006) "Purification and biochemical characterization of an alpha-glucosidase from *Xanthophyllomyces dendrorhous*". *Yeast* 23:117-125
- Martinez C., Hermosilla G., Leon R., Pincheira G. y Cifuentes V.** (1998) "Genetic transformation of astaxanthin mutants of *Phaffia rhodozyma*". *Antonie Van Leeuwenhoek* 73:147-153.
- McCammon M.T.** (1996) "Mutants of *Saccharomyces cerevisiae* with defects in acetate metabolism: isolation and characterization of Acn- mutants". *Genetics* 144:57-69.
- McColl D., Valencia C.A. y Vierula P.J.** (2003) "Characterization and expression of the *Neurospora crassa* nmt-1 gene". *Curr. Genet.* 44:216-23.
- Misawa N. y Shimada H.** (1997) "Metabolic engineering for the production of carotenoids in non-carotenogenic bacteria and yeasts". *J. Biotechnol.* 59:169-181.
- Morton B.R.** (1993) "Chloroplast DNA codon use: evidence for selection at the psbA locus based on tRNA availability". *J. Mol. Evol.* 37:273-280.
- Mueller O., Hahnenberger K., Dittmann M., Yee H., Dubrow R., Nagle R. y Ilsley D.** (2000) "A microfluidic system for highspeed reproducible DNA sizing and quantitation". *Electrophoresis* 21:128-34.
- Mueller O., Lightfoot S. y Schroeder A.** (2006) "RNA Integrity Number (RIN) - Standardization of RNA Quality Control". Agilent Application Note, Publication Number 5989-1165EN.
- Niklitschek M.** (2007) "Caracterización funcional y estructural del gen crtE de *X. dendrorhous*". Seminario de Título de Ingeniería en Biotecnología Molecular, Facultad de Ciencias, Universidad de Chile.
- Novembre J.** (2002) "Accounting for Background Nucleotide Composition When Measuring Codon Usage Bias". *Mol Biol. Evol.* 19:1390-1394.
- Ojima K., Breitenbach J., Visser H., Setoguchi Y., Tabata K., Hoshino T., van den Berg J. y Sandmann G.** (2006) "Cloning of the astaxanthin synthase gene from *Xanthophyllomyces dendrorhous* (*Phaffia rhodozyma*) and its assignment as a beta-carotene 3-hydroxylase/4-ketolase". *Mol. Genet. Genomics.* 275:148-58.
- Patel A. y Steitz J.** (2003) "Splicing double: insights from the second spliceosome" *Nat. Rev. Mol. Cell Biol.* 4:960-970.

- Phaff H.J., Miller M.W., Yoneyama M. y Soneda M.** (1972) "A comparative study of the yeast floras associated with trees on the Japanese Island and on the West Coast of North America". Proc. IV IFS: Ferment. Technol. Today, pp. 759-774.
- Reese M., Eeckman F., Kulp D. y Haussler D.** (1997) "Improved splice site detection in genie," J. Comp. Biol. 4:311-23.
- Retamales P., Hermosilla G., León R., Martínez C., Jiménez A. y Cifuentes V.** (2002) "Development of the sexual reproductive cycle of *Xanthophyllomyces dendrorhous*". J. Microbiol. Met 48:87-93.
- Saccharomyces genome database.** <http://www.yeastgenome.org/>
- Sambrook, J. y Russel, D. W.** (2001). "Molecular Cloning: A laboratory Manual". Third edition, CSHL Press.
- Schweingruber A. M., Dlugonski J., Edenharter E. y Schweingruber M. E.** (1991). "Thiamine in *Schizosaccharomyces pombe*: dephosphorylation, intracellular pool, biosynthesis and transport". Curr. Genet. 19, 249-254.
- Sharp P., Tuohy T. y Mosurski K.** (1986) "Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes". Nucl. Acids Res. 14:5125-5143.
- Sharp P. y Li W.** (1987) "The codon adaptation index – a measure of directional synonymous codon usage bias, and its potential applications". Nucleic Acids Res. 15:1281-1295.
- Sharp P. y Cowe E.** (1991) "Synonymous codon usage in *Saccharomyces cerevisiae*". Yeast 7:657-678.
- Stajich J. y Dietrich F.** (2005) "Fungal Intron evolution: Loss, Gain and Retention" Poster disponible en <http://fungal.genome.duke.edu/~jes12/talks/> (2005-05-16-AsilomarPoster.pdf)
- Stajich J. y Dietrich F.** (2006) "Evidence of mRNA-Mediated Intron Loss in the Human-Pathogenic Fungus *Cryptococcus neoformans*". Eukaryotic Cell 5:789-793.
- Supek F. y Vlahovicek K.** (2005) "Comparison of codon usage measures and their applicability in prediction of microbial gene expressivity". Bioinformatics 6:182.
- Thanaraj T. y Robinson A.** (2000) "Prediction of exact exon boundaries" Brief. Bioinf. 1:343-356.
- Vazquez M., Santos V. y Parajo J.C.** (1997) "Effect of the carbon source on the carotenoid profiles of *Phaffia rhodozyma* strains". J. Ind. Microbiol. Biotechnol. 19:263-268.
- Vazquez M. y Martin A.M.** (1998) "Optimization of *Phaffia Rhodozyma* continuous culture through response surface methodology". Biotechnol. Bioeng. 57:314-320.
- Verdoes J.C., Misawa N. y van Ooyen A.J.J.** (1999 a) "Cloning and characterization of the astaxanthin biosynthetic gene encoding phytoene desaturase of *Xanthophyllomyces dendrorhous*". Biotechnol. Bioeng. 63:750-755.
- Verdoes J.C., Krubasik P., Sandmann G. y van Ooyen, A.J.J.** (1999 b) "Isolation and functional characterization of a novel type of carotenoid biosynthetic gene from *Xanthophyllomyces dendrorhous*". Mol. Gen. Genet. 262:453-461.

Verdoes J. C. y van Ooyen A. J. (2000) "Codon usage in *Xanthophyllomyces dendrorhous* (formerly *Phaffia rhodozyma*)". *Biotechnol. lett.* 22:9-13.

Wery J., Gutker D, Renniers A.C., Verdoes J.C. y van Ooyen A.J. (1997) "High copy number integration into the ribosomal DNA of the yeast *Phaffia rhodozyma*". *Gene* 184:89-97.

Wightman R. y Meacock P.A. (2003) "The *THI5* gene family of *Saccharomyces cerevisiae*: distribution of homologues among the hemiascomycetes and functional redundancy in the aerobic biosynthesis of thiamin from pyridoxine". *Microbiology* 149:1447-1460.

Wozniak A., Alcaíno J., Carmona M., Marcoleta A., Lodato P., Baeza M. y Cifuentes V. (2005) "Effect of aerobic metabolism in the expression of carotenogenesis genes from *X. dendrorhous*". *Biol. Res.* 38 R1-R157.

Wright F. (1990) "The effective number of codons used in a gene". *Gene* 87:23-29.

Xing C., Bitzer D., Alexander W. y Stomp A. (2006) "Thermodynamics Exploration to Identify Donor Sites for Yeast" Proceedings of the IEEE International Workshop on Genomic Signal Processing and Statistics, disponible en <http://www4.ncsu.edu/~cxing/publications.htm>.