

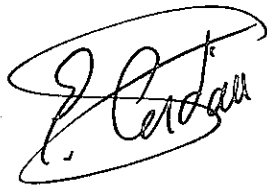
**ANÁLISIS DE LA TRANSCRIPCIÓN DE 103 ORFs
LOCALIZADAS EN LOS CROMOSOMAS VII y XIV
DE *Saccharomyces cerevisiae*.**

Luis José Lombardia Ferreira

El presente trabajo: **Análisis de la transcripción de 103 ORFs localizadas en los cromosomas VII y XIV de *Saccharomyces cerevisiae*** presentado por D. Luis José Lombardía Ferreira para aspirar al grado de Doctor en Biología, ha sido realizado bajo nuestra dirección en el Departamento de Biología Celular y Molecular de la Universidad de La Coruña.

Revisado el texto, estamos conformes con su presentación para ser juzgado.

La Coruña a 9 de Junio de 1997



VºBº

LAS DIRECTORAS DEL TRABAJO

Dra. María Esperanza Cerdán Villanueva

Dra. Ana María Rodríguez Torres

Profesoras de Bioquímica y Biología Molecular

Este trabajo forma parte del proyecto EUROFAN I subvencionado por los programas **Biotech IV** de la Comunidad Económica Europea y **CICYT BIO96-2022-CE**.

Por las numerosas imperfecciones persistentes, tanto en este manuscrito como en mi papel de investigador en Biología Molecular, soy consciente del camino que aún me queda por recorrer; sin embargo, al recordar mi trayecto a lo largo de estos casi cuatro años en el laboratorio de Bioquímica de la Facultad de Ciencias de La Coruña, me agrada reconocer que los acontecimientos superaran mis expectativas más optimistas. Sin duda debo esta "buena racha" a personas que aceptaron que caminase junto a ellas y a las que deseo expresar mis agradecimientos.

Quisiera agradecerle a la **Doctora Ana Rodríguez Torres**, no sólo haberme permitido emprender este camino sino también su complicidad al tolerar alguna de mis irreverencias.

También deseo expresar mis agradecimientos y admiración a la **Doctora Esperanza Cerdán**: agradecimientos por su confianza y ayuda en la redacción de este trabajo y en mi formación de investigador, Admiración por su capacidad en conciliar dos aspectos muy a menudo distantes, humanidad y racionalidad.

Deseo citar a cada uno de mis compañeros de laboratorio; **Isabel, Marián, Elvira, Esther, Belén, Mónica, Manuel, José Luis** y mas recientemente **Mónica "II"** y agradecerles su acogida, buen humor y esfuerzos para que nuestras variadas personalidades logren cohabitar.

Le agradezco a **Yvon Van Delft** de la Universidad Libre de Amsterdam en Holanda su ayuda durante la ejecución del proyecto EUROFAN y por proporcionarme los datos obtenidos por todos los laboratorios miembros del consorcio B2.

No quisiera olvidar a **María Elena** por su refrescante espíritu, al campeón de Europa **Federico "Agassi"** así como a los (las) demás compañeros(as) de la Facultad de Ciencias y a algún "gabacho" que no me atrevo a citar por temor a un olvido.

Finalmente un recuerdo "especial" para mi punto de referencia y fuente de equilibrio, **Elvira**, que me enseñó y sigue enseñando que, como una investigación, una relación sólida se "curra" todos los días.

Porque, a su manera, colaboraron más que ninguno en ella, dedico esta Tesis de Doctorado a mis padres y familia.

*" En essayant continuellement , on finit par reussir.
DONC : Plus ça rate, plus on a de chances que ça marche !"*

Profesor Rémi SHADOKO

(“intentándolo continuamente, acabamos consiguiéndolo.
Así: cuanto más falla algo, más posibilidades tenemos que acabe funcionando.”)

ÍNDICE

1. INTRODUCCIÓN	1
1.1. Secuenciación del genoma de <i>S. cerevisiae</i>	4
1.2. Características del genoma de <i>S. cerevisiae</i>	5
1.3. Estudio de la función de los genes desconocidos	10
1.4. Regulación de la expresión génica en la levadura	12
1.4.1. Regulación por la fuente de carbono	13
1.4.2. Privación de aminoácidos	16
1.4.3. Regulación por la fuente de nitrógeno	17
1.4.4. Regulación génica en la fase estacionaria	19
1.4.5. Choque osmótico	21
1.4.6. Choque térmico	22
1.4.7. Regulación por oxígeno	23
2. OBJETIVOS	27
3. MATERIALES Y MÉTODOS	31
3.1. Cepas celulares	33
3.1.1. Levaduras	33
3.1.2. Bacterias	33
3.2. Medios y condiciones de cultivo	34
3.2.1. Composición de los medios de cultivo	34
3.2.2. Condiciones de cultivo	36
3.3 Extracción del RNA	39
3.3.1. Condiciones de extracción	39
3.3.2. El protocolo de extracción	40
3.3.3. Medida del rendimiento de la extracción	41

3.4. <i>Northern blot</i>	42
3.4.1. El gel de agarosa	42
3.4.2. Las muestras de RNA	42
3.4.3. La electroforesis	42
3.4.4. La transferencia	43
3.5. Obtención de sondas por PCR	44
3.5.1. Diseño de los cebadores	44
3.5.2. El DNA molde utilizado	51
3.5.3. La reacción en cadena de la polimerasa	55
3.5.4. Análisis de los productos de PCR	57
3.6. Reacciones de hibridación	62
3.6.1. Marcaje de las sondas	62
3.6.2. Separación y cuantificación de la incorporación de isótopo	64
3.6.3. Reacciones de hibridación y lavado	65
3.7. Cuantificación de las señales de hibridación	66
3.7.1. Nivel relativo	68
3.7.2. Nivel absoluto	69
3.8. Búsqueda de características deducidas de las secuencias de DNA	70
3.8.1. Determinación del uso de codones	70
3.8.2. Estudio de los promotores de las ORFs	71
3.8.3. Otras características deducidas de las secuencias	72
4. RESULTADOS Y DISCUSIÓN	75
4.1. Selección de las condiciones de análisis	77
4.2. Verificación del método de análisis con sondas de regulación conocida	82
4.3. Análisis mediante <i>Northern blot</i> de la expresión de las ORFs objeto de estudio	87
4.4. Niveles absolutos de expresión	119
4.4.1. Clasificación de las ORFs en función de su nivel absoluto	119
4.4.1.1. ORFs cuya expresión no fue detectada	123

4.4.1.2. ORFs para las que fueron detectados varios transcritos	124
4.4.1.3. ORFs de alta expresión	125
4.4.1.4. ORFs de expresión media	126
4.4.1.5. ORFs de baja expresión	126
4.4.2. Relación de los niveles absolutos de expresión con otros parámetros génicos y genómicos	129
4.4.2.1. Comparación de los niveles de expresión con los índices de CAI	129
4.4.2.2. Comparación de los niveles de expresión con el tamaño de las ORFs	132
4.4.2.3. Análisis de los niveles absolutos de expresión en relación con la posición en el cromosoma	134
4.4.2.4. Expresión y densidad génica	139
4.5. Niveles relativos de expresión	140
4.5.1. Niveles de expresión en las distintas condiciones de cultivo	140
4.5.2. Clasificación de las ORFs en función de su regulación	143
4.5.3. Cuantificación del factor de regulación	148
4.5.4. Regulación y función	152
4.6. Búsqueda de sitios de unión a factores transcripcionales en el promotor de las ORFs analizadas	156
4.7. Regulación de la expresión génica en mutantes <i>hap1</i> , <i>rox1</i> y <i>rox6</i>	172
5. CONCLUSIONES	179
6. BIBLIOGRAFÍA	183
7. APÉNDICE	201

ÍNDICE DE TABLAS Y FIGURAS

TABLAS

Tabla 1.1.	6
Tabla 1.2.	15
Tabla 1.3.	18
Tabla 3.1.	33
Tabla 3.2.	33
Tabla 3.3.	46
Tabla 4.1.	77
Tabla 4.2.	81
Tabla 4.3.	82
Tabla 4.4.	115
Tabla 4.5.	120
Tabla 4.6.	121
Tabla 4.7.	131
Tabla 4.8.	134
Tabla 4.9.	139
Tabla 4.10.	143
Tabla 4.11.	145
Tabla 4.12.	146
Tabla 4.13.	149
Tabla 4.14.	152
Tabla 4.15.	157
Tabla 4.16.	159
Tabla 4.17.	161
Tabla 4.18.	163
Tabla 4.19.	169
Tabla 4.20.	179

FIGURAS

Figura 3.1.	38
Figura 3.2.	55
Figura 3.5.1	59
Figura 3.5.2.	60
Figura 4.1.	83
Figura 4.2.1.	89
a 4.2.26	114
Figura 4.3.	128
Figura 4.4.	132
Figura 4.5.	133
Figura 4.6.	136
Figura 4.7.	138
Figura 4.8.	141
Figura 4.9.	142
Figura 4.10.	147
Figura 4.11.	151
Figura 4.12.	173

Apéndice

Tabla Apéndice	203
----------------	-----

ABREVIATURAS

A₆₀₀: absorbancia a 600 nm

ATP: adenosin-trifosfato

CAI: índice de adaptación de codones

crom.: cromosoma

DNA: ácido desoxirribonucleico

dNTPs: desoxirribonucleósidos trifosfato

GTP: Guanosin-trifosfato

Kb: kilobase

min: minutos

MIPS: *Munich Information Center for Protein Sequences*

mRNA: RNA mensajero

nt: nucleótidos.

ORF: pauta abierta de lectura (*Open Reading Frame*)

pb: pares de bases

PCR: reacción en cadena de la polimerasa

RNA: ácido ribonucleico

rpm: revoluciones por minuto

rRNA: RNA ribosómico

s: segundos

sn RNA: *small nuclear RNA*

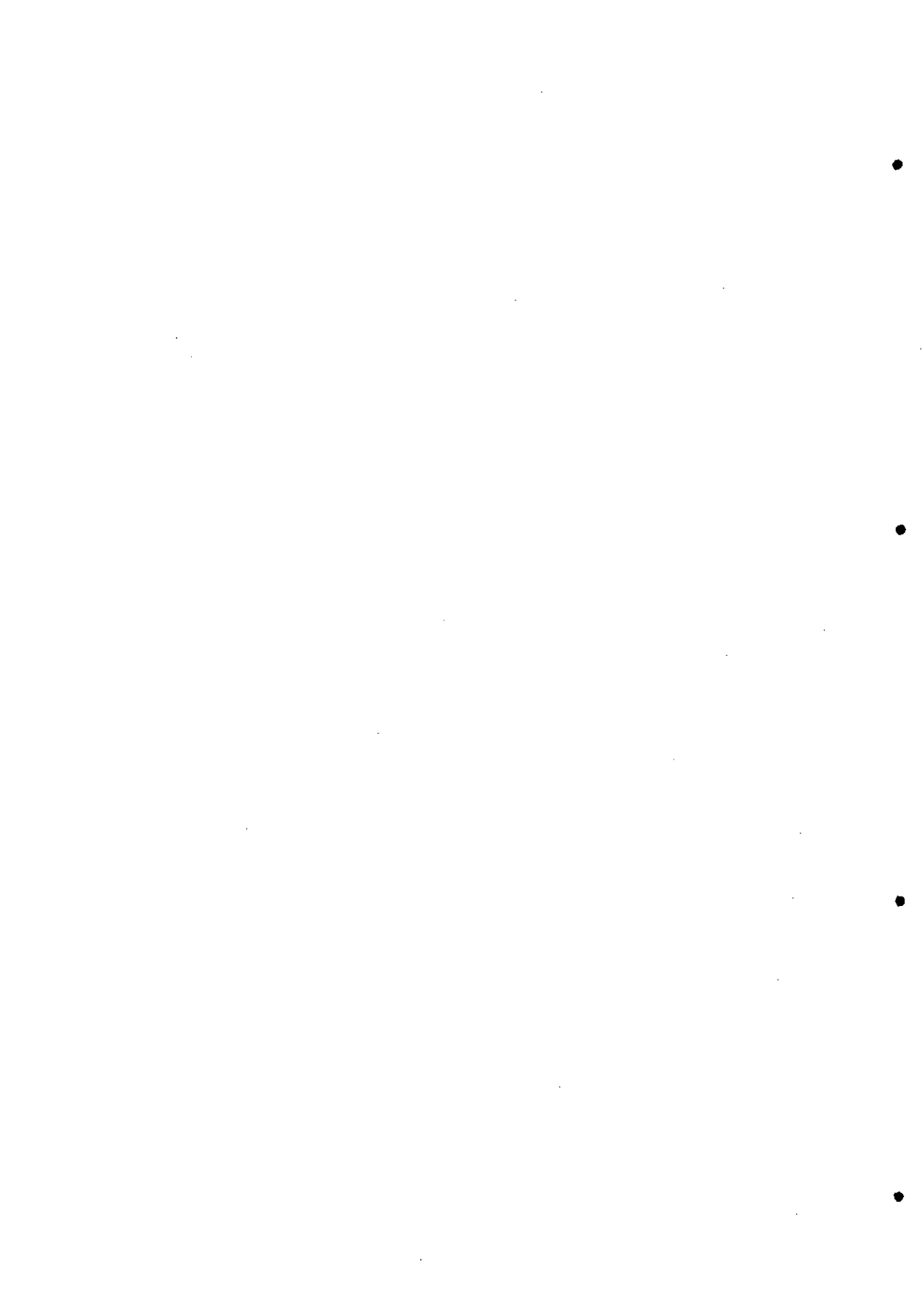
T_m: temperatura de fusión del DNA (temperatura de *melting*)

U: unidades

UAS: *Upstream Activating Sequence*

UV: ultravioleta

1. INTRODUCCIÓN



Durante miles de años el hombre ha utilizado las levaduras para la producción y conservación de alimentos por su capacidad de fermentar glucosa a etanol y dióxido de carbono. Así, las industrias panificadoras, vitivinícolas y cerveceras han utilizado una levadura conocida por todos como la del pan o de la cerveza: *Saccharomyces cerevisiae*. Esta levadura ha permitido además, a lo largo de este último siglo, otras utilidades que le conceden el título de organismo modelo para la mayor parte de la investigación en genética molecular.

S. cerevisiae, además de ser de fácil y barato cultivo, por su corto tiempo de generación, posee un genoma de tamaño pequeño, 200 veces menor que el humano, y compacto, alrededor de 6000 genes con pocos intrones y secuencias repetidas. Sus mecanismos básicos de replicación, recombinación, división celular y metabolismo han sido ampliamente estudiados a lo largo de los últimos decenios. Por otro lado, su organización celular eucariótica es semejante a la de los organismos superiores; esto nos permite proyectar conclusiones obtenidas de estudios en levadura a genomas mucho más complejos como es el del ser humano. Por ejemplo, el estudio del genoma de la levadura puede contribuir a una mayor comprensión a nivel molecular de las regiones reguladoras de la transcripción, a la descripción de nuevos motivos funcionales de DNA, a la búsqueda del origen y función de las secuencias no codificantes del DNA y de las causas de la multiplicidad del RNA de transferencia. También nos permite acercarnos a la comprensión de ciertos mecanismos patológicos ampliamente extendidos en organismos superiores, merced a la similitud de secuencias entre genes humanos responsables de enfermedades y genes de levadura (Decottignies y Goffeau, 1997; Goffeau *et al.*, 1997; Foury, 1997).

1.1. Secuenciación del genoma de *S. cerevisiae*

Los antecedentes del estudio del genoma de la levadura pueden encontrarse en los años 1970, cuando Lauer *et al.* (1977) estimaron el tamaño del genoma en aproximadamente 14.000 Kb y Kaback *et al.* (1979), a partir de su análisis "R-loop", postularon la existencia de más de 5.000 genes que se expresaban durante la vida vegetativa de *S. cerevisiae*. En 1985 se sabía que el núcleo de levadura contiene 16 cromosomas, y se conocía el mapa genético de 769 genes (Mortimer y Schild, 1989) de los cuales, sólo 17 presentaban intrones (Fink, 1987).

A finales de los años 1980 los métodos tradicionales de la Genética Molecular, cuya estrategia principal para el estudio del genoma consistía en caracterizar primero la función de un gen para después secuenciarlo, aceleraron el descubrimiento de nuevos genes de tal manera que la identificación del mismo gen por distintos investigadores era cada vez más frecuente; sin embargo, amplias regiones permanecían inexploradas.

Entre Enero de 1989 y Julio de 1990, un proyecto piloto europeo permitió, por primera vez, la secuenciación completa del cromosoma III (315 Kb) de *S. cerevisiae* (Oliver *et al.*, 1992). Este éxito permitió ampliar el estudio a los 15 cromosomas restantes de la levadura (Goffeau y Vassaroti, 1991), gracias a la colaboración internacional de más de 100 laboratorios en Europa, América del Norte y Japón. Nuestro laboratorio se incorporó a este proyecto en 1994 participando en la secuenciación de 17 Kb del cromosoma VII (Rodríguez Belmonte *et al.*, 1996; Tizón *et al.*, 1996; Tettelin *et al.*, 1997).

El 24 de Abril de 1996, se hizo pública la primera secuencia completa de un genoma eucariota. Hasta la fecha sólo habían sido secuenciados completamente los genomas de 6 organismos procariotas: *Haemophilus influenzae*, *Mycoplasma genitalium*, *Methanococcus jannaschii*, *Synechocystis*, *Mycoplasma pneumoniae* y *Escherichia coli*, pero en la actualidad se están secuenciando los genomas de organismos eucariotas superiores como los de *Arabidopsis thaliana* y *Drosophila melanogaster* así como el del ser humano.

La secuenciación de *S. cerevisiae* no se consideró completa hasta que fue determinada en las dos hebras sin ningún hueco, incluyendo a los telómeros. El grado de precisión también fue elevado, estimando la tasa de error en menos de un fallo cada 10 Kb; por ejemplo, únicamente 14 diferencias de lectura fueron detectadas en un fragmento de 172 Kb secuenciado independientemente en Europa y Estados Unidos (Johnston, 1996).

El genoma mitocondrial no fue objeto de la secuenciación sistemática sino de la labor de Zamaroczy y Bernardi (1986). Reunieron los datos de diferentes cepas de levaduras, originados en varios laboratorios, para ensamblar las secuencias respectivas en un genoma mitocondrial único.

1.2. Características del genoma de *S. cerevisiae*

La secuenciación completa ha dado a conocer que el genoma, presente en el núcleo de una cepa haploide de *S. cerevisiae* habitualmente usada en el laboratorio, FY279, consiste en 12.147 Kb de DNA único que codifica aproximadamente 6.000 proteínas (6.758 ORFs encontradas de las cuales 431 son cuestionables), excluyendo las agrupaciones de DNA ribosómico o DNAr

(1-2Mb) y otras familias de repeticiones incluyendo los genes *ENA2* y *CUP1*, regiones LTRs (Long Terminations Repeats) y retrotransposones Ty de varias clases (Tabla 1.1) (Dujon, 1996).

Tabla 1.1. Composición del genoma de *S. cerevisiae*. Datos extraídos de la base de datos MIPS (*Martinsride Institute for Protein Sequences*), cuya dirección en Internet es: <http://www.mips.biochem.mpg.de/>

	CROMOSOMA															
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI
Tamaño de la secuencia total (Kb)	230	813	315	1554	577	271	1091	589	440	745	666	2352	924	784	1091	948
REPETICIONES Tamaño (Kb)				<i>ENA2</i> & <i>Y'</i> 8+14				<i>CUP1</i> 26				<i>rDNA</i> & <i>Y'</i> 1260 + 14				
Proteínas Hipotéticas	103	393	160	752	279	130	516	280	220	356	315	509	458	399	534	463
Proteínas cuestionables	3	31	11	66	13	6	57	11	11	32	20	41	30	23	37	37
Intrones	3	18	4	32	12	5	16	16	8	13	11	18	20	15	13	19
Elementos Ty	1	3	2	9	2	1	6	2	1	3	0	6	4	3	4	5
Genes tRNA	4	13	10	28	20	10	36	11	10	24	16	21	21	14	20	17
Genes snRNA	1	2	3	2	2	0	6	1	1	4	2	5	9	4	7	3

Para evitar el análisis de ORFs fortuitas (sin capacidad codificadora real), los autores de la secuenciación del primer cromosoma de levadura adoptaron un umbral de 100 codones por debajo del cual se rechazaban todas las secuencias de DNA con la estructura de una ORF (Oliver *et al.*, 1992). Estimaron que una ORF mayor de 100 codones tenía menos del 0,2% de probabilidad de ser fortuita. Como no había garantía de que las ORFs de tamaño superior a 100 codones no fueran fortuitas, Dujon *et al.* (1994) propusieron 2 reglas adicionales: Una ORF menor de 150 codones y con un índice de uso de codones bajo (CAI < 0,11) era declarada cuestionable hasta su futuro análisis. Sin embargo, cabe señalar que se conocen genes con un

tamaño inferior a 100 codones y que codifican proteínas de levadura implicadas en diferentes e importantes funciones; de 58 genes conocidos con estas características, algunos codifican proteínas ribosomales, otros proteínas asociadas a snRNA, subunidades de citocromo o proteínas implicadas en el metabolismo del ATP (Barry *et al.*, 1996). Esto sugiere que el genoma de la levadura puede contener otros genes funcionalmente importantes que se obviaron por el método anteriormente descrito. Se desarrollaron modelos matemáticos que permiten pronosticar, tanto ORFs de tamaño inferior a 100 codones y potencialmente codificadoras de proteínas funcionales, como ORFs mayores de 100 codones cuya capacidad codificadora resulta cuestionable (Termier y Kalogeropoulos, 1996; Barry *et al.*, 1996). Para no descartar ORFs biológicamente funcionales y de tamaño inferior a 100 codones, se compararon sus secuencias con las que codificaban proteínas ya conocidas en otros organismos. Esto permitió incluir a 87 ORFs menores de 100 codones como potencialmente funcionales (Mewes *et al.*, 1997b).

Una de las sorpresas surgidas del análisis de la secuencia del genoma de levadura fue su alto contenido de información (densidad génica); en torno al 70 % del DNA cromosómico codifica proteínas, existiendo un gen aproximadamente cada 2 Kb. Dada la distribución de tamaños de las ORFs, la elevada compactación del genoma de *S. cerevisiae* resulta de dos factores: el tamaño reducido de las regiones intergénicas y la escasez de los intrones (Kalogeropoulos, 1995). En efecto, sólo el 4,5% de sus genes poseen intrones de tamaño reducido y localizados en los extremos 5' de ORFs cortas con una gran desviación en el uso de codones y, por lo tanto, posiblemente asociados a genes de alta expresión (Dujon, 1996).

Los genes se distribuyen globalmente de igual modo en cada una de las dos hebras de DNA y parecen estar orientados entre sí aleatoriamente, aunque

existen agrupaciones de ORFs en la misma orientación y con un posible significado funcional. Sin embargo, la densidad de genes no es uniforme a lo largo de los cromosomas, alcanzando valores superiores al 85% en ciertas zonas y descendiendo por debajo de la media (50-55 %) en otras. Las regiones pericentroméricas y subteloméricas siempre muestran una baja densidad en genes. Se apreció, además, una correlación entre el contenido elevado en guanina y citosina (GC) y las regiones de alta densidad génica, como se suele observar en genomas más complejos (Dujon, 1996).

El genoma de la levadura, comparándolo con el de organismos superiores, es particularmente pobre en secuencias repetidas (Tabla 1.1). Estas agrupan, básicamente, las repeticiones teloméricas, a algunas repeticiones de poli-A, poli-T y poli-AT, y a las repeticiones de genes (Dujon, 1996). Estas últimas son el reflejo de una alta redundancia del genoma de *S. cerevisiae*. Se observaron distintos casos de duplicaciones de secuencias:

- En las regiones subteloméricas existen segmentos largos, con varias ORFs, que son parecidos cuando se comparan varios cromosomas y que sugieren cambios de información genética recientes y/o continuos (Dujon, 1996),
- Aleatoriamente esparcidas a lo largo de la secuencia de distintos cromosomas aparecen duplicaciones de genes aislados, en tándem o con una orientación invertida (Dujon, 1996).
- También se han encontrado algunas duplicaciones “fósiles” de conjuntos de genes, cuyo orden se conservó, en las zonas pericentroméricas (Johnston *et al.*, 1994; Galibert *et al.*, 1996).

La duplicación de buena parte de los genes y su ordenación en los cromosomas llevó a sugerir que el genoma de *S. cerevisiae* es el resultado de

una duplicación ancestral de todo el genoma y de una evolución posterior con pérdida de parte de los genes duplicados y algunas reorganizaciones cromosómicas (Wolfe y Shields, 1996). Si bien es obvio que una cierta parte de la redundancia génica es en realidad sólo aparente (varios genes homólogos desempeñan finalmente funciones distintas), también parece lógico pensar que la duplicación completa de genes a nivel funcional resulta una ventaja para la célula; ésta albergaría así una reserva de genes repetidos para contrarrestar cualquier tipo de mutación que resultara peligrosa para la célula. Además, se sabe que la duplicación génica es un método evolutivo para la adquisición de nuevas funciones sin afectar las que ya están presentes; su existencia en la levadura podría implicar un mecanismo de adaptación evolutiva de este microorganismo al medio ambiente.

Tal vez, el mayor hallazgo obtenido en la secuenciación completa del genoma de *S. cerevisiae* es que menos de la mitad (alrededor del 30%) de todos los genes habían sido previamente identificados y estudiados por métodos convencionales bioquímicos y genéticos. Las ORFs restantes de función desconocida, excluyendo las que presentan un carácter dudoso (7%), pueden repartirse en dos grupos proporcionales (Dujon, 1996):

- Aproximadamente un 30% son ORFs que codifican proteínas estructuralmente relacionadas con productos de genes cuya función está caracterizada en levaduras o en otros organismos, o con motivos o dominios presentes en familias de proteínas funcionalmente caracterizadas.
- El 30 a 35% restante, son ORFs que no presentan homología estructural significativa con ninguna secuencia existente en las bases de datos (huérfanos solitarios o simples) o que muestran una

homología estructural pero con productos de ORFs que no fueron caracterizados funcionalmente y que en la mayoría de los casos son ORFs de levadura (parejas o familias de huérfanos).

No obstante, y aunque no se conozcan homólogos, el análisis informático proporciona datos sobre la naturaleza de algunos huérfanos. Por ejemplo, se pueden pronosticar los segmentos transmembrana de las proteínas resultantes y se ha determinado que de un 35% a un 40% de las proteínas poseen hélices transmembrana. Alrededor del 10% de las ORFs contienen más de cuatro hélices transmembrana y podrían codificar para transportadores localizados en la membrana plasmática.

También se puede conocer mediante análisis informático la presencia de motivos típicos como son los dedos de zinc, las cremalleras de leucina, los sitios de fosforilación, *etc.* No obstante, este tipo de predicciones sólo aporta pistas para la búsqueda de la función real del gen.

1.3. Estudio de la función de los genes desconocidos

La disponibilidad de la secuencia completa del genoma de *S. cerevisiae* en las bases de datos públicas proporcionó a los biólogos e informáticos una fuente incomparable para el análisis sistemático de las claves genéticas de la función biológica. También provocó la puesta en marcha de proyectos de investigación con el objetivo de enriquecer los resultados obtenidos en el proyecto de secuenciación.

La pauta común a estos proyectos era obtener una información biológica rápida y comprensible usando tecnologías efectivas (buena relación coste/información obtenida y aplicabilidad a la totalidad del genoma) y

relacionarla con la información *in silico* resultante de la fase de secuenciación. La meta prioritaria de muchos de estos proyectos es encaminarse a nuevas estrategias globales para el estudio de fenómenos biológicos importantes para la salud humana.

El proyecto EUROFAN (*EURO*pean *F*unctional *A*nalysis *N*etwork) aprobado por la Unión Europea en Diciembre de 1995, tiene como finalidad el análisis sistemático de la función de los genes no identificados previamente. EUROFAN implica a 144 laboratorios de 14 estados europeos en universidades, institutos de investigación e industrias. Está organizado de manera jerárquica, lo que permitirá acercarse a la función biológica específica de cada gen a medida que este avanza en la ruta del análisis (Oliver, 1996).

EUROFAN está permitiendo acelerar el progreso de la investigación biológica tradicional, creando una fuente de información y de material para la ciencia básica, la investigación médica y la bioindustria. Estos recursos se materializan en la creación de:

- Una base de datos que recopila la función de los genes, e información de las repercusiones de la organización del genoma sobre las funciones celulares.
- Una colección centralizada formada por:
 - Las líneas de levadura que contienen los genes interrumpidos.
 - Los plásmidos que contienen los genes clonados.
 - Los plásmidos que contienen los casetes de interrupción.

Con este apoyo, se podrá asignar rápidamente funciones a nuevos genes encontrados en otros genomas (Tugendreich *et al.*, 1994), identificando la equivalencia entre los productos de los genes de levadura y los de eucariotas

superiores tanto a nivel estructural, comparando las secuencias, como funcional, vía la complementación de una mutación específica en la levadura con cDNA derivado de organismos superiores: plantas, animales o humanos.

Una de las colaboraciones de nuestro laboratorio a este proyecto se concreta en la realización de un análisis de transcripción detallado de 1.000 ORFs de función desconocida en colaboración con otros cuatro laboratorios europeos. Esta red de trabajo constituye el denominado consorcio B2. Los objetivos de esta investigación abarcan los puntos siguientes:

- Determinar si las secuencias codificadoras identificadas *in silico* se transcriben *in vivo*.
- Proporcionar información en cuanto al papel que el producto del gen puede desempeñar en la célula.
- Aportar datos sobre los elementos genómicos no codificantes (esencialmente en las zonas promotoras).
- Desarrollar nuevas técnicas que mejoren la tecnología actual.

El trabajo aquí presentado es una parte del citado proyecto. También participamos en un análisis básico de identificación de funciones basado en la obtención de deleciones seguida de un análisis básico de los fenotipos resultantes (consorcio B0).

1.4. Regulación de la expresión génica en la levadura

Las células vivas, ya sean procariontas o eucariotas, emplean sistemas sensores y de señalización específicos para obtener y transmitir información procedente de su entorno con el objeto de ajustar su metabolismo celular, su crecimiento y su desarrollo a las alteraciones medio ambientales.

La adaptación del metabolismo de la levadura a cambios en las condiciones de crecimiento implica, además de un control a nivel post-transcripcional y de un control alostérico de las proteínas pre-existentes en el citoplasma (respuesta temprana o rápida), un control de la expresión génica (respuesta retardada). La combinación de los controles rápidos y retardados permite a las células adaptarse y crecer bajo condiciones de estrés suave, o sobrevivir bajo condiciones más severas (resistencia inducida por el estrés).

1.4.1. Regulación por la fuente de carbono (glucosa):

- Represión e inactivación por glucosa:

S. cerevisiae es una levadura con una gran capacidad fermentadora, incluso si las condiciones de crecimiento son aeróbicas (Lagunas, 1986). Se pensó que este predominio de la fermentación se debía a que numerosas enzimas de la levadura, implicadas en la respiración (ciclo de Krebs, cadena mitocondrial y fosforilación oxidativa) o en la utilización de otras fuentes de carbono, se veían negativamente afectadas por la presencia de glucosa u otros azúcares fácilmente fermentables como la fructosa o la manosa (Mahler *et al.*, 1981). La amplitud de este efecto varía, según la enzima implicada, desde un factor de más de 100 para enzimas gluconeogénicas y del ciclo del glioxilato (Haarasilta y Oura, 1975) hasta un factor de 3-10 para enzimas mitocondriales (Polakis y Bartley, 1966; Polakis *et al.*, 1964). Dos mecanismos distintos pueden actuar a este nivel: por una parte la “inactivación” por glucosa inhibe rápidamente la actividad de algunas proteínas enzimáticas, que son modificadas y/o degradadas; por otro lado la “represión” por glucosa restringe la expresión de muchos genes a nivel transcripcional.

La represión por glucosa afecta mayoritariamente a los genes que codifican proteínas innecesarias durante el crecimiento en glucosa; por ejemplo los genes para la utilización de otras fuentes de carbono como, la galactosa (genes *GAL*, Adams, 1972), la maltosa (*MAL*, Federoff *et al.*, 1983), la sacarosa (*SUC*, Perlman y Halvorson, 1981; Carlson y Botstein, 1982), el glicerol (*GUT*, Sprague y Cronan, 1977) y el etanol (*ADH*, Denis *et al.*, 1981). También está afectada la expresión de las proteínas que no se requieren a altos niveles durante la fermentación, es decir los genes que codifican para las enzimas del ciclo del ácido cítrico y del ciclo del glioxilato (Lombardo *et al.*, 1990), las proteínas mitocondriales de la cadena respiratoria (Repetto y Tzagoloff, 1990; Wright y Poyton, 1990), y las enzimas responsables de la gluconeogénesis (Valdés-Hevia *et al.*, 1989).

Otros genes que presentan represión por glucosa son los relacionados con diversas funciones como la esporulación (Salmon *et al.*, 1989), el transporte de la glucosa (Bisson, 1988) y varias actividades enzimáticas: proteasas (Moehle y Jones, 1990), enzimas peroxisomales (Einerhand *et al.*, 1991; Igual *et al.*, 1991) y la acetil-CoA hidrolasa (Lee *et al.*, 1990).

Se conocen más de una docena de genes implicados en la represión por glucosa (Gancedo, 1992; Trumbly, 1992), algunos de ellos fueron redundantemente caracterizados en distintos laboratorios y su papel mejor definido (Tabla 1.2.).

Tabla 1.2. Principales factores *trans* implicados en la represión por glucosa.

GEN	Producto	Características y/o función pronosticada	Referencia
<i>HXK2</i>	Hexokinasa isozima PII	Al fosforilar la glucosa podría generar la señal para la represión.	Lobo y Maitra, 1977
<i>SNF1</i>	Serín-treonin-kinasa	Desrepresión de los genes reprimidos por glucosa activándolos directamente y/o inhibiendo a Mig1p	Schüller y Entian, 1987
<i>TUP1/SSN6</i>	Tup1p/Ssn6p	Complejo corepresor general que sería reclutado por Mig1p en la represión por glucosa.	Rothstein y Sherman, 1980 Trumbly, 1986
<i>MIG1</i>	Mig1p	Proteína de unión a DNA. Represor que se une a los promotores (cajas G-C) de muchos genes reprimidos por glucosa.	Nehlin <i>et al.</i> , 1991 Treitel y Carlson, 1995 Ronne, 1995

El sitio de unión de Mig1p está representado por la secuencia consenso SYGGGG (donde S representa a los nucleótidos C ó G, e Y: C ó T) flanqueada por una zona rica en A-T (Lundin *et al.*, 1994).

- Activación por glucosa:

La expresión de algunos genes aumenta cuando las células de levadura se cambian de un medio de cultivo con una fuente de carbono no fermentable (etanol/glicerol) a un cultivo con glucosa. Estos genes codifican probablemente proteínas necesarias para el crecimiento en glucosa, como son las enzimas glucolíticas, pero también experimentan activación otros genes no relacionados como pueden ser los que codifican para proteínas ribosomales (Herruer *et al.*, 1989). El nivel de inducción de la expresión de estos genes varía de un factor de 20 a 30 para *ENO2* (Cohen *et al.*, 1986) a un factor de 2 a 10 para *ADHI* (Denis *et al.*, 1983).

En la mayoría de los casos, la regulación parece llevarse a cabo a nivel transcripcional. Aunque se han observado sitios UAS (*Upstream Activating Sequence*), en las regiones promotoras de los genes inducidos por glucosa que contienen zonas de unión para el activador transcripcional codificado por el gen *RAP1*, este parece ser un factor general de transcripción no específicamente relacionado con la inducción por glucosa (Capieaux *et al.*, 1989).

Otra proteína, necesaria para el crecimiento de *S. cerevisiae* en medios cuya fuente de carbono es la glucosa, parece activar la expresión de los genes glucolíticos y está codificada por *GCR1* (Baker, 1986). Es una proteína de unión a DNA, mediante la secuencia consenso GMWTCCW, donde M representa a los nucleótidos A ó C y W a las bases pirimidínicas A ó T (Svetlov y Cooper, 1995). En muchos promotores estudiados, los sitios de unión a Gcr1p están cercanos a los de unión a Rap1p, y probablemente se produce una acción conjunta de los dos factores en la inducción por glucosa de dichos genes (Johnston y Carlson, 1992).

1.4.2. Privación de aminoácidos:

S. cerevisiae es capaz de biosintetizar todos los aminoácidos que necesita cuando estos no se encuentran en el medio de cultivo mediante la puesta en marcha de varios mecanismos, tanto generales como específicos, que controlan la síntesis de las enzimas de las vías biosintéticas de los aminoácidos.

La expresión de al menos 35 genes, que codifican enzimas de 12 vías de biosíntesis de aminoácidos, está co-regulada en la levadura. La transcripción de esos genes, en respuesta a una privación en cualquiera de los

aminoácidos estudiados (histidina, arginina, lisina, isoleucina, valina, leucina, serina, fenilalanina, triptófano, metionina y prolina), se incrementa de 2 a 10 veces debido a su desrepresión directa por Gcn4p, el principal activador transcripcional de este sistema de regulación, que se une en el promotor de cada gen a la secuencia consenso ATGASTCAT (S = C ó G) (Svetlov y Cooper, 1995). La activación transcripcional de Gcn4p y su unión a los promotores se integra con las funciones de otras proteínas reguladoras responsables de la expresión basal y de la represión de los genes que codifican enzimas de vías específicas de biosíntesis de aminoácidos.

1.4.3. Regulación por la fuente de nitrógeno:

S. cerevisiae es capaz de ajustar su composición enzimática adaptándose a la fuente de nitrógeno disponible. La presencia de fuentes de nitrógeno que permiten un crecimiento óptimo (y usadas preferentemente por *S. cerevisiae*) como la glutamina, la asparragina y el amonio, provocan la disminución del nivel de enzimas necesarias para la utilización de fuentes de nitrógeno pobres (glutamato, prolina, alantoína, nitrato, urea, aminoácidos, proteínas, etc.). La represión catabólica por la fuente de nitrógeno o *NCR* (*Nitrogen Catabolic Repression*) es el fruto de dos mecanismos generales (Magasanik, 1992):

- La inutilización del sistema de transporte a la célula de las fuentes de nitrógeno “pobres”.
- La represión directa de la expresión de los genes implicados en la utilización de estas fuentes de nitrógeno.

Las respuestas a la fuente de nitrógeno disponible están mediadas por factores reguladores de la transcripción. Muchas de estas proteínas

reguladoras resultaron pertenecer a la familia de los “factores de transcripción GATA” que intervienen en distintos tipos de funciones reguladoras en varios organismos. El nombre de la familia se debe a su unión, mediante un motivo “dedo de zinc” muy conservado, a la secuencia GATAA repetida dos o más veces en los promotores de los genes regulados (Marzluf, 1997). *S. cerevisiae* posee, por lo menos, 4 factores GATA que intervienen en la regulación de la expresión génica relacionada con el metabolismo del nitrógeno (Tabla 1.3.).

Tabla 1.3. Los factores GATA que intervienen en la regulación de la transcripción de los genes implicados en el metabolismo del nitrógeno. (*NCR*: represión catabólica por la fuente de nitrógeno).

GEN	FACTOR	REGULACION y CARACTERISTICAS	Referencia
<i>GLN3</i>	Gln3p	Activación de los genes sensibles a la <i>NCR</i> para la utilización de fuentes de nitrógeno que no son óptimas.	Stanbrough <i>et al.</i> , 1996
<i>GAT1/NIL1</i>	Gat1p		Coffman <i>et al.</i> , 1996
<i>DAL80/UGA43</i>	Dal80p	Regulación negativa de los genes sensibles a <i>NCR</i> . Dal80p (activo cuando el nitrógeno es limitante) está fuertemente reprimido y Deh1p activado cuando la fuente de nitrógeno es la glutamina (exceso de nitrógeno).	Coffman <i>et al.</i> , 1997
<i>DAL80/DEH1</i> (YJL110c)	Deh1p		

Intervienen además otros genes que, si bien no pertenecen a la familia de los factores GATA, sí participan en este tipo de regulación:

- *GAP1* codifica una permeasa general, de baja afinidad y alta capacidad, para aminoácidos utilizados como fuente de nitrógeno. Este gen, es reprimido en cultivos con fuente rica de nitrógeno y

activado, cuando la fuente de nitrógeno es pobre, por Gln3p y Gat1p/Nil1p (Stanbrough y Magasanik, 1996).

- *URE2*, cuyo producto inhabilita post-transcripcionalmente al factor Gln3p es, por consiguiente, responsable de la baja expresión de varios genes inutilizados en fuentes de nitrógeno óptimas. Se ha postulado que Ure2p no sólo actuaría sobre Gln3p, sino también sobre los otros factores GATA (Blinder *et al.*, 1996; Coffman *et al.*, 1997).

1.4.4. Regulación génica en la fase estacionaria:

Cuando llega la fase estacionaria, los nutrientes del medio de cultivo se hacen limitantes para el crecimiento, la mayoría de las células se encuentran en la fase G₀ del ciclo celular y no se observan divisiones; por otro lado, las células de levadura sufren cambios fisiológicos, bioquímicos y morfológicos al llegar a esta fase de cultivo (Werner-Washburne *et al.*, 1993). A pesar de que los niveles celulares de mRNA (Sogin y Saunders, 1980) y la tasa de síntesis proteica se reducen significativamente (Fuge *et al.*, 1994), las células siguen produciendo y transformando varios compuestos, demostrando así que el nivel de transcripción y/o la estabilidad de los mRNA aumenta para ciertos genes durante la fase estacionaria (Werner-Washburne *et al.*, 1996).

Se sabe que numerosos genes (*CTT1*, *HSP104*, *HSP12*, *SOD2*, *GSY2*, *etc.*) inducidos durante la fase estacionaria, contienen elementos STRE (*Stress Response Element*) situados en las zonas promotoras y formados por la secuencia consenso AGGGG (Schüller *et al.*, 1994). La inducción transcripcional mediada por estos elementos está relacionada con la baja actividad de la protein-quinasa A (PKA), dependiente de los niveles de AMPc intracelular (Schüller *et al.*, 1994; Flattery-O'Brien *et al.*, 1997).

Los elementos STRE participan en la respuesta a diferentes tipos de estrés; así se observó que:

- Son activados por un amplio espectro de condiciones estresantes, como la privación de nitrógeno, la fase estacionaria del cultivo o el estrés osmótico y oxidativo (Marchler *et al.*, 1993; Flattery-O'Brien *et al.*, 1997).
- Controlan la transcripción de varios genes que codifican proteínas de estrés (Marchler *et al.*, 1993).
- Los productos de alguno de estos genes contribuyen a incrementar la resistencia de las células de levadura que se encuentran sometidas a estrés. Además, unas condiciones estresantes particulares, que activan los STRE, también proporcionan protección contra otros tipos de estrés. Se sabe, por ejemplo, que la privación en nutrientes esenciales provoca una transición de la célula a un estado que la hace más resistente, no solamente a la falta del nutriente sino también a un choque térmico severo (Pringle y Hartwell, 1981).

Un elemento de control como el STRE, capaz de integrar los efectos de diferentes señales de estrés y dirigir así una transcripción óptima, permite a la célula adaptarse a numerosas condiciones medioambientales desfavorables y por lo tanto constituye una gran ventaja selectiva.

Conjuntamente con las secuencias STRE, se han indentificado otros factores que podrían influir en la regulación transcripcional durante la fase estacionaria. Se conocen, por ejemplo, los elementos *PDS* (*Post-Diauxic Shift*), parecidos a las secuencias STRE, que actuarían en respuesta a los niveles de AMPc (Ruis y Schüller, 1995). Sin embargo, hasta el momento no

ha sido posible caracterizar ningún factor regulador específico relacionado con la inducción de la expresión génica en la fase estacionaria.

1.4.5. Choque osmótico:

El estrés provocado por el aumento de la osmolaridad externa reduce el crecimiento y la viabilidad de las células de levadura. Estas se adaptan rápidamente en el medio de cultivo al aporte de altas concentraciones de NaCl, expulsando el sodio intracelular y acumulando glicerol para aumentar su osmolaridad interna (Blomberg y Adler, 1989). La acumulación de glicerol resulta del aumento de la actividad de la glicerol-3-fosfato deshidrogenasa codificada por *GPD1* (Larson *et al.*, 1993) y de la expresión de dos genes adicionales, *HOG1* y *PBS2* que codifican protein-quinasas presentes en la ruta de traducción de señales de osmosensibilidad (Boguslawski, 1992; Brewster *et al.*, 1993) de la ruta HOG (*High Osmolarity Glycerol response*).

Si bien *S. cerevisiae* exhibe respuestas características a otros tipos de condiciones estresantes, los mecanismos y factores de traducción de las señales del estrés osmótico son prácticamente desconocidos. No obstante, se relacionó la presencia de secuencias STRE en los promotores de genes cuya transcripción se inducía por choque osmótico, y la ruta HOG. Se demostró así que, inmediatamente después de la fosforilación de Hog1p en respuesta a un incremento en la osmolaridad, (Brewster *et al.*, 1993), se activaba la transcripción mediante la unión específica de un factor de transcripción a los elementos STRE (Schüller *et al.*, 1994).

1.4.6. Choque térmico:

La respuesta al choque térmico es un sistema protector inducible en todas las células vivas. De las diferentes condiciones estresantes, el choque térmico ha sido quizás el más ampliamente estudiado (revisado por Mager y Moradas Ferreira, 1993). Un incremento de la temperatura (pero también otros tipos de estrés como, por ejemplo, altas concentraciones de etanol en el medio de cultivo o la llegada a la fase estacionaria) induce la acumulación de un conjunto reducido de proteínas de choque térmico o hsp (*heat shock proteins*). Esta acumulación se debe a la activación de la transcripción de los genes de choque térmico (*HSP*). Algunas de estas proteínas facilitan el crecimiento de las células a temperaturas supranormales y las habilitan así para sobrevivir a estas temperaturas haciéndolas termotolerantes (Piper, 1993).

Por otra parte, la síntesis de muchas proteínas (300 de 500 analizadas) decrece transitoriamente después de un choque térmico (Miller *et al.*, 1982) y varias proteínas inducibles térmicamente también se sintetizan en condiciones de crecimiento óptimo, como por ejemplo Hsp104p (Lindquist y Craig, 1988). Estas últimas podrían estar relacionadas con la resistencia de las células a breves, pero severos, choques térmicos que precisan de una respuesta inmediata.

Las temperaturas elevadas inducen la transcripción por lo menos mediante dos tipos diferentes de secuencias UAS:

- Secuencias HSE (*H*eat *S*hock *E*lements) formadas por tres o más repeticiones de GAA con una orientación alternada y separadas entre sí por dos nucleótidos (NTTCNNGAAN)₂ >

(NGAANNTTCN)₂ (Svetlov y Cooper, 1995). Estas secuencias unen específicamente el factor Hsfp (*Heat shock factor protein*), un activador transcripcional de respuesta al choque térmico capaz de unirse a DNA (Sorger, 1991). Se observó que, no solamente después de un choque térmico sino también durante el crecimiento exponencial, la secuencia HSE era necesaria para la expresión basal constitutiva de algunos genes inducibles térmicamente (Park y Craig, 1989). El análisis de mutantes *hsf* mostró que Hsfp poseía por lo menos dos dominios de activación cuya actividad estaba reprimida en ausencia de choque térmico (Nieto-Sotelo *et al.*, 1990).

- Secuencias STRE, no relacionadas con las secuencias HSE, pueden ser suficientes para la activación del gen CTT1 (que codifica la catalasa T) en respuesta a un choque térmico (Marchler *et al.*, 1993). Sin embargo, como ya se ha comentado, este elemento está también relacionado con respuestas a otros tipos de estrés (ver los apartados 1.4.5. y 1.4.6.) y podría estar relacionado con una regulación de tipo global por parte de las células, llevándolas a un mayor grado de resistencia bajo condiciones adversas.

1.4.7. Regulación por oxígeno:

S. cerevisiae es un microorganismo anaerobio facultativo y puede crecer fácilmente en anaerobiosis si el medio de cultivo aporta los metabolitos que sólo se sintetizan en aerobiosis (*i.e.* ergosterol). Durante la adaptación al crecimiento anaerobio, cuando los niveles de oxígeno son bajos (hipoxia), la levadura expresa genes, denominados genes hipóxicos, que son homólogos a otros genes de alta expresión en aerobiosis y cuya regulación es hemo-dependiente (Zitomer *et al.*, 1997).

Por otro lado la glucosa, fuente de carbono preferida por *S. cerevisiae*, induce la transcripción de varios genes relacionados con la fermentación y reprime la expresión de los genes utilizados en la respiración y la utilización de otras fuentes de carbono (ver apartado 1.4.2.). Estos últimos, sin embargo, son activados por otras fuentes de carbono; así, el etanol y/o el glicerol inducen la expresión de los genes que codifican las enzimas mitocondriales implicadas en el ciclo de Krebs y en la cadena respiratoria. Al mismo tiempo, la capacidad respiratoria de las mitocondrias depende de la presencia de oxígeno molecular como aceptor final de la cadena de transporte de electrones. Los componentes de ésta son mayoritariamente codificados por genes, a su vez regulados a nivel transcripcional por oxígeno (Zitomer y Lowry, 1992).

La señal intracelular de los niveles de oxígeno está mediada por hemo y los mecanismos que permiten la activación transcripcional implican la intervención de las proteínas reguladoras específicas HAP (*Heme Activator Protein*) entre las que se encuentran Hap1p y el complejo heteromérico Hap2/3/4/5p. El factor Hap1p se une a sitios UAS caracterizados por la secuencia CGGNNNTANCGG (Ha *et al.*, 1996; Zhang *et al.*, 1998), y Hap2/3/4/5p, que también interviene en la regulación por fuentes de carbono y de nitrógeno (Dang *et al.*, 1996), es capaz de reconocer cajas CCAAT situadas en los promotores de los genes regulados por hemo (Xing *et al.*, 1993).

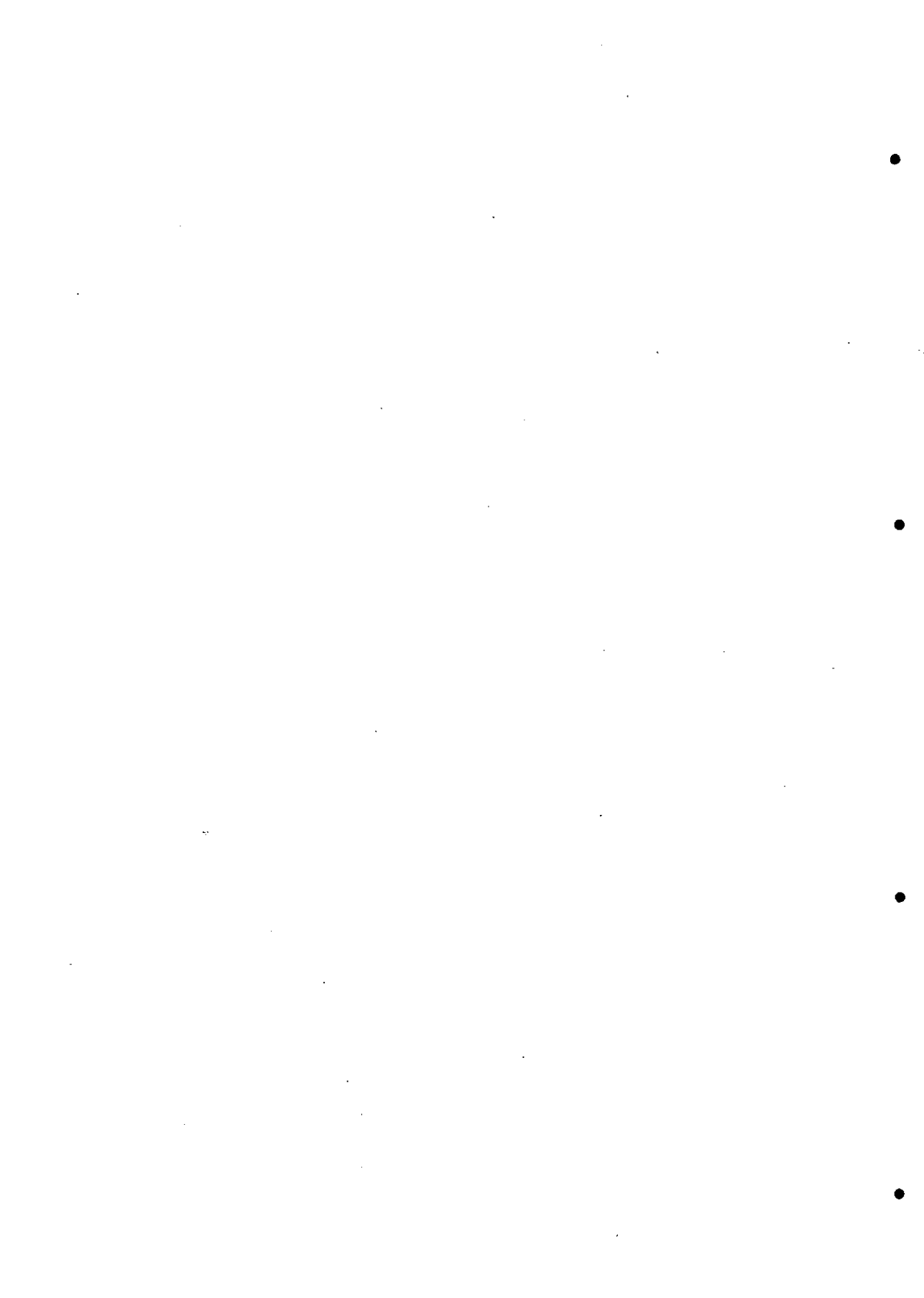
La proteína Rox1p presenta una función represora específica en la regulación por hemo. Es un represor transcripcional de genes hipóxicos y de genes anaeróbicos en condiciones aerobias (Lowry y Zitomer, 1984). Rox1p se encuentra regulado a su vez por Hap1p (Keng, 1992), y actúa junto con el complejo corepresor Tup1p/Ssn6p (Tzamarias y Struhl, 1995).

La proteína Rox6p se caracteriza porque su mutación provoca un incremento de la expresión del gen hipóxico *CYC7* y esta mutación es complementada por Srb10p (Cadahía, Tesis de licenciatura, 1996).

Podemos concluir que *S. cerevisiae* es un organismo capaz de adaptarse a un amplio abanico de condiciones del medio, gracias al papel fundamental desempeñado por la regulación de la transcripción génica. Mediante ésta, y en función de las condiciones externas, se activarán o inhibirán muchos de los cerca de 6.000 genes que componen el genoma de la levadura. El estudio de la regulación transcripcional de las ORFs de función desconocida es un primer nivel de conocimiento de su función en la célula.



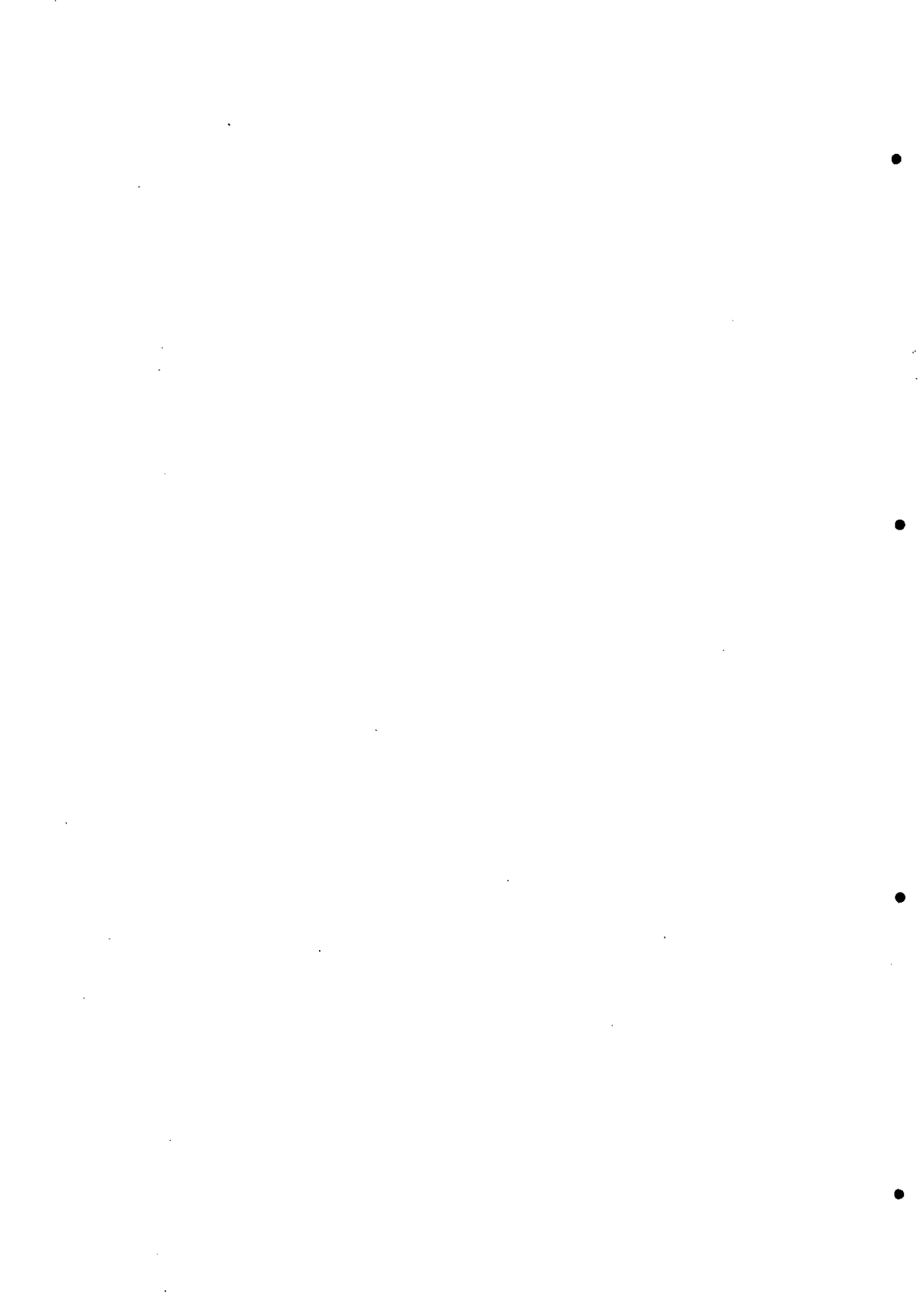
2. OBJETIVOS



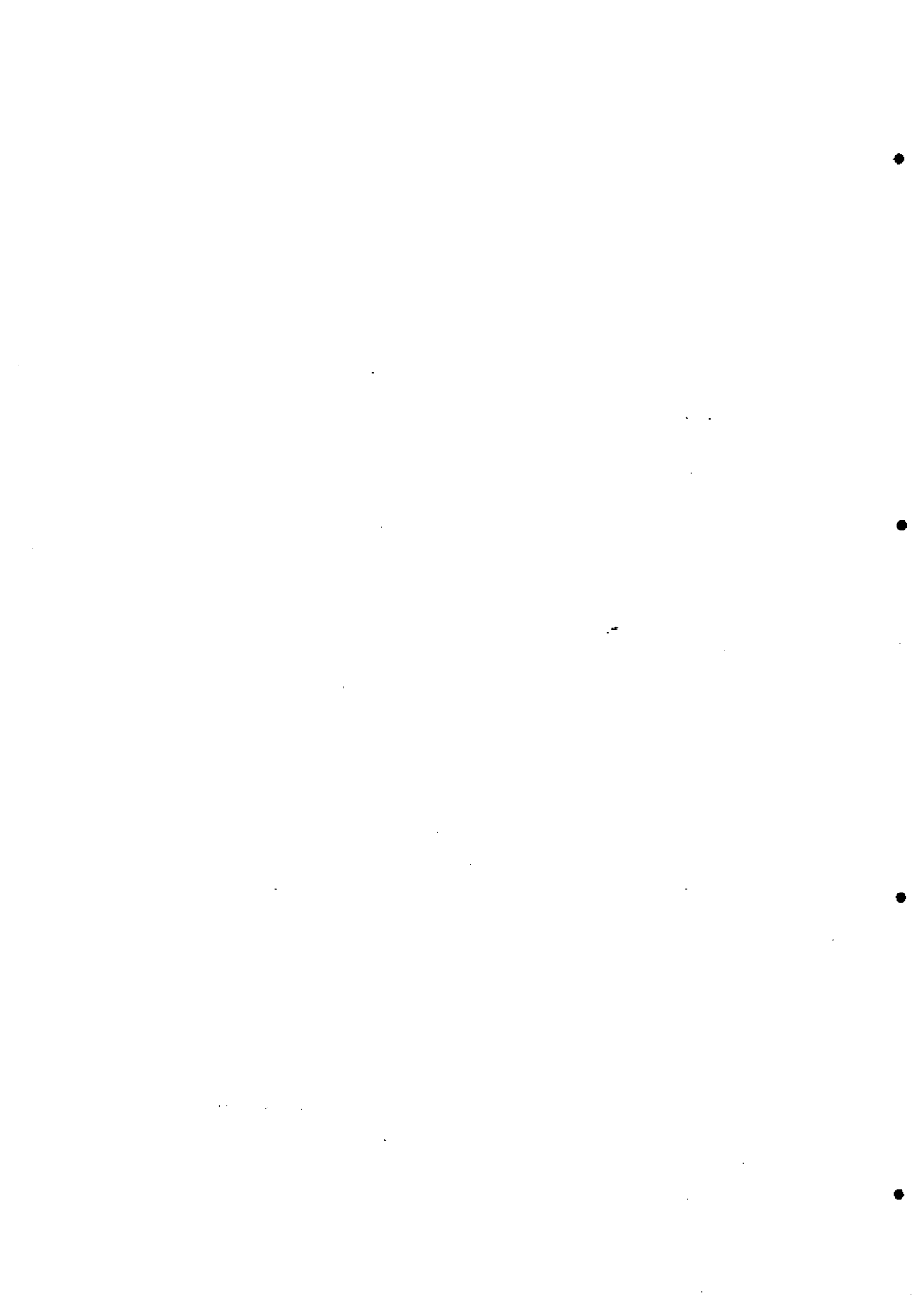
En este trabajo se analiza la regulación de la expresión transcripcional de una serie de ORFs de *S. cerevisiae* de función desconocida situadas en los cromosomas VII y XIV.

Este estudio tiene como objetivos:

- Comprobar si las ORFs, determinadas mediante análisis “*in silico*” de la secuencia, se transcriben.
- Analizar la regulación transcripcional de la expresión de estas ORFs ante diversas condiciones en el medio de cultivo.
- Relacionar la regulación de dichas ORFs con su posible función y con la presencia de elementos reguladores específicos en la región promotora.



3. MATERIALES Y MÉTODOS



3.1. Cepas celulares:

3.1.1. Levaduras:

Las extracciones de RNA se realizaron a partir de las cepas de *S. cerevisiae* cuyas características se resumen en la tabla 3.1.

Tabla 3.1. Cepas de levadura utilizadas en las extracciones de RNA.

Nombre de la cepa	Genotipo	Características/Referencias
FY73	<i>MATα, ura3-52, his3Δ200</i>	Es el progenitor α de la cepa FY1679 que se utilizó en el proyecto de secuenciación (Winston <i>et al.</i> , 1995)
aGH1- <i>rox1::LEU2</i>	<i>MATα, trp1-289, leu2-3, leu2-112, rox1::LEU2</i>	Balasubramanian <i>et al.</i> , 1993
α -LRI-217 (<i>rox6</i>)	<i>MATα-LRI, trp1-289, his3-532, gal1Δ152, rox6-217</i>	Roseblum-Vos, 1988
LPY22 <i>hap1::LEU2</i>	<i>MATα, adel-100, his4-519, leu2-3, leu2-112, ura3-52, Δhap1::LEU2, CYC1ΔUAS2</i>	Pfiefer, 1988

3.1.2. Bacterias:

Para los procesos de amplificación de DNA se utilizaron dos cepas bacterianas de laboratorio (Tabla 3.2.).

Tabla 3.2. Cepas bacterianas usadas.

Nombre de la cepa	Organismo	Genotipo	Referencias
HB101	<i>Escherichia coli</i>	<i>supE44, hsdS20 (rB,mB), recA13, ara-14, proA2, lacY1, galK2, rpsL20, xyl-5, mtl-1</i>	Bólvivar y Backman, 1979
JM101	<i>Escherichia coli</i>	<i>supE, thiΔ(lac-proAB), F'[traΔ36proAB⁺, lacP⁺lacZAM15]</i>	Messing, 1979

3.2. Medios y condiciones de cultivo:

3.2.1. Composición de los medios de cultivo:

3.2.1.1. Crecimiento de levaduras:

Los medios de cultivo, preparados con agua destilada, fueron esterilizados en autoclave por espacio de 20 minutos a 121 °C y 2 Bares de presión atmosférica. Sus composiciones son las siguientes:

- YPGE (medio rico con etanol/glicerol):

- Peptona bacteriológica 10 g/l
- Extracto de levadura 10 g/l
- Glicerol 20 g/l
- Etanol 10 g/l

- YPD (medio rico con Glucosa):

- Peptona bacteriológica 10 g/l
- Extracto de levadura 10 g/l
- Glucosa 20 g/l

- YNBglu (medio mínimo con Glucosa)

- Glucosa 20 g/l
- Uracilo 20 mg/l⁽¹⁾
- Histidina 20 mg/l⁽¹⁾
- YNB 6,7 g/l

(1) El medio mínimo se suplementó con los aminoácidos para los cuales la cepa de *S. cerevisiae* utilizada era auxotrofa.

El YNB (*Yeast Nitrogen Base*) se añade después de la esterilización, cuando la temperatura del matraz alcanza unos 60°C. La composición de un litro de YNB se describe a continuación:

- Vitaminas (300x)	50,0 ml
- Sales traza (150x)	100,0 ml
- KH ₂ PO ₄	15,0 g
- MgSO ₄	7,5 g
- NaCl	1,5 g
- CaCl ₂	1,5 g
- SO ₄ (NH ₄) ₂ ⁽¹⁾	75,0 g

Vitaminas 300x: (para 1 litro)	- Biotina	0,6 mg
	- Pantotenato cálcico	120,0 mg
	- Ácido fólico	0,6 mg
	- Inositol	600,0 mg
	- Niacina	120,0 mg
	- Ácido p-Aminobenzóico	60,0 mg
	- Piridoxina	120,0 mg
	- Riboflavina	60,0 mg
- Tiamina	120,0 mg	

Sales traza 150x: (para 1 litro)	- Ácido bórico	75,0 mg
	- Sulfato cúprico	6,0 mg
	- Ioduro potásico	15,0 mg
	- Cloruro férrico	30,0 mg
	- Molibdato sódico	30,0 mg
	- Sulfato de zinc	60,0 mg

⁽¹⁾En la condición de cultivo nº5 (privación de nitrógeno) se preparó una solución de YNB que no contenía sulfato amónico.

3.2.1.2. Crecimiento de bacterias:

Para seleccionar las bacterias que habían sido transformadas con un plásmido que contiene el gen de resistencia a la ampicilina, se utilizó el medio LB (Luria-Bertani) suplementado con el antibiótico (Sambrook *et al.*, 1989). Su composición es:

- Triptona bacteriológica 1%
- Extracto de levadura 0,5%
- Cloruro sódico 0,5%
- D-glucosa 0,1%

Una vez esterilizada la solución, cuando su temperatura alcanzaba los 60°C se añadía el antibiótico a una concentración de 40-80 µg/ml. Para la conservación de las bacterias en medio sólido se añadió a la solución 1,5% de agar bacteriológico y se vertió en placas *Petri*.

3.2.2. Condiciones de cultivo:

En las 11 condiciones de estudio utilizadas se partió de pre-cultivos en fase de crecimiento exponencial. Con ellos se inocularon los correspondientes medios de cultivo (1:200) y cuando en estos se alcanzó una $A_{600} = 0,8$ (3×10^7 células/ml), se centrifugaron los medios de cultivo en botellas de 125 ml a 5.000 rpm y a 4°C durante 5 min. El sedimento celular se mantuvo en frío para evitar la degradación del RNA.

Las diferentes condiciones de cultivo se establecieron como se indica a continuación. Se cultivaron las células FY73 en YPGE a 30°C hasta que $A_{600}=0,8$ y se recogieron mediante centrifugación. La tercera parte de las

células constituyeron la **muestra 1**. Los 2/3 restantes se lavaron en agua bidestilada estéril y se resuspendieron en YPD. Se dejó crecer el cultivo agitándolo a 30°C y después de una hora se recogió la mitad de las células (**muestra 2**) continuando la otra mitad del cultivo durante 23 horas más a 30°C hasta llegar a la fase estacionaria ($>2 \times 10^8$ células/ml cultivo) (**muestra 3**).

Paralelamente se inició otro cultivo de células FY73 en YNBglu a 30°C hasta que $A_{600} = 0,8$. Un tercio de las células obtenidas constituyó la **muestra 4**; Otro tercio se precipitó, se lavó con agua bidestilada estéril y se resuspendió en dos volúmenes de YNBglu sin sulfato amónico, dejando las células privadas de fuente de nitrógeno durante dos horas a 30°C (**muestra 5**). Al volumen restante del cultivo inicial se le añadió una solución estéril de cloruro sódico 5M hasta obtener una concentración final de 0,7M de NaCl y se incubó durante una hora a 30°C (**muestra 6**).

También se inició otro cultivo de células FY73 en YNBglu a 23°C hasta $A_{600} = 0,8$. La mitad de las células obtenidas constituyó la **muestra 7** mientras que la **muestra 8** resultó de la incubación de la mitad restante a 36°C durante 30 minutos en un baño con agitación.

También se realizaron cultivos de tres cepas mutantes de *S. cerevisiae* en las que se encuentran mutados o deletionados diversos factores que afectan la regulación transcripcional mediada por oxígeno (*aGH1-rox1::LEU2*, α -LRI-217*rox6* y LPY22*hap1::LEU2*); los cultivos se realizaron en YPGE a 30°C hasta $A_{600}=0,8$; así se obtuvieron, las muestras A, B y C respectivamente.

Las diferentes condiciones de cultivo se pueden resumir en el esquema de la figura 3.1.

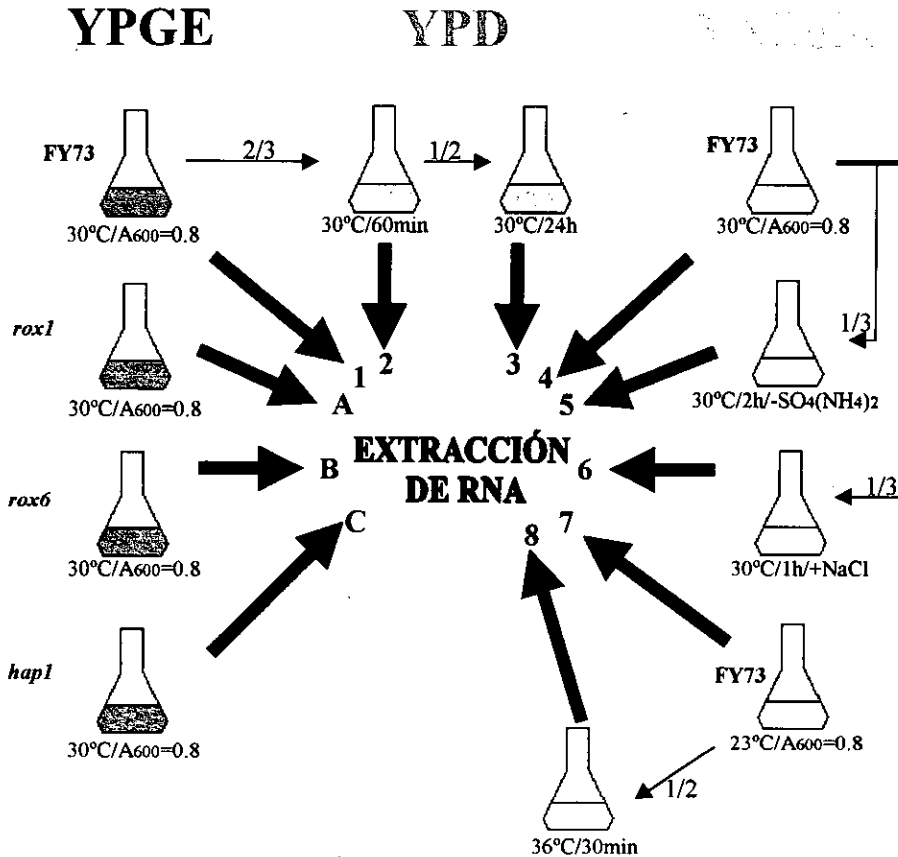


Figura 3.1. Esquema de las 11 condiciones de crecimiento a las cuales fueron sometidas las células de levadura para la extracción posterior de su RNA.

3.3. Extracción del RNA:

3.3.1. Condiciones de extracción:

La presencia de RNAsa disminuye el rendimiento de la extracción; para prevenir este problema se tomaron las siguientes precauciones :

- El agua y las disoluciones salinas usadas se trataron con dietilpircarbonato (DEPC) al 0,1 % (v/v), se agitaron toda la noche a 37 °C y se esterilizaron en autoclave al día siguiente para eliminar el DEPC remanente. Alguna solución, como el tampón Tris⁽¹⁾-EDTA⁽²⁾, no se trató con DEPC ya que éste modificaría covalentemente los grupos amino que contiene. Se preparó directamente usando agua tratada y reactivos libres de RNAsas. El SDS⁽³⁾, que no puede ser autoclavado, se preparó de igual modo.
- El material de vidrio se esterilizó durante una noche en un horno a 180°C.
- El material de plástico se enjuagó en cloroformo aclarándolo posteriormente con agua tratada.
- Se usaron guantes en las operaciones de extracción de RNA, ya que las manos son la mayor fuente de RNAsas contaminantes.
- Todas las operaciones de extracción se realizaron a 4°C.

⁽¹⁾ Tris-HCl: 2-Amino-2(hidroxiometil)-1-3-propanodiol equilibrado con ácido clorhídrico.

⁽²⁾ EDTA: Ácido etilenodiaminotetraacético.

⁽³⁾ SDS: Dodecil Sulfato Sódico.

3.3.2. El protocolo de extracción:

La extracción de RNA se realizó siguiendo un protocolo común acordado por todos los laboratorios participantes en el consorcio B2 (Lindquist, 1981; Brown, 1995). Las cantidades señaladas se corresponden con un volumen de cultivo inicial de 400 ml.

- a) Se sedimentaron las células, centrifugando a 5.000 rpm durante 5 minutos a 4°C y eliminando el sobrenadante.
- b) Se lavó el sedimento en 3,5 ml de tampón de extracción (LiCl 0,1M, DTT 0,01 M, Tris-HCl 0,1 M pH 7,5), se centrifugó de nuevo y se retiró el sobrenadante. En este paso las células pueden conservarse varios días en un congelador de -80°C sin que el RNA se vea afectado.
- c) Se resuspendió el precipitado en 3,5 ml de tampón de extracción e inmediatamente se añadió la suspensión celular a un tubo que contenía 10 g de perlas de vidrio (1 mm de diámetro), 0,5 ml de SDS 10 %, 5 ml de PCIA-1xRE⁽¹⁾ y 2 ml de cloroformo. Con el fin de romper las células, se agitó vigorosamente y de manera continua durante 5 minutos en un agitador *Vortex*. Se volvió a centrifugar durante 5 minutos a 5.000 rpm, recogiendo la fase acuosa.

(1) El PCIA-1xRE (*Phenol-Chloroform-Isoamyl Alcohol*) se obtiene mezclando un volumen de fenol fundido a 65°C con un volumen de cloroformo y dos volúmenes de tampón 1xRE (Tris 0.1M, LiCl 0.1M, EDTA 0.1 mM, β-mercaptoetanol 10 mM, pH 7.4). Se agita vigorosamente durante varios minutos y se deja en la oscuridad a temperatura ambiente hasta que se separe completamente la fase acuosa. Se retira ésta por aspiración y se repite el proceso dos o tres veces más. Tras el último equilibrado se añaden 1/25 volúmenes (8ml) de alcohol isoamílico y se conserva en la oscuridad a 4°C.

(2) DTT: ditiotreitól.

- d) Para eliminar las proteínas se hicieron dos extracciones con 7ml de PCIA-1xRE y dos lavados sólo con cloroformo (5ml), centrifugando cada vez a 5.000 rpm y recogiendo cuidadosamente la fase acuosa.
- e) A continuación se añadió 5 ml de etanol absoluto y acetato sódico (concentración final de 0,2 M) y se dejó precipitar el RNA a -20°C toda la noche. Al día siguiente se centrifugó a 10.000 rpm durante 10 minutos, se eliminó el sobrenadante y se secó al vacío durante 15 minutos para eliminar los residuos de etanol.
- f) Se realizó una nueva precipitación resuspendiendo el precipitado en 2 ml de tampón TE (Tris-HCl pH 7,5 10 mM, EDTA 1 mM) y mezclándolo con 140 μl de acetato sódico 3M y 4,5 ml de etanol absoluto. Se dejó toda la noche en el congelador a -20°C .
- g) Después de centrifugar, de eliminar el sobrenadante y de secar el precipitado al vacío, se eliminó el exceso de sales mediante lavados con etanol al 70%. Se precipitó, se secó y resuspendió en 200 μl de tampón TE para almacenarlo a -80°C .

3.3.3. Medida del rendimiento de la extracción:

Se tomaron 10 μl de la preparación de RNA y se llevaron a un volumen de 1 ml con agua tratada. La medida se realizó en un espectrofotómetro *Cary* (*Varian*) utilizando cubetas de cuarzo de 1 ml. Se midieron las absorbancias de cada muestra a 260 y 280 nm. La concentración real se calculó sabiendo que una $A_{260}=1$ corresponde a una concentración de 40 $\mu\text{g/ml}$ de RNA (una extracción correcta proporciona aproximadamente 100 μg de RNA total/ 2×10^8 células). La relación A_{260}/A_{280} indica la calidad de la extracción y debe aproximarse a 2,0 para ser óptima (Sambrook *et al.*, 1989).

3. 4. Northern blot:

3.4.1. El gel de agarosa:

Para preparar el gel de agarosa al 1,5% se pesaron 1,8 g de agarosa y se mezclaron con 6 ml de tampón 20xMOPS (ácido 3-[N-Morfolino]-propano-sulfónico 0,2M, acetato sódico 0,05 M pH7, EDTA 0,01M) y 94 ml de agua tratada con DEPC. Se disolvió la agarosa en un microondas y se dejó enfriar en un baño a 65 °C. Justo antes de verterla en la base de la electroforesis, se añadieron 20 ml de formaldehído (37% v/v) y tras verter, se dejó media hora en una campana de extracción para eliminar los vapores tóxicos del formaldehído.

3.4.2. Las muestras de RNA:

Se mezclaron 20 µg de RNA (en un volumen final de 7µl), 12,5 µl de formamida desionizada, 1,25 µl de Tampón 20xMOPS y 4 µl de formaldehído (37%). Se desnaturalizó la mezcla, incubando 5 minutos a 65°C, se le añadió 2,5 µl de tampón “azul de carga” (glicerol 50 % (v/v) + azul de bromofenol 0,25 % (w/v) + xileno cianol 0,25 (w/v)) y se mantuvo en hielo hasta cargar el gel.

3.4.3. La electroforesis:

La electroforesis se llevó a cabo en una cubeta de electroforesis (*HE100 Super SubTM Horizontal Unit de Hoefer*) que posee un circuito de refrigeración para evitar la degradación térmica del RNA y un sistema de

mezcla del tampón de electroforesis que permite una separación uniforme de las muestras.

Después de cargar las muestras, se conectó la fuente a 140 V durante unas 3 horas, hasta que el primer frente (azul de bromofenol) recorrió aproximadamente 10 centímetros. A continuación se lavó el gel en agua tratada para eliminar el formaldehído que puede interferir en la transferencia, se dejó en una solución de 5xSSC (NaCl 3M, citrato sódico 0,3M) durante 30 minutos y se transfirió a una membrana de nylon mediante vacío.

3.4.4. La transferencia:

En las transferencias, se utilizaron membranas de nylon (de porosidad 0,45 μm) con cargas positivas (*Boehringer Mannheim*) que se cortaron a medida y que, inmediatamente después de humedecerlas en agua tratada, se sumergieron en una disolución de 5xSSC antes de realizar la transferencia.

Se transfirió el RNA a la membrana utilizando una solución de 5xSSC a una presión de vacío de 50 a 60 mBAR durante una hora, se retiró el gel (comprobándose en el transiluminador que todo el RNA había sido transferido) y tras un lavado rápido de la membrana en 5xSSC para eliminar los fragmentos de agarosa y los excesos de sales, se fijó el RNA a la membrana con luz UV, aplicando mediante el dispositivo *UV Cross Linker 500* de *Hoefler*, una energía de 120mJ/cm² de membrana. Finalmente se etiquetó la membrana, se selló en una bolsa de plástico y se conservó en la nevera hasta la hibridación.

3.5. Obtención de sondas por PCR:

La PCR (*Polymerase Chain Reaction*) (Mullis y Faloona, 1987) es el método más efectivo y más rápido para amplificar un DNA diana que se va a utilizar en cantidades relativamente importantes y cuya especificidad es imprescindible.

3.5.1. Diseño de los cebadores:

La selección de los cebadores (*primers*) es particularmente importante para la especificidad de la amplificación por PCR:

- Su tamaño, generalmente próximo a 20 nucleótidos, debe permitir una hibridación con una secuencia única,
- los dos cebadores deben poseer similar T_m (*melting temperature* o temperatura de fusión):

$$T_m = (4\text{ }^\circ\text{C} \times n(\text{G} + \text{C})) + (2\text{ }^\circ\text{C} \times n(\text{A} + \text{T}))$$

Donde n representa el número de pares de bases (GC o AT). De modo que los cebadores tienen que tener un tamaño y un porcentaje de GC similar.

- los oligonucleótidos no pueden presentar ninguna complementariedad interna que origine estructuras secundarias, las cuales dificultarían la hibridación con el molde. Tampoco deben ser complementarios entre sí.

Una selección exacta, teniendo en cuenta todos estos factores, resulta pesada y complicada. Afortunadamente la informática alivia esta tarea; el programa informático *OLIGO* (*National Biosciences Inc. 725 Tower Drive. Harvard, MN 55340 U.S.A.*) permitió seleccionar oligonucleótidos cebadores

para realizar las reacciones de amplificación del DNA. El ordenador define automáticamente varias características independientes que hay que tener en cuenta en el momento de elegir el cebador. Estos factores son:

- La T_m que el programa determina con la fórmula:

$$T_m = \Delta H / \Delta S + R \ln (C/4)$$

(ΔH : variación de entalpía, ΔS : variación de entropía de formación del duplex, R : constante de los gases, C : concentración de la muestra),

- La posibilidad de que se formen horquillas en el DNA monohebra.
- La estabilidad del duplex molde/cebador.
- La autocomplementaridad del cebador.
- La presencia de regiones complementarias en los dos cebadores.

Un adecuado diseño de los cebadores asistido por ordenador fué imprescindible para asegurar que los productos obtenidos mediante PCR fueran únicos ya que el elevado tamaño de los moldes (DNA genómico o procedente de cósmidos con insertos de 30-40 Kb) y la pequeña longitud de algún oligonucleótido utilizado (13 a 16 nt) potenciaba la aparición de productos no adecuados y debidos a anillamientos inespecíficos.

Como un control adicional, la especificidad del producto obtenido mediante PCR se comprobó siempre mediante análisis de restricción; por tanto en el diseño de los cebadores se tuvo en cuenta que el producto amplificado incluyese un sitio de reconocimiento para una enzima de restricción.

En la tabla 3.3 se recopilan los datos del diseño de los cebadores y de las reacciones de comprobación de las 102 ORFs analizadas. La posición de

inicio y final de cada oligonucleótido en dirección 5'→3' es coincidente con su localización en las secuencia del cromosoma VII (ORFs denominadas G) y XIV (ORFs N). También incluye el diseño de cebadores de algunas sondas correspondientes a genes de regulación conocida utilizados como controles.

Tabla 3.3. Diseño de los cebadores para las reacciones de amplificación y selección de las enzimas utilizadas en la comprobación de cada producto obtenido. Tm: temperatura de *melting*, L.P.: Longitud del producto de PCR; L: longitud del cebador.

Nº ORF	Nombre del Cebador	Posición del Cebador Inicio final	SECUENCIA	Tm (°C)	L (nt)	L.P. (pb)	Enzima	Longitud Fragmentos (pb)																																																																																																																																																																																																														
ACT1	ACT-5'	51151 51138	TGGATTCGGGTGAT	36,3	14	250	<i>BglII</i>	115+135																																																																																																																																																																																																														
	ACT-3'	50902 50916	ATFGAAGAAGATTGA	25,1	15				PCK1	PCK-5'	631283 631295	CATCAAGCGGTGC	35,1	13	270	<i>BglII</i>	45+225	PCK-3'	631552 631538	AAGCGTGGTAAGCCC	42,7	15	RPL25	RPL-5'	81615 81627	AGTTGGCTAGAGC	24,3	13	301	<i>AhlI</i>	153+168	RPL-3'	81915 81901	GATTAAATGTAACCG	27,4	15	HSP12	HPSK-5'	107250 107267	CAATGTCTGACGCAGGTA	44,0	18	327	<i>AluI</i>	230+97	HPSK-3'	107576 107559	CTTCTGGTTGGGTCTTC	43,6	18	ARG4	ARG4-T1	139842 139826	AACGGAGACAGAAATGG	50,0	17	271	<i>NdeII</i>	149+89+21+12	ARG4-T2c	139572 139588	AGAACTCAACCAAGCC	50,0	17	CARI	CARI-T1	339977 339993	TCGCGAATTGTCCATCG	52,0	17	268	<i>HaeIII</i>	117+86+65	CARI-T2c	340244 340228	AACTTGGTGGCTTACC	52,0	17	G2850	2850-5'	263774 263759	AAATCGCCTGCCATA	46,4	16	396	<i>KpnI</i>	307+89	2850-3'	263379 263394	CGTCTCCGGGTGAAAC	45,6	16	G2856	2856-5'	268630 268615	GAGAGCCAGGGTTCGG	50,4	16	356	<i>HindIII</i>	185+140+30	2856-3'	268275 268290	CGATTATGCCGGTGGG	49,8	16	G2861	2861-3'	270185 270170	GCAAGTCCATGCGCGT	49,3	16	293	<i>DraI</i>	163+130	2861-3'	269893 269908	AGCGGTGTAATGGGG	48,4	16	G2882	2882-5'	272898 272913	TCCCCTGAAGGTGGC	49,1	16	306	<i>HindIII</i>	176+130	2882-3'	273203 273188	CTCAAGAAGGCCGCGT	48,9	16	G2889	2889-5'	275637 275622	GCTTGATAAATGCGCC	44,0	16	441	<i>KpnI</i>	302+139	2889-3'	275197 275212	CGAAGTCCGGCATAAC	44,6	16	G2913	2913-5'	281119 281104	GGGGCTGAACCAACTT	44,3	16	259	<i>PvuI</i>	164+95	2913-3'	280861 280876	CGTCGATGTCCCAGTC	42,9	16	G2916	2916-5'	283298 283278	TGGATGAAGCGCACGAACGTA	59,8	21	551	<i>AluI</i>	367+158+25	2916-3'	282748 282769	TGAAAACTCTGGGTCAACGACG	60,3	22	G3080	3080-5'	306606 306624	AAGGGTCAAGGCCCATTTG	58,8	19	325	<i>Avall</i>	190+134	3080-3'	306930 306909	GCTCGTCTACTTCAGCGTGGT	64,0	22	G3090	3090-5'	309669 309648	AGGAAGTTGCTCCAAGCGGTTG	62,1	22	276	<i>AluI</i>	137+109+29	3090-3'	309394 309415	CGGCTTCATCAAAAGTCACGCC	62,1	22	G3107	3107-5'	312396 312411	CCATGCTGATTACGGA	49,2	16	429	<i>PstI</i>	261+177	3107-3'
PCK1	PCK-5'	631283 631295	CATCAAGCGGTGC	35,1	13	270	<i>BglII</i>	45+225																																																																																																																																																																																																														
	PCK-3'	631552 631538	AAGCGTGGTAAGCCC	42,7	15				RPL25	RPL-5'	81615 81627	AGTTGGCTAGAGC	24,3	13	301	<i>AhlI</i>	153+168	RPL-3'	81915 81901	GATTAAATGTAACCG	27,4	15	HSP12	HPSK-5'	107250 107267	CAATGTCTGACGCAGGTA	44,0	18	327	<i>AluI</i>	230+97	HPSK-3'	107576 107559	CTTCTGGTTGGGTCTTC	43,6	18	ARG4	ARG4-T1	139842 139826	AACGGAGACAGAAATGG	50,0	17	271	<i>NdeII</i>	149+89+21+12	ARG4-T2c	139572 139588	AGAACTCAACCAAGCC	50,0	17	CARI	CARI-T1	339977 339993	TCGCGAATTGTCCATCG	52,0	17	268	<i>HaeIII</i>	117+86+65	CARI-T2c	340244 340228	AACTTGGTGGCTTACC	52,0	17	G2850	2850-5'	263774 263759	AAATCGCCTGCCATA	46,4	16	396	<i>KpnI</i>	307+89	2850-3'	263379 263394	CGTCTCCGGGTGAAAC	45,6	16	G2856	2856-5'	268630 268615	GAGAGCCAGGGTTCGG	50,4	16	356	<i>HindIII</i>	185+140+30	2856-3'	268275 268290	CGATTATGCCGGTGGG	49,8	16	G2861	2861-3'	270185 270170	GCAAGTCCATGCGCGT	49,3	16	293	<i>DraI</i>	163+130	2861-3'	269893 269908	AGCGGTGTAATGGGG	48,4	16	G2882	2882-5'	272898 272913	TCCCCTGAAGGTGGC	49,1	16	306	<i>HindIII</i>	176+130	2882-3'	273203 273188	CTCAAGAAGGCCGCGT	48,9	16	G2889	2889-5'	275637 275622	GCTTGATAAATGCGCC	44,0	16	441	<i>KpnI</i>	302+139	2889-3'	275197 275212	CGAAGTCCGGCATAAC	44,6	16	G2913	2913-5'	281119 281104	GGGGCTGAACCAACTT	44,3	16	259	<i>PvuI</i>	164+95	2913-3'	280861 280876	CGTCGATGTCCCAGTC	42,9	16	G2916	2916-5'	283298 283278	TGGATGAAGCGCACGAACGTA	59,8	21	551	<i>AluI</i>	367+158+25	2916-3'	282748 282769	TGAAAACTCTGGGTCAACGACG	60,3	22	G3080	3080-5'	306606 306624	AAGGGTCAAGGCCCATTTG	58,8	19	325	<i>Avall</i>	190+134	3080-3'	306930 306909	GCTCGTCTACTTCAGCGTGGT	64,0	22	G3090	3090-5'	309669 309648	AGGAAGTTGCTCCAAGCGGTTG	62,1	22	276	<i>AluI</i>	137+109+29	3090-3'	309394 309415	CGGCTTCATCAAAAGTCACGCC	62,1	22	G3107	3107-5'	312396 312411	CCATGCTGATTACGGA	49,2	16	429	<i>PstI</i>	261+177	3107-3'	312824 312809	TGTCGAAAAAGGCTTG	46,6	16										
RPL25	RPL-5'	81615 81627	AGTTGGCTAGAGC	24,3	13	301	<i>AhlI</i>	153+168																																																																																																																																																																																																														
	RPL-3'	81915 81901	GATTAAATGTAACCG	27,4	15				HSP12	HPSK-5'	107250 107267	CAATGTCTGACGCAGGTA	44,0	18	327	<i>AluI</i>	230+97	HPSK-3'	107576 107559	CTTCTGGTTGGGTCTTC	43,6	18	ARG4	ARG4-T1	139842 139826	AACGGAGACAGAAATGG	50,0	17	271	<i>NdeII</i>	149+89+21+12	ARG4-T2c	139572 139588	AGAACTCAACCAAGCC	50,0	17	CARI	CARI-T1	339977 339993	TCGCGAATTGTCCATCG	52,0	17	268	<i>HaeIII</i>	117+86+65	CARI-T2c	340244 340228	AACTTGGTGGCTTACC	52,0	17	G2850	2850-5'	263774 263759	AAATCGCCTGCCATA	46,4	16	396	<i>KpnI</i>	307+89	2850-3'	263379 263394	CGTCTCCGGGTGAAAC	45,6	16	G2856	2856-5'	268630 268615	GAGAGCCAGGGTTCGG	50,4	16	356	<i>HindIII</i>	185+140+30	2856-3'	268275 268290	CGATTATGCCGGTGGG	49,8	16	G2861	2861-3'	270185 270170	GCAAGTCCATGCGCGT	49,3	16	293	<i>DraI</i>	163+130	2861-3'	269893 269908	AGCGGTGTAATGGGG	48,4	16	G2882	2882-5'	272898 272913	TCCCCTGAAGGTGGC	49,1	16	306	<i>HindIII</i>	176+130	2882-3'	273203 273188	CTCAAGAAGGCCGCGT	48,9	16	G2889	2889-5'	275637 275622	GCTTGATAAATGCGCC	44,0	16	441	<i>KpnI</i>	302+139	2889-3'	275197 275212	CGAAGTCCGGCATAAC	44,6	16	G2913	2913-5'	281119 281104	GGGGCTGAACCAACTT	44,3	16	259	<i>PvuI</i>	164+95	2913-3'	280861 280876	CGTCGATGTCCCAGTC	42,9	16	G2916	2916-5'	283298 283278	TGGATGAAGCGCACGAACGTA	59,8	21	551	<i>AluI</i>	367+158+25	2916-3'	282748 282769	TGAAAACTCTGGGTCAACGACG	60,3	22	G3080	3080-5'	306606 306624	AAGGGTCAAGGCCCATTTG	58,8	19	325	<i>Avall</i>	190+134	3080-3'	306930 306909	GCTCGTCTACTTCAGCGTGGT	64,0	22	G3090	3090-5'	309669 309648	AGGAAGTTGCTCCAAGCGGTTG	62,1	22	276	<i>AluI</i>	137+109+29	3090-3'	309394 309415	CGGCTTCATCAAAAGTCACGCC	62,1	22	G3107	3107-5'	312396 312411	CCATGCTGATTACGGA	49,2	16	429	<i>PstI</i>	261+177	3107-3'	312824 312809	TGTCGAAAAAGGCTTG	46,6	16																								
HSP12	HPSK-5'	107250 107267	CAATGTCTGACGCAGGTA	44,0	18	327	<i>AluI</i>	230+97																																																																																																																																																																																																														
	HPSK-3'	107576 107559	CTTCTGGTTGGGTCTTC	43,6	18				ARG4	ARG4-T1	139842 139826	AACGGAGACAGAAATGG	50,0	17	271	<i>NdeII</i>	149+89+21+12	ARG4-T2c	139572 139588	AGAACTCAACCAAGCC	50,0	17	CARI	CARI-T1	339977 339993	TCGCGAATTGTCCATCG	52,0	17	268	<i>HaeIII</i>	117+86+65	CARI-T2c	340244 340228	AACTTGGTGGCTTACC	52,0	17	G2850	2850-5'	263774 263759	AAATCGCCTGCCATA	46,4	16	396	<i>KpnI</i>	307+89	2850-3'	263379 263394	CGTCTCCGGGTGAAAC	45,6	16	G2856	2856-5'	268630 268615	GAGAGCCAGGGTTCGG	50,4	16	356	<i>HindIII</i>	185+140+30	2856-3'	268275 268290	CGATTATGCCGGTGGG	49,8	16	G2861	2861-3'	270185 270170	GCAAGTCCATGCGCGT	49,3	16	293	<i>DraI</i>	163+130	2861-3'	269893 269908	AGCGGTGTAATGGGG	48,4	16	G2882	2882-5'	272898 272913	TCCCCTGAAGGTGGC	49,1	16	306	<i>HindIII</i>	176+130	2882-3'	273203 273188	CTCAAGAAGGCCGCGT	48,9	16	G2889	2889-5'	275637 275622	GCTTGATAAATGCGCC	44,0	16	441	<i>KpnI</i>	302+139	2889-3'	275197 275212	CGAAGTCCGGCATAAC	44,6	16	G2913	2913-5'	281119 281104	GGGGCTGAACCAACTT	44,3	16	259	<i>PvuI</i>	164+95	2913-3'	280861 280876	CGTCGATGTCCCAGTC	42,9	16	G2916	2916-5'	283298 283278	TGGATGAAGCGCACGAACGTA	59,8	21	551	<i>AluI</i>	367+158+25	2916-3'	282748 282769	TGAAAACTCTGGGTCAACGACG	60,3	22	G3080	3080-5'	306606 306624	AAGGGTCAAGGCCCATTTG	58,8	19	325	<i>Avall</i>	190+134	3080-3'	306930 306909	GCTCGTCTACTTCAGCGTGGT	64,0	22	G3090	3090-5'	309669 309648	AGGAAGTTGCTCCAAGCGGTTG	62,1	22	276	<i>AluI</i>	137+109+29	3090-3'	309394 309415	CGGCTTCATCAAAAGTCACGCC	62,1	22	G3107	3107-5'	312396 312411	CCATGCTGATTACGGA	49,2	16	429	<i>PstI</i>	261+177	3107-3'	312824 312809	TGTCGAAAAAGGCTTG	46,6	16																																						
ARG4	ARG4-T1	139842 139826	AACGGAGACAGAAATGG	50,0	17	271	<i>NdeII</i>	149+89+21+12																																																																																																																																																																																																														
	ARG4-T2c	139572 139588	AGAACTCAACCAAGCC	50,0	17				CARI	CARI-T1	339977 339993	TCGCGAATTGTCCATCG	52,0	17	268	<i>HaeIII</i>	117+86+65	CARI-T2c	340244 340228	AACTTGGTGGCTTACC	52,0	17	G2850	2850-5'	263774 263759	AAATCGCCTGCCATA	46,4	16	396	<i>KpnI</i>	307+89	2850-3'	263379 263394	CGTCTCCGGGTGAAAC	45,6	16	G2856	2856-5'	268630 268615	GAGAGCCAGGGTTCGG	50,4	16	356	<i>HindIII</i>	185+140+30	2856-3'	268275 268290	CGATTATGCCGGTGGG	49,8	16	G2861	2861-3'	270185 270170	GCAAGTCCATGCGCGT	49,3	16	293	<i>DraI</i>	163+130	2861-3'	269893 269908	AGCGGTGTAATGGGG	48,4	16	G2882	2882-5'	272898 272913	TCCCCTGAAGGTGGC	49,1	16	306	<i>HindIII</i>	176+130	2882-3'	273203 273188	CTCAAGAAGGCCGCGT	48,9	16	G2889	2889-5'	275637 275622	GCTTGATAAATGCGCC	44,0	16	441	<i>KpnI</i>	302+139	2889-3'	275197 275212	CGAAGTCCGGCATAAC	44,6	16	G2913	2913-5'	281119 281104	GGGGCTGAACCAACTT	44,3	16	259	<i>PvuI</i>	164+95	2913-3'	280861 280876	CGTCGATGTCCCAGTC	42,9	16	G2916	2916-5'	283298 283278	TGGATGAAGCGCACGAACGTA	59,8	21	551	<i>AluI</i>	367+158+25	2916-3'	282748 282769	TGAAAACTCTGGGTCAACGACG	60,3	22	G3080	3080-5'	306606 306624	AAGGGTCAAGGCCCATTTG	58,8	19	325	<i>Avall</i>	190+134	3080-3'	306930 306909	GCTCGTCTACTTCAGCGTGGT	64,0	22	G3090	3090-5'	309669 309648	AGGAAGTTGCTCCAAGCGGTTG	62,1	22	276	<i>AluI</i>	137+109+29	3090-3'	309394 309415	CGGCTTCATCAAAAGTCACGCC	62,1	22	G3107	3107-5'	312396 312411	CCATGCTGATTACGGA	49,2	16	429	<i>PstI</i>	261+177	3107-3'	312824 312809	TGTCGAAAAAGGCTTG	46,6	16																																																				
CARI	CARI-T1	339977 339993	TCGCGAATTGTCCATCG	52,0	17	268	<i>HaeIII</i>	117+86+65																																																																																																																																																																																																														
	CARI-T2c	340244 340228	AACTTGGTGGCTTACC	52,0	17				G2850	2850-5'	263774 263759	AAATCGCCTGCCATA	46,4	16	396	<i>KpnI</i>	307+89	2850-3'	263379 263394	CGTCTCCGGGTGAAAC	45,6	16	G2856	2856-5'	268630 268615	GAGAGCCAGGGTTCGG	50,4	16	356	<i>HindIII</i>	185+140+30	2856-3'	268275 268290	CGATTATGCCGGTGGG	49,8	16	G2861	2861-3'	270185 270170	GCAAGTCCATGCGCGT	49,3	16	293	<i>DraI</i>	163+130	2861-3'	269893 269908	AGCGGTGTAATGGGG	48,4	16	G2882	2882-5'	272898 272913	TCCCCTGAAGGTGGC	49,1	16	306	<i>HindIII</i>	176+130	2882-3'	273203 273188	CTCAAGAAGGCCGCGT	48,9	16	G2889	2889-5'	275637 275622	GCTTGATAAATGCGCC	44,0	16	441	<i>KpnI</i>	302+139	2889-3'	275197 275212	CGAAGTCCGGCATAAC	44,6	16	G2913	2913-5'	281119 281104	GGGGCTGAACCAACTT	44,3	16	259	<i>PvuI</i>	164+95	2913-3'	280861 280876	CGTCGATGTCCCAGTC	42,9	16	G2916	2916-5'	283298 283278	TGGATGAAGCGCACGAACGTA	59,8	21	551	<i>AluI</i>	367+158+25	2916-3'	282748 282769	TGAAAACTCTGGGTCAACGACG	60,3	22	G3080	3080-5'	306606 306624	AAGGGTCAAGGCCCATTTG	58,8	19	325	<i>Avall</i>	190+134	3080-3'	306930 306909	GCTCGTCTACTTCAGCGTGGT	64,0	22	G3090	3090-5'	309669 309648	AGGAAGTTGCTCCAAGCGGTTG	62,1	22	276	<i>AluI</i>	137+109+29	3090-3'	309394 309415	CGGCTTCATCAAAAGTCACGCC	62,1	22	G3107	3107-5'	312396 312411	CCATGCTGATTACGGA	49,2	16	429	<i>PstI</i>	261+177	3107-3'	312824 312809	TGTCGAAAAAGGCTTG	46,6	16																																																																		
G2850	2850-5'	263774 263759	AAATCGCCTGCCATA	46,4	16	396	<i>KpnI</i>	307+89																																																																																																																																																																																																														
	2850-3'	263379 263394	CGTCTCCGGGTGAAAC	45,6	16				G2856	2856-5'	268630 268615	GAGAGCCAGGGTTCGG	50,4	16	356	<i>HindIII</i>	185+140+30	2856-3'	268275 268290	CGATTATGCCGGTGGG	49,8	16	G2861	2861-3'	270185 270170	GCAAGTCCATGCGCGT	49,3	16	293	<i>DraI</i>	163+130	2861-3'	269893 269908	AGCGGTGTAATGGGG	48,4	16	G2882	2882-5'	272898 272913	TCCCCTGAAGGTGGC	49,1	16	306	<i>HindIII</i>	176+130	2882-3'	273203 273188	CTCAAGAAGGCCGCGT	48,9	16	G2889	2889-5'	275637 275622	GCTTGATAAATGCGCC	44,0	16	441	<i>KpnI</i>	302+139	2889-3'	275197 275212	CGAAGTCCGGCATAAC	44,6	16	G2913	2913-5'	281119 281104	GGGGCTGAACCAACTT	44,3	16	259	<i>PvuI</i>	164+95	2913-3'	280861 280876	CGTCGATGTCCCAGTC	42,9	16	G2916	2916-5'	283298 283278	TGGATGAAGCGCACGAACGTA	59,8	21	551	<i>AluI</i>	367+158+25	2916-3'	282748 282769	TGAAAACTCTGGGTCAACGACG	60,3	22	G3080	3080-5'	306606 306624	AAGGGTCAAGGCCCATTTG	58,8	19	325	<i>Avall</i>	190+134	3080-3'	306930 306909	GCTCGTCTACTTCAGCGTGGT	64,0	22	G3090	3090-5'	309669 309648	AGGAAGTTGCTCCAAGCGGTTG	62,1	22	276	<i>AluI</i>	137+109+29	3090-3'	309394 309415	CGGCTTCATCAAAAGTCACGCC	62,1	22	G3107	3107-5'	312396 312411	CCATGCTGATTACGGA	49,2	16	429	<i>PstI</i>	261+177	3107-3'	312824 312809	TGTCGAAAAAGGCTTG	46,6	16																																																																																
G2856	2856-5'	268630 268615	GAGAGCCAGGGTTCGG	50,4	16	356	<i>HindIII</i>	185+140+30																																																																																																																																																																																																														
	2856-3'	268275 268290	CGATTATGCCGGTGGG	49,8	16				G2861	2861-3'	270185 270170	GCAAGTCCATGCGCGT	49,3	16	293	<i>DraI</i>	163+130	2861-3'	269893 269908	AGCGGTGTAATGGGG	48,4	16	G2882	2882-5'	272898 272913	TCCCCTGAAGGTGGC	49,1	16	306	<i>HindIII</i>	176+130	2882-3'	273203 273188	CTCAAGAAGGCCGCGT	48,9	16	G2889	2889-5'	275637 275622	GCTTGATAAATGCGCC	44,0	16	441	<i>KpnI</i>	302+139	2889-3'	275197 275212	CGAAGTCCGGCATAAC	44,6	16	G2913	2913-5'	281119 281104	GGGGCTGAACCAACTT	44,3	16	259	<i>PvuI</i>	164+95	2913-3'	280861 280876	CGTCGATGTCCCAGTC	42,9	16	G2916	2916-5'	283298 283278	TGGATGAAGCGCACGAACGTA	59,8	21	551	<i>AluI</i>	367+158+25	2916-3'	282748 282769	TGAAAACTCTGGGTCAACGACG	60,3	22	G3080	3080-5'	306606 306624	AAGGGTCAAGGCCCATTTG	58,8	19	325	<i>Avall</i>	190+134	3080-3'	306930 306909	GCTCGTCTACTTCAGCGTGGT	64,0	22	G3090	3090-5'	309669 309648	AGGAAGTTGCTCCAAGCGGTTG	62,1	22	276	<i>AluI</i>	137+109+29	3090-3'	309394 309415	CGGCTTCATCAAAAGTCACGCC	62,1	22	G3107	3107-5'	312396 312411	CCATGCTGATTACGGA	49,2	16	429	<i>PstI</i>	261+177	3107-3'	312824 312809	TGTCGAAAAAGGCTTG	46,6	16																																																																																														
G2861	2861-3'	270185 270170	GCAAGTCCATGCGCGT	49,3	16	293	<i>DraI</i>	163+130																																																																																																																																																																																																														
	2861-3'	269893 269908	AGCGGTGTAATGGGG	48,4	16				G2882	2882-5'	272898 272913	TCCCCTGAAGGTGGC	49,1	16	306	<i>HindIII</i>	176+130	2882-3'	273203 273188	CTCAAGAAGGCCGCGT	48,9	16	G2889	2889-5'	275637 275622	GCTTGATAAATGCGCC	44,0	16	441	<i>KpnI</i>	302+139	2889-3'	275197 275212	CGAAGTCCGGCATAAC	44,6	16	G2913	2913-5'	281119 281104	GGGGCTGAACCAACTT	44,3	16	259	<i>PvuI</i>	164+95	2913-3'	280861 280876	CGTCGATGTCCCAGTC	42,9	16	G2916	2916-5'	283298 283278	TGGATGAAGCGCACGAACGTA	59,8	21	551	<i>AluI</i>	367+158+25	2916-3'	282748 282769	TGAAAACTCTGGGTCAACGACG	60,3	22	G3080	3080-5'	306606 306624	AAGGGTCAAGGCCCATTTG	58,8	19	325	<i>Avall</i>	190+134	3080-3'	306930 306909	GCTCGTCTACTTCAGCGTGGT	64,0	22	G3090	3090-5'	309669 309648	AGGAAGTTGCTCCAAGCGGTTG	62,1	22	276	<i>AluI</i>	137+109+29	3090-3'	309394 309415	CGGCTTCATCAAAAGTCACGCC	62,1	22	G3107	3107-5'	312396 312411	CCATGCTGATTACGGA	49,2	16	429	<i>PstI</i>	261+177	3107-3'	312824 312809	TGTCGAAAAAGGCTTG	46,6	16																																																																																																												
G2882	2882-5'	272898 272913	TCCCCTGAAGGTGGC	49,1	16	306	<i>HindIII</i>	176+130																																																																																																																																																																																																														
	2882-3'	273203 273188	CTCAAGAAGGCCGCGT	48,9	16				G2889	2889-5'	275637 275622	GCTTGATAAATGCGCC	44,0	16	441	<i>KpnI</i>	302+139	2889-3'	275197 275212	CGAAGTCCGGCATAAC	44,6	16	G2913	2913-5'	281119 281104	GGGGCTGAACCAACTT	44,3	16	259	<i>PvuI</i>	164+95	2913-3'	280861 280876	CGTCGATGTCCCAGTC	42,9	16	G2916	2916-5'	283298 283278	TGGATGAAGCGCACGAACGTA	59,8	21	551	<i>AluI</i>	367+158+25	2916-3'	282748 282769	TGAAAACTCTGGGTCAACGACG	60,3	22	G3080	3080-5'	306606 306624	AAGGGTCAAGGCCCATTTG	58,8	19	325	<i>Avall</i>	190+134	3080-3'	306930 306909	GCTCGTCTACTTCAGCGTGGT	64,0	22	G3090	3090-5'	309669 309648	AGGAAGTTGCTCCAAGCGGTTG	62,1	22	276	<i>AluI</i>	137+109+29	3090-3'	309394 309415	CGGCTTCATCAAAAGTCACGCC	62,1	22	G3107	3107-5'	312396 312411	CCATGCTGATTACGGA	49,2	16	429	<i>PstI</i>	261+177	3107-3'	312824 312809	TGTCGAAAAAGGCTTG	46,6	16																																																																																																																										
G2889	2889-5'	275637 275622	GCTTGATAAATGCGCC	44,0	16	441	<i>KpnI</i>	302+139																																																																																																																																																																																																														
	2889-3'	275197 275212	CGAAGTCCGGCATAAC	44,6	16				G2913	2913-5'	281119 281104	GGGGCTGAACCAACTT	44,3	16	259	<i>PvuI</i>	164+95	2913-3'	280861 280876	CGTCGATGTCCCAGTC	42,9	16	G2916	2916-5'	283298 283278	TGGATGAAGCGCACGAACGTA	59,8	21	551	<i>AluI</i>	367+158+25	2916-3'	282748 282769	TGAAAACTCTGGGTCAACGACG	60,3	22	G3080	3080-5'	306606 306624	AAGGGTCAAGGCCCATTTG	58,8	19	325	<i>Avall</i>	190+134	3080-3'	306930 306909	GCTCGTCTACTTCAGCGTGGT	64,0	22	G3090	3090-5'	309669 309648	AGGAAGTTGCTCCAAGCGGTTG	62,1	22	276	<i>AluI</i>	137+109+29	3090-3'	309394 309415	CGGCTTCATCAAAAGTCACGCC	62,1	22	G3107	3107-5'	312396 312411	CCATGCTGATTACGGA	49,2	16	429	<i>PstI</i>	261+177	3107-3'	312824 312809	TGTCGAAAAAGGCTTG	46,6	16																																																																																																																																								
G2913	2913-5'	281119 281104	GGGGCTGAACCAACTT	44,3	16	259	<i>PvuI</i>	164+95																																																																																																																																																																																																														
	2913-3'	280861 280876	CGTCGATGTCCCAGTC	42,9	16				G2916	2916-5'	283298 283278	TGGATGAAGCGCACGAACGTA	59,8	21	551	<i>AluI</i>	367+158+25	2916-3'	282748 282769	TGAAAACTCTGGGTCAACGACG	60,3	22	G3080	3080-5'	306606 306624	AAGGGTCAAGGCCCATTTG	58,8	19	325	<i>Avall</i>	190+134	3080-3'	306930 306909	GCTCGTCTACTTCAGCGTGGT	64,0	22	G3090	3090-5'	309669 309648	AGGAAGTTGCTCCAAGCGGTTG	62,1	22	276	<i>AluI</i>	137+109+29	3090-3'	309394 309415	CGGCTTCATCAAAAGTCACGCC	62,1	22	G3107	3107-5'	312396 312411	CCATGCTGATTACGGA	49,2	16	429	<i>PstI</i>	261+177	3107-3'	312824 312809	TGTCGAAAAAGGCTTG	46,6	16																																																																																																																																																						
G2916	2916-5'	283298 283278	TGGATGAAGCGCACGAACGTA	59,8	21	551	<i>AluI</i>	367+158+25																																																																																																																																																																																																														
	2916-3'	282748 282769	TGAAAACTCTGGGTCAACGACG	60,3	22				G3080	3080-5'	306606 306624	AAGGGTCAAGGCCCATTTG	58,8	19	325	<i>Avall</i>	190+134	3080-3'	306930 306909	GCTCGTCTACTTCAGCGTGGT	64,0	22	G3090	3090-5'	309669 309648	AGGAAGTTGCTCCAAGCGGTTG	62,1	22	276	<i>AluI</i>	137+109+29	3090-3'	309394 309415	CGGCTTCATCAAAAGTCACGCC	62,1	22	G3107	3107-5'	312396 312411	CCATGCTGATTACGGA	49,2	16	429	<i>PstI</i>	261+177	3107-3'	312824 312809	TGTCGAAAAAGGCTTG	46,6	16																																																																																																																																																																				
G3080	3080-5'	306606 306624	AAGGGTCAAGGCCCATTTG	58,8	19	325	<i>Avall</i>	190+134																																																																																																																																																																																																														
	3080-3'	306930 306909	GCTCGTCTACTTCAGCGTGGT	64,0	22				G3090	3090-5'	309669 309648	AGGAAGTTGCTCCAAGCGGTTG	62,1	22	276	<i>AluI</i>	137+109+29	3090-3'	309394 309415	CGGCTTCATCAAAAGTCACGCC	62,1	22	G3107	3107-5'	312396 312411	CCATGCTGATTACGGA	49,2	16	429	<i>PstI</i>	261+177	3107-3'	312824 312809	TGTCGAAAAAGGCTTG	46,6	16																																																																																																																																																																																		
G3090	3090-5'	309669 309648	AGGAAGTTGCTCCAAGCGGTTG	62,1	22	276	<i>AluI</i>	137+109+29																																																																																																																																																																																																														
	3090-3'	309394 309415	CGGCTTCATCAAAAGTCACGCC	62,1	22				G3107	3107-5'	312396 312411	CCATGCTGATTACGGA	49,2	16	429	<i>PstI</i>	261+177	3107-3'	312824 312809	TGTCGAAAAAGGCTTG	46,6	16																																																																																																																																																																																																
G3107	3107-5'	312396 312411	CCATGCTGATTACGGA	49,2	16	429	<i>PstI</i>	261+177																																																																																																																																																																																																														
	3107-3'	312824 312809	TGTCGAAAAAGGCTTG	46,6	16																																																																																																																																																																																																																	

Tabla 3.3. Diseño de los cebadores para las reacciones de amplificación y selección de las enzimas utilizadas en la comprobación de cada producto obtenido Tm: temperatura de *melting*; L.P.: Longitud del producto de PCR; L: longitud del cebador. (continuación).

Nº ORF	Nombre del Cebador	Posición del Cebador		SECUENCIA	Tm (°C)	L (nt)	L.P. (pb)	Enzima	Longitud Fragmentos (pb)
		Inicio	final						
G3110	3110-5'	313509	313528	GCTCTGGCCGTCGTTGGAAC	63,5	20	655	<i>PvuII</i>	376+278
	3110-3'	314163	314142	GACACGGACCAAACTCGCCAT	62,1	22			
G3113	3113-5'	315037	315058	CCAAGGAGGCTGAATGGAACG	64,0	22	588	<i>BglII</i>	356+231
	3113-3'	315624	315603	ACCGGGCAGAGGTGGCAATAAA	62,1	22			
G3116	YGL98-183	317522	317542	GCAAACAAAACATGACGCCCA	57,9	21	336	<i>Clal</i>	253+82
	YGL98-498	317857	317837	GCTCCTTGCCTCAAACCTGCC	61,8	21			
G3146	3146-5'	325369	325383	GCCCCAATCCAGGT	47,4	15	333	<i>HpaI</i>	202+131
	3146-3'	325701	325687	CGGGCTGCCAAGAAG	47,4	15			
G3168	3168-5'	334895	334915	TGAAAGAAGCCGGATTGGGGG	61,8	21	610	<i>Clal</i>	364+213+40
	3168-3'	335504	335483	TGTTTCTCTCGCACTGCCTCG	64,0	22			
G3175	3175-5'	342661	342646	CCGTCCATGAATCGAA	49,2	16	479	<i>DraI</i>	336+142
	3175-3'	342183	342197	CGGAACAGACCCCAA	50,6	15			
G3179	3179-5'	343707	343722	TTGGAGTCAATCGGCGA	47,8	16	374	<i>HindIII</i>	204+170
	3179-3'	344080	344065	GGCTTTTGCATCTCCG	46,6	16			
G3189	3189-5'	346762	346747	AAGGGTTTGGTCCAGA	41,5	16	341	<i>NcoI</i>	267+73
	3189-3'	346422	346436	CTTCGGTTGACGCAA	42,3	15			
G3193	3193-5'	350106	350121	AACCCAGCTCAACCGT	44,8	16	313	<i>HindIII</i>	158+130+25
	3193-3'	350418	350403	CACCAGGCGTTTCAAA	44,0	16			
G3210	3210-5'	361452	361437	ACGCCTATTCAAGCCG	51,7	16	488	<i>EcoRI</i>	271+217
	3210-3'	360965	360980	GGCCAGGTAGCGGTAA	54,3	16			
G3406	3406-5'	379058	379073	CCAGGGCCAGTACCCAA	46,3	16	433	<i>EcoRI</i>	369+63
	3406-3'	379490	379475	TACACCTTGCCCGGAG	46,5	16			
G3432	YGL061c5'	389597	389577	CAGGAGGTGGAGCAAGCGATA	61,8	21	551	<i>AluI</i>	272+235+43
	YGL061c3'	389047	389066	CGAGGCAGGAACATGGGTCT	61,4	20			
G3435	3435-5'	390456	390471	CTCCGGAGATTGGGCA	49,2	16	581	<i>Clal</i>	443+138
	3435-3'	391036	391021	GGCCGGTTTTCATGAGA	48,9	16			
G3474	3474-5'	401283	401268	TGGGTGCTTGGTTATT	46,6	16	318	<i>HpaI</i>	190+128
	3474-3'	400966	400981	CCCTTTTATGTTTGCC	46,6	16			
G3483	3483-5'	402635	402650	AACCCGGTGCATATTT	46,6	16	421	<i>NdeI</i>	250+171
	3483-3'	403055	403041	CGGTGGCAATCTCA	47,8	15			
G3489	3489-5'	404038	404053	CCGGCAGTAGAAGGGA	54,3	16	330	<i>Clal</i>	321+9
	3489-3'	404367	404352	CGCATCGATATTGGGA	49,2	16			
G3505	3505-5'	405776	405791	GCTGCGCAGAATGCCT	50,0	16	542	<i>MboI</i>	281+147+114
	3505-3'	406317	406302	CCGAGCGGTACGTCCA	50,5	16			
G3517	3517-5'	411580	411595	TTACGTGTGGGGCAAC	44,3	16	329	<i>EcoRI</i>	273+56
	3517-3'	411908	411893	CGTTCACGCTGTGAT	44,1	16			
G3535	3535-5'	414541	414556	CCACGAGGCAGAGGAA	46,3	16	330	<i>XhoI</i>	172+158
	3535-3'	414870	414855	GGGGGACGCTTGGAG	47,3	16			
G3540	3540-5'	415140	415155	GCCCCTTGATGCATT	45,6	16	535	<i>AluI</i>	320+215
	3540-3'	415527	415512	TGAGGATTGGGATGGG	45,6	16			
G3679	3679-5'	451694	451679	AGCTCGGTGAAGGGCA	49,0	16	472	<i>aLUI</i>	385+87
	3679-3'	451223	451237	TGATGCGGGCGGTTTC	48,4	15			
G3690	3690-5'	457856	457841	CAGCAGCGGTAGCGAT	47,2	16	499	<i>EcoRI</i>	350+149
	3690-3'	457358	457373	CCAAATTTGCCCTCCA	46,6	16			
G3719	3719-5'	468300	468315	GGGTGTTGGGTTTGC	51,7	16	440	<i>BamHI</i>	342+98
	3719-3'	468739	468724	ACCAGTAAAGGCCCA	51,7	16			
G3759	3759-5'	490476	490461	GTTCCCAATGCATTCG	46,6	16	510	<i>HindIII</i>	258+252
	3759-3'	489967	489982	ATCTGTCCCCGTGCTG	45,9	16			

Tabla 3.3. Diseño de los cebadores para las reacciones de amplificación y selección de las enzimas utilizadas en la comprobación de cada producto obtenido Tm: temperatura de *melting*, L.P.: Longitud del producto de PCR; L: longitud del cebador. (continuación).

Nº ORF	Nombre del Cebador	Posición del Cebador Inicio final	SECUENCIA	Tm (°C)	L (nt)	L.P. (pb)	Enzima	Longitud Fragmentos (pb)																																																																																																																																																																																																																																																																																																		
G4332	4332-5'	595595 595580	AAATGTTACGGCCGCA	49,2	16	320	<i>EcoRI</i>	248+72																																																																																																																																																																																																																																																																																																		
	4332-3'	595276 595291	TTGGTGGCTTTCCTGG	51,7	16				G7538	7538-5'	871246 871231	GGATCCAGAACACGGC	45,3	16	469	<i>EcoRI</i>	238+231	7538-3'	870778 870793	AGGAACAACCCATCG	45,0	16	G7553	7553-5'	878076 878061	TGCTTCGTGTAACCCG	44,8	16	445	<i>KpnI</i>	241+203	7553-3'	877632 877647	GTCCATCGGCCAATCT	44,9	16	G7584	7584-5'	887654 887639	CGAATGTCCCGTAGCC	46,1	16	302	<i>SacI</i>	198+104	7584-3'	887353 887368	TTTTTCAGGCCACC	45,7	16	G7587	7587-5'	888978 888993	GCTGCAGACGGTTCAT	43,2	16	390	<i>EcoRV</i>	264+126	7587-3'	889367 889352	CCCGACGGATATACCA	43,4	16	G7589	7589-5'	891425 891405	TCCGAAAGAGTACCTTGC	59,8	21	756	<i>EcoRV</i>	537+218	7589-3'	890670 890691	CTGTTTCCAATTCAGAACCGC	58,4	22	G7594	7594-5'	895806 895821	AACTCTCCAAAACCG	41,7	16	350	<i>EcoRI</i>	276+74	7594-3'	896155 896140	AATGCCTTCCTCCT	40,0	16	G7725	7725-5'	900799 900778	TGCCAGACTCTGCTACCGTCC	64,0	22	568	<i>HindIII</i>	376+191	7725-3'	900232 900252	GTTTTATCCCTGGACGCGGTG	61,8	21	G7737	7737-5'	909312 909326	TTCTCGGGCCACAA	47,6	15	325	<i>AluI</i>	166+159	7737-3'	909636 909622	GGCAACGGTCCCTT	47,9	15	G7748	7748-5'	914303 914283	ATCAGGCTTTGGTTGAGGGG	61,8	21	624	<i>ClaI</i>	527+96	7748-3'	913680 913701	AAATACGTTGGTTCGGTGCCA	58,4	22	G7801	7801-5'	915316 915337	AACAAAGCTAATGACGGGGTCA	58,4	22	620	<i>ClaI</i>	446+174	7801-3'	915935 915914	TGAACGTTTGTTCGTCGGTTC	58,4	22	G7806	7806-5'	917116 917137	TCAAACCAACGAATGCCAA	56,5	22	611	<i>HindIII</i>	583+27	7806-3'	917726 917705	GCTCGTATTATCATCCAGCAGG	62,1	22	G7826	7826-5'	923809 923794	CTCGGCAATTGTCAG	44,0	16	358	<i>ClaI</i>	339+19	7826-3'	923452 923467	GCCAAGGGGGTTATTT	44,0	16	G8523	8523-5'	938207 938192	TCTTCGATACGGCGTG	44,9	16	298	<i>SacI</i>	240+56	8523-3'	937910 937924	TCGGTGGTGGGAATG	44,2	15	G8530	8530-5'	941410 941396	GGTGGACAAGGCCGA	46,7	15	503	<i>HindIII</i>	468+34	8530-3'	940908 940923	CACCCGCTGATTCGT	46,5	16	G8537	8537-5'	942944 942959	GCTCGAGGACAACGCC	56,9	16	363	<i>EcoRV</i>	262+100	8537-3'	943306 943291	CAACCGCCACTCCAG	56,9	16	G8541	8541-5'	945197 945212	TCGAGCGGAAATCTA	40,9	16	440	<i>EcoRI</i>	290+150	8541-3'	945636 945622	CTGGTCCAAAGCAA	40,9	15	N1696	N1696-5'	318762 318749	TGGGCGTAATTACG	36,0	14	370	<i>XhoI</i>	286+84	N1696-3'	318393 318405	CCAAGGTAGCCCC	36,0	13	N1706	N1706-5'	322860 322848	CCGCAAAACCATAG	31,5	13	246	<i>EcoRI</i>	82+164	N1706-3'	322615 322627	GCGTCGTGTCCTCG	34,5	13	N1710	N1710-5'	324349 324364	TAGACTCCCAATGCGA	40,9	16	330	<i>EcoRI</i>	80+250	N1710-3'	324678 324663	TGTAGGATGCCATCGT	40,3	16	N1714	N1714-5'	326260 326245	GAGGATGGACCTATGC	38,1	16	267	<i>HindIII</i>	119+56+92	N1714-3'	325994 326009	CTCTGCGGAATCAATG	40,9	16	N1718	N1718-5'	329634 329620	CGGAGAAGCTGAAGC	39,9	15	280	<i>HindIII</i>	199+81	N1718-3'
G7538	7538-5'	871246 871231	GGATCCAGAACACGGC	45,3	16	469	<i>EcoRI</i>	238+231																																																																																																																																																																																																																																																																																																		
	7538-3'	870778 870793	AGGAACAACCCATCG	45,0	16				G7553	7553-5'	878076 878061	TGCTTCGTGTAACCCG	44,8	16	445	<i>KpnI</i>	241+203	7553-3'	877632 877647	GTCCATCGGCCAATCT	44,9	16	G7584	7584-5'	887654 887639	CGAATGTCCCGTAGCC	46,1	16	302	<i>SacI</i>	198+104	7584-3'	887353 887368	TTTTTCAGGCCACC	45,7	16	G7587	7587-5'	888978 888993	GCTGCAGACGGTTCAT	43,2	16	390	<i>EcoRV</i>	264+126	7587-3'	889367 889352	CCCGACGGATATACCA	43,4	16	G7589	7589-5'	891425 891405	TCCGAAAGAGTACCTTGC	59,8	21	756	<i>EcoRV</i>	537+218	7589-3'	890670 890691	CTGTTTCCAATTCAGAACCGC	58,4	22	G7594	7594-5'	895806 895821	AACTCTCCAAAACCG	41,7	16	350	<i>EcoRI</i>	276+74	7594-3'	896155 896140	AATGCCTTCCTCCT	40,0	16	G7725	7725-5'	900799 900778	TGCCAGACTCTGCTACCGTCC	64,0	22	568	<i>HindIII</i>	376+191	7725-3'	900232 900252	GTTTTATCCCTGGACGCGGTG	61,8	21	G7737	7737-5'	909312 909326	TTCTCGGGCCACAA	47,6	15	325	<i>AluI</i>	166+159	7737-3'	909636 909622	GGCAACGGTCCCTT	47,9	15	G7748	7748-5'	914303 914283	ATCAGGCTTTGGTTGAGGGG	61,8	21	624	<i>ClaI</i>	527+96	7748-3'	913680 913701	AAATACGTTGGTTCGGTGCCA	58,4	22	G7801	7801-5'	915316 915337	AACAAAGCTAATGACGGGGTCA	58,4	22	620	<i>ClaI</i>	446+174	7801-3'	915935 915914	TGAACGTTTGTTCGTCGGTTC	58,4	22	G7806	7806-5'	917116 917137	TCAAACCAACGAATGCCAA	56,5	22	611	<i>HindIII</i>	583+27	7806-3'	917726 917705	GCTCGTATTATCATCCAGCAGG	62,1	22	G7826	7826-5'	923809 923794	CTCGGCAATTGTCAG	44,0	16	358	<i>ClaI</i>	339+19	7826-3'	923452 923467	GCCAAGGGGGTTATTT	44,0	16	G8523	8523-5'	938207 938192	TCTTCGATACGGCGTG	44,9	16	298	<i>SacI</i>	240+56	8523-3'	937910 937924	TCGGTGGTGGGAATG	44,2	15	G8530	8530-5'	941410 941396	GGTGGACAAGGCCGA	46,7	15	503	<i>HindIII</i>	468+34	8530-3'	940908 940923	CACCCGCTGATTCGT	46,5	16	G8537	8537-5'	942944 942959	GCTCGAGGACAACGCC	56,9	16	363	<i>EcoRV</i>	262+100	8537-3'	943306 943291	CAACCGCCACTCCAG	56,9	16	G8541	8541-5'	945197 945212	TCGAGCGGAAATCTA	40,9	16	440	<i>EcoRI</i>	290+150	8541-3'	945636 945622	CTGGTCCAAAGCAA	40,9	15	N1696	N1696-5'	318762 318749	TGGGCGTAATTACG	36,0	14	370	<i>XhoI</i>	286+84	N1696-3'	318393 318405	CCAAGGTAGCCCC	36,0	13	N1706	N1706-5'	322860 322848	CCGCAAAACCATAG	31,5	13	246	<i>EcoRI</i>	82+164	N1706-3'	322615 322627	GCGTCGTGTCCTCG	34,5	13	N1710	N1710-5'	324349 324364	TAGACTCCCAATGCGA	40,9	16	330	<i>EcoRI</i>	80+250	N1710-3'	324678 324663	TGTAGGATGCCATCGT	40,3	16	N1714	N1714-5'	326260 326245	GAGGATGGACCTATGC	38,1	16	267	<i>HindIII</i>	119+56+92	N1714-3'	325994 326009	CTCTGCGGAATCAATG	40,9	16	N1718	N1718-5'	329634 329620	CGGAGAAGCTGAAGC	39,9	15	280	<i>HindIII</i>	199+81	N1718-3'	329355 329368	CCATCAATTGCGGA	38,9	14										
G7553	7553-5'	878076 878061	TGCTTCGTGTAACCCG	44,8	16	445	<i>KpnI</i>	241+203																																																																																																																																																																																																																																																																																																		
	7553-3'	877632 877647	GTCCATCGGCCAATCT	44,9	16				G7584	7584-5'	887654 887639	CGAATGTCCCGTAGCC	46,1	16	302	<i>SacI</i>	198+104	7584-3'	887353 887368	TTTTTCAGGCCACC	45,7	16	G7587	7587-5'	888978 888993	GCTGCAGACGGTTCAT	43,2	16	390	<i>EcoRV</i>	264+126	7587-3'	889367 889352	CCCGACGGATATACCA	43,4	16	G7589	7589-5'	891425 891405	TCCGAAAGAGTACCTTGC	59,8	21	756	<i>EcoRV</i>	537+218	7589-3'	890670 890691	CTGTTTCCAATTCAGAACCGC	58,4	22	G7594	7594-5'	895806 895821	AACTCTCCAAAACCG	41,7	16	350	<i>EcoRI</i>	276+74	7594-3'	896155 896140	AATGCCTTCCTCCT	40,0	16	G7725	7725-5'	900799 900778	TGCCAGACTCTGCTACCGTCC	64,0	22	568	<i>HindIII</i>	376+191	7725-3'	900232 900252	GTTTTATCCCTGGACGCGGTG	61,8	21	G7737	7737-5'	909312 909326	TTCTCGGGCCACAA	47,6	15	325	<i>AluI</i>	166+159	7737-3'	909636 909622	GGCAACGGTCCCTT	47,9	15	G7748	7748-5'	914303 914283	ATCAGGCTTTGGTTGAGGGG	61,8	21	624	<i>ClaI</i>	527+96	7748-3'	913680 913701	AAATACGTTGGTTCGGTGCCA	58,4	22	G7801	7801-5'	915316 915337	AACAAAGCTAATGACGGGGTCA	58,4	22	620	<i>ClaI</i>	446+174	7801-3'	915935 915914	TGAACGTTTGTTCGTCGGTTC	58,4	22	G7806	7806-5'	917116 917137	TCAAACCAACGAATGCCAA	56,5	22	611	<i>HindIII</i>	583+27	7806-3'	917726 917705	GCTCGTATTATCATCCAGCAGG	62,1	22	G7826	7826-5'	923809 923794	CTCGGCAATTGTCAG	44,0	16	358	<i>ClaI</i>	339+19	7826-3'	923452 923467	GCCAAGGGGGTTATTT	44,0	16	G8523	8523-5'	938207 938192	TCTTCGATACGGCGTG	44,9	16	298	<i>SacI</i>	240+56	8523-3'	937910 937924	TCGGTGGTGGGAATG	44,2	15	G8530	8530-5'	941410 941396	GGTGGACAAGGCCGA	46,7	15	503	<i>HindIII</i>	468+34	8530-3'	940908 940923	CACCCGCTGATTCGT	46,5	16	G8537	8537-5'	942944 942959	GCTCGAGGACAACGCC	56,9	16	363	<i>EcoRV</i>	262+100	8537-3'	943306 943291	CAACCGCCACTCCAG	56,9	16	G8541	8541-5'	945197 945212	TCGAGCGGAAATCTA	40,9	16	440	<i>EcoRI</i>	290+150	8541-3'	945636 945622	CTGGTCCAAAGCAA	40,9	15	N1696	N1696-5'	318762 318749	TGGGCGTAATTACG	36,0	14	370	<i>XhoI</i>	286+84	N1696-3'	318393 318405	CCAAGGTAGCCCC	36,0	13	N1706	N1706-5'	322860 322848	CCGCAAAACCATAG	31,5	13	246	<i>EcoRI</i>	82+164	N1706-3'	322615 322627	GCGTCGTGTCCTCG	34,5	13	N1710	N1710-5'	324349 324364	TAGACTCCCAATGCGA	40,9	16	330	<i>EcoRI</i>	80+250	N1710-3'	324678 324663	TGTAGGATGCCATCGT	40,3	16	N1714	N1714-5'	326260 326245	GAGGATGGACCTATGC	38,1	16	267	<i>HindIII</i>	119+56+92	N1714-3'	325994 326009	CTCTGCGGAATCAATG	40,9	16	N1718	N1718-5'	329634 329620	CGGAGAAGCTGAAGC	39,9	15	280	<i>HindIII</i>	199+81	N1718-3'	329355 329368	CCATCAATTGCGGA	38,9	14																								
G7584	7584-5'	887654 887639	CGAATGTCCCGTAGCC	46,1	16	302	<i>SacI</i>	198+104																																																																																																																																																																																																																																																																																																		
	7584-3'	887353 887368	TTTTTCAGGCCACC	45,7	16				G7587	7587-5'	888978 888993	GCTGCAGACGGTTCAT	43,2	16	390	<i>EcoRV</i>	264+126	7587-3'	889367 889352	CCCGACGGATATACCA	43,4	16	G7589	7589-5'	891425 891405	TCCGAAAGAGTACCTTGC	59,8	21	756	<i>EcoRV</i>	537+218	7589-3'	890670 890691	CTGTTTCCAATTCAGAACCGC	58,4	22	G7594	7594-5'	895806 895821	AACTCTCCAAAACCG	41,7	16	350	<i>EcoRI</i>	276+74	7594-3'	896155 896140	AATGCCTTCCTCCT	40,0	16	G7725	7725-5'	900799 900778	TGCCAGACTCTGCTACCGTCC	64,0	22	568	<i>HindIII</i>	376+191	7725-3'	900232 900252	GTTTTATCCCTGGACGCGGTG	61,8	21	G7737	7737-5'	909312 909326	TTCTCGGGCCACAA	47,6	15	325	<i>AluI</i>	166+159	7737-3'	909636 909622	GGCAACGGTCCCTT	47,9	15	G7748	7748-5'	914303 914283	ATCAGGCTTTGGTTGAGGGG	61,8	21	624	<i>ClaI</i>	527+96	7748-3'	913680 913701	AAATACGTTGGTTCGGTGCCA	58,4	22	G7801	7801-5'	915316 915337	AACAAAGCTAATGACGGGGTCA	58,4	22	620	<i>ClaI</i>	446+174	7801-3'	915935 915914	TGAACGTTTGTTCGTCGGTTC	58,4	22	G7806	7806-5'	917116 917137	TCAAACCAACGAATGCCAA	56,5	22	611	<i>HindIII</i>	583+27	7806-3'	917726 917705	GCTCGTATTATCATCCAGCAGG	62,1	22	G7826	7826-5'	923809 923794	CTCGGCAATTGTCAG	44,0	16	358	<i>ClaI</i>	339+19	7826-3'	923452 923467	GCCAAGGGGGTTATTT	44,0	16	G8523	8523-5'	938207 938192	TCTTCGATACGGCGTG	44,9	16	298	<i>SacI</i>	240+56	8523-3'	937910 937924	TCGGTGGTGGGAATG	44,2	15	G8530	8530-5'	941410 941396	GGTGGACAAGGCCGA	46,7	15	503	<i>HindIII</i>	468+34	8530-3'	940908 940923	CACCCGCTGATTCGT	46,5	16	G8537	8537-5'	942944 942959	GCTCGAGGACAACGCC	56,9	16	363	<i>EcoRV</i>	262+100	8537-3'	943306 943291	CAACCGCCACTCCAG	56,9	16	G8541	8541-5'	945197 945212	TCGAGCGGAAATCTA	40,9	16	440	<i>EcoRI</i>	290+150	8541-3'	945636 945622	CTGGTCCAAAGCAA	40,9	15	N1696	N1696-5'	318762 318749	TGGGCGTAATTACG	36,0	14	370	<i>XhoI</i>	286+84	N1696-3'	318393 318405	CCAAGGTAGCCCC	36,0	13	N1706	N1706-5'	322860 322848	CCGCAAAACCATAG	31,5	13	246	<i>EcoRI</i>	82+164	N1706-3'	322615 322627	GCGTCGTGTCCTCG	34,5	13	N1710	N1710-5'	324349 324364	TAGACTCCCAATGCGA	40,9	16	330	<i>EcoRI</i>	80+250	N1710-3'	324678 324663	TGTAGGATGCCATCGT	40,3	16	N1714	N1714-5'	326260 326245	GAGGATGGACCTATGC	38,1	16	267	<i>HindIII</i>	119+56+92	N1714-3'	325994 326009	CTCTGCGGAATCAATG	40,9	16	N1718	N1718-5'	329634 329620	CGGAGAAGCTGAAGC	39,9	15	280	<i>HindIII</i>	199+81	N1718-3'	329355 329368	CCATCAATTGCGGA	38,9	14																																						
G7587	7587-5'	888978 888993	GCTGCAGACGGTTCAT	43,2	16	390	<i>EcoRV</i>	264+126																																																																																																																																																																																																																																																																																																		
	7587-3'	889367 889352	CCCGACGGATATACCA	43,4	16				G7589	7589-5'	891425 891405	TCCGAAAGAGTACCTTGC	59,8	21	756	<i>EcoRV</i>	537+218	7589-3'	890670 890691	CTGTTTCCAATTCAGAACCGC	58,4	22	G7594	7594-5'	895806 895821	AACTCTCCAAAACCG	41,7	16	350	<i>EcoRI</i>	276+74	7594-3'	896155 896140	AATGCCTTCCTCCT	40,0	16	G7725	7725-5'	900799 900778	TGCCAGACTCTGCTACCGTCC	64,0	22	568	<i>HindIII</i>	376+191	7725-3'	900232 900252	GTTTTATCCCTGGACGCGGTG	61,8	21	G7737	7737-5'	909312 909326	TTCTCGGGCCACAA	47,6	15	325	<i>AluI</i>	166+159	7737-3'	909636 909622	GGCAACGGTCCCTT	47,9	15	G7748	7748-5'	914303 914283	ATCAGGCTTTGGTTGAGGGG	61,8	21	624	<i>ClaI</i>	527+96	7748-3'	913680 913701	AAATACGTTGGTTCGGTGCCA	58,4	22	G7801	7801-5'	915316 915337	AACAAAGCTAATGACGGGGTCA	58,4	22	620	<i>ClaI</i>	446+174	7801-3'	915935 915914	TGAACGTTTGTTCGTCGGTTC	58,4	22	G7806	7806-5'	917116 917137	TCAAACCAACGAATGCCAA	56,5	22	611	<i>HindIII</i>	583+27	7806-3'	917726 917705	GCTCGTATTATCATCCAGCAGG	62,1	22	G7826	7826-5'	923809 923794	CTCGGCAATTGTCAG	44,0	16	358	<i>ClaI</i>	339+19	7826-3'	923452 923467	GCCAAGGGGGTTATTT	44,0	16	G8523	8523-5'	938207 938192	TCTTCGATACGGCGTG	44,9	16	298	<i>SacI</i>	240+56	8523-3'	937910 937924	TCGGTGGTGGGAATG	44,2	15	G8530	8530-5'	941410 941396	GGTGGACAAGGCCGA	46,7	15	503	<i>HindIII</i>	468+34	8530-3'	940908 940923	CACCCGCTGATTCGT	46,5	16	G8537	8537-5'	942944 942959	GCTCGAGGACAACGCC	56,9	16	363	<i>EcoRV</i>	262+100	8537-3'	943306 943291	CAACCGCCACTCCAG	56,9	16	G8541	8541-5'	945197 945212	TCGAGCGGAAATCTA	40,9	16	440	<i>EcoRI</i>	290+150	8541-3'	945636 945622	CTGGTCCAAAGCAA	40,9	15	N1696	N1696-5'	318762 318749	TGGGCGTAATTACG	36,0	14	370	<i>XhoI</i>	286+84	N1696-3'	318393 318405	CCAAGGTAGCCCC	36,0	13	N1706	N1706-5'	322860 322848	CCGCAAAACCATAG	31,5	13	246	<i>EcoRI</i>	82+164	N1706-3'	322615 322627	GCGTCGTGTCCTCG	34,5	13	N1710	N1710-5'	324349 324364	TAGACTCCCAATGCGA	40,9	16	330	<i>EcoRI</i>	80+250	N1710-3'	324678 324663	TGTAGGATGCCATCGT	40,3	16	N1714	N1714-5'	326260 326245	GAGGATGGACCTATGC	38,1	16	267	<i>HindIII</i>	119+56+92	N1714-3'	325994 326009	CTCTGCGGAATCAATG	40,9	16	N1718	N1718-5'	329634 329620	CGGAGAAGCTGAAGC	39,9	15	280	<i>HindIII</i>	199+81	N1718-3'	329355 329368	CCATCAATTGCGGA	38,9	14																																																				
G7589	7589-5'	891425 891405	TCCGAAAGAGTACCTTGC	59,8	21	756	<i>EcoRV</i>	537+218																																																																																																																																																																																																																																																																																																		
	7589-3'	890670 890691	CTGTTTCCAATTCAGAACCGC	58,4	22				G7594	7594-5'	895806 895821	AACTCTCCAAAACCG	41,7	16	350	<i>EcoRI</i>	276+74	7594-3'	896155 896140	AATGCCTTCCTCCT	40,0	16	G7725	7725-5'	900799 900778	TGCCAGACTCTGCTACCGTCC	64,0	22	568	<i>HindIII</i>	376+191	7725-3'	900232 900252	GTTTTATCCCTGGACGCGGTG	61,8	21	G7737	7737-5'	909312 909326	TTCTCGGGCCACAA	47,6	15	325	<i>AluI</i>	166+159	7737-3'	909636 909622	GGCAACGGTCCCTT	47,9	15	G7748	7748-5'	914303 914283	ATCAGGCTTTGGTTGAGGGG	61,8	21	624	<i>ClaI</i>	527+96	7748-3'	913680 913701	AAATACGTTGGTTCGGTGCCA	58,4	22	G7801	7801-5'	915316 915337	AACAAAGCTAATGACGGGGTCA	58,4	22	620	<i>ClaI</i>	446+174	7801-3'	915935 915914	TGAACGTTTGTTCGTCGGTTC	58,4	22	G7806	7806-5'	917116 917137	TCAAACCAACGAATGCCAA	56,5	22	611	<i>HindIII</i>	583+27	7806-3'	917726 917705	GCTCGTATTATCATCCAGCAGG	62,1	22	G7826	7826-5'	923809 923794	CTCGGCAATTGTCAG	44,0	16	358	<i>ClaI</i>	339+19	7826-3'	923452 923467	GCCAAGGGGGTTATTT	44,0	16	G8523	8523-5'	938207 938192	TCTTCGATACGGCGTG	44,9	16	298	<i>SacI</i>	240+56	8523-3'	937910 937924	TCGGTGGTGGGAATG	44,2	15	G8530	8530-5'	941410 941396	GGTGGACAAGGCCGA	46,7	15	503	<i>HindIII</i>	468+34	8530-3'	940908 940923	CACCCGCTGATTCGT	46,5	16	G8537	8537-5'	942944 942959	GCTCGAGGACAACGCC	56,9	16	363	<i>EcoRV</i>	262+100	8537-3'	943306 943291	CAACCGCCACTCCAG	56,9	16	G8541	8541-5'	945197 945212	TCGAGCGGAAATCTA	40,9	16	440	<i>EcoRI</i>	290+150	8541-3'	945636 945622	CTGGTCCAAAGCAA	40,9	15	N1696	N1696-5'	318762 318749	TGGGCGTAATTACG	36,0	14	370	<i>XhoI</i>	286+84	N1696-3'	318393 318405	CCAAGGTAGCCCC	36,0	13	N1706	N1706-5'	322860 322848	CCGCAAAACCATAG	31,5	13	246	<i>EcoRI</i>	82+164	N1706-3'	322615 322627	GCGTCGTGTCCTCG	34,5	13	N1710	N1710-5'	324349 324364	TAGACTCCCAATGCGA	40,9	16	330	<i>EcoRI</i>	80+250	N1710-3'	324678 324663	TGTAGGATGCCATCGT	40,3	16	N1714	N1714-5'	326260 326245	GAGGATGGACCTATGC	38,1	16	267	<i>HindIII</i>	119+56+92	N1714-3'	325994 326009	CTCTGCGGAATCAATG	40,9	16	N1718	N1718-5'	329634 329620	CGGAGAAGCTGAAGC	39,9	15	280	<i>HindIII</i>	199+81	N1718-3'	329355 329368	CCATCAATTGCGGA	38,9	14																																																																		
G7594	7594-5'	895806 895821	AACTCTCCAAAACCG	41,7	16	350	<i>EcoRI</i>	276+74																																																																																																																																																																																																																																																																																																		
	7594-3'	896155 896140	AATGCCTTCCTCCT	40,0	16				G7725	7725-5'	900799 900778	TGCCAGACTCTGCTACCGTCC	64,0	22	568	<i>HindIII</i>	376+191	7725-3'	900232 900252	GTTTTATCCCTGGACGCGGTG	61,8	21	G7737	7737-5'	909312 909326	TTCTCGGGCCACAA	47,6	15	325	<i>AluI</i>	166+159	7737-3'	909636 909622	GGCAACGGTCCCTT	47,9	15	G7748	7748-5'	914303 914283	ATCAGGCTTTGGTTGAGGGG	61,8	21	624	<i>ClaI</i>	527+96	7748-3'	913680 913701	AAATACGTTGGTTCGGTGCCA	58,4	22	G7801	7801-5'	915316 915337	AACAAAGCTAATGACGGGGTCA	58,4	22	620	<i>ClaI</i>	446+174	7801-3'	915935 915914	TGAACGTTTGTTCGTCGGTTC	58,4	22	G7806	7806-5'	917116 917137	TCAAACCAACGAATGCCAA	56,5	22	611	<i>HindIII</i>	583+27	7806-3'	917726 917705	GCTCGTATTATCATCCAGCAGG	62,1	22	G7826	7826-5'	923809 923794	CTCGGCAATTGTCAG	44,0	16	358	<i>ClaI</i>	339+19	7826-3'	923452 923467	GCCAAGGGGGTTATTT	44,0	16	G8523	8523-5'	938207 938192	TCTTCGATACGGCGTG	44,9	16	298	<i>SacI</i>	240+56	8523-3'	937910 937924	TCGGTGGTGGGAATG	44,2	15	G8530	8530-5'	941410 941396	GGTGGACAAGGCCGA	46,7	15	503	<i>HindIII</i>	468+34	8530-3'	940908 940923	CACCCGCTGATTCGT	46,5	16	G8537	8537-5'	942944 942959	GCTCGAGGACAACGCC	56,9	16	363	<i>EcoRV</i>	262+100	8537-3'	943306 943291	CAACCGCCACTCCAG	56,9	16	G8541	8541-5'	945197 945212	TCGAGCGGAAATCTA	40,9	16	440	<i>EcoRI</i>	290+150	8541-3'	945636 945622	CTGGTCCAAAGCAA	40,9	15	N1696	N1696-5'	318762 318749	TGGGCGTAATTACG	36,0	14	370	<i>XhoI</i>	286+84	N1696-3'	318393 318405	CCAAGGTAGCCCC	36,0	13	N1706	N1706-5'	322860 322848	CCGCAAAACCATAG	31,5	13	246	<i>EcoRI</i>	82+164	N1706-3'	322615 322627	GCGTCGTGTCCTCG	34,5	13	N1710	N1710-5'	324349 324364	TAGACTCCCAATGCGA	40,9	16	330	<i>EcoRI</i>	80+250	N1710-3'	324678 324663	TGTAGGATGCCATCGT	40,3	16	N1714	N1714-5'	326260 326245	GAGGATGGACCTATGC	38,1	16	267	<i>HindIII</i>	119+56+92	N1714-3'	325994 326009	CTCTGCGGAATCAATG	40,9	16	N1718	N1718-5'	329634 329620	CGGAGAAGCTGAAGC	39,9	15	280	<i>HindIII</i>	199+81	N1718-3'	329355 329368	CCATCAATTGCGGA	38,9	14																																																																																
G7725	7725-5'	900799 900778	TGCCAGACTCTGCTACCGTCC	64,0	22	568	<i>HindIII</i>	376+191																																																																																																																																																																																																																																																																																																		
	7725-3'	900232 900252	GTTTTATCCCTGGACGCGGTG	61,8	21				G7737	7737-5'	909312 909326	TTCTCGGGCCACAA	47,6	15	325	<i>AluI</i>	166+159	7737-3'	909636 909622	GGCAACGGTCCCTT	47,9	15	G7748	7748-5'	914303 914283	ATCAGGCTTTGGTTGAGGGG	61,8	21	624	<i>ClaI</i>	527+96	7748-3'	913680 913701	AAATACGTTGGTTCGGTGCCA	58,4	22	G7801	7801-5'	915316 915337	AACAAAGCTAATGACGGGGTCA	58,4	22	620	<i>ClaI</i>	446+174	7801-3'	915935 915914	TGAACGTTTGTTCGTCGGTTC	58,4	22	G7806	7806-5'	917116 917137	TCAAACCAACGAATGCCAA	56,5	22	611	<i>HindIII</i>	583+27	7806-3'	917726 917705	GCTCGTATTATCATCCAGCAGG	62,1	22	G7826	7826-5'	923809 923794	CTCGGCAATTGTCAG	44,0	16	358	<i>ClaI</i>	339+19	7826-3'	923452 923467	GCCAAGGGGGTTATTT	44,0	16	G8523	8523-5'	938207 938192	TCTTCGATACGGCGTG	44,9	16	298	<i>SacI</i>	240+56	8523-3'	937910 937924	TCGGTGGTGGGAATG	44,2	15	G8530	8530-5'	941410 941396	GGTGGACAAGGCCGA	46,7	15	503	<i>HindIII</i>	468+34	8530-3'	940908 940923	CACCCGCTGATTCGT	46,5	16	G8537	8537-5'	942944 942959	GCTCGAGGACAACGCC	56,9	16	363	<i>EcoRV</i>	262+100	8537-3'	943306 943291	CAACCGCCACTCCAG	56,9	16	G8541	8541-5'	945197 945212	TCGAGCGGAAATCTA	40,9	16	440	<i>EcoRI</i>	290+150	8541-3'	945636 945622	CTGGTCCAAAGCAA	40,9	15	N1696	N1696-5'	318762 318749	TGGGCGTAATTACG	36,0	14	370	<i>XhoI</i>	286+84	N1696-3'	318393 318405	CCAAGGTAGCCCC	36,0	13	N1706	N1706-5'	322860 322848	CCGCAAAACCATAG	31,5	13	246	<i>EcoRI</i>	82+164	N1706-3'	322615 322627	GCGTCGTGTCCTCG	34,5	13	N1710	N1710-5'	324349 324364	TAGACTCCCAATGCGA	40,9	16	330	<i>EcoRI</i>	80+250	N1710-3'	324678 324663	TGTAGGATGCCATCGT	40,3	16	N1714	N1714-5'	326260 326245	GAGGATGGACCTATGC	38,1	16	267	<i>HindIII</i>	119+56+92	N1714-3'	325994 326009	CTCTGCGGAATCAATG	40,9	16	N1718	N1718-5'	329634 329620	CGGAGAAGCTGAAGC	39,9	15	280	<i>HindIII</i>	199+81	N1718-3'	329355 329368	CCATCAATTGCGGA	38,9	14																																																																																														
G7737	7737-5'	909312 909326	TTCTCGGGCCACAA	47,6	15	325	<i>AluI</i>	166+159																																																																																																																																																																																																																																																																																																		
	7737-3'	909636 909622	GGCAACGGTCCCTT	47,9	15				G7748	7748-5'	914303 914283	ATCAGGCTTTGGTTGAGGGG	61,8	21	624	<i>ClaI</i>	527+96	7748-3'	913680 913701	AAATACGTTGGTTCGGTGCCA	58,4	22	G7801	7801-5'	915316 915337	AACAAAGCTAATGACGGGGTCA	58,4	22	620	<i>ClaI</i>	446+174	7801-3'	915935 915914	TGAACGTTTGTTCGTCGGTTC	58,4	22	G7806	7806-5'	917116 917137	TCAAACCAACGAATGCCAA	56,5	22	611	<i>HindIII</i>	583+27	7806-3'	917726 917705	GCTCGTATTATCATCCAGCAGG	62,1	22	G7826	7826-5'	923809 923794	CTCGGCAATTGTCAG	44,0	16	358	<i>ClaI</i>	339+19	7826-3'	923452 923467	GCCAAGGGGGTTATTT	44,0	16	G8523	8523-5'	938207 938192	TCTTCGATACGGCGTG	44,9	16	298	<i>SacI</i>	240+56	8523-3'	937910 937924	TCGGTGGTGGGAATG	44,2	15	G8530	8530-5'	941410 941396	GGTGGACAAGGCCGA	46,7	15	503	<i>HindIII</i>	468+34	8530-3'	940908 940923	CACCCGCTGATTCGT	46,5	16	G8537	8537-5'	942944 942959	GCTCGAGGACAACGCC	56,9	16	363	<i>EcoRV</i>	262+100	8537-3'	943306 943291	CAACCGCCACTCCAG	56,9	16	G8541	8541-5'	945197 945212	TCGAGCGGAAATCTA	40,9	16	440	<i>EcoRI</i>	290+150	8541-3'	945636 945622	CTGGTCCAAAGCAA	40,9	15	N1696	N1696-5'	318762 318749	TGGGCGTAATTACG	36,0	14	370	<i>XhoI</i>	286+84	N1696-3'	318393 318405	CCAAGGTAGCCCC	36,0	13	N1706	N1706-5'	322860 322848	CCGCAAAACCATAG	31,5	13	246	<i>EcoRI</i>	82+164	N1706-3'	322615 322627	GCGTCGTGTCCTCG	34,5	13	N1710	N1710-5'	324349 324364	TAGACTCCCAATGCGA	40,9	16	330	<i>EcoRI</i>	80+250	N1710-3'	324678 324663	TGTAGGATGCCATCGT	40,3	16	N1714	N1714-5'	326260 326245	GAGGATGGACCTATGC	38,1	16	267	<i>HindIII</i>	119+56+92	N1714-3'	325994 326009	CTCTGCGGAATCAATG	40,9	16	N1718	N1718-5'	329634 329620	CGGAGAAGCTGAAGC	39,9	15	280	<i>HindIII</i>	199+81	N1718-3'	329355 329368	CCATCAATTGCGGA	38,9	14																																																																																																												
G7748	7748-5'	914303 914283	ATCAGGCTTTGGTTGAGGGG	61,8	21	624	<i>ClaI</i>	527+96																																																																																																																																																																																																																																																																																																		
	7748-3'	913680 913701	AAATACGTTGGTTCGGTGCCA	58,4	22				G7801	7801-5'	915316 915337	AACAAAGCTAATGACGGGGTCA	58,4	22	620	<i>ClaI</i>	446+174	7801-3'	915935 915914	TGAACGTTTGTTCGTCGGTTC	58,4	22	G7806	7806-5'	917116 917137	TCAAACCAACGAATGCCAA	56,5	22	611	<i>HindIII</i>	583+27	7806-3'	917726 917705	GCTCGTATTATCATCCAGCAGG	62,1	22	G7826	7826-5'	923809 923794	CTCGGCAATTGTCAG	44,0	16	358	<i>ClaI</i>	339+19	7826-3'	923452 923467	GCCAAGGGGGTTATTT	44,0	16	G8523	8523-5'	938207 938192	TCTTCGATACGGCGTG	44,9	16	298	<i>SacI</i>	240+56	8523-3'	937910 937924	TCGGTGGTGGGAATG	44,2	15	G8530	8530-5'	941410 941396	GGTGGACAAGGCCGA	46,7	15	503	<i>HindIII</i>	468+34	8530-3'	940908 940923	CACCCGCTGATTCGT	46,5	16	G8537	8537-5'	942944 942959	GCTCGAGGACAACGCC	56,9	16	363	<i>EcoRV</i>	262+100	8537-3'	943306 943291	CAACCGCCACTCCAG	56,9	16	G8541	8541-5'	945197 945212	TCGAGCGGAAATCTA	40,9	16	440	<i>EcoRI</i>	290+150	8541-3'	945636 945622	CTGGTCCAAAGCAA	40,9	15	N1696	N1696-5'	318762 318749	TGGGCGTAATTACG	36,0	14	370	<i>XhoI</i>	286+84	N1696-3'	318393 318405	CCAAGGTAGCCCC	36,0	13	N1706	N1706-5'	322860 322848	CCGCAAAACCATAG	31,5	13	246	<i>EcoRI</i>	82+164	N1706-3'	322615 322627	GCGTCGTGTCCTCG	34,5	13	N1710	N1710-5'	324349 324364	TAGACTCCCAATGCGA	40,9	16	330	<i>EcoRI</i>	80+250	N1710-3'	324678 324663	TGTAGGATGCCATCGT	40,3	16	N1714	N1714-5'	326260 326245	GAGGATGGACCTATGC	38,1	16	267	<i>HindIII</i>	119+56+92	N1714-3'	325994 326009	CTCTGCGGAATCAATG	40,9	16	N1718	N1718-5'	329634 329620	CGGAGAAGCTGAAGC	39,9	15	280	<i>HindIII</i>	199+81	N1718-3'	329355 329368	CCATCAATTGCGGA	38,9	14																																																																																																																										
G7801	7801-5'	915316 915337	AACAAAGCTAATGACGGGGTCA	58,4	22	620	<i>ClaI</i>	446+174																																																																																																																																																																																																																																																																																																		
	7801-3'	915935 915914	TGAACGTTTGTTCGTCGGTTC	58,4	22				G7806	7806-5'	917116 917137	TCAAACCAACGAATGCCAA	56,5	22	611	<i>HindIII</i>	583+27	7806-3'	917726 917705	GCTCGTATTATCATCCAGCAGG	62,1	22	G7826	7826-5'	923809 923794	CTCGGCAATTGTCAG	44,0	16	358	<i>ClaI</i>	339+19	7826-3'	923452 923467	GCCAAGGGGGTTATTT	44,0	16	G8523	8523-5'	938207 938192	TCTTCGATACGGCGTG	44,9	16	298	<i>SacI</i>	240+56	8523-3'	937910 937924	TCGGTGGTGGGAATG	44,2	15	G8530	8530-5'	941410 941396	GGTGGACAAGGCCGA	46,7	15	503	<i>HindIII</i>	468+34	8530-3'	940908 940923	CACCCGCTGATTCGT	46,5	16	G8537	8537-5'	942944 942959	GCTCGAGGACAACGCC	56,9	16	363	<i>EcoRV</i>	262+100	8537-3'	943306 943291	CAACCGCCACTCCAG	56,9	16	G8541	8541-5'	945197 945212	TCGAGCGGAAATCTA	40,9	16	440	<i>EcoRI</i>	290+150	8541-3'	945636 945622	CTGGTCCAAAGCAA	40,9	15	N1696	N1696-5'	318762 318749	TGGGCGTAATTACG	36,0	14	370	<i>XhoI</i>	286+84	N1696-3'	318393 318405	CCAAGGTAGCCCC	36,0	13	N1706	N1706-5'	322860 322848	CCGCAAAACCATAG	31,5	13	246	<i>EcoRI</i>	82+164	N1706-3'	322615 322627	GCGTCGTGTCCTCG	34,5	13	N1710	N1710-5'	324349 324364	TAGACTCCCAATGCGA	40,9	16	330	<i>EcoRI</i>	80+250	N1710-3'	324678 324663	TGTAGGATGCCATCGT	40,3	16	N1714	N1714-5'	326260 326245	GAGGATGGACCTATGC	38,1	16	267	<i>HindIII</i>	119+56+92	N1714-3'	325994 326009	CTCTGCGGAATCAATG	40,9	16	N1718	N1718-5'	329634 329620	CGGAGAAGCTGAAGC	39,9	15	280	<i>HindIII</i>	199+81	N1718-3'	329355 329368	CCATCAATTGCGGA	38,9	14																																																																																																																																								
G7806	7806-5'	917116 917137	TCAAACCAACGAATGCCAA	56,5	22	611	<i>HindIII</i>	583+27																																																																																																																																																																																																																																																																																																		
	7806-3'	917726 917705	GCTCGTATTATCATCCAGCAGG	62,1	22				G7826	7826-5'	923809 923794	CTCGGCAATTGTCAG	44,0	16	358	<i>ClaI</i>	339+19	7826-3'	923452 923467	GCCAAGGGGGTTATTT	44,0	16	G8523	8523-5'	938207 938192	TCTTCGATACGGCGTG	44,9	16	298	<i>SacI</i>	240+56	8523-3'	937910 937924	TCGGTGGTGGGAATG	44,2	15	G8530	8530-5'	941410 941396	GGTGGACAAGGCCGA	46,7	15	503	<i>HindIII</i>	468+34	8530-3'	940908 940923	CACCCGCTGATTCGT	46,5	16	G8537	8537-5'	942944 942959	GCTCGAGGACAACGCC	56,9	16	363	<i>EcoRV</i>	262+100	8537-3'	943306 943291	CAACCGCCACTCCAG	56,9	16	G8541	8541-5'	945197 945212	TCGAGCGGAAATCTA	40,9	16	440	<i>EcoRI</i>	290+150	8541-3'	945636 945622	CTGGTCCAAAGCAA	40,9	15	N1696	N1696-5'	318762 318749	TGGGCGTAATTACG	36,0	14	370	<i>XhoI</i>	286+84	N1696-3'	318393 318405	CCAAGGTAGCCCC	36,0	13	N1706	N1706-5'	322860 322848	CCGCAAAACCATAG	31,5	13	246	<i>EcoRI</i>	82+164	N1706-3'	322615 322627	GCGTCGTGTCCTCG	34,5	13	N1710	N1710-5'	324349 324364	TAGACTCCCAATGCGA	40,9	16	330	<i>EcoRI</i>	80+250	N1710-3'	324678 324663	TGTAGGATGCCATCGT	40,3	16	N1714	N1714-5'	326260 326245	GAGGATGGACCTATGC	38,1	16	267	<i>HindIII</i>	119+56+92	N1714-3'	325994 326009	CTCTGCGGAATCAATG	40,9	16	N1718	N1718-5'	329634 329620	CGGAGAAGCTGAAGC	39,9	15	280	<i>HindIII</i>	199+81	N1718-3'	329355 329368	CCATCAATTGCGGA	38,9	14																																																																																																																																																						
G7826	7826-5'	923809 923794	CTCGGCAATTGTCAG	44,0	16	358	<i>ClaI</i>	339+19																																																																																																																																																																																																																																																																																																		
	7826-3'	923452 923467	GCCAAGGGGGTTATTT	44,0	16				G8523	8523-5'	938207 938192	TCTTCGATACGGCGTG	44,9	16	298	<i>SacI</i>	240+56	8523-3'	937910 937924	TCGGTGGTGGGAATG	44,2	15	G8530	8530-5'	941410 941396	GGTGGACAAGGCCGA	46,7	15	503	<i>HindIII</i>	468+34	8530-3'	940908 940923	CACCCGCTGATTCGT	46,5	16	G8537	8537-5'	942944 942959	GCTCGAGGACAACGCC	56,9	16	363	<i>EcoRV</i>	262+100	8537-3'	943306 943291	CAACCGCCACTCCAG	56,9	16	G8541	8541-5'	945197 945212	TCGAGCGGAAATCTA	40,9	16	440	<i>EcoRI</i>	290+150	8541-3'	945636 945622	CTGGTCCAAAGCAA	40,9	15	N1696	N1696-5'	318762 318749	TGGGCGTAATTACG	36,0	14	370	<i>XhoI</i>	286+84	N1696-3'	318393 318405	CCAAGGTAGCCCC	36,0	13	N1706	N1706-5'	322860 322848	CCGCAAAACCATAG	31,5	13	246	<i>EcoRI</i>	82+164	N1706-3'	322615 322627	GCGTCGTGTCCTCG	34,5	13	N1710	N1710-5'	324349 324364	TAGACTCCCAATGCGA	40,9	16	330	<i>EcoRI</i>	80+250	N1710-3'	324678 324663	TGTAGGATGCCATCGT	40,3	16	N1714	N1714-5'	326260 326245	GAGGATGGACCTATGC	38,1	16	267	<i>HindIII</i>	119+56+92	N1714-3'	325994 326009	CTCTGCGGAATCAATG	40,9	16	N1718	N1718-5'	329634 329620	CGGAGAAGCTGAAGC	39,9	15	280	<i>HindIII</i>	199+81	N1718-3'	329355 329368	CCATCAATTGCGGA	38,9	14																																																																																																																																																																				
G8523	8523-5'	938207 938192	TCTTCGATACGGCGTG	44,9	16	298	<i>SacI</i>	240+56																																																																																																																																																																																																																																																																																																		
	8523-3'	937910 937924	TCGGTGGTGGGAATG	44,2	15				G8530	8530-5'	941410 941396	GGTGGACAAGGCCGA	46,7	15	503	<i>HindIII</i>	468+34	8530-3'	940908 940923	CACCCGCTGATTCGT	46,5	16	G8537	8537-5'	942944 942959	GCTCGAGGACAACGCC	56,9	16	363	<i>EcoRV</i>	262+100	8537-3'	943306 943291	CAACCGCCACTCCAG	56,9	16	G8541	8541-5'	945197 945212	TCGAGCGGAAATCTA	40,9	16	440	<i>EcoRI</i>	290+150	8541-3'	945636 945622	CTGGTCCAAAGCAA	40,9	15	N1696	N1696-5'	318762 318749	TGGGCGTAATTACG	36,0	14	370	<i>XhoI</i>	286+84	N1696-3'	318393 318405	CCAAGGTAGCCCC	36,0	13	N1706	N1706-5'	322860 322848	CCGCAAAACCATAG	31,5	13	246	<i>EcoRI</i>	82+164	N1706-3'	322615 322627	GCGTCGTGTCCTCG	34,5	13	N1710	N1710-5'	324349 324364	TAGACTCCCAATGCGA	40,9	16	330	<i>EcoRI</i>	80+250	N1710-3'	324678 324663	TGTAGGATGCCATCGT	40,3	16	N1714	N1714-5'	326260 326245	GAGGATGGACCTATGC	38,1	16	267	<i>HindIII</i>	119+56+92	N1714-3'	325994 326009	CTCTGCGGAATCAATG	40,9	16	N1718	N1718-5'	329634 329620	CGGAGAAGCTGAAGC	39,9	15	280	<i>HindIII</i>	199+81	N1718-3'	329355 329368	CCATCAATTGCGGA	38,9	14																																																																																																																																																																																		
G8530	8530-5'	941410 941396	GGTGGACAAGGCCGA	46,7	15	503	<i>HindIII</i>	468+34																																																																																																																																																																																																																																																																																																		
	8530-3'	940908 940923	CACCCGCTGATTCGT	46,5	16				G8537	8537-5'	942944 942959	GCTCGAGGACAACGCC	56,9	16	363	<i>EcoRV</i>	262+100	8537-3'	943306 943291	CAACCGCCACTCCAG	56,9	16	G8541	8541-5'	945197 945212	TCGAGCGGAAATCTA	40,9	16	440	<i>EcoRI</i>	290+150	8541-3'	945636 945622	CTGGTCCAAAGCAA	40,9	15	N1696	N1696-5'	318762 318749	TGGGCGTAATTACG	36,0	14	370	<i>XhoI</i>	286+84	N1696-3'	318393 318405	CCAAGGTAGCCCC	36,0	13	N1706	N1706-5'	322860 322848	CCGCAAAACCATAG	31,5	13	246	<i>EcoRI</i>	82+164	N1706-3'	322615 322627	GCGTCGTGTCCTCG	34,5	13	N1710	N1710-5'	324349 324364	TAGACTCCCAATGCGA	40,9	16	330	<i>EcoRI</i>	80+250	N1710-3'	324678 324663	TGTAGGATGCCATCGT	40,3	16	N1714	N1714-5'	326260 326245	GAGGATGGACCTATGC	38,1	16	267	<i>HindIII</i>	119+56+92	N1714-3'	325994 326009	CTCTGCGGAATCAATG	40,9	16	N1718	N1718-5'	329634 329620	CGGAGAAGCTGAAGC	39,9	15	280	<i>HindIII</i>	199+81	N1718-3'	329355 329368	CCATCAATTGCGGA	38,9	14																																																																																																																																																																																																
G8537	8537-5'	942944 942959	GCTCGAGGACAACGCC	56,9	16	363	<i>EcoRV</i>	262+100																																																																																																																																																																																																																																																																																																		
	8537-3'	943306 943291	CAACCGCCACTCCAG	56,9	16				G8541	8541-5'	945197 945212	TCGAGCGGAAATCTA	40,9	16	440	<i>EcoRI</i>	290+150	8541-3'	945636 945622	CTGGTCCAAAGCAA	40,9	15	N1696	N1696-5'	318762 318749	TGGGCGTAATTACG	36,0	14	370	<i>XhoI</i>	286+84	N1696-3'	318393 318405	CCAAGGTAGCCCC	36,0	13	N1706	N1706-5'	322860 322848	CCGCAAAACCATAG	31,5	13	246	<i>EcoRI</i>	82+164	N1706-3'	322615 322627	GCGTCGTGTCCTCG	34,5	13	N1710	N1710-5'	324349 324364	TAGACTCCCAATGCGA	40,9	16	330	<i>EcoRI</i>	80+250	N1710-3'	324678 324663	TGTAGGATGCCATCGT	40,3	16	N1714	N1714-5'	326260 326245	GAGGATGGACCTATGC	38,1	16	267	<i>HindIII</i>	119+56+92	N1714-3'	325994 326009	CTCTGCGGAATCAATG	40,9	16	N1718	N1718-5'	329634 329620	CGGAGAAGCTGAAGC	39,9	15	280	<i>HindIII</i>	199+81	N1718-3'	329355 329368	CCATCAATTGCGGA	38,9	14																																																																																																																																																																																																														
G8541	8541-5'	945197 945212	TCGAGCGGAAATCTA	40,9	16	440	<i>EcoRI</i>	290+150																																																																																																																																																																																																																																																																																																		
	8541-3'	945636 945622	CTGGTCCAAAGCAA	40,9	15				N1696	N1696-5'	318762 318749	TGGGCGTAATTACG	36,0	14	370	<i>XhoI</i>	286+84	N1696-3'	318393 318405	CCAAGGTAGCCCC	36,0	13	N1706	N1706-5'	322860 322848	CCGCAAAACCATAG	31,5	13	246	<i>EcoRI</i>	82+164	N1706-3'	322615 322627	GCGTCGTGTCCTCG	34,5	13	N1710	N1710-5'	324349 324364	TAGACTCCCAATGCGA	40,9	16	330	<i>EcoRI</i>	80+250	N1710-3'	324678 324663	TGTAGGATGCCATCGT	40,3	16	N1714	N1714-5'	326260 326245	GAGGATGGACCTATGC	38,1	16	267	<i>HindIII</i>	119+56+92	N1714-3'	325994 326009	CTCTGCGGAATCAATG	40,9	16	N1718	N1718-5'	329634 329620	CGGAGAAGCTGAAGC	39,9	15	280	<i>HindIII</i>	199+81	N1718-3'	329355 329368	CCATCAATTGCGGA	38,9	14																																																																																																																																																																																																																												
N1696	N1696-5'	318762 318749	TGGGCGTAATTACG	36,0	14	370	<i>XhoI</i>	286+84																																																																																																																																																																																																																																																																																																		
	N1696-3'	318393 318405	CCAAGGTAGCCCC	36,0	13				N1706	N1706-5'	322860 322848	CCGCAAAACCATAG	31,5	13	246	<i>EcoRI</i>	82+164	N1706-3'	322615 322627	GCGTCGTGTCCTCG	34,5	13	N1710	N1710-5'	324349 324364	TAGACTCCCAATGCGA	40,9	16	330	<i>EcoRI</i>	80+250	N1710-3'	324678 324663	TGTAGGATGCCATCGT	40,3	16	N1714	N1714-5'	326260 326245	GAGGATGGACCTATGC	38,1	16	267	<i>HindIII</i>	119+56+92	N1714-3'	325994 326009	CTCTGCGGAATCAATG	40,9	16	N1718	N1718-5'	329634 329620	CGGAGAAGCTGAAGC	39,9	15	280	<i>HindIII</i>	199+81	N1718-3'	329355 329368	CCATCAATTGCGGA	38,9	14																																																																																																																																																																																																																																										
N1706	N1706-5'	322860 322848	CCGCAAAACCATAG	31,5	13	246	<i>EcoRI</i>	82+164																																																																																																																																																																																																																																																																																																		
	N1706-3'	322615 322627	GCGTCGTGTCCTCG	34,5	13				N1710	N1710-5'	324349 324364	TAGACTCCCAATGCGA	40,9	16	330	<i>EcoRI</i>	80+250	N1710-3'	324678 324663	TGTAGGATGCCATCGT	40,3	16	N1714	N1714-5'	326260 326245	GAGGATGGACCTATGC	38,1	16	267	<i>HindIII</i>	119+56+92	N1714-3'	325994 326009	CTCTGCGGAATCAATG	40,9	16	N1718	N1718-5'	329634 329620	CGGAGAAGCTGAAGC	39,9	15	280	<i>HindIII</i>	199+81	N1718-3'	329355 329368	CCATCAATTGCGGA	38,9	14																																																																																																																																																																																																																																																								
N1710	N1710-5'	324349 324364	TAGACTCCCAATGCGA	40,9	16	330	<i>EcoRI</i>	80+250																																																																																																																																																																																																																																																																																																		
	N1710-3'	324678 324663	TGTAGGATGCCATCGT	40,3	16				N1714	N1714-5'	326260 326245	GAGGATGGACCTATGC	38,1	16	267	<i>HindIII</i>	119+56+92	N1714-3'	325994 326009	CTCTGCGGAATCAATG	40,9	16	N1718	N1718-5'	329634 329620	CGGAGAAGCTGAAGC	39,9	15	280	<i>HindIII</i>	199+81	N1718-3'	329355 329368	CCATCAATTGCGGA	38,9	14																																																																																																																																																																																																																																																																						
N1714	N1714-5'	326260 326245	GAGGATGGACCTATGC	38,1	16	267	<i>HindIII</i>	119+56+92																																																																																																																																																																																																																																																																																																		
	N1714-3'	325994 326009	CTCTGCGGAATCAATG	40,9	16				N1718	N1718-5'	329634 329620	CGGAGAAGCTGAAGC	39,9	15	280	<i>HindIII</i>	199+81	N1718-3'	329355 329368	CCATCAATTGCGGA	38,9	14																																																																																																																																																																																																																																																																																				
N1718	N1718-5'	329634 329620	CGGAGAAGCTGAAGC	39,9	15	280	<i>HindIII</i>	199+81																																																																																																																																																																																																																																																																																																		
	N1718-3'	329355 329368	CCATCAATTGCGGA	38,9	14																																																																																																																																																																																																																																																																																																					

Tabla 3.3. Diseño de los cebadores para las reacciones de amplificación y selección de las enzimas utilizadas en la comprobación de cada producto obtenido Tm: temperatura de *melting*; L.P.: Longitud del producto de PCR; L: longitud del cebador. (continuación).

Nº ORF	Nombre del Cebador	Posición del Cebador Inicio final	SECUENCIA	Tm (°C)	L (nt)	L.P. (pb)	Enzima	Longitud Fragmentos (pb)																																																																																																																																																																																																																																																																																																																														
N1727	N1727-5'	332990 333005	GGGAGTGGGAATGCAC	45,1	16	280	<i>HindIII</i>	161+119																																																																																																																																																																																																																																																																																																																														
	N1727-3'	333271 333256	TTGTTGCTGCTGCTGG	45,9	16				N1735	N1735-5'	338786 338769	GGGTTGCTTGGAAAGTTT	46,5	18	281	<i>HindIII</i>	154+127	N1735-3'	338506 338523	CTCACCAACATTCCGATGG	46,0	18	N1739	N1739-5'	339611 339626	GTCCGGCTCAGAATG	45,6	16	271	<i>AfuI</i>	127+143	N1739-3'	339881 339866	TCGCTGTAAGTCCGCC	46,9	16	N1743	N1743-5'	340469 340484	ATGTACGGAAAGCTGC	40,0	16	252	<i>NdeII</i>	111+141	N1743-3'	340720 340705	TTGTATCAACGTCGCA	39,8	16	N1747	N1747-5'	342081 342066	GTTGGCGTGCAAAGG	47,4	16	351	<i>BamHI</i>	140+211	N1747-3'	341730 341745	GTGTTCCITGGCAGCG	47,1	16	N1751	N1751-5'	342664 342679	GGTGGTTGCTAGAGCA	40,4	16	348	<i>AfuI</i>	182+165	N1751-3'	343011 342996	GAATCTTGGGGTCACC	40,9	16	N1761	N1761-5'	346012 345997	CGTTGGAAGATCAGCA	41,5	16	308	<i>AfuI</i>	97+211	N1761-3'	345705 345720	CAGGATCGCCTCAGA	43,3	16	N1765	N1765-5'	347284 347299	CTATAACCAACCGCCC	42,7	16	252	<i>EcoRI</i>	147+44+36+24	N1765-3'	347535 347520	CTGAGCCTTGGAGGTG	42,7	16	N1773	N1773-5'	349261 349276	CAACATCGGTAATGGT	36,8	16	242	<i>DdeI</i>	75+166	N1773-3'	349502 349487	CGGTTGGGGTAAGAAA	43,0	16	N1774	N1774-5'	349733 349718	ATGGTGGATACCTGG	41,5	16	248	<i>NdeII</i>	177+58+13	N1774-3'	349486 349501	TTTTCTTACCCCAACC	39,0	16	N1777	N1777-5'	350301 350286	GGAAGATGGCTCACGG	45,6	16	282	<i>EcoRI</i>	221+61	N1777-3'	350020 350035	CCACAGGCCCAAACAG	46,7	16	N1780	N1780-5'	351011 351026	ACCAAAAAAATGCA	38,2	16	270	<i>AfuI</i>	187+83	N1780-3'	351280 351265	AGCCTTTTAGGACACC	38,0	16	N1785	N1785-5'	351705 351720	TCATCACATGGAAGTC	32,5	16	265	<i>EcoRI</i>	138+126	N1785-3'	351969 351954	CGGTATAGTAATGATG	27,7	16	N1790	N1790-5'	354074 354059	AAGGAGACATCCGAGG	40,9	16	277	<i>HindIII</i>	200+76	N1790-3'	353798 353813	ACTCTCCCCTGCCAAT	43,5	16	N1795	N1795-5'	357208 357193	TCTTACGGCGCTTAG	44,5	16	351	<i>HindIII</i>	305+45	N1795-3'	356858 356873	TCGGCCATCGAAATCA	47,9	16	N1825	N1825-5'	360133 360148	CCCCACATTTATTTGT	36,5	16	337	<i>Claf</i>	135+201	N1825-3'	360466 360454	ACCACAAAATAGGCG	43,0	16	N1843	N1843-5'	370645 370660	CACGGAATGGCAATGG	47,7	16	334	<i>XhoI</i>	171+162	N1843-3'	370978 370963	TGGACGCTATGGGGC	46,0	16	N1847	N1847-5'	372965 372950	TCGTCGTTGCTTCTCG	44,8	16	312	<i>EcoRV</i>	230+82	N1847-3'	372654 372669	CAAATGGGACGTCGTT	43,2	16	N1850	N1850-5'	374723 374708	GACTGTGGGCTCTGTT	38,3	16	309	<i>HaeIII</i>	114+195	N1850-3'	374415 374430	ATCTTGCTCTCTG	35,4	16	N1858	N1858-5'	375406 375421	GGACAGGCGTACGTA	46,2	16	308	<i>HindIII</i>	95+213	N1858-3'	375713 375698	TGGGGTTAATGCCTCG	46,7	16	N1870	N1870-5'	381918 381933	GTGCGTGCTCCCTATG	44,0	16	291	<i>BamHI</i>	77+213	N1870-3'	382208 382193	GGGCTTTCCTAATCC	44,9	16	N1872	N1872-5'	383059 383074	GCGTCACCATACCTC	42,3	16	272	<i>HindIII</i>	123+148	N1872-3'	383330 383315	TACCACCGCATTCATG	42,1	16	N1875	N1875-5'	385301 385316	AATCTCCATGGCTFAA	36,3	16	293	<i>HindIII</i>	39+253	N1875-3'	385593 385578	CCTTCACCTAAATGCA	37,7	16	N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290	N1879-3'
N1735	N1735-5'	338786 338769	GGGTTGCTTGGAAAGTTT	46,5	18	281	<i>HindIII</i>	154+127																																																																																																																																																																																																																																																																																																																														
	N1735-3'	338506 338523	CTCACCAACATTCCGATGG	46,0	18				N1739	N1739-5'	339611 339626	GTCCGGCTCAGAATG	45,6	16	271	<i>AfuI</i>	127+143	N1739-3'	339881 339866	TCGCTGTAAGTCCGCC	46,9	16	N1743	N1743-5'	340469 340484	ATGTACGGAAAGCTGC	40,0	16	252	<i>NdeII</i>	111+141	N1743-3'	340720 340705	TTGTATCAACGTCGCA	39,8	16	N1747	N1747-5'	342081 342066	GTTGGCGTGCAAAGG	47,4	16	351	<i>BamHI</i>	140+211	N1747-3'	341730 341745	GTGTTCCITGGCAGCG	47,1	16	N1751	N1751-5'	342664 342679	GGTGGTTGCTAGAGCA	40,4	16	348	<i>AfuI</i>	182+165	N1751-3'	343011 342996	GAATCTTGGGGTCACC	40,9	16	N1761	N1761-5'	346012 345997	CGTTGGAAGATCAGCA	41,5	16	308	<i>AfuI</i>	97+211	N1761-3'	345705 345720	CAGGATCGCCTCAGA	43,3	16	N1765	N1765-5'	347284 347299	CTATAACCAACCGCCC	42,7	16	252	<i>EcoRI</i>	147+44+36+24	N1765-3'	347535 347520	CTGAGCCTTGGAGGTG	42,7	16	N1773	N1773-5'	349261 349276	CAACATCGGTAATGGT	36,8	16	242	<i>DdeI</i>	75+166	N1773-3'	349502 349487	CGGTTGGGGTAAGAAA	43,0	16	N1774	N1774-5'	349733 349718	ATGGTGGATACCTGG	41,5	16	248	<i>NdeII</i>	177+58+13	N1774-3'	349486 349501	TTTTCTTACCCCAACC	39,0	16	N1777	N1777-5'	350301 350286	GGAAGATGGCTCACGG	45,6	16	282	<i>EcoRI</i>	221+61	N1777-3'	350020 350035	CCACAGGCCCAAACAG	46,7	16	N1780	N1780-5'	351011 351026	ACCAAAAAAATGCA	38,2	16	270	<i>AfuI</i>	187+83	N1780-3'	351280 351265	AGCCTTTTAGGACACC	38,0	16	N1785	N1785-5'	351705 351720	TCATCACATGGAAGTC	32,5	16	265	<i>EcoRI</i>	138+126	N1785-3'	351969 351954	CGGTATAGTAATGATG	27,7	16	N1790	N1790-5'	354074 354059	AAGGAGACATCCGAGG	40,9	16	277	<i>HindIII</i>	200+76	N1790-3'	353798 353813	ACTCTCCCCTGCCAAT	43,5	16	N1795	N1795-5'	357208 357193	TCTTACGGCGCTTAG	44,5	16	351	<i>HindIII</i>	305+45	N1795-3'	356858 356873	TCGGCCATCGAAATCA	47,9	16	N1825	N1825-5'	360133 360148	CCCCACATTTATTTGT	36,5	16	337	<i>Claf</i>	135+201	N1825-3'	360466 360454	ACCACAAAATAGGCG	43,0	16	N1843	N1843-5'	370645 370660	CACGGAATGGCAATGG	47,7	16	334	<i>XhoI</i>	171+162	N1843-3'	370978 370963	TGGACGCTATGGGGC	46,0	16	N1847	N1847-5'	372965 372950	TCGTCGTTGCTTCTCG	44,8	16	312	<i>EcoRV</i>	230+82	N1847-3'	372654 372669	CAAATGGGACGTCGTT	43,2	16	N1850	N1850-5'	374723 374708	GACTGTGGGCTCTGTT	38,3	16	309	<i>HaeIII</i>	114+195	N1850-3'	374415 374430	ATCTTGCTCTCTG	35,4	16	N1858	N1858-5'	375406 375421	GGACAGGCGTACGTA	46,2	16	308	<i>HindIII</i>	95+213	N1858-3'	375713 375698	TGGGGTTAATGCCTCG	46,7	16	N1870	N1870-5'	381918 381933	GTGCGTGCTCCCTATG	44,0	16	291	<i>BamHI</i>	77+213	N1870-3'	382208 382193	GGGCTTTCCTAATCC	44,9	16	N1872	N1872-5'	383059 383074	GCGTCACCATACCTC	42,3	16	272	<i>HindIII</i>	123+148	N1872-3'	383330 383315	TACCACCGCATTCATG	42,1	16	N1875	N1875-5'	385301 385316	AATCTCCATGGCTFAA	36,3	16	293	<i>HindIII</i>	39+253	N1875-3'	385593 385578	CCTTCACCTAAATGCA	37,7	16	N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290	N1879-3'	387681 387666	GGCGAAGATACCATGG	42,9	16										
N1739	N1739-5'	339611 339626	GTCCGGCTCAGAATG	45,6	16	271	<i>AfuI</i>	127+143																																																																																																																																																																																																																																																																																																																														
	N1739-3'	339881 339866	TCGCTGTAAGTCCGCC	46,9	16				N1743	N1743-5'	340469 340484	ATGTACGGAAAGCTGC	40,0	16	252	<i>NdeII</i>	111+141	N1743-3'	340720 340705	TTGTATCAACGTCGCA	39,8	16	N1747	N1747-5'	342081 342066	GTTGGCGTGCAAAGG	47,4	16	351	<i>BamHI</i>	140+211	N1747-3'	341730 341745	GTGTTCCITGGCAGCG	47,1	16	N1751	N1751-5'	342664 342679	GGTGGTTGCTAGAGCA	40,4	16	348	<i>AfuI</i>	182+165	N1751-3'	343011 342996	GAATCTTGGGGTCACC	40,9	16	N1761	N1761-5'	346012 345997	CGTTGGAAGATCAGCA	41,5	16	308	<i>AfuI</i>	97+211	N1761-3'	345705 345720	CAGGATCGCCTCAGA	43,3	16	N1765	N1765-5'	347284 347299	CTATAACCAACCGCCC	42,7	16	252	<i>EcoRI</i>	147+44+36+24	N1765-3'	347535 347520	CTGAGCCTTGGAGGTG	42,7	16	N1773	N1773-5'	349261 349276	CAACATCGGTAATGGT	36,8	16	242	<i>DdeI</i>	75+166	N1773-3'	349502 349487	CGGTTGGGGTAAGAAA	43,0	16	N1774	N1774-5'	349733 349718	ATGGTGGATACCTGG	41,5	16	248	<i>NdeII</i>	177+58+13	N1774-3'	349486 349501	TTTTCTTACCCCAACC	39,0	16	N1777	N1777-5'	350301 350286	GGAAGATGGCTCACGG	45,6	16	282	<i>EcoRI</i>	221+61	N1777-3'	350020 350035	CCACAGGCCCAAACAG	46,7	16	N1780	N1780-5'	351011 351026	ACCAAAAAAATGCA	38,2	16	270	<i>AfuI</i>	187+83	N1780-3'	351280 351265	AGCCTTTTAGGACACC	38,0	16	N1785	N1785-5'	351705 351720	TCATCACATGGAAGTC	32,5	16	265	<i>EcoRI</i>	138+126	N1785-3'	351969 351954	CGGTATAGTAATGATG	27,7	16	N1790	N1790-5'	354074 354059	AAGGAGACATCCGAGG	40,9	16	277	<i>HindIII</i>	200+76	N1790-3'	353798 353813	ACTCTCCCCTGCCAAT	43,5	16	N1795	N1795-5'	357208 357193	TCTTACGGCGCTTAG	44,5	16	351	<i>HindIII</i>	305+45	N1795-3'	356858 356873	TCGGCCATCGAAATCA	47,9	16	N1825	N1825-5'	360133 360148	CCCCACATTTATTTGT	36,5	16	337	<i>Claf</i>	135+201	N1825-3'	360466 360454	ACCACAAAATAGGCG	43,0	16	N1843	N1843-5'	370645 370660	CACGGAATGGCAATGG	47,7	16	334	<i>XhoI</i>	171+162	N1843-3'	370978 370963	TGGACGCTATGGGGC	46,0	16	N1847	N1847-5'	372965 372950	TCGTCGTTGCTTCTCG	44,8	16	312	<i>EcoRV</i>	230+82	N1847-3'	372654 372669	CAAATGGGACGTCGTT	43,2	16	N1850	N1850-5'	374723 374708	GACTGTGGGCTCTGTT	38,3	16	309	<i>HaeIII</i>	114+195	N1850-3'	374415 374430	ATCTTGCTCTCTG	35,4	16	N1858	N1858-5'	375406 375421	GGACAGGCGTACGTA	46,2	16	308	<i>HindIII</i>	95+213	N1858-3'	375713 375698	TGGGGTTAATGCCTCG	46,7	16	N1870	N1870-5'	381918 381933	GTGCGTGCTCCCTATG	44,0	16	291	<i>BamHI</i>	77+213	N1870-3'	382208 382193	GGGCTTTCCTAATCC	44,9	16	N1872	N1872-5'	383059 383074	GCGTCACCATACCTC	42,3	16	272	<i>HindIII</i>	123+148	N1872-3'	383330 383315	TACCACCGCATTCATG	42,1	16	N1875	N1875-5'	385301 385316	AATCTCCATGGCTFAA	36,3	16	293	<i>HindIII</i>	39+253	N1875-3'	385593 385578	CCTTCACCTAAATGCA	37,7	16	N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290	N1879-3'	387681 387666	GGCGAAGATACCATGG	42,9	16																								
N1743	N1743-5'	340469 340484	ATGTACGGAAAGCTGC	40,0	16	252	<i>NdeII</i>	111+141																																																																																																																																																																																																																																																																																																																														
	N1743-3'	340720 340705	TTGTATCAACGTCGCA	39,8	16				N1747	N1747-5'	342081 342066	GTTGGCGTGCAAAGG	47,4	16	351	<i>BamHI</i>	140+211	N1747-3'	341730 341745	GTGTTCCITGGCAGCG	47,1	16	N1751	N1751-5'	342664 342679	GGTGGTTGCTAGAGCA	40,4	16	348	<i>AfuI</i>	182+165	N1751-3'	343011 342996	GAATCTTGGGGTCACC	40,9	16	N1761	N1761-5'	346012 345997	CGTTGGAAGATCAGCA	41,5	16	308	<i>AfuI</i>	97+211	N1761-3'	345705 345720	CAGGATCGCCTCAGA	43,3	16	N1765	N1765-5'	347284 347299	CTATAACCAACCGCCC	42,7	16	252	<i>EcoRI</i>	147+44+36+24	N1765-3'	347535 347520	CTGAGCCTTGGAGGTG	42,7	16	N1773	N1773-5'	349261 349276	CAACATCGGTAATGGT	36,8	16	242	<i>DdeI</i>	75+166	N1773-3'	349502 349487	CGGTTGGGGTAAGAAA	43,0	16	N1774	N1774-5'	349733 349718	ATGGTGGATACCTGG	41,5	16	248	<i>NdeII</i>	177+58+13	N1774-3'	349486 349501	TTTTCTTACCCCAACC	39,0	16	N1777	N1777-5'	350301 350286	GGAAGATGGCTCACGG	45,6	16	282	<i>EcoRI</i>	221+61	N1777-3'	350020 350035	CCACAGGCCCAAACAG	46,7	16	N1780	N1780-5'	351011 351026	ACCAAAAAAATGCA	38,2	16	270	<i>AfuI</i>	187+83	N1780-3'	351280 351265	AGCCTTTTAGGACACC	38,0	16	N1785	N1785-5'	351705 351720	TCATCACATGGAAGTC	32,5	16	265	<i>EcoRI</i>	138+126	N1785-3'	351969 351954	CGGTATAGTAATGATG	27,7	16	N1790	N1790-5'	354074 354059	AAGGAGACATCCGAGG	40,9	16	277	<i>HindIII</i>	200+76	N1790-3'	353798 353813	ACTCTCCCCTGCCAAT	43,5	16	N1795	N1795-5'	357208 357193	TCTTACGGCGCTTAG	44,5	16	351	<i>HindIII</i>	305+45	N1795-3'	356858 356873	TCGGCCATCGAAATCA	47,9	16	N1825	N1825-5'	360133 360148	CCCCACATTTATTTGT	36,5	16	337	<i>Claf</i>	135+201	N1825-3'	360466 360454	ACCACAAAATAGGCG	43,0	16	N1843	N1843-5'	370645 370660	CACGGAATGGCAATGG	47,7	16	334	<i>XhoI</i>	171+162	N1843-3'	370978 370963	TGGACGCTATGGGGC	46,0	16	N1847	N1847-5'	372965 372950	TCGTCGTTGCTTCTCG	44,8	16	312	<i>EcoRV</i>	230+82	N1847-3'	372654 372669	CAAATGGGACGTCGTT	43,2	16	N1850	N1850-5'	374723 374708	GACTGTGGGCTCTGTT	38,3	16	309	<i>HaeIII</i>	114+195	N1850-3'	374415 374430	ATCTTGCTCTCTG	35,4	16	N1858	N1858-5'	375406 375421	GGACAGGCGTACGTA	46,2	16	308	<i>HindIII</i>	95+213	N1858-3'	375713 375698	TGGGGTTAATGCCTCG	46,7	16	N1870	N1870-5'	381918 381933	GTGCGTGCTCCCTATG	44,0	16	291	<i>BamHI</i>	77+213	N1870-3'	382208 382193	GGGCTTTCCTAATCC	44,9	16	N1872	N1872-5'	383059 383074	GCGTCACCATACCTC	42,3	16	272	<i>HindIII</i>	123+148	N1872-3'	383330 383315	TACCACCGCATTCATG	42,1	16	N1875	N1875-5'	385301 385316	AATCTCCATGGCTFAA	36,3	16	293	<i>HindIII</i>	39+253	N1875-3'	385593 385578	CCTTCACCTAAATGCA	37,7	16	N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290	N1879-3'	387681 387666	GGCGAAGATACCATGG	42,9	16																																						
N1747	N1747-5'	342081 342066	GTTGGCGTGCAAAGG	47,4	16	351	<i>BamHI</i>	140+211																																																																																																																																																																																																																																																																																																																														
	N1747-3'	341730 341745	GTGTTCCITGGCAGCG	47,1	16				N1751	N1751-5'	342664 342679	GGTGGTTGCTAGAGCA	40,4	16	348	<i>AfuI</i>	182+165	N1751-3'	343011 342996	GAATCTTGGGGTCACC	40,9	16	N1761	N1761-5'	346012 345997	CGTTGGAAGATCAGCA	41,5	16	308	<i>AfuI</i>	97+211	N1761-3'	345705 345720	CAGGATCGCCTCAGA	43,3	16	N1765	N1765-5'	347284 347299	CTATAACCAACCGCCC	42,7	16	252	<i>EcoRI</i>	147+44+36+24	N1765-3'	347535 347520	CTGAGCCTTGGAGGTG	42,7	16	N1773	N1773-5'	349261 349276	CAACATCGGTAATGGT	36,8	16	242	<i>DdeI</i>	75+166	N1773-3'	349502 349487	CGGTTGGGGTAAGAAA	43,0	16	N1774	N1774-5'	349733 349718	ATGGTGGATACCTGG	41,5	16	248	<i>NdeII</i>	177+58+13	N1774-3'	349486 349501	TTTTCTTACCCCAACC	39,0	16	N1777	N1777-5'	350301 350286	GGAAGATGGCTCACGG	45,6	16	282	<i>EcoRI</i>	221+61	N1777-3'	350020 350035	CCACAGGCCCAAACAG	46,7	16	N1780	N1780-5'	351011 351026	ACCAAAAAAATGCA	38,2	16	270	<i>AfuI</i>	187+83	N1780-3'	351280 351265	AGCCTTTTAGGACACC	38,0	16	N1785	N1785-5'	351705 351720	TCATCACATGGAAGTC	32,5	16	265	<i>EcoRI</i>	138+126	N1785-3'	351969 351954	CGGTATAGTAATGATG	27,7	16	N1790	N1790-5'	354074 354059	AAGGAGACATCCGAGG	40,9	16	277	<i>HindIII</i>	200+76	N1790-3'	353798 353813	ACTCTCCCCTGCCAAT	43,5	16	N1795	N1795-5'	357208 357193	TCTTACGGCGCTTAG	44,5	16	351	<i>HindIII</i>	305+45	N1795-3'	356858 356873	TCGGCCATCGAAATCA	47,9	16	N1825	N1825-5'	360133 360148	CCCCACATTTATTTGT	36,5	16	337	<i>Claf</i>	135+201	N1825-3'	360466 360454	ACCACAAAATAGGCG	43,0	16	N1843	N1843-5'	370645 370660	CACGGAATGGCAATGG	47,7	16	334	<i>XhoI</i>	171+162	N1843-3'	370978 370963	TGGACGCTATGGGGC	46,0	16	N1847	N1847-5'	372965 372950	TCGTCGTTGCTTCTCG	44,8	16	312	<i>EcoRV</i>	230+82	N1847-3'	372654 372669	CAAATGGGACGTCGTT	43,2	16	N1850	N1850-5'	374723 374708	GACTGTGGGCTCTGTT	38,3	16	309	<i>HaeIII</i>	114+195	N1850-3'	374415 374430	ATCTTGCTCTCTG	35,4	16	N1858	N1858-5'	375406 375421	GGACAGGCGTACGTA	46,2	16	308	<i>HindIII</i>	95+213	N1858-3'	375713 375698	TGGGGTTAATGCCTCG	46,7	16	N1870	N1870-5'	381918 381933	GTGCGTGCTCCCTATG	44,0	16	291	<i>BamHI</i>	77+213	N1870-3'	382208 382193	GGGCTTTCCTAATCC	44,9	16	N1872	N1872-5'	383059 383074	GCGTCACCATACCTC	42,3	16	272	<i>HindIII</i>	123+148	N1872-3'	383330 383315	TACCACCGCATTCATG	42,1	16	N1875	N1875-5'	385301 385316	AATCTCCATGGCTFAA	36,3	16	293	<i>HindIII</i>	39+253	N1875-3'	385593 385578	CCTTCACCTAAATGCA	37,7	16	N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290	N1879-3'	387681 387666	GGCGAAGATACCATGG	42,9	16																																																				
N1751	N1751-5'	342664 342679	GGTGGTTGCTAGAGCA	40,4	16	348	<i>AfuI</i>	182+165																																																																																																																																																																																																																																																																																																																														
	N1751-3'	343011 342996	GAATCTTGGGGTCACC	40,9	16				N1761	N1761-5'	346012 345997	CGTTGGAAGATCAGCA	41,5	16	308	<i>AfuI</i>	97+211	N1761-3'	345705 345720	CAGGATCGCCTCAGA	43,3	16	N1765	N1765-5'	347284 347299	CTATAACCAACCGCCC	42,7	16	252	<i>EcoRI</i>	147+44+36+24	N1765-3'	347535 347520	CTGAGCCTTGGAGGTG	42,7	16	N1773	N1773-5'	349261 349276	CAACATCGGTAATGGT	36,8	16	242	<i>DdeI</i>	75+166	N1773-3'	349502 349487	CGGTTGGGGTAAGAAA	43,0	16	N1774	N1774-5'	349733 349718	ATGGTGGATACCTGG	41,5	16	248	<i>NdeII</i>	177+58+13	N1774-3'	349486 349501	TTTTCTTACCCCAACC	39,0	16	N1777	N1777-5'	350301 350286	GGAAGATGGCTCACGG	45,6	16	282	<i>EcoRI</i>	221+61	N1777-3'	350020 350035	CCACAGGCCCAAACAG	46,7	16	N1780	N1780-5'	351011 351026	ACCAAAAAAATGCA	38,2	16	270	<i>AfuI</i>	187+83	N1780-3'	351280 351265	AGCCTTTTAGGACACC	38,0	16	N1785	N1785-5'	351705 351720	TCATCACATGGAAGTC	32,5	16	265	<i>EcoRI</i>	138+126	N1785-3'	351969 351954	CGGTATAGTAATGATG	27,7	16	N1790	N1790-5'	354074 354059	AAGGAGACATCCGAGG	40,9	16	277	<i>HindIII</i>	200+76	N1790-3'	353798 353813	ACTCTCCCCTGCCAAT	43,5	16	N1795	N1795-5'	357208 357193	TCTTACGGCGCTTAG	44,5	16	351	<i>HindIII</i>	305+45	N1795-3'	356858 356873	TCGGCCATCGAAATCA	47,9	16	N1825	N1825-5'	360133 360148	CCCCACATTTATTTGT	36,5	16	337	<i>Claf</i>	135+201	N1825-3'	360466 360454	ACCACAAAATAGGCG	43,0	16	N1843	N1843-5'	370645 370660	CACGGAATGGCAATGG	47,7	16	334	<i>XhoI</i>	171+162	N1843-3'	370978 370963	TGGACGCTATGGGGC	46,0	16	N1847	N1847-5'	372965 372950	TCGTCGTTGCTTCTCG	44,8	16	312	<i>EcoRV</i>	230+82	N1847-3'	372654 372669	CAAATGGGACGTCGTT	43,2	16	N1850	N1850-5'	374723 374708	GACTGTGGGCTCTGTT	38,3	16	309	<i>HaeIII</i>	114+195	N1850-3'	374415 374430	ATCTTGCTCTCTG	35,4	16	N1858	N1858-5'	375406 375421	GGACAGGCGTACGTA	46,2	16	308	<i>HindIII</i>	95+213	N1858-3'	375713 375698	TGGGGTTAATGCCTCG	46,7	16	N1870	N1870-5'	381918 381933	GTGCGTGCTCCCTATG	44,0	16	291	<i>BamHI</i>	77+213	N1870-3'	382208 382193	GGGCTTTCCTAATCC	44,9	16	N1872	N1872-5'	383059 383074	GCGTCACCATACCTC	42,3	16	272	<i>HindIII</i>	123+148	N1872-3'	383330 383315	TACCACCGCATTCATG	42,1	16	N1875	N1875-5'	385301 385316	AATCTCCATGGCTFAA	36,3	16	293	<i>HindIII</i>	39+253	N1875-3'	385593 385578	CCTTCACCTAAATGCA	37,7	16	N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290	N1879-3'	387681 387666	GGCGAAGATACCATGG	42,9	16																																																																		
N1761	N1761-5'	346012 345997	CGTTGGAAGATCAGCA	41,5	16	308	<i>AfuI</i>	97+211																																																																																																																																																																																																																																																																																																																														
	N1761-3'	345705 345720	CAGGATCGCCTCAGA	43,3	16				N1765	N1765-5'	347284 347299	CTATAACCAACCGCCC	42,7	16	252	<i>EcoRI</i>	147+44+36+24	N1765-3'	347535 347520	CTGAGCCTTGGAGGTG	42,7	16	N1773	N1773-5'	349261 349276	CAACATCGGTAATGGT	36,8	16	242	<i>DdeI</i>	75+166	N1773-3'	349502 349487	CGGTTGGGGTAAGAAA	43,0	16	N1774	N1774-5'	349733 349718	ATGGTGGATACCTGG	41,5	16	248	<i>NdeII</i>	177+58+13	N1774-3'	349486 349501	TTTTCTTACCCCAACC	39,0	16	N1777	N1777-5'	350301 350286	GGAAGATGGCTCACGG	45,6	16	282	<i>EcoRI</i>	221+61	N1777-3'	350020 350035	CCACAGGCCCAAACAG	46,7	16	N1780	N1780-5'	351011 351026	ACCAAAAAAATGCA	38,2	16	270	<i>AfuI</i>	187+83	N1780-3'	351280 351265	AGCCTTTTAGGACACC	38,0	16	N1785	N1785-5'	351705 351720	TCATCACATGGAAGTC	32,5	16	265	<i>EcoRI</i>	138+126	N1785-3'	351969 351954	CGGTATAGTAATGATG	27,7	16	N1790	N1790-5'	354074 354059	AAGGAGACATCCGAGG	40,9	16	277	<i>HindIII</i>	200+76	N1790-3'	353798 353813	ACTCTCCCCTGCCAAT	43,5	16	N1795	N1795-5'	357208 357193	TCTTACGGCGCTTAG	44,5	16	351	<i>HindIII</i>	305+45	N1795-3'	356858 356873	TCGGCCATCGAAATCA	47,9	16	N1825	N1825-5'	360133 360148	CCCCACATTTATTTGT	36,5	16	337	<i>Claf</i>	135+201	N1825-3'	360466 360454	ACCACAAAATAGGCG	43,0	16	N1843	N1843-5'	370645 370660	CACGGAATGGCAATGG	47,7	16	334	<i>XhoI</i>	171+162	N1843-3'	370978 370963	TGGACGCTATGGGGC	46,0	16	N1847	N1847-5'	372965 372950	TCGTCGTTGCTTCTCG	44,8	16	312	<i>EcoRV</i>	230+82	N1847-3'	372654 372669	CAAATGGGACGTCGTT	43,2	16	N1850	N1850-5'	374723 374708	GACTGTGGGCTCTGTT	38,3	16	309	<i>HaeIII</i>	114+195	N1850-3'	374415 374430	ATCTTGCTCTCTG	35,4	16	N1858	N1858-5'	375406 375421	GGACAGGCGTACGTA	46,2	16	308	<i>HindIII</i>	95+213	N1858-3'	375713 375698	TGGGGTTAATGCCTCG	46,7	16	N1870	N1870-5'	381918 381933	GTGCGTGCTCCCTATG	44,0	16	291	<i>BamHI</i>	77+213	N1870-3'	382208 382193	GGGCTTTCCTAATCC	44,9	16	N1872	N1872-5'	383059 383074	GCGTCACCATACCTC	42,3	16	272	<i>HindIII</i>	123+148	N1872-3'	383330 383315	TACCACCGCATTCATG	42,1	16	N1875	N1875-5'	385301 385316	AATCTCCATGGCTFAA	36,3	16	293	<i>HindIII</i>	39+253	N1875-3'	385593 385578	CCTTCACCTAAATGCA	37,7	16	N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290	N1879-3'	387681 387666	GGCGAAGATACCATGG	42,9	16																																																																																
N1765	N1765-5'	347284 347299	CTATAACCAACCGCCC	42,7	16	252	<i>EcoRI</i>	147+44+36+24																																																																																																																																																																																																																																																																																																																														
	N1765-3'	347535 347520	CTGAGCCTTGGAGGTG	42,7	16				N1773	N1773-5'	349261 349276	CAACATCGGTAATGGT	36,8	16	242	<i>DdeI</i>	75+166	N1773-3'	349502 349487	CGGTTGGGGTAAGAAA	43,0	16	N1774	N1774-5'	349733 349718	ATGGTGGATACCTGG	41,5	16	248	<i>NdeII</i>	177+58+13	N1774-3'	349486 349501	TTTTCTTACCCCAACC	39,0	16	N1777	N1777-5'	350301 350286	GGAAGATGGCTCACGG	45,6	16	282	<i>EcoRI</i>	221+61	N1777-3'	350020 350035	CCACAGGCCCAAACAG	46,7	16	N1780	N1780-5'	351011 351026	ACCAAAAAAATGCA	38,2	16	270	<i>AfuI</i>	187+83	N1780-3'	351280 351265	AGCCTTTTAGGACACC	38,0	16	N1785	N1785-5'	351705 351720	TCATCACATGGAAGTC	32,5	16	265	<i>EcoRI</i>	138+126	N1785-3'	351969 351954	CGGTATAGTAATGATG	27,7	16	N1790	N1790-5'	354074 354059	AAGGAGACATCCGAGG	40,9	16	277	<i>HindIII</i>	200+76	N1790-3'	353798 353813	ACTCTCCCCTGCCAAT	43,5	16	N1795	N1795-5'	357208 357193	TCTTACGGCGCTTAG	44,5	16	351	<i>HindIII</i>	305+45	N1795-3'	356858 356873	TCGGCCATCGAAATCA	47,9	16	N1825	N1825-5'	360133 360148	CCCCACATTTATTTGT	36,5	16	337	<i>Claf</i>	135+201	N1825-3'	360466 360454	ACCACAAAATAGGCG	43,0	16	N1843	N1843-5'	370645 370660	CACGGAATGGCAATGG	47,7	16	334	<i>XhoI</i>	171+162	N1843-3'	370978 370963	TGGACGCTATGGGGC	46,0	16	N1847	N1847-5'	372965 372950	TCGTCGTTGCTTCTCG	44,8	16	312	<i>EcoRV</i>	230+82	N1847-3'	372654 372669	CAAATGGGACGTCGTT	43,2	16	N1850	N1850-5'	374723 374708	GACTGTGGGCTCTGTT	38,3	16	309	<i>HaeIII</i>	114+195	N1850-3'	374415 374430	ATCTTGCTCTCTG	35,4	16	N1858	N1858-5'	375406 375421	GGACAGGCGTACGTA	46,2	16	308	<i>HindIII</i>	95+213	N1858-3'	375713 375698	TGGGGTTAATGCCTCG	46,7	16	N1870	N1870-5'	381918 381933	GTGCGTGCTCCCTATG	44,0	16	291	<i>BamHI</i>	77+213	N1870-3'	382208 382193	GGGCTTTCCTAATCC	44,9	16	N1872	N1872-5'	383059 383074	GCGTCACCATACCTC	42,3	16	272	<i>HindIII</i>	123+148	N1872-3'	383330 383315	TACCACCGCATTCATG	42,1	16	N1875	N1875-5'	385301 385316	AATCTCCATGGCTFAA	36,3	16	293	<i>HindIII</i>	39+253	N1875-3'	385593 385578	CCTTCACCTAAATGCA	37,7	16	N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290	N1879-3'	387681 387666	GGCGAAGATACCATGG	42,9	16																																																																																														
N1773	N1773-5'	349261 349276	CAACATCGGTAATGGT	36,8	16	242	<i>DdeI</i>	75+166																																																																																																																																																																																																																																																																																																																														
	N1773-3'	349502 349487	CGGTTGGGGTAAGAAA	43,0	16				N1774	N1774-5'	349733 349718	ATGGTGGATACCTGG	41,5	16	248	<i>NdeII</i>	177+58+13	N1774-3'	349486 349501	TTTTCTTACCCCAACC	39,0	16	N1777	N1777-5'	350301 350286	GGAAGATGGCTCACGG	45,6	16	282	<i>EcoRI</i>	221+61	N1777-3'	350020 350035	CCACAGGCCCAAACAG	46,7	16	N1780	N1780-5'	351011 351026	ACCAAAAAAATGCA	38,2	16	270	<i>AfuI</i>	187+83	N1780-3'	351280 351265	AGCCTTTTAGGACACC	38,0	16	N1785	N1785-5'	351705 351720	TCATCACATGGAAGTC	32,5	16	265	<i>EcoRI</i>	138+126	N1785-3'	351969 351954	CGGTATAGTAATGATG	27,7	16	N1790	N1790-5'	354074 354059	AAGGAGACATCCGAGG	40,9	16	277	<i>HindIII</i>	200+76	N1790-3'	353798 353813	ACTCTCCCCTGCCAAT	43,5	16	N1795	N1795-5'	357208 357193	TCTTACGGCGCTTAG	44,5	16	351	<i>HindIII</i>	305+45	N1795-3'	356858 356873	TCGGCCATCGAAATCA	47,9	16	N1825	N1825-5'	360133 360148	CCCCACATTTATTTGT	36,5	16	337	<i>Claf</i>	135+201	N1825-3'	360466 360454	ACCACAAAATAGGCG	43,0	16	N1843	N1843-5'	370645 370660	CACGGAATGGCAATGG	47,7	16	334	<i>XhoI</i>	171+162	N1843-3'	370978 370963	TGGACGCTATGGGGC	46,0	16	N1847	N1847-5'	372965 372950	TCGTCGTTGCTTCTCG	44,8	16	312	<i>EcoRV</i>	230+82	N1847-3'	372654 372669	CAAATGGGACGTCGTT	43,2	16	N1850	N1850-5'	374723 374708	GACTGTGGGCTCTGTT	38,3	16	309	<i>HaeIII</i>	114+195	N1850-3'	374415 374430	ATCTTGCTCTCTG	35,4	16	N1858	N1858-5'	375406 375421	GGACAGGCGTACGTA	46,2	16	308	<i>HindIII</i>	95+213	N1858-3'	375713 375698	TGGGGTTAATGCCTCG	46,7	16	N1870	N1870-5'	381918 381933	GTGCGTGCTCCCTATG	44,0	16	291	<i>BamHI</i>	77+213	N1870-3'	382208 382193	GGGCTTTCCTAATCC	44,9	16	N1872	N1872-5'	383059 383074	GCGTCACCATACCTC	42,3	16	272	<i>HindIII</i>	123+148	N1872-3'	383330 383315	TACCACCGCATTCATG	42,1	16	N1875	N1875-5'	385301 385316	AATCTCCATGGCTFAA	36,3	16	293	<i>HindIII</i>	39+253	N1875-3'	385593 385578	CCTTCACCTAAATGCA	37,7	16	N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290	N1879-3'	387681 387666	GGCGAAGATACCATGG	42,9	16																																																																																																												
N1774	N1774-5'	349733 349718	ATGGTGGATACCTGG	41,5	16	248	<i>NdeII</i>	177+58+13																																																																																																																																																																																																																																																																																																																														
	N1774-3'	349486 349501	TTTTCTTACCCCAACC	39,0	16				N1777	N1777-5'	350301 350286	GGAAGATGGCTCACGG	45,6	16	282	<i>EcoRI</i>	221+61	N1777-3'	350020 350035	CCACAGGCCCAAACAG	46,7	16	N1780	N1780-5'	351011 351026	ACCAAAAAAATGCA	38,2	16	270	<i>AfuI</i>	187+83	N1780-3'	351280 351265	AGCCTTTTAGGACACC	38,0	16	N1785	N1785-5'	351705 351720	TCATCACATGGAAGTC	32,5	16	265	<i>EcoRI</i>	138+126	N1785-3'	351969 351954	CGGTATAGTAATGATG	27,7	16	N1790	N1790-5'	354074 354059	AAGGAGACATCCGAGG	40,9	16	277	<i>HindIII</i>	200+76	N1790-3'	353798 353813	ACTCTCCCCTGCCAAT	43,5	16	N1795	N1795-5'	357208 357193	TCTTACGGCGCTTAG	44,5	16	351	<i>HindIII</i>	305+45	N1795-3'	356858 356873	TCGGCCATCGAAATCA	47,9	16	N1825	N1825-5'	360133 360148	CCCCACATTTATTTGT	36,5	16	337	<i>Claf</i>	135+201	N1825-3'	360466 360454	ACCACAAAATAGGCG	43,0	16	N1843	N1843-5'	370645 370660	CACGGAATGGCAATGG	47,7	16	334	<i>XhoI</i>	171+162	N1843-3'	370978 370963	TGGACGCTATGGGGC	46,0	16	N1847	N1847-5'	372965 372950	TCGTCGTTGCTTCTCG	44,8	16	312	<i>EcoRV</i>	230+82	N1847-3'	372654 372669	CAAATGGGACGTCGTT	43,2	16	N1850	N1850-5'	374723 374708	GACTGTGGGCTCTGTT	38,3	16	309	<i>HaeIII</i>	114+195	N1850-3'	374415 374430	ATCTTGCTCTCTG	35,4	16	N1858	N1858-5'	375406 375421	GGACAGGCGTACGTA	46,2	16	308	<i>HindIII</i>	95+213	N1858-3'	375713 375698	TGGGGTTAATGCCTCG	46,7	16	N1870	N1870-5'	381918 381933	GTGCGTGCTCCCTATG	44,0	16	291	<i>BamHI</i>	77+213	N1870-3'	382208 382193	GGGCTTTCCTAATCC	44,9	16	N1872	N1872-5'	383059 383074	GCGTCACCATACCTC	42,3	16	272	<i>HindIII</i>	123+148	N1872-3'	383330 383315	TACCACCGCATTCATG	42,1	16	N1875	N1875-5'	385301 385316	AATCTCCATGGCTFAA	36,3	16	293	<i>HindIII</i>	39+253	N1875-3'	385593 385578	CCTTCACCTAAATGCA	37,7	16	N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290	N1879-3'	387681 387666	GGCGAAGATACCATGG	42,9	16																																																																																																																										
N1777	N1777-5'	350301 350286	GGAAGATGGCTCACGG	45,6	16	282	<i>EcoRI</i>	221+61																																																																																																																																																																																																																																																																																																																														
	N1777-3'	350020 350035	CCACAGGCCCAAACAG	46,7	16				N1780	N1780-5'	351011 351026	ACCAAAAAAATGCA	38,2	16	270	<i>AfuI</i>	187+83	N1780-3'	351280 351265	AGCCTTTTAGGACACC	38,0	16	N1785	N1785-5'	351705 351720	TCATCACATGGAAGTC	32,5	16	265	<i>EcoRI</i>	138+126	N1785-3'	351969 351954	CGGTATAGTAATGATG	27,7	16	N1790	N1790-5'	354074 354059	AAGGAGACATCCGAGG	40,9	16	277	<i>HindIII</i>	200+76	N1790-3'	353798 353813	ACTCTCCCCTGCCAAT	43,5	16	N1795	N1795-5'	357208 357193	TCTTACGGCGCTTAG	44,5	16	351	<i>HindIII</i>	305+45	N1795-3'	356858 356873	TCGGCCATCGAAATCA	47,9	16	N1825	N1825-5'	360133 360148	CCCCACATTTATTTGT	36,5	16	337	<i>Claf</i>	135+201	N1825-3'	360466 360454	ACCACAAAATAGGCG	43,0	16	N1843	N1843-5'	370645 370660	CACGGAATGGCAATGG	47,7	16	334	<i>XhoI</i>	171+162	N1843-3'	370978 370963	TGGACGCTATGGGGC	46,0	16	N1847	N1847-5'	372965 372950	TCGTCGTTGCTTCTCG	44,8	16	312	<i>EcoRV</i>	230+82	N1847-3'	372654 372669	CAAATGGGACGTCGTT	43,2	16	N1850	N1850-5'	374723 374708	GACTGTGGGCTCTGTT	38,3	16	309	<i>HaeIII</i>	114+195	N1850-3'	374415 374430	ATCTTGCTCTCTG	35,4	16	N1858	N1858-5'	375406 375421	GGACAGGCGTACGTA	46,2	16	308	<i>HindIII</i>	95+213	N1858-3'	375713 375698	TGGGGTTAATGCCTCG	46,7	16	N1870	N1870-5'	381918 381933	GTGCGTGCTCCCTATG	44,0	16	291	<i>BamHI</i>	77+213	N1870-3'	382208 382193	GGGCTTTCCTAATCC	44,9	16	N1872	N1872-5'	383059 383074	GCGTCACCATACCTC	42,3	16	272	<i>HindIII</i>	123+148	N1872-3'	383330 383315	TACCACCGCATTCATG	42,1	16	N1875	N1875-5'	385301 385316	AATCTCCATGGCTFAA	36,3	16	293	<i>HindIII</i>	39+253	N1875-3'	385593 385578	CCTTCACCTAAATGCA	37,7	16	N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290	N1879-3'	387681 387666	GGCGAAGATACCATGG	42,9	16																																																																																																																																								
N1780	N1780-5'	351011 351026	ACCAAAAAAATGCA	38,2	16	270	<i>AfuI</i>	187+83																																																																																																																																																																																																																																																																																																																														
	N1780-3'	351280 351265	AGCCTTTTAGGACACC	38,0	16				N1785	N1785-5'	351705 351720	TCATCACATGGAAGTC	32,5	16	265	<i>EcoRI</i>	138+126	N1785-3'	351969 351954	CGGTATAGTAATGATG	27,7	16	N1790	N1790-5'	354074 354059	AAGGAGACATCCGAGG	40,9	16	277	<i>HindIII</i>	200+76	N1790-3'	353798 353813	ACTCTCCCCTGCCAAT	43,5	16	N1795	N1795-5'	357208 357193	TCTTACGGCGCTTAG	44,5	16	351	<i>HindIII</i>	305+45	N1795-3'	356858 356873	TCGGCCATCGAAATCA	47,9	16	N1825	N1825-5'	360133 360148	CCCCACATTTATTTGT	36,5	16	337	<i>Claf</i>	135+201	N1825-3'	360466 360454	ACCACAAAATAGGCG	43,0	16	N1843	N1843-5'	370645 370660	CACGGAATGGCAATGG	47,7	16	334	<i>XhoI</i>	171+162	N1843-3'	370978 370963	TGGACGCTATGGGGC	46,0	16	N1847	N1847-5'	372965 372950	TCGTCGTTGCTTCTCG	44,8	16	312	<i>EcoRV</i>	230+82	N1847-3'	372654 372669	CAAATGGGACGTCGTT	43,2	16	N1850	N1850-5'	374723 374708	GACTGTGGGCTCTGTT	38,3	16	309	<i>HaeIII</i>	114+195	N1850-3'	374415 374430	ATCTTGCTCTCTG	35,4	16	N1858	N1858-5'	375406 375421	GGACAGGCGTACGTA	46,2	16	308	<i>HindIII</i>	95+213	N1858-3'	375713 375698	TGGGGTTAATGCCTCG	46,7	16	N1870	N1870-5'	381918 381933	GTGCGTGCTCCCTATG	44,0	16	291	<i>BamHI</i>	77+213	N1870-3'	382208 382193	GGGCTTTCCTAATCC	44,9	16	N1872	N1872-5'	383059 383074	GCGTCACCATACCTC	42,3	16	272	<i>HindIII</i>	123+148	N1872-3'	383330 383315	TACCACCGCATTCATG	42,1	16	N1875	N1875-5'	385301 385316	AATCTCCATGGCTFAA	36,3	16	293	<i>HindIII</i>	39+253	N1875-3'	385593 385578	CCTTCACCTAAATGCA	37,7	16	N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290	N1879-3'	387681 387666	GGCGAAGATACCATGG	42,9	16																																																																																																																																																						
N1785	N1785-5'	351705 351720	TCATCACATGGAAGTC	32,5	16	265	<i>EcoRI</i>	138+126																																																																																																																																																																																																																																																																																																																														
	N1785-3'	351969 351954	CGGTATAGTAATGATG	27,7	16				N1790	N1790-5'	354074 354059	AAGGAGACATCCGAGG	40,9	16	277	<i>HindIII</i>	200+76	N1790-3'	353798 353813	ACTCTCCCCTGCCAAT	43,5	16	N1795	N1795-5'	357208 357193	TCTTACGGCGCTTAG	44,5	16	351	<i>HindIII</i>	305+45	N1795-3'	356858 356873	TCGGCCATCGAAATCA	47,9	16	N1825	N1825-5'	360133 360148	CCCCACATTTATTTGT	36,5	16	337	<i>Claf</i>	135+201	N1825-3'	360466 360454	ACCACAAAATAGGCG	43,0	16	N1843	N1843-5'	370645 370660	CACGGAATGGCAATGG	47,7	16	334	<i>XhoI</i>	171+162	N1843-3'	370978 370963	TGGACGCTATGGGGC	46,0	16	N1847	N1847-5'	372965 372950	TCGTCGTTGCTTCTCG	44,8	16	312	<i>EcoRV</i>	230+82	N1847-3'	372654 372669	CAAATGGGACGTCGTT	43,2	16	N1850	N1850-5'	374723 374708	GACTGTGGGCTCTGTT	38,3	16	309	<i>HaeIII</i>	114+195	N1850-3'	374415 374430	ATCTTGCTCTCTG	35,4	16	N1858	N1858-5'	375406 375421	GGACAGGCGTACGTA	46,2	16	308	<i>HindIII</i>	95+213	N1858-3'	375713 375698	TGGGGTTAATGCCTCG	46,7	16	N1870	N1870-5'	381918 381933	GTGCGTGCTCCCTATG	44,0	16	291	<i>BamHI</i>	77+213	N1870-3'	382208 382193	GGGCTTTCCTAATCC	44,9	16	N1872	N1872-5'	383059 383074	GCGTCACCATACCTC	42,3	16	272	<i>HindIII</i>	123+148	N1872-3'	383330 383315	TACCACCGCATTCATG	42,1	16	N1875	N1875-5'	385301 385316	AATCTCCATGGCTFAA	36,3	16	293	<i>HindIII</i>	39+253	N1875-3'	385593 385578	CCTTCACCTAAATGCA	37,7	16	N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290	N1879-3'	387681 387666	GGCGAAGATACCATGG	42,9	16																																																																																																																																																																				
N1790	N1790-5'	354074 354059	AAGGAGACATCCGAGG	40,9	16	277	<i>HindIII</i>	200+76																																																																																																																																																																																																																																																																																																																														
	N1790-3'	353798 353813	ACTCTCCCCTGCCAAT	43,5	16				N1795	N1795-5'	357208 357193	TCTTACGGCGCTTAG	44,5	16	351	<i>HindIII</i>	305+45	N1795-3'	356858 356873	TCGGCCATCGAAATCA	47,9	16	N1825	N1825-5'	360133 360148	CCCCACATTTATTTGT	36,5	16	337	<i>Claf</i>	135+201	N1825-3'	360466 360454	ACCACAAAATAGGCG	43,0	16	N1843	N1843-5'	370645 370660	CACGGAATGGCAATGG	47,7	16	334	<i>XhoI</i>	171+162	N1843-3'	370978 370963	TGGACGCTATGGGGC	46,0	16	N1847	N1847-5'	372965 372950	TCGTCGTTGCTTCTCG	44,8	16	312	<i>EcoRV</i>	230+82	N1847-3'	372654 372669	CAAATGGGACGTCGTT	43,2	16	N1850	N1850-5'	374723 374708	GACTGTGGGCTCTGTT	38,3	16	309	<i>HaeIII</i>	114+195	N1850-3'	374415 374430	ATCTTGCTCTCTG	35,4	16	N1858	N1858-5'	375406 375421	GGACAGGCGTACGTA	46,2	16	308	<i>HindIII</i>	95+213	N1858-3'	375713 375698	TGGGGTTAATGCCTCG	46,7	16	N1870	N1870-5'	381918 381933	GTGCGTGCTCCCTATG	44,0	16	291	<i>BamHI</i>	77+213	N1870-3'	382208 382193	GGGCTTTCCTAATCC	44,9	16	N1872	N1872-5'	383059 383074	GCGTCACCATACCTC	42,3	16	272	<i>HindIII</i>	123+148	N1872-3'	383330 383315	TACCACCGCATTCATG	42,1	16	N1875	N1875-5'	385301 385316	AATCTCCATGGCTFAA	36,3	16	293	<i>HindIII</i>	39+253	N1875-3'	385593 385578	CCTTCACCTAAATGCA	37,7	16	N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290	N1879-3'	387681 387666	GGCGAAGATACCATGG	42,9	16																																																																																																																																																																																		
N1795	N1795-5'	357208 357193	TCTTACGGCGCTTAG	44,5	16	351	<i>HindIII</i>	305+45																																																																																																																																																																																																																																																																																																																														
	N1795-3'	356858 356873	TCGGCCATCGAAATCA	47,9	16				N1825	N1825-5'	360133 360148	CCCCACATTTATTTGT	36,5	16	337	<i>Claf</i>	135+201	N1825-3'	360466 360454	ACCACAAAATAGGCG	43,0	16	N1843	N1843-5'	370645 370660	CACGGAATGGCAATGG	47,7	16	334	<i>XhoI</i>	171+162	N1843-3'	370978 370963	TGGACGCTATGGGGC	46,0	16	N1847	N1847-5'	372965 372950	TCGTCGTTGCTTCTCG	44,8	16	312	<i>EcoRV</i>	230+82	N1847-3'	372654 372669	CAAATGGGACGTCGTT	43,2	16	N1850	N1850-5'	374723 374708	GACTGTGGGCTCTGTT	38,3	16	309	<i>HaeIII</i>	114+195	N1850-3'	374415 374430	ATCTTGCTCTCTG	35,4	16	N1858	N1858-5'	375406 375421	GGACAGGCGTACGTA	46,2	16	308	<i>HindIII</i>	95+213	N1858-3'	375713 375698	TGGGGTTAATGCCTCG	46,7	16	N1870	N1870-5'	381918 381933	GTGCGTGCTCCCTATG	44,0	16	291	<i>BamHI</i>	77+213	N1870-3'	382208 382193	GGGCTTTCCTAATCC	44,9	16	N1872	N1872-5'	383059 383074	GCGTCACCATACCTC	42,3	16	272	<i>HindIII</i>	123+148	N1872-3'	383330 383315	TACCACCGCATTCATG	42,1	16	N1875	N1875-5'	385301 385316	AATCTCCATGGCTFAA	36,3	16	293	<i>HindIII</i>	39+253	N1875-3'	385593 385578	CCTTCACCTAAATGCA	37,7	16	N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290	N1879-3'	387681 387666	GGCGAAGATACCATGG	42,9	16																																																																																																																																																																																																
N1825	N1825-5'	360133 360148	CCCCACATTTATTTGT	36,5	16	337	<i>Claf</i>	135+201																																																																																																																																																																																																																																																																																																																														
	N1825-3'	360466 360454	ACCACAAAATAGGCG	43,0	16				N1843	N1843-5'	370645 370660	CACGGAATGGCAATGG	47,7	16	334	<i>XhoI</i>	171+162	N1843-3'	370978 370963	TGGACGCTATGGGGC	46,0	16	N1847	N1847-5'	372965 372950	TCGTCGTTGCTTCTCG	44,8	16	312	<i>EcoRV</i>	230+82	N1847-3'	372654 372669	CAAATGGGACGTCGTT	43,2	16	N1850	N1850-5'	374723 374708	GACTGTGGGCTCTGTT	38,3	16	309	<i>HaeIII</i>	114+195	N1850-3'	374415 374430	ATCTTGCTCTCTG	35,4	16	N1858	N1858-5'	375406 375421	GGACAGGCGTACGTA	46,2	16	308	<i>HindIII</i>	95+213	N1858-3'	375713 375698	TGGGGTTAATGCCTCG	46,7	16	N1870	N1870-5'	381918 381933	GTGCGTGCTCCCTATG	44,0	16	291	<i>BamHI</i>	77+213	N1870-3'	382208 382193	GGGCTTTCCTAATCC	44,9	16	N1872	N1872-5'	383059 383074	GCGTCACCATACCTC	42,3	16	272	<i>HindIII</i>	123+148	N1872-3'	383330 383315	TACCACCGCATTCATG	42,1	16	N1875	N1875-5'	385301 385316	AATCTCCATGGCTFAA	36,3	16	293	<i>HindIII</i>	39+253	N1875-3'	385593 385578	CCTTCACCTAAATGCA	37,7	16	N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290	N1879-3'	387681 387666	GGCGAAGATACCATGG	42,9	16																																																																																																																																																																																																														
N1843	N1843-5'	370645 370660	CACGGAATGGCAATGG	47,7	16	334	<i>XhoI</i>	171+162																																																																																																																																																																																																																																																																																																																														
	N1843-3'	370978 370963	TGGACGCTATGGGGC	46,0	16				N1847	N1847-5'	372965 372950	TCGTCGTTGCTTCTCG	44,8	16	312	<i>EcoRV</i>	230+82	N1847-3'	372654 372669	CAAATGGGACGTCGTT	43,2	16	N1850	N1850-5'	374723 374708	GACTGTGGGCTCTGTT	38,3	16	309	<i>HaeIII</i>	114+195	N1850-3'	374415 374430	ATCTTGCTCTCTG	35,4	16	N1858	N1858-5'	375406 375421	GGACAGGCGTACGTA	46,2	16	308	<i>HindIII</i>	95+213	N1858-3'	375713 375698	TGGGGTTAATGCCTCG	46,7	16	N1870	N1870-5'	381918 381933	GTGCGTGCTCCCTATG	44,0	16	291	<i>BamHI</i>	77+213	N1870-3'	382208 382193	GGGCTTTCCTAATCC	44,9	16	N1872	N1872-5'	383059 383074	GCGTCACCATACCTC	42,3	16	272	<i>HindIII</i>	123+148	N1872-3'	383330 383315	TACCACCGCATTCATG	42,1	16	N1875	N1875-5'	385301 385316	AATCTCCATGGCTFAA	36,3	16	293	<i>HindIII</i>	39+253	N1875-3'	385593 385578	CCTTCACCTAAATGCA	37,7	16	N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290	N1879-3'	387681 387666	GGCGAAGATACCATGG	42,9	16																																																																																																																																																																																																																												
N1847	N1847-5'	372965 372950	TCGTCGTTGCTTCTCG	44,8	16	312	<i>EcoRV</i>	230+82																																																																																																																																																																																																																																																																																																																														
	N1847-3'	372654 372669	CAAATGGGACGTCGTT	43,2	16				N1850	N1850-5'	374723 374708	GACTGTGGGCTCTGTT	38,3	16	309	<i>HaeIII</i>	114+195	N1850-3'	374415 374430	ATCTTGCTCTCTG	35,4	16	N1858	N1858-5'	375406 375421	GGACAGGCGTACGTA	46,2	16	308	<i>HindIII</i>	95+213	N1858-3'	375713 375698	TGGGGTTAATGCCTCG	46,7	16	N1870	N1870-5'	381918 381933	GTGCGTGCTCCCTATG	44,0	16	291	<i>BamHI</i>	77+213	N1870-3'	382208 382193	GGGCTTTCCTAATCC	44,9	16	N1872	N1872-5'	383059 383074	GCGTCACCATACCTC	42,3	16	272	<i>HindIII</i>	123+148	N1872-3'	383330 383315	TACCACCGCATTCATG	42,1	16	N1875	N1875-5'	385301 385316	AATCTCCATGGCTFAA	36,3	16	293	<i>HindIII</i>	39+253	N1875-3'	385593 385578	CCTTCACCTAAATGCA	37,7	16	N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290	N1879-3'	387681 387666	GGCGAAGATACCATGG	42,9	16																																																																																																																																																																																																																																										
N1850	N1850-5'	374723 374708	GACTGTGGGCTCTGTT	38,3	16	309	<i>HaeIII</i>	114+195																																																																																																																																																																																																																																																																																																																														
	N1850-3'	374415 374430	ATCTTGCTCTCTG	35,4	16				N1858	N1858-5'	375406 375421	GGACAGGCGTACGTA	46,2	16	308	<i>HindIII</i>	95+213	N1858-3'	375713 375698	TGGGGTTAATGCCTCG	46,7	16	N1870	N1870-5'	381918 381933	GTGCGTGCTCCCTATG	44,0	16	291	<i>BamHI</i>	77+213	N1870-3'	382208 382193	GGGCTTTCCTAATCC	44,9	16	N1872	N1872-5'	383059 383074	GCGTCACCATACCTC	42,3	16	272	<i>HindIII</i>	123+148	N1872-3'	383330 383315	TACCACCGCATTCATG	42,1	16	N1875	N1875-5'	385301 385316	AATCTCCATGGCTFAA	36,3	16	293	<i>HindIII</i>	39+253	N1875-3'	385593 385578	CCTTCACCTAAATGCA	37,7	16	N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290	N1879-3'	387681 387666	GGCGAAGATACCATGG	42,9	16																																																																																																																																																																																																																																																								
N1858	N1858-5'	375406 375421	GGACAGGCGTACGTA	46,2	16	308	<i>HindIII</i>	95+213																																																																																																																																																																																																																																																																																																																														
	N1858-3'	375713 375698	TGGGGTTAATGCCTCG	46,7	16				N1870	N1870-5'	381918 381933	GTGCGTGCTCCCTATG	44,0	16	291	<i>BamHI</i>	77+213	N1870-3'	382208 382193	GGGCTTTCCTAATCC	44,9	16	N1872	N1872-5'	383059 383074	GCGTCACCATACCTC	42,3	16	272	<i>HindIII</i>	123+148	N1872-3'	383330 383315	TACCACCGCATTCATG	42,1	16	N1875	N1875-5'	385301 385316	AATCTCCATGGCTFAA	36,3	16	293	<i>HindIII</i>	39+253	N1875-3'	385593 385578	CCTTCACCTAAATGCA	37,7	16	N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290	N1879-3'	387681 387666	GGCGAAGATACCATGG	42,9	16																																																																																																																																																																																																																																																																						
N1870	N1870-5'	381918 381933	GTGCGTGCTCCCTATG	44,0	16	291	<i>BamHI</i>	77+213																																																																																																																																																																																																																																																																																																																														
	N1870-3'	382208 382193	GGGCTTTCCTAATCC	44,9	16				N1872	N1872-5'	383059 383074	GCGTCACCATACCTC	42,3	16	272	<i>HindIII</i>	123+148	N1872-3'	383330 383315	TACCACCGCATTCATG	42,1	16	N1875	N1875-5'	385301 385316	AATCTCCATGGCTFAA	36,3	16	293	<i>HindIII</i>	39+253	N1875-3'	385593 385578	CCTTCACCTAAATGCA	37,7	16	N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290	N1879-3'	387681 387666	GGCGAAGATACCATGG	42,9	16																																																																																																																																																																																																																																																																																				
N1872	N1872-5'	383059 383074	GCGTCACCATACCTC	42,3	16	272	<i>HindIII</i>	123+148																																																																																																																																																																																																																																																																																																																														
	N1872-3'	383330 383315	TACCACCGCATTCATG	42,1	16				N1875	N1875-5'	385301 385316	AATCTCCATGGCTFAA	36,3	16	293	<i>HindIII</i>	39+253	N1875-3'	385593 385578	CCTTCACCTAAATGCA	37,7	16	N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290	N1879-3'	387681 387666	GGCGAAGATACCATGG	42,9	16																																																																																																																																																																																																																																																																																																		
N1875	N1875-5'	385301 385316	AATCTCCATGGCTFAA	36,3	16	293	<i>HindIII</i>	39+253																																																																																																																																																																																																																																																																																																																														
	N1875-3'	385593 385578	CCTTCACCTAAATGCA	37,7	16				N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290	N1879-3'	387681 387666	GGCGAAGATACCATGG	42,9	16																																																																																																																																																																																																																																																																																																																
N1879	N1879-5'	387337 387352	CACCAAATCGGCAACT	43,5	16	345	<i>HindIII</i>	54+290																																																																																																																																																																																																																																																																																																																														
	N1879-3'	387681 387666	GGCGAAGATACCATGG	42,9	16																																																																																																																																																																																																																																																																																																																																	

Tabla 3.3. Diseño de los cebadores para las reacciones de amplificación y selección de las enzimas utilizadas en la comprobación de cada producto obtenido Tm: temperatura de *melting*; L.P.: Longitud del producto de PCR; L: longitud del cebador. (continuación).

Nº ORF	Nombre del Cebador	Posición del Cebador Inicio final	SECUENCIA	Tm (°C)	L (nt)	L.P. (pb)	Enzima	Longitud Fragmentos (pb)
N1882	N1882-5'	391860 391845	CCGATGCCATTAGGCT	45,9	16	305	<i>EcoRI</i>	57+247
	N1882-3'	391556 391571	CAGCTGCCCGGTAGAC	46,1	16			
N1888	N1888-5'	393198 393213	GACCCITCACCTTCCG	44,6	16	332	<i>XhoI</i>	204+127
	N1888-3'	393529 393514	TCACCCAAACGCACCT	43,9	16			
N1897	N1897-5'	395373 395388	TTGTGGCGGTCAAGT	43,1	16	309	<i>EcoRI</i>	110+198
	N1897-3'	395681 395666	CCTGGGCGTCTCTTA	42,0	16			
N1913	N1913-5'	401065 401080	CCAGAAATCCAGCGAC	42,8	16	280	<i>BamHI</i>	149+130
	N1913-3'	401344 401329	TCGACGGTAAAGCCTG	44,0	16			
N1917	N1917-5'	404858 404843	AGGCATCAGAGGCCAAA	43,1	16	353	<i>XhoI</i>	165+187
	N1917-3'	404506 404521	ACGTTAGCCATTGGCA	44,4	16			
N1925	N1925-5'	409091 409106	TTGAAGCAAAGGAAAG	37,4	16	341	<i>BamHI</i>	120+222
	N1925-3'	409433 409418	TCGATAAAGAGGAAGC	36,2	16			
N1929	N1929-5'	411427 411412	TGAGCCATTTGGAGGAA	43,7	16	300	<i>AluI</i>	165+135
	N1929-3'	411128 411143	GGCGACTGAACGGTTA	43,7	16			
N1954	N1954-5'	418156 418141	TTTTGCTCAATTCGGT	40,6	16	320	<i>EcoRI</i>	232+88
	N1954-3'	417837 417852	CGATACCTGACTTGGC	39,8	16			
N1958	N1958-5'	418931 418946	ACACAGCTTCATCATG	33,4	16	169	<i>EcoRI</i>	57+111
	N1958-3'	419099 419084	GGACGTTTATGAATGC	37,0	16			
N1962	N1962-5'	419637 419622	TCCTCCCATTITACC	41,5	16	337	<i>PvuI</i>	260+76
	N1962-3'	419301 419316	GAGCGGATTTAGTGGC	42,6	16			
N1966	N1966-5'	420355 420370	AGGGTGGGGTGAATTT	43,5	16	252	<i>EcoRI</i>	115+137
	N1966-3'	420606 420591	CCTGGTGGGCTCATT	43,6	16			
N2160	N2160-5'	423097 423082	TGAAAACGGCCAGTGA	44,9	16	288	<i>HindIII</i>	215+72
	N2160-3'	423384 423399	CGAGGCGATACAAGGA	43,8	16			
N2185	N2185-5'	435166 435181	ATGTGGAAGCGTTGAT	39,7	16	309	<i>EcoRI</i>	178+131
	N2185-3'	435474 435459	AGCCGTTTAGGATCCT	40,7	16			
N2190	N2190-5'	437972 437987	CAGGTCTAATTGGGGA	39,2	16	315	<i>XhoI</i>	53+262
	N2190-3'	438286 438271	TGTTCCCTTGCATACC	40,2	16			
N2194	N2194-5'	439015 439000	TGAATTTCCCTTTTIT	38,9	16	321	<i>PvuI</i>	117+204
	N2194-3'	438695 438710	CTTCGGATACAATGGC	39,8	16			
N2205	N2205-5'	442428 442413	CCGGAAGGAGAAAAGT	41,0	16	328	<i>EcoRI</i>	192+135
	N2205-3'	442101 442116	CTCTCCAGCGATGGC	48,8	16			
N2215	N2215-5'	446156 446141	ACAAAAGGGTGAANAAG	39,7	16	300	<i>EcoRI</i>	208+92
	N2215-3'	445857 445872	ATGTGAAGCTCTCGGC	39,2	16			
N2219	N2219-5'	447671 447686	GGAGACCCACGAATGG	45,2	16	275	<i>EcoRI</i>	125+149
	N2219-3'	447945 447930	GCACATCTGACAGTAG	29,3	16			
N2227	N2227-5'	451593 451608	TCAAGGCACAAAATC	37,5	16	332	<i>BamHI</i>	112+219
	N2227-3'	451924 451909	TGGCAATGCTTATGAT	38,0	16			
N2231	N2231-5'	452547 452562	CGGATTACCCAATTC	39,1	16	316	<i>XbaI</i>	151+165
	N2231-3'	452862 452847	CGCTCCATTCTTGAT	39,7	16			
N2250	N2250-5'	464225 464240	GATAGAACAGAAGTGA	24,7	16	284	<i>EcoRV</i>	85+198
	N2250-3'	464508 464493	ATTGGTCCAGATAGG	35,1	16			
N2254	N2254-5'	466338 466353	GGGATAGTATAAGCGC	36,4	16	284	<i>Clai</i>	91+192
	N2254-3'	466621 466606	AGCATTCCCCTGAAGC	44,6	16			
N2312	N2312-5'	472351 472366	CGAGGTCAACCAAGAAA	40,1	16	312	<i>XhoI</i>	51+261
	N2312-3'	472662 472647	CAGAAAAGTCCCCAA	39,7	16			

3.5.2. El DNA molde utilizado:

Se utilizó, o bien DNA genómico extraído de la línea FY73 de *S.cerevisiae*, o bien DNA de diferentes cósmidos ó plásmidos portadores de distintos clones del genoma de la levadura.

3.5.2.1. Obtención del DNA genómico:

Para purificar el DNA genómico a gran escala se utilizó el protocolo que comprende los pasos siguientes:

- a) Se hicieron crecer células de la línea FY73 hasta la fase estacionaria en YPD y se centrifugaron 100 ml de cultivo a 5.000 rpm durante 3 min.
- b) Se lavó el sedimento con 10 ml de disolución I (sorbitol 1M, EDTA 0,1M) y se resuspendió en 10 ml de disolución II de extracción (Sorbitol 0,9M, EDTA 50mM, NaH₂PO₄ 50mM pH7,5, β-mercaptoetanol 14 mM).
- c) Se añadió 0,5 ml de liticasa (5 g/ml) y se agitó suavemente durante 30 min a 30°C.
- d) Se añadió SDS 1% , se incubó 20 min a 60°C sin agitación y se dejó enfriar a temperatura ambiente.
- e) Se hicieron 2 ó 3 extracciones sucesivas con 1,5 volúmenes de PCIA saturado con tampón 1xTE (Tris-HCl 10 mM; EDTA 1 mM, pH 8), agitando suavemente para evitar que se produjesen roturas en el DNA. Las fases se separaron por centrifugación a 5.000 rpm durante 10 min, hasta conseguir una fase acuosa lo más limpia posible de proteínas.

- f) Para completar la extracción de proteínas, la fase acuosa se incubó en hielo con acetato potásico (pH 4,8) durante 2-4 horas en una proporción 1:5 y se centrifugó 15 min a 10.000 rpm. Se recuperó el sobrenadante, descartando el posible precipitado que se produce cuando quedan restos de proteína.
- g) Una vez purificado, el DNA genómico se precipitó con 1,5 volúmenes de etanol al 95%, mezclándolo suavemente por inversión del tubo, hasta que se observó la formación del ovillo típico de DNA y se centrifugó durante 5 min a 5.000 rpm.
- h) El DNA precipitado se lavó una o dos veces con etanol al 70% para eliminar restos de sales, se secó a vacío durante unos 5 min y se resuspendió en 1xTE.

3.5.2.2. Obtención del DNA plasmídico o cosmídico:

El DNA plasmídico o cosmídico se purificó por el método de lisis alcalina (Sambrook *et al.*, 1989) a partir de transformantes de *Escherichia coli*:

- a) Se inocularon las células bacterianas en 30 ml de LBA. Tras su crecimiento, en agitación durante toda la noche a 37 °C, se centrifugaron las células durante tres minutos a 5.000 rpm.
- b) El sedimento se resuspendió en 1,5 ml de solución 1 (glucosa 50mM; EDTA 10 mM; Tris-HCl 25 mM, pH 8) previamente enfriada en hielo y se incubó cinco minutos a temperatura ambiente. Posteriormente, se añadieron 3 ml de la solución 2 (NaOH 0,2 N; SDS 1 %), se mezcló por inversión dos o tres veces hasta conseguir una viscosidad homogénea y se mantuvo en hielo durante cinco minutos. Transcurrido este tiempo se añadieron 2,25 ml de solución

3 (60 ml de acetato potásico 5M; 11,5 ml de ácido acético glacial; 28,5 ml de agua destilada) que se mezclaron por inversión suave del tubo, incubando otros cinco minutos en hielo. Los tubos se centrifugaron a 10.000 rpm durante cinco minutos.

- c) Como con este procedimiento, además del DNA, se obtiene una considerable cantidad de RNA; este último se eliminó con un tratamiento con RNAsa A, a una concentración final de 1 µg/ml, durante 30 minutos a 37 °C.
- d) Para eliminar restos de proteínas que podrían dificultar posteriores manipulaciones del DNA, el sobrenadante se sometió a una extracción con igual volumen de PCIA y un posterior lavado con igual volumen de CIA (cloroformo-álcohol isoamílico en proporción 24:1).
- e) A cada muestra se le añadió 10 ml de etanol al 95 %, se incubó cinco minutos a - 20 °C y posteriormente el DNA se precipitó por centrifugación a 12.000 rpm durante cinco minutos. Para eliminar restos de sales, el sedimento se lavó con 10 ml de etanol al 70 % y se precipitó como en el paso anterior. Las muestras se secaron al vacío y posteriormente se resuspendieron en 200 µl de 1xTE.

3.5.2.3. Sondas de ORFs controles:

Diversos plásmidos se usaron en la amplificación por PCR del DNA correspondiente a una serie de sondas controles relacionadas con genes de *S.cerevisiae* de función y regulación conocida:

- **pSPACT9** es un subclón del fragmento *Bam*HI/*Hind*III (1,5 Kb) del gen *ACT1* en el plásmido pSP64 (Bettany *et al.*, 1989)

- **pUC-PCK** fue construido amplificando, por PCR, la región codificadora de *PCK1*, y tras introducirle sitios *XbaI* en ambos extremos, se clonó en el plásmido pUC18 (amablemente cedido por A. Brown, *University of Aberdeen*, Escocia)
- **yCPlac111-rpL25** que contiene el fragmento *SacI/EcoRI* (0,3 Kb) de *rpL25* fue amablemente cedido por R. Planta (*Vrije Universiteit*, Amsterdam, Holanda)
- **pUC19-HSP12** donde está insertado el fragmento *EcoRI* (0,6 Kb) de *HSP12* fue también facilitado por R. Planta (*Vrije Universiteit*, Amsterdam, Holanda)
- **pYcARG4** con el fragmento *EcoRI/ClaI* (1,0 Kb) de *ARG4* fue proporcionado por F. Messenguy (*Institut de Recherches du CERIA*, Bruselas, Bélgica).

La amplificación por PCR del control correspondiente al gen *CAR1* se hizo, como para las ORFs del cromosoma VII, a partir de DNA genómico.

Para amplificar las ORFs correspondientes al cromosoma XIV de *S. cerevisiae* se utilizaron cósmidos generados en el proyecto de secuenciación. Estos fragmentos comprenden regiones del cromosoma XIV que se solapan entre sí de manera que abarcan la totalidad de su material genético (784 Kb). Los cósmidos necesarios fueron amablemente cedidos por el coordinador del proyecto de secuenciación del cromosoma XIV (P. Philippsen, *Biocentrum*, Basel, Suiza). Una representación esquemática nos permite situar, en el cromosoma XIV cada cósmido utilizado (Figura 3.2.).

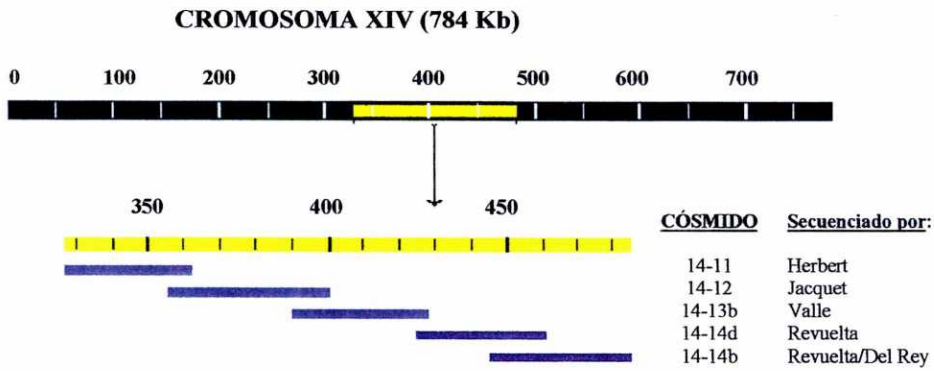


Figura 3.2. Ubicación de los cósmidos utilizados para las amplificaciones en el cromosoma XIV de *S. cerevisiae*.

3.5.3. La Reacción en Cadena de la Polimerasa:

Para la realización de este trabajo se utilizaron los reactivos y la *Taq* DNA polimerasa suministrada por *PERKIN Elmer* bajo la denominación “*Gene Amp® PCR Core reagents*” y los cebadores obtenidos a través del servicio de síntesis de oligonucleótidos de *Boehringer Manheim* según el diseño recogido en la tabla 3.3.

Se mezcló el DNA molde a una concentración final de 1nM (evitando contaminar la reacción con DNA extraño que también podría amplificarse) con los otros componentes de la reacción que vienen descritos a continuación:

- 10 μ l de Tampón de PCR 10x (Tris-HCl 10 mM pH 8,3, KCl 50 mM),
- 10 μ l de $MgCl_2$ 2,5 mM,
- 10 μ l de mezcla de dNTPs (desoxi-nucleósidos trifosfato) 2mM,
- los cebadores oligonucleotídicos adecuados (1 μ M de concentración final).

Se enrasó con agua bidestilada estéril para obtener un volumen final de 100 μ l y se cubrió con dos gotas de aceite mineral para impedir la evaporación durante la síntesis. A continuación, se desnaturalizó el DNA durante 5 minutos a 92 °C, se enfrió en hielo y se le añadieron 2,5 U de *Taq* DNA polimerasa cuya temperatura óptima de trabajo es de 72 °C.

Se colocó el tubo con la mezcla en el termociclador, *Gene ATAQ Controller* de *Pharmacia*; este instrumento que proporciona automáticamente variaciones rápidas y cíclicas de temperatura se programó para realizar 30 ciclos de amplificación.

En teoría, después de 20 ciclos, una molécula de DNA se amplifica al rededor de un millón de veces, pero en la práctica, la amplificación no es exponencial, especialmente después de 20 ciclos, debido al consumo parcial de los cebadores y a una cierta desnaturalización térmica de la *Taq* polimerasa.

La temperatura de anillamiento se calculó mediante el programa *OLIGO*. Todos los oligonucleótidos se seleccionaron de forma que las condiciones de anillamiento fuesen óptimas alrededor de 45-50 °C. A esta temperatura cerca del 90 % de los cebadores se unen al DNA molde. El paso siguiente a 72 °C (alargamiento) interrumpe las hibridaciones inespecíficas accidentales. La duración de la fase de alargamiento (72 °C) en el último ciclo se extendió a 120 segundos para darle tiempo suficiente a la *Taq* polimerasa de finalizar la síntesis de todos los fragmentos.

Finalizada la reacción de PCR, se transfirió la muestra resultante (sin el aceite mineral) a un tubo *Eppendorf*, se le añadió 5 μ l de Acetato Sódico 3 M, 120 μ l de etanol absoluto y se dejó precipitando toda una noche en el

congelador a -20°C . Al día siguiente se centrifugó a 13.000 rpm durante 20 minutos, se lavó el precipitado con etanol al 70 %, se secó durante 10 minutos al vacío y se resuspendió en 30 μl de 1xTE.

3.5.4. Análisis de los productos de PCR:

Finalizada la PCR, necesitábamos corroborar que el DNA amplificado se correspondía con el fragmento deseado y medir la concentración de producto obtenido para poder utilizarlo posteriormente.

3.5.4.1. Análisis de restricción:

Los productos obtenidos por PCR fueron verificados en dos fases. En una primera etapa se analizó en geles de agarosa el funcionamiento adecuado de la PCR, comprobando que se obtenía un producto único y de un tamaño cercano al esperado (datos no mostrados). Una vez conseguidos todos los productos, se sometieron a los análisis de restricción previstos en la Tabla 3.3.

Los fragmentos de DNA amplificados por PCR se eligieron de tal forma que todos poseyesen de uno a tres sitios de corte para una endonucleasa específica, lo que proporciona después de la digestión de dos a cuatro fragmentos de tamaño conocido; esto nos permitió garantizar que habíamos obtenido el fragmento esperado. El DNA se digirió con endonucleasas de restricción de *Boehringer-Manheim*; para cada enzima se utilizó el tampón recomendado y suministrado por los proveedores. Las digestiones (1 UE/ μg de DNA en un volumen final de 10 μl) se realizaron durante dos a tres horas a la temperatura indicada y los productos de la digestión se analizaron en geles de agarosa o poli-acrilamida según el tamaño de las bandas a detectar:

- Las electroforesis en geles de agarosa al 2 % se llevaron a cabo durante 60 minutos a 70V y los geles se tiñeron en una disolución de bromuro de etidio (5µg/ml) (Figura 3.5.1.)
- los geles de poliacrilamida al 6%, de mayor resolución en la separación electroforética de fragmentos pequeños, se tiñeron con nitrato de plata; la tinción se hizo bañando el gel durante media hora en una solución de ácido acético al 0,1 % y de etanol al 10 %, se lavó con agua, se dejó de nuevo 30 minutos en una solución de tinción con nitrato de plata (0,2 g de NO₃Ag en 200 ml de agua) y, después de lavarlo otra vez en agua, se sumergió en la solución de revelado (3 g de NaOH, 200 ml de agua, 1 ml de formaldehído al 37%) unos 10 minutos hasta la aparición de las bandas teñidas. Finalmente se lavó el gel con agua y se secó al vacío sobre un papel de filtro (Figura 3.5.2.).

En algunos casos, los fragmentos de DNA que se iban a utilizar como sondas se rescataron del gel utilizando el *Kit de Gene Clean BIO-101* siguiendo el protocolo indicado por los vendedores. El proceso se basa en disolver la agarosa con una disolución de yoduro sódico a 55°C y añadir una matriz de sílice, denominada “*glassmilk*”, a la que se adhiere el DNA a 0°C. Tras un proceso de lavados sucesivos con una disolución de cloruro sódico, etanol y agua, el DNA se eluye (con tampón 1xTE o agua bidestilada) de la matriz incubando a 55°C.

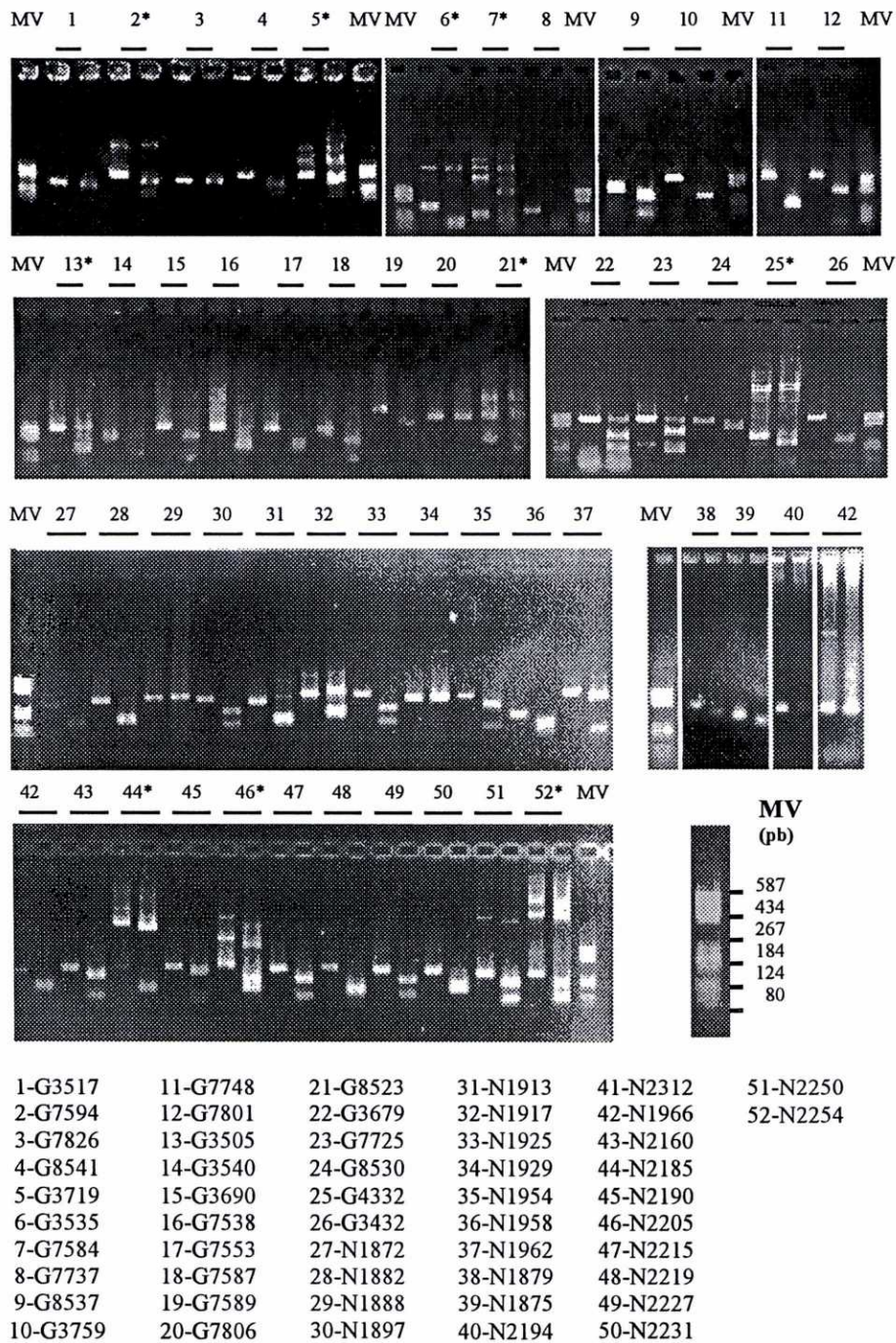


Figura 3.5.1. Geles de agarosa teñidos con bromuro de etidio mostrando, en un transiluminador de luz UV, el análisis de restricción de los productos de PCR de 52 Orfs. (MV: Marcador V de pesos moleculares de *Boehringer-Manheim*)

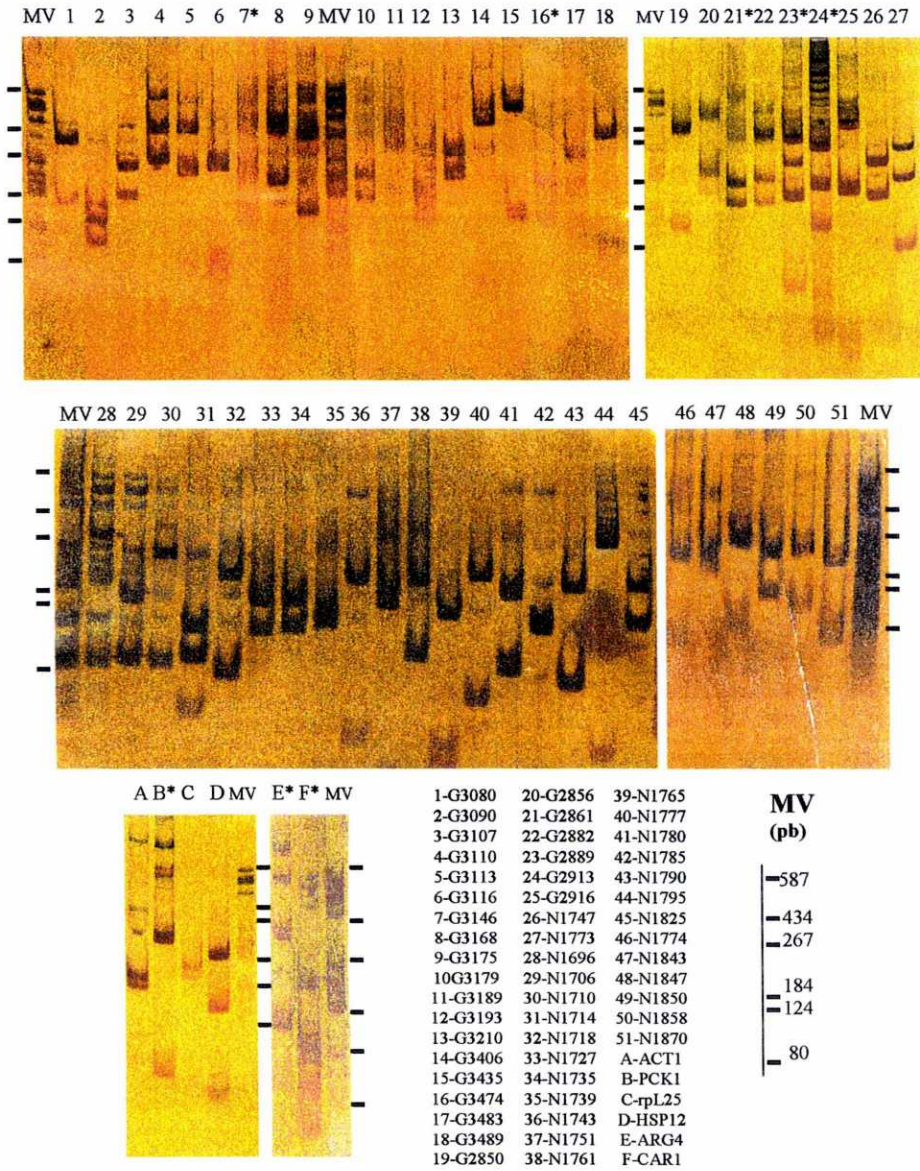


Figura 3.5.2. Geles de poliacrilamida teñidos con nitrato de plata mostrando el análisis de restricción de los productos de PCR de 51 Orfs. (MV: Marcador V de pesos moleculares de *Boehringer-Manheim*)

La mayor parte de los análisis de restricción mostraron el patrón adecuado. Sin embargo para 16 ORFs (marcadas con un asterisco en las figuras 3.5.1. y 3.5.2.), además del producto esperado aparece otra banda, o bandas mayores inespecíficas. En este caso se rescató de un gel de agarosa la banda del tamaño adecuado antes de marcar la sonda.

Se presentaron también algunos casos (Figuras 3.5.1. y 3.5.2.) en los que, aunque el producto es de tamaño adecuado, no se produce la digestión esperada (N1743, N1888, N1929, N2194 y N1966), o ésta es parcial (G2856, G2882, G3080, G3193, G3505 y N1773). Esta situación podría interpretarse debido al hecho de que la región del sitio de restricción para la enzima utilizada estuviese metilada afectando al patrón de restricción, sin que ello indicase que el producto obtenido no es correcto. En estos casos, como el tamaño del fragmento no digerido era correcto y el producto de la *PCR* único, se optó por considerarlos adecuados, aunque posteriormente fuese necesario secuenciar parcialmente estos productos para tener total certeza sobre su idoneidad.

3.5.4.2. Cuantificación de los niveles de DNA:

Para poder marcar una cantidad definida de sonda en los pasos siguientes, se necesita averiguar la concentración de DNA resultante de cada *PCR*. Para ello se preparó un gel de agarosa al 0,8 % en 1xTBE (Tris-HCl 90 mM, ácido bórico 90 mM, EDTA 4 mM, pH 8,3) con Bromuro de Etidio (1µg/ml) sobre el cual se depositaron gotas de 1 µl del producto de cada *PCR* y una gota de 1 µl para cada uno de los siete patrones de concentración

conocida: 250, 125, 63, 32, 16, 8 y 4 ng/μl. Estos patrones se prepararon diluyendo en forma seriada el marcador V de *Boehringer-Manheim*.

Después de 20 minutos, cuando el gel había absorbido todas las gotas, se visualizó el DNA en el transiluminador de rayos UV, se tomó una fotografía y se estimó la concentración de cada producto de PCR comparando la intensidad de los productos resultantes de la PCR con la de los estándares.

3.6. Reacciones de hibridación:

3.6.1. Marcaje de las sondas:

Se realizó por el método de marcaje “*random-primed*” (cebadores aleatorios) (Feinberg y Vogelstein, 1983) utilizando los reactivos suministrados en el kit *Prime-a-Gene® Labelling System* de *Promega*.

La mezcla de reacción contiene:

- 300 μl de tampón 5x (Tris-HCl 250mM pH 8,0, MgCl₂ 25 mM, DTT 10 mM, HEPES⁽¹⁾ 1M pH 6,6, 26 unidades de A₂₆₀/ml de hexadesoxirribonucleótidos aleatorios),
- 20 μl de dCTP (1,5 mM), 20 μl de dGTP (1,5 mM) y 20 μl de dTTP (1,5 mM),
- 60 μl de seroalbumina bovina (10 mg/ml).

(1) Ácido (4-(2-Hidroxi-etil)-1-piperazina-etano-sulfónico

La cantidad "P" (en ng) de DNA que iba a ser marcado se calculó usando la concentración anteriormente estimada (apartado 3.5.4.2.) y se tomaron 30 ng de DNA. Después de desnaturalizar dos minutos a 100 °C en 14 µl de agua, se añadió 7 µl de la mezcla de reacción, 3 µl de $\alpha^{32}\text{P}$ -dATP (con una actividad específica de 3.000 Ci/mmol) y 1 µl de Klenow (5-10 unidades). Se incubó la reacción a 37 °C durante un mínimo de una hora.

Para rectificar la pequeña heterogeneidad resultante de defectos de carga del RNA entre los distintos pocillos del gel, se hibridaron posteriormente los filtros con un control de carga representado por el rRNA 25S. Se sintetizó una sonda oligonucleotídica (5'-CTCCGCTTATT GATATGC-3') a partir de la secuencia de DNA que codifica el rRNA 25S y se marcó mediante marcaje terminal con $\gamma^{32}\text{P}$ -ATP (Lillehaug y Kleppe, 1975).

En este caso se mezclaron en un volumen final de reacción de 15 µl:

- 5-10 pmol del oligonucleótido,
- 1 µl de tampón 10x (Tris-HCl 500 mM, MgCl_2 100 mM, EDTA 1mM, DTT 50 mM, espermidina 1 mM, pH 8,2),
- 1 µl de $\gamma^{32}\text{P}$ ATP,
- 1 µl de T4-polinucleótido quinasa (*Boehringer-Manheim*)

Se dejó incubando la reacción a 37°C durante 30 minutos y se añadió posteriormente a la solución de hibridación.

3.6.2. Separación y cuantificación de la incorporación de isótopo:

Para poder valorar la incorporación de isótopo en el DNA, se prepararon columnas de exclusión con Sephadex G-50 saturado con STE (Tris-HCl 10mM pH 7,5, EDTA 0,1 mM, SDS 0,1%). Una vez empaquetada, se equilibró la columna con STE, se añadió el DNA marcado y se recolectaron 6 fracciones, haciendo pasar por la columna 400 μ l de STE para la primera fracción y luego cinco veces 200 μ l de STE para las siguientes. El isótopo no incorporado se quedó retenido en las columnas.

A continuación, se localizaron las dos fracciones con mayor incorporación de radioactividad con el sensor de un contador Geiger y se mezcló el contenido de estos tubos (volumen V_1). Luego se tomaron 2 μ l (volumen V_2) de esta mezcla, se añadieron a 10 ml de líquido de centelleo (cóctel de centelleo biodegradable para muestras acuosas *NBCS 104* de *Amersham*) y en un contador de radiaciones β (*Wallac System 1410TM liquid scintillation counter* de *Pharmacia*) se determinaron las cuentas por minuto (cpm) de cada muestra (actividad A').

Se calculó la actividad específica de las diferentes sondas usando la siguiente relación:

$$\text{A.E. (cpm}/\mu\text{g DNA)} = \frac{A' \text{ (cpm)} \times V_1 \text{ (\mu l)} \times 1000}{V_2 \text{ (\mu l)} \times P \text{ (ng)}}$$

Una incorporación de isótopo adecuada suele dar actividades específicas comprendidas entre 0,5 y 2×10^9 cpm/ μ g DNA.

3.6.3. Reacciones de hibridación y lavado:

3.6.3.1. La hibridación:

Se introdujeron las réplicas en botellas de hibridación (una por botella) e inmediatamente después de añadirle 15 ml de una mezcla de hibridación (Na_2HPO_4 0,5 M pH 7,2, SDS 7%, EDTA 1 mM) (Church y Gilbert, 1984), se dejó prehibridando una hora a 65 °C en un horno de hibridación (*Minihybridisation oven* de APPLIGENE).

Una vez terminada la prehibridación, se renovó la mezcla de hibridación (10 ml) y se añadió la sonda marcada (400 μl) cuidadosamente para que no tocara la membrana antes de que se diluyese en la mezcla de hibridación. Se dejó hibridando toda una noche a 65 °C. Al día siguiente, por la mañana, se retiró la solución de hibridación y se procedió al lavado.

Para las hibridaciones con la sonda de rRNA 25S se procedió de la misma manera excepto que la temperatura de hibridación fue de 45°C.

3.6.3.2. Los lavados:

La membrana se lavó con 20 ml de 2xSSC (NaCl 0,3 M, citrato sódico 30 mM) y se dejó unos minutos a 65 °C. A continuación se cambió el 2xSSC por 30 ml de solución de lavado (SDS 0,1%, 2xSSC) precalentado a 65 °C y se dejó de 15 a 20 minutos a esa temperatura en el horno de hibridación. Para la sonda oligonucleotídica de rRNA 25S los lavados se hicieron a 45°C.

Después de retirar la solución de lavado, las membranas se introdujeron en una bolsa de plástico para poder exponerlas 48 horas a la pantalla del *PhosphorImager* (ver apartado 3.7.).

3.7. Cuantificación de las señales de hibridación:

Para analizar las señales radioactivas de hibridación sobre la membrana de nylon se usa un escaner (*PhosphorImager Scanning Instrument™* 425 de *Molecular Dynamics*). El fundamento de este sistema se describe a continuación.

El dispositivo posee una pantalla compuesta de cristales de BaFBr:Eu^{+2} que tienen la facultad de sufrir reacciones redox al ser excitados por radiaciones de alta energía (emitidas en este caso por el fósforo radioactivo) y según su estado de oxidación, emitir o absorber energía de diferente longitud de onda. Durante la exposición de la muestra a la pantalla de fósforo el Eu^{+2} es oxidado a Eu^{+3} y el BaFBr reducido a BaFBr^- .

Al escanear la pantalla de fósforo con un láser de Neón-Helio que emite una luz roja a una longitud de onda de 633 nm, el BaFBr^- absorbe la energía del láser provocando la liberación de sus electrones; BaFBr^- es oxidado y Eu^{+3} , captando estos electrones, se reduce a Eu^{+2} . Esta nueva reacción redox va a producir una liberación de energía a una longitud de onda diferente (luz azul) y proporcional a la radioactividad que estaba presente en la muestra. Esta energía es recolectada y posteriormente medida y procesada por un ordenador conectado al escaner. El procesamiento consiste en relacionar las unidades de radioactividad (cpm) de cada señal de hibridación con un número

de pixels (de diferentes intensidades y abarcando un rango de 0 a 10700) de imagen digital; se suele trabajar con un tamaño de pixel de 200 μm .

La cuantificación y el estudio de las señales de hibridación se hace gracias a un programa informático llamado *Image QuANT™*. Permite realizar la imagen ajustando el rango de análisis a los 256 niveles de grises y colores de la pantalla (elimina las altas intensidades de pixel inutilizadas y las bajas intensidades que aumentan el fondo contaminante). Además usando una de sus opciones (*Peak Finder*) se puede calcular la intensidad de las señales a lo largo de un área trazada en la imagen y que contiene la(s) señal(es).

Para poder distinguir correctamente los picos de la línea base (la cual representa el ruido de fondo) se necesita antes de realizar el análisis predeterminar una serie de parámetros:

- Ruido de fondo: permite destacar los picos correspondientes a señales de hibridación por encima del fondo contaminante. Su valor por defecto está comprendido entre 0,01 y 0,05, pero en caso de señales intensas (por ejemplo en la rehibridación con rRNA25S) se puede aumentar hasta un 1,00.
- Sensibilidad: toma valores comprendidos entre 0,03125 y 256 (niveles de grises y colores de la pantalla). Cuando hay muchos picos (señal débil o membrana insuficientemente lavada) se fija un valor de sensibilidad bajo (0,5 a 1) y se aumenta (hasta 8) cuando hay pocos picos sobresalientes.
- Kernel: se fijó a un valor de 20 que permitía excluir los picos estrechos y de escasa intensidad.

La elección de estos valores se realizó de modo empírico, teniendo en cuenta la naturaleza de las señales que se iban a cuantificar.

Los valores de área Sm (Señal medida) equivalen al número de pixels por su intensidad correspondientes al área delimitada entre la curva que define el pico y la línea base. Estos son proporcionales a las cpm de la señal y se utilizan para calcular los niveles relativo y absoluto de cada señal después de haber corregido la heterogeneidad debida a las diferencias de carga gracias a las hibridaciones con el rRNA 25S:

$$\text{SmORF}_2 = \text{SmORF}_1 \times \frac{\text{Sm25S}_2}{\text{Sm25S}_1}$$

Dónde SmORF_1 y Sm25S_1 son los valores medidos por el *Phosphorimager* para la ORF y el control respectivamente, Sm25S_2 es el valor teórico esperado para el control si la carga fuera totalmente homogénea y SmORF_2 el valor corregido de la señal para cada ORF.

Para cada transcrito cuantificado se calcularon dos valores: los niveles absolutos y relativos de mRNA.

3.7.1. Nivel relativo:

3.7.1.1. Nivel relativo a *ACT1*:

Se calculó usando la relación:

$$\text{N.R.}_{\text{ACT1}} (\text{Unidades arbitrarias}) = \frac{\text{SmORF}}{\text{SmACT1}_{\text{YPD}}}$$

Donde SmORF es el valor corregido del área para una determinada ORF, y SmACT1 es el valor corregido para la actina en la condición 2 de cultivo (YPD). Se usaron los valores obtenidos para calcular el nivel absoluto de mRNA de cada ORF.

3.7.1.2. Nivel relativo de las ORFs:

Los niveles relativos de mRNA permiten comparar las señales de cada transcrito bajo las diferentes condiciones de crecimiento analizadas. Se calcularon en relación a la señal máxima obtenida para este transcrito en un determinado medio, usando la relación:

$$\text{N.R.}_{\text{ORF}} (\%) = \frac{\text{SmORF}}{\text{SmORF}_{\text{max}}} \times 100$$

Donde SmORF es el valor corregido del área para una determinada ORF y SmORF_{max} es el valor máximo corregido obtenido para esta misma ORF en uno de los medios de cultivo.

3.7.2. Nivel Absoluto:

El nivel absoluto de mRNA fue calculado usando como base el nivel máximo observado para este transcrito en las condiciones analizadas en este estudio. *ACT1* es un estandar de referencia en el *Northern blot* para la medida de la abundancia de mRNAs, sin embargo los niveles de *ACT1* cambian durante el crecimiento o la privación de nutrientes (Delbruck y Ernst, 1993). Por este motivo los niveles absolutos de mRNA fueron calculados

comparándolos con el nivel de mRNA de *ACT1* observado durante el crecimiento exponencial en glucosa (condición 2) usando la relación:

$$\text{N.A. (unidades arbitrarias)} = \frac{\text{A.E.}_{ACT1} \times \text{N.R.}_{ACT1 \text{ max}}}{\text{A.E.}_{ORF}}$$

Donde A.E. es la actividad específica de cada sonda utilizada (ver apartado 3.6.2.) y $\text{N.R.}_{ACT1 \text{ max}}$ el nivel relativo máximo de una ORF cuando se compara al nivel de expresión de *ACT1* en YPD.

El nivel absoluto de mRNA calculado por el procedimiento descrito nos permitió clasificar las ORFs por su tasa de expresión:

- Alta expresión: $\text{N.A.} > \cong 0.5$
- Mediana expresión: $\cong 0.1 < \text{N.A.} < 0.5$
- Baja expresión: $\text{N.A.} < 0.1$
- No detectable: ausencia de señal.

3.8. Búsqueda de características deducidas de las secuencias de ADN:

3.8.1. Determinación del uso de codones:

La desviación en el uso de codones se cuantificó mediante el cálculo del CAI (*Codon Adaptation Index*) propuesto por Sharp y Li (1987).

El CAI está basado en la determinación de un valor w para cada codon de la ORF y representa la frecuencia relativa de uso del codon, utilizando

como grupo de referencia un conjunto de 24 genes de alta expresión. En posesión del w , se puede calcular el CAI mediante la relación:

$$\text{CAI} = \exp 1/L \sum \ln w_i$$

donde, L es la longitud del gen en codones y w_i el valor calculado para los diferentes codones, excluyendo Met, Trp y los codones de terminación.

Los valores de CAI oscilan de 0 a 1, los valores más elevados indican mayor desviación en el uso de codones.

Los valores de CAI que se recogen en este trabajo, para las distintas ORFs analizadas, proceden de la página *Web* de MIPS (Mewes *et al.*, 1997a) con acceso "<http://www.mips.biochem.mpg.de>" y están calculados mediante el programa CODONS (Lloyd y Sharp, 1992). Este programa calcula una serie de datos e índices relativos a la composición en aminoácidos y al uso de codones.

3.8.2. Estudio de los promotores de las ORFs:

Una región de 800 pb en posición 5' respecto del ATG de inicio de cada ORF estudiada fue sometida a una búsqueda de sitios de unión de factores reguladores de la transcripción conocidos, utilizando para ello los programas asociados a la base de datos *TRANSFAC-The Transcription factor database* (Wingender *et al.*, 1997) de acceso "<http://transfac.gbf.de/>":

- *MatInspector 2.1* (Quandt *et al.*, 1995) compara la secuencia analizada a las secuencias matrices de unión a factores de transcripción de la librería *TRANSFAC3.2* para *S. cerevisiae*. Identifica los sitios de unión a Abf1p, Gal4p, Gcn4p, Gcr1p, Hsfp, Matalp, Mcm1p, Mig1p, Pho4p, Rap1p, y secuencias STRE y para el represor de *CARI*. El programa escanea las dos hebras de la secuencia de manera simultánea.
- *Pattern Search 1.1(GBF-Braunschweig)* que, además de utilizar *TRANSFAC 3.2*, también utiliza la base de datos *TRRD 3.4*; ésta se compone de sitios de unión a factores de transcripción (75 en *S.cerevisiae*) comprobados experimentalmente en numerosos genes. El programa escanea únicamente la hebra analizada y no la complementaria.

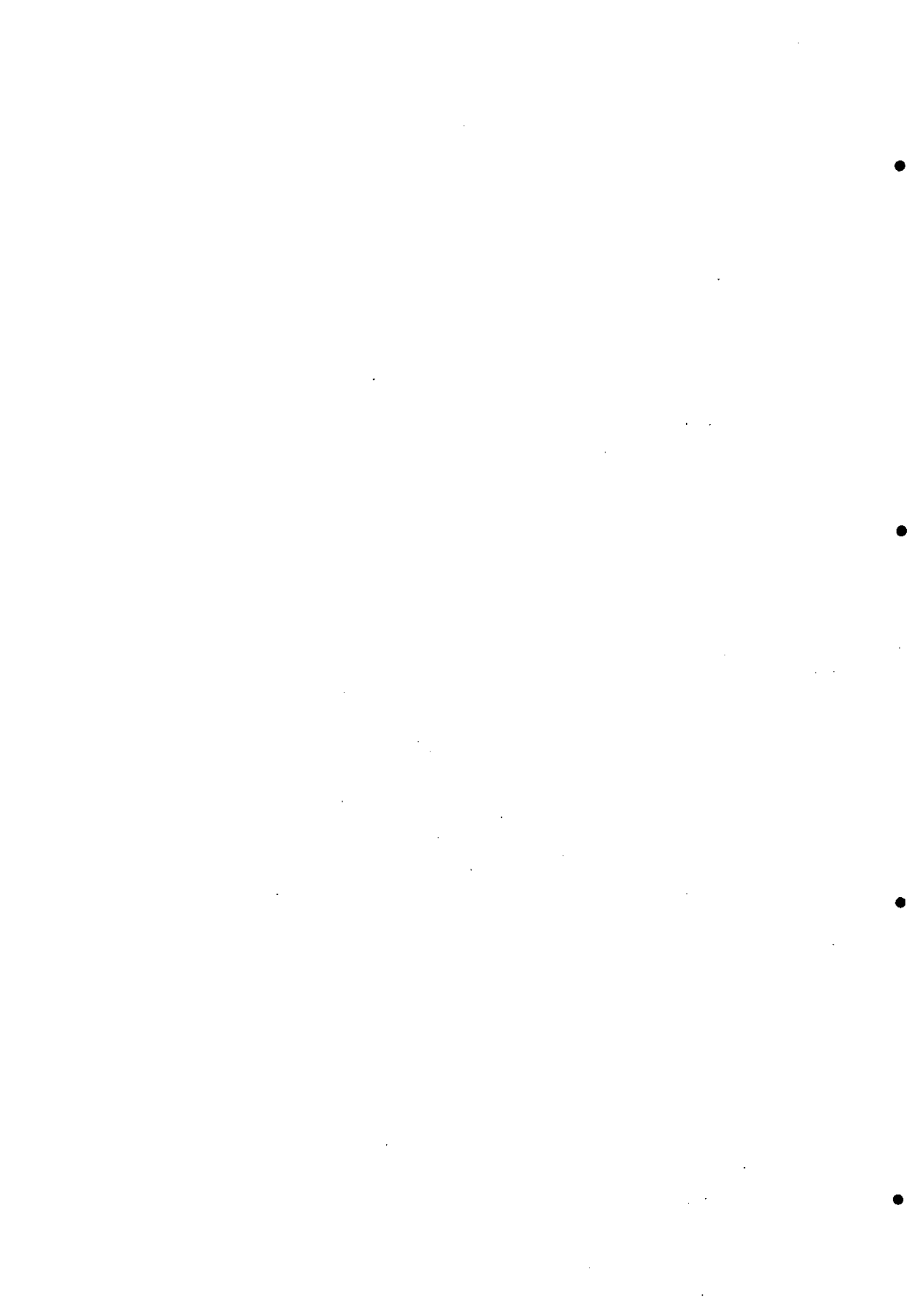
La búsqueda se hace, en los dos casos, con parámetros predefinidos en cuanto a homología de secuencias; en este trabajo el umbral utilizado fue del 70% excluyéndose del resultado de la búsqueda todos los sitios con una homología menor.

3.8.3. Otras características deducidas de las secuencias:

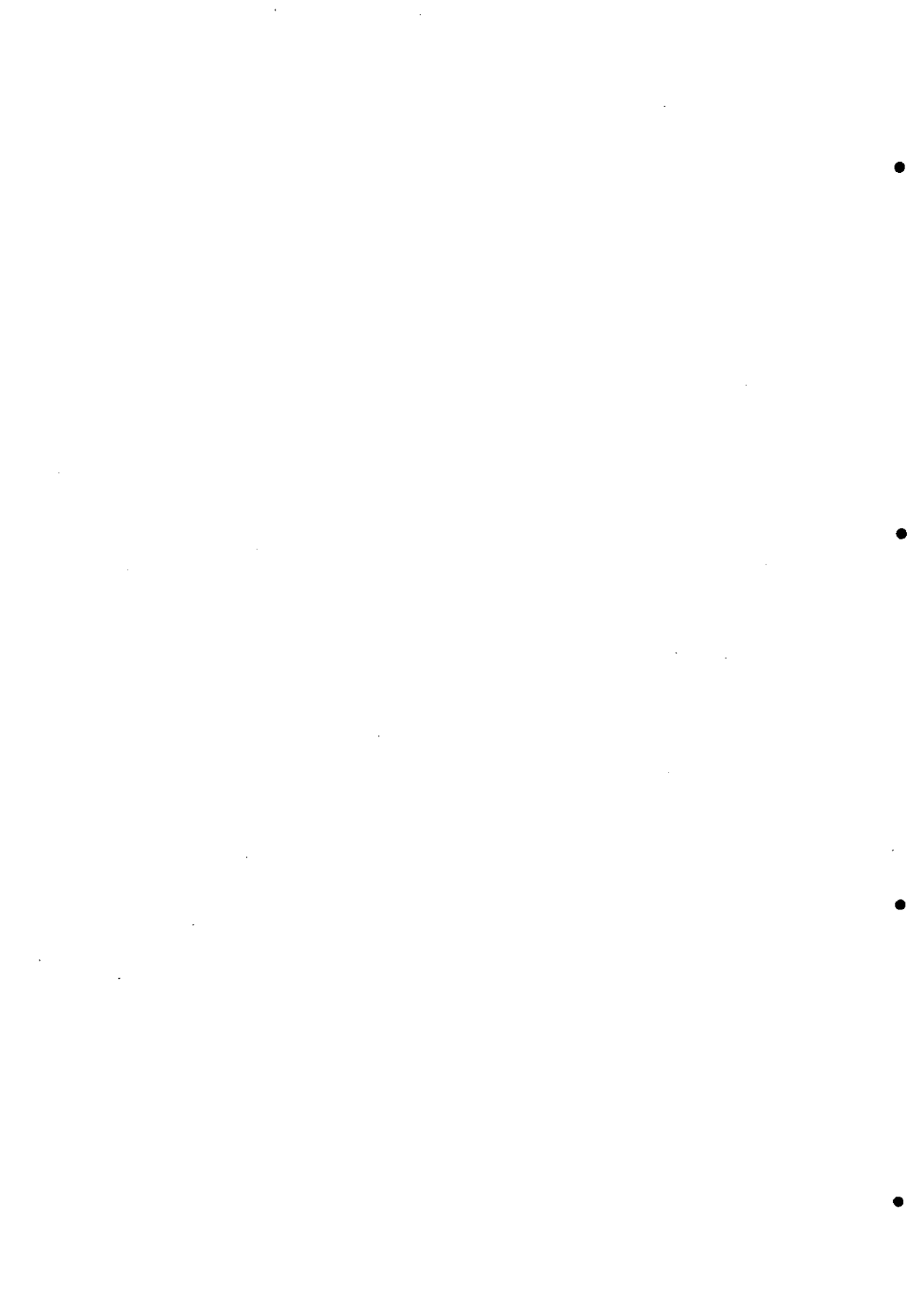
La base de datos *MIPS* proporciona otros parámetros a partir de la secuencia de las ORFs desconocidas:

- Homologías con otros genes o proteínas, procariotas o eucariotas, encontradas mediante programas de comparación con bases de datos de ácidos nucleicos y de proteínas

- Predicciones de regiones transmembrana mediante los programas *TMAP* (Persson y Argos, 1994) y *ALOM* que tiene en cuenta el algoritmo *KKD* (Klein *et al.*, 1985)
- Localización de intrones a través del programa *EXPLORA* (Kalogeropoulos, 1995)
- Predicciones de localización subcelular llevadas a cabo según el método diseñado por Nakai y Kanehisa (1992)
- Fenotipos provocados por la disrupción del gen y obtenidos en el consorcio B0
- Otros valores como la predicción del punto isoelectrico o el peso molecular del precursor.



4. RESULTADOS Y DISCUSION



4.1. Selección de las condiciones de análisis:

En este trabajo se exponen los resultados obtenidos en el estudio de la expresión de 51 y 52 ORFs, respectivamente de los cromosomas VII y XIV de *S. cerevisiae*. Al estudiar la secuencia nucleotídica de cada uno de los cromosomas se pudieron deducir numerosas características que se resumen en la siguiente tabla (Tabla 4.1).

Tabla 4.1. Características de los cromosomas VII y XIV de *S. cerevisiae* deducidas del estudio de sus secuencias nucleotídicas respectivas (ND: No Determinado)

CARACTERÍSTICAS		CROMOSOMA VII (Tettelin <i>et al.</i> , 1997)	CROMOSOMA XIV (Philippsen <i>et al.</i> , 1997)
Tamaño cromosoma (pb)		1.090.936	784.328
Nº genes potenciales		574	408
Tamaño medio ORF (codones)		468	500
Genes caracterizados experimentalmente		256 (44,6%)	138 (33,8%)
ORFs similares a genes conocidos		261 (45,5%)	43 (10,5%)
Huérfanos			204 (50%)
ORFs cuestionables		57 (9,9%)	23 (5,6%)
ORFs con intrones		19	16
ORFs redundantes		166 (29%)	116 (29,7%)
ORFs con hélices transmembrana		79 (14%)	68 (17%)
Similitud con proteínas humanas	Genes	79 (13,8%)	ND
	Huérfanos	16 (2,8%)	12 (3%)

Para estudiar en que medida se comportaban como “genes” las restantes 341 y 247 ORFs desconocidas de cada cromosoma respectivo, era preciso observar si estos genes potenciales se expresaban realmente ya que los procesamientos informáticos son insuficientes en este tipo de determinación. Existen varias técnicas de análisis de expresión para este tipo de estudio:

- La hibridación *in situ* (Schul *et al.*, 1996) es un método de detección citoimmunoquímica de la expresión génica a través de hibridaciones con sondas marcadas de células previamente fijadas. Este método da una información directa y precisa de la transcripción génica, sin embargo no es adecuado para estudios a gran escala y no informa sobre el tamaño y número de transcritos.
- La cartografía mediante PCR, *PCR-mapping* (Crauwels *et al.*, 1997), que consiste en amplificar extractos de mRNA por RT-PCR (*Reverse Transcriptase-PCR*) y estudiar la expresión génica mediante la separación del cDNA obtenido en un gel; permite detectar incluso los genes de muy baja expresión y proporciona información sobre el tamaño del transcrito.
- Las fusiones a GFP (*Green Fluorescent Protein*) consisten en introducir el gen *GFP*, que codifica una proteína fluorescente, en el locus del gen estudiado y cuantificar la producción de GFP mediante citometría de flujo (Niedenthal *et al.*, 1996). Es un método ventajoso porque permite el estudio de la expresión génica *in vivo* y proporciona datos sobre la localización subcelular de la proteína estudiada. Pese a todo, no informa sobre el tamaño y número de transcritos.
- El método de *Hybridisation array* (Fodor *et al.*, 1993, Shena *et al.*, 1995; Hoheisel, 1997) permite, a partir de mRNA, sintetizar cDNA para marcarlo e hibridarlo con oligonucleótidos específicos, de las ORFs que se quieren estudiar, fijados a una plantilla o panel (*array*). Esta técnica permite la elaboración rápida de mapas transcripcionales, pero al inicio del proyecto EUROFAN I no estaba

optimizada para levadura y además no proporciona información sobre la integridad y el tamaño de los transcritos. Recientemente esta técnica se ha aplicado al análisis del “transcriptoma” de levadura (De Risi *et al.*, 1997; Wodicka *et al.*, 1997) y se está convirtiendo en una herramienta potente y efectiva en el estudio funcional de transcriptomas a gran escala.

- El método *SAGE* (*Serial Analysis of Gene Expression*) consiste en sintetizar cDNA a partir de una extracción de RNA de células en diferentes condiciones de cultivo. Este cDNA es digerido específicamente en fragmentos (*Tags*) de unos 10 pb específicos de los transcritos. Los *Tags* son ligados posteriormente entre sí, y la hebra de DNA así obtenida es finalmente secuenciada (Velculescu *et al.*, 1995). La secuenciación del DNA proporciona el número real de transcritos para cada ORF estudiada. Este método fue aplicado a la totalidad del genoma de *S. cerevisiae* (Velculescu *et al.*, 1997) y se pudo así obtener datos en cuanto a la expresión de 4.665 genes (76%) de los 6.121 que predice el análisis de la secuencia total del genoma de levadura. Esta técnica permitió detectar la expresión de ORFs de tamaño inferior a 100pb no consideradas inicialmente como genes reales pero, por otro lado, no tiene suficiente sensibilidad para detectar los transcritos de ORFs de muy baja expresión (inferior a 0,3 transcritos por célula). Además no proporciona información sobre el tamaño de los transcritos.
- La técnica de *Dot/Slot-blot* (Kafatos *et al.*, 1979) consiste en depositar gotas (*dots*) o líneas (*slots*) de extractos totales de mRNA sobre una membrana de nylon o de nitrocelulosa para posteriormente hibridarlas con sondas específicas marcadas. A pesar

de no dar información sobre el tamaño y número de transcritos, se adapta a estudios a gran escala, por el menor coste y la rapidez de la información obtenida. Además es compatible con el estudio de la regulación de transcritos abundantes (Mager *et al.*, 1998) pero se comprobó que no tenía sensibilidad suficiente para ser aplicable a las ORFs analizadas en este estudio debido a que son de baja expresión.

La técnica elegida en este trabajo, a pesar de ser pesada y costosa para estudios a gran escala, fue la del *Northern blot* porque, además de investigar la funcionalidad de las ORFs desconocidas, también es importante determinar el tamaño real de cada transcrito con el fin de comprobar si el tamaño de la secuencia de la ORF estudiada se adecua al de su transcrito y, de este modo confirmar la ausencia de errores durante el proyecto de secuenciación del genoma de la levadura.

Para asignar rigurosamente cada transcrito a su respectiva ORF y minimizar las fluctuaciones en la precisión de la señal obtenida, se emplearon como sondas, productos de PCR de tamaño uniforme (alrededor de 300pb) amplificados con cebadores específicos de cada una de las 103 ORFs estudiadas.

Se estudió la expresión de las 103 ORFs en distintas condiciones de cultivo, ya que el estudio de su regulación podía aportar algunos datos sobre la función de las proteínas desconocidas.

En una primera fase se estudiaron las 52 ORFs del cromosoma XIV en tres condiciones de cultivo: medio rico con glucosa o etanol/glicerol (YPD y YPGE respectivamente) y medio mínimo (YNB) recogiendo las células en la

fase temprana de crecimiento exponencial. Dado el escaso número de respuestas reguladoras obtenidas (Lombardia, tesis de licenciatura, 1996), se decidió en una segunda etapa estudiar todas las ORFs en experimentos de transición de unas condiciones a otras. Estos se concretaron cultivando las células bajo 8 condiciones distintas que se describen detalladamente en Materiales y Métodos (apartado 3.2.2. página 36) y que hacen referencia a 5 tipos de regulación (Tabla 4.2.).

Tabla 4.2. Descripción de los cinco tipos de regulación estudiada. (1-8 hace referencia a las condiciones descritas en las páginas 36 a 38).

REGULACIÓN	Comparación de la intensidad de las señales en los diferentes medios	DENOMINACIÓN
Por fuente de carbono (glucosa)	1<2 -1>2	G+ (activación) G- (represión)
Por fuente de nitrógeno	4>5 4<5	N+ (activación) N- (represión)
Activación en fase estacionaria	3>2/1	S+
Activación por choque osmótico	6>4	O+
Activación por choque térmico	8>7	H+

Teniendo en cuenta los intereses de investigación particulares de nuestro laboratorio, decidimos incluir 3 condiciones adicionales y externas al proyecto EUROFAN I. Se proyectó así, estudiar el comportamiento de las ORFs desconocidas utilizando tres cepas de *S. cerevisiae* donde habían sido mutados los genes *HAP1*, *ROX1* o *ROX6* respectivamente que son genes reguladores de la expresión génica en relación con las condiciones de aerobiosis y anaerobiosis.

4.2. Verificación del método de análisis con sondas de regulación conocida:

Antes de analizar la expresión de las ORFs de función desconocida se verificó el método de análisis a fin de comprobar su correcta respuesta respecto a varios genes de regulación conocida. Los *Northern blots* se realizaron según se ha descrito en Materiales y Métodos. La preparación de las sondas de hibridación se realizó mediante *PCR*, usando cebadores (Tabla 3.3.) y DNA moldes cuyas características fueron descritas en Materiales y Métodos.

Los resultados de los cálculos de los niveles de expresión relativo y absoluto (posteriores a la corrección con el rRNA 25S) correspondientes a las hibridaciones con las sondas control (Figura 4.1.) se recopilan en la Tabla 4.3.

Tabla 4.3. Valores de Nivel Relativo (N.R.) y Absoluto (N.A.) correspondientes a las hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo: 1-YPGE, 2-YPD, 3-YPD/fase estacionaria, 4-YNBglu/privación de aminoácidos, 5-YNBglu/privación de nitrógeno, 6-YNBglu/choque osmótico, 7-YNBglu/23°C, 8-YNBglu/36°C, con sondas de regulación conocida. (A.E.: Actividad específica de las sondas $\times 10^6$ cpm/ μ g de DNA).

GEN	A.E.	NIVELES RELATIVOS (%)								N.A.
		1	2	3	4	5	6	7	8	
<i>PCK1</i>	1886	100	4	64	2	1	<1	<1	<1	1,27
<i>RpL25</i>	1648	64	100	12	19	15	30	45	26	3,20
<i>ARG4</i>	511	5	3	2	69	14	86	70	100	1,90
<i>HSP12</i>	1251	95	4	64	10	98	95	7	100	2,61
<i>CARI</i>	1353	26	100	9	7	38	4	2	2	0,26

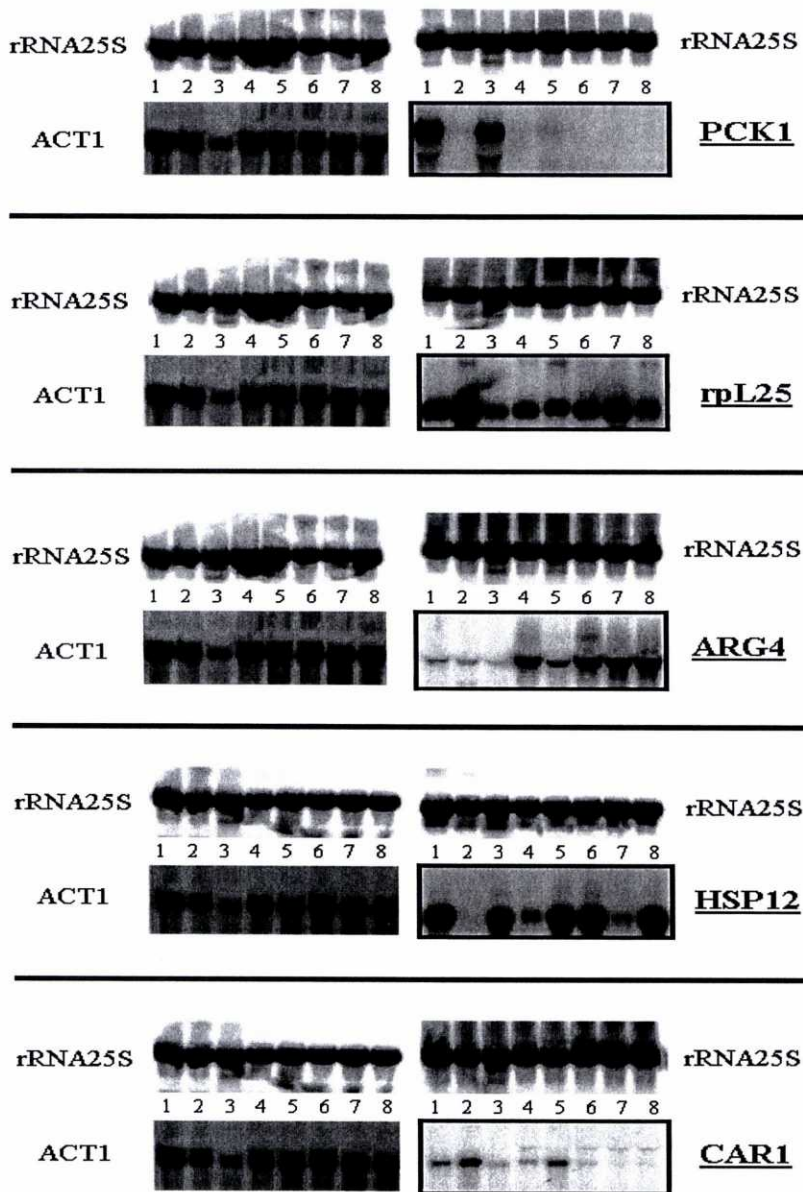


Figura 4.1. Control de hibridación de los RNAs extraídos en diferentes condiciones de cultivo (1-YPGE, 2-YPD, 3-YPD/fase estacionaria, 4-YNBglu/privación de aminoácidos, 5-YNBglu/privación de nitrógeno, 6-YNBglu/choque osmótico, 7-YNBglu/23°C, YNBglu/36°C) con las sondas controles correspondientes a los genes *PCK1*, *rpL25*, *ARG4*, *HSP12* y *CAR1*. Cada membrana fue hibridada de nuevo con una sonda oligonucleotídica específica del rRNA25S para poder realizar las correcciones de la heterogeneidad de carga.

Los niveles absolutos, tanto de las hibridaciones con sondas controles como con las sondas específicas de las ORFs de función desconocida, se calcularon usando el nivel de expresión del gen *ACT1* en la condición de cultivo número 2 (YPD). *ACT1* codifica para una proteína, la actina, que constituye el citoesqueleto de todas las células eucariotas, además está implicada en los procesos de polarización celular y de endocitosis. Su expresión se usa habitualmente también como control de carga en los experimentos de *Northern blot*; sin embargo en las condiciones probadas en este trabajo, se comprobó que el nivel de expresión de *ACT1* decaía significativamente en las condiciones de cultivo 3 (fase estacionaria) y 5 (choque osmótico) y por consiguiente sólo se usó su correspondiente señal en el medio 2 para calcular los niveles absolutos de cada ORF.

Los genes utilizados como controles de regulación se eligieron porque presentaban respuestas típicas, caracterizadas previamente por otros investigadores, en las condiciones de cultivo que se utilizaron en EUROFAN:

- *PCK1*: se usó como control para la represión por glucosa porque codifica la fosfoenolpiruvato carboxikinasa, una enzima clave en la gluconeogenesis (vía metabólica que se activa cuando las células crecen en ausencia de glucosa con una fuente de carbono como el etanol o el acetato y que disminuye su actividad en presencia de glucosa). Se conocen tres niveles de regulación para esta enzima: la inactivación de la enzima presente en el citoplasma por fosforilación reversible o interconversión (Burlini *et al.*, 1989), la degradación proteolítica o inactivación por glucosa (Haarasilta y Oura, 1975; Gancedo y Schwezmann, 1976), y la represión de la transcripción por glucosa (Haarasilta y Oura, 1975, Yin *et al.*, 1996). Además, en *S. cerevisiae*, los niveles de mRNA, transcrito a

partir del gen *PCK1*, son muy elevados en una fuente de carbono gluconeogénica (Mercado *et al.*, 1994). En la figura 4.1 podemos comprobar el elevado nivel de expresión de *PCK1* en etanol-glicerol (condición 1) y su represión en el medio con glucosa (condición 2).

- *rpl25*: codifica una de las proteínas ribosomales, la subunidad mayor de la proteína ribosómica 25 que forma los ribosomas, y se utilizó como control de la inducción por glucosa y regulación en función de la tasa de crecimiento (Kraakman *et al.*, 1993). Los ribosomas participan activamente en la síntesis de proteínas y representan una parte importante del peso seco de las células (Nomura *et al.*, 1984). Los genes codificadores de proteínas ribosómicas son de alta expresión y además la transcripción de estos genes es proporcional a la tasa de síntesis proteica, la cual varía en función de las condiciones de crecimiento (la síntesis de proteínas disminuye cuando el medio de cultivo no es óptimo). Por ejemplo, en *S. cerevisiae* se observó que una privación en aminoácidos del medio de cultivo provocaba una reducción significativa (por un factor de 2 a 4) de los mRNAs procedentes de genes que codifican proteínas ribosómicas (Moehle y Hinnebusch, 1991). En las hibridaciones realizadas (Figura 4.1.) se confirma tanto el alto nivel de expresión de este gen en glucosa, como la disminución de la expresión cuando las condiciones son menos óptimas.
- *ARG4*: codifica la argino-succinato liasa, una enzima que cataliza la reacción final en la síntesis de la arginina. Su expresión es muy baja, tanto en YPD como en YPGE (Penn *et al.*, 1983); pero como muchos otros genes que codifican para enzimas que participan en la biosíntesis de los aminoácidos, el gen *ARG4* se desreprime cuando se priva a la levadura del aminoácido correspondiente en el medio de cultivo (Niederberger *et al.*, 1981; Greer *et al.*, 1982). Esta pauta de expresión se observó en las

hibridaciones realizadas (Figura 4.1.); cabe señalar la bajada del nivel de expresión en la condición 5 debida sin duda a la privación de fuente de nitrógeno en el medio de cultivo.

- *HSP12*: codifica para una proteína de choque térmico (*Heat Shock Protein*) Hsp12p, cuya síntesis es baja en condiciones normales, se acumula considerablemente en células de levadura expuestas a un choque térmico, a un estrés osmótico u oxidativo, a altas concentraciones de alcohol y glicerol y en células en fase estacionaria temprana (Piper *et al.*, 1986; Praekelt y Meacock, 1990; Varela *et al.*, 1995). Este patrón de regulación concuerda con los datos obtenidos en las hibridaciones; incluso se observó una fuerte expresión de *HSP12* con la privación de nitrógeno en el medio de cultivo, cuyo efecto puede ser traducido como un factor de estrés por parte de la célula. En nuestros análisis se usó como control de las respuestas al estrés (choques osmóticos y térmicos) y de la fase estacionaria.
- *CARI*: codifica la arginasa, una enzima que degrada la arginina para abastecer a la célula de nitrógeno cuando éste escasea. La expresión de *CARI* se induce con la acumulación intracelular de arginina, y se reprime en presencia de fuentes de nitrógeno óptimas como el amonio (Smart *et al.*, 1996; Dubois y Messenguy, 1997). Los resultados de las hibridaciones confirman que *CARI* se expresa en medio rico (condiciones 1 y 2) y cuando se priva a la célula de fuente de nitrógeno (condición 5). Se utilizó en este proyecto como control para la privación de nitrógeno.

Por consiguiente, se puede considerar que los patrones de regulación obtenidos para cada control concuerdan con los que vienen dados en la bibliografía.

4.3. Análisis mediante Northern blot de la expresión de las ORFs objeto de estudio :

Una vez obtenidas las sondas y realizadas las transferencias del RNA total a membrana se procedió a realizar las hibridaciones y cuantificaciones según se ha descrito en Materiales y Métodos. Las imágenes resultantes de las hibridaciones de las ORFs de función desconocida de los cromosomas VII y XIV vienen representadas por orden creciente en la numeración de trabajo en las Figuras 4.2.1 a 4.2.26. Del mismo modo, se recopilaron los valores de los niveles de mRNA relativo y absoluto de cada ORF hibridada (Tabla 4.4).

En las figuras que acompañan a las tablas de cuantificación se incluye un dato relativo al tamaño mínimo del transcrito, éste es un valor de referencia que se corresponde exactamente con el tamaño de la ORF expresado en nucleótidos. Teniendo en cuenta que los puntos de inicio de la transcripción en levaduras pueden estar situados, respecto al ATG, entre las posiciones -60 y -400, y que las señales de terminación, más las colas poli-A características del mensajero, pueden suponer un incremento adicional entre 150 y 400 nucleótidos, el tamaño real del transcrito puede verse aumentado entre 210 y 800 nucleótidos. La mayoría de los valores obtenidos de manera experimental se adecuaron a los valores teóricos calculados, si bien la determinación exacta de su longitud queda pendiente de la localización concreta de los puntos de inicio y terminación de la transcripción.

Para el cálculo de los niveles relativos de expresión se eligió como referencia la condición de cultivo en la que se obtuvo el mayor nivel de expresión. El nivel absoluto se calculó por referencia al de actina en la condición de cultivo número 2. En los casos en que se obtuvieron múltiples transcritos (MT), no se realizó la cuantificación.

A continuación se muestran los resultados obtenidos (Figuras 4.2.1 a 4.2.26.) en las hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo: 1-YPGE, 2-YPD, 3-YPD/fase estacionaria, 4-YNBglu/privación de aminoácidos, 5-YNBglu/privación de nitrógeno, 6-YNBglu/choque osmótico, 7-YNBglu/23°C, 8-YNBglu/36°C, A-YPGE/*Sc.rox1*, B-YPGE/*Sc.rox6*, C-YPGE/*Sc.hap1*, con sondas específicas de 103 ORFs de función desconocida de los cromosomas VII y XIV de *S. cerevisiae*. Cada hibridación está acompañada con su respectiva hibridación con *ACT1* para el cálculo del nivel absoluto y con rRNA25S para la corrección de carga. Entre paréntesis se señala la nomenclatura estándar de cada ORF. También se indica el tamaño mínimo esperado del transcrito.

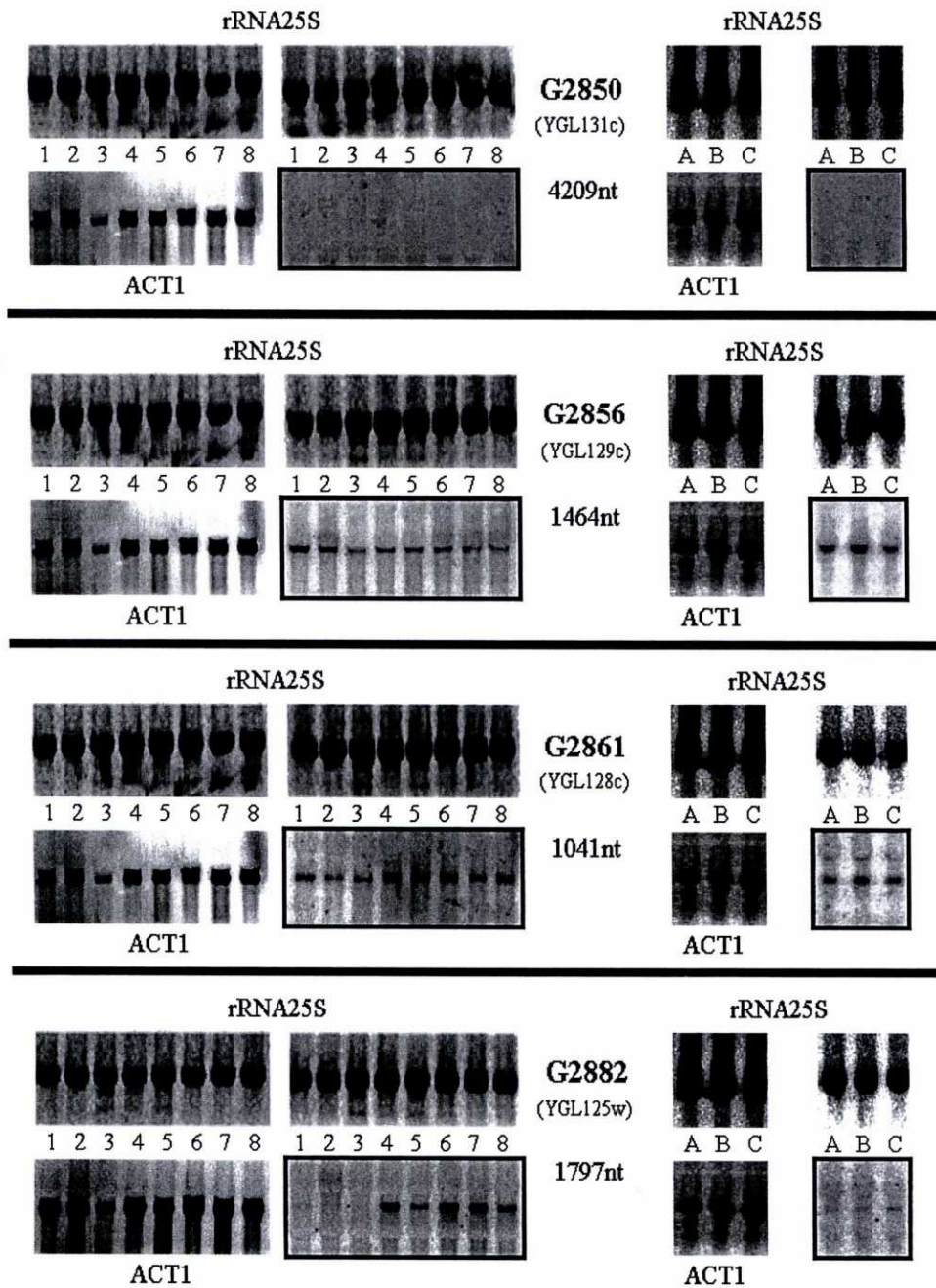


Figura 4.2.1. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida G2850, G2856, G2861 y G2882 del cromosoma VII de *S. cerevisiae*.

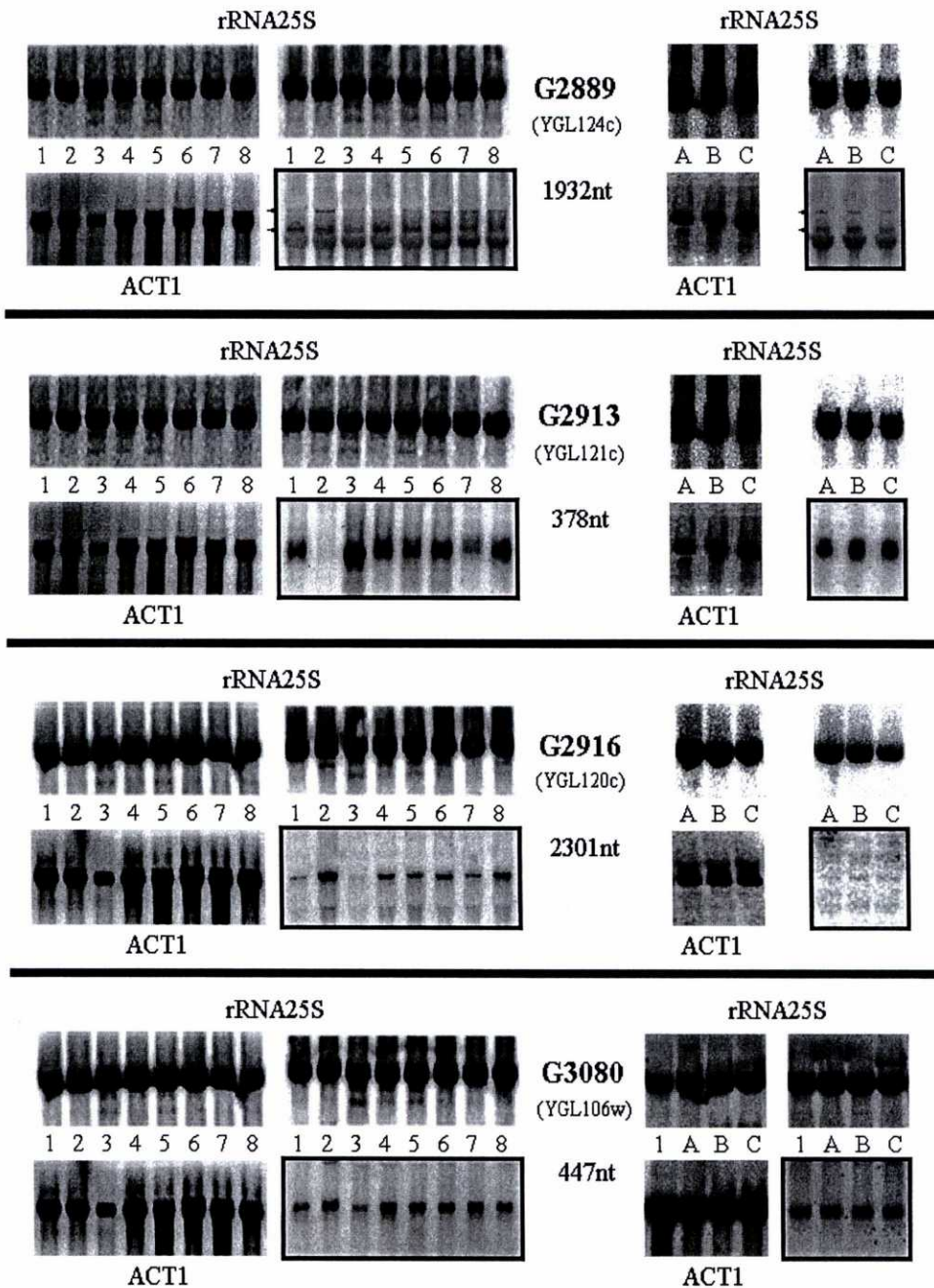


Figura 4.2.2. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida G2889, G2913, G2916 y G3080 del cromosoma VII de *S. cerevisiae*.

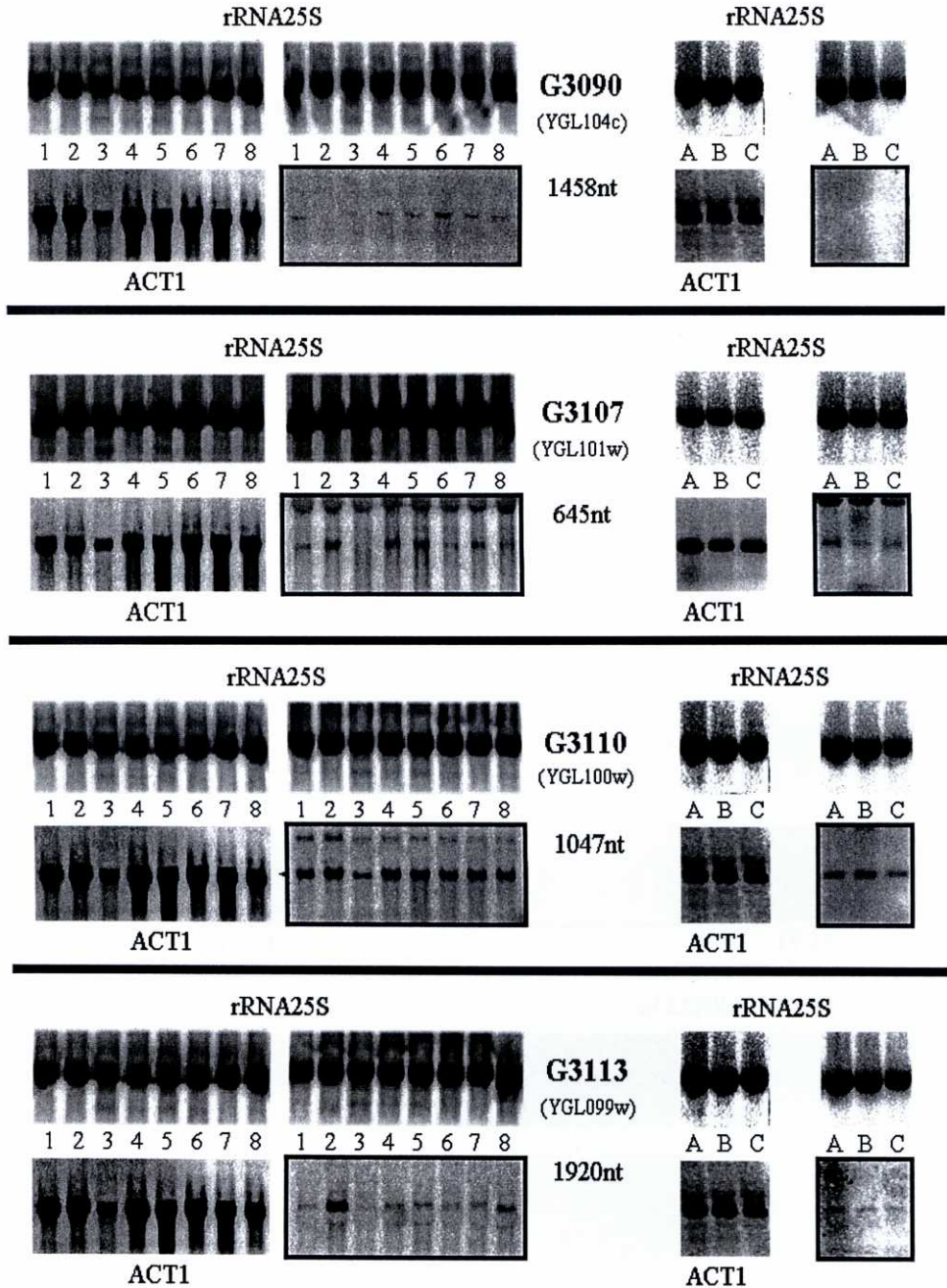


Figura 4.2.3. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida G3090, G3107, G3110 y G3113 del cromosoma VII de *S. cerevisiae*.

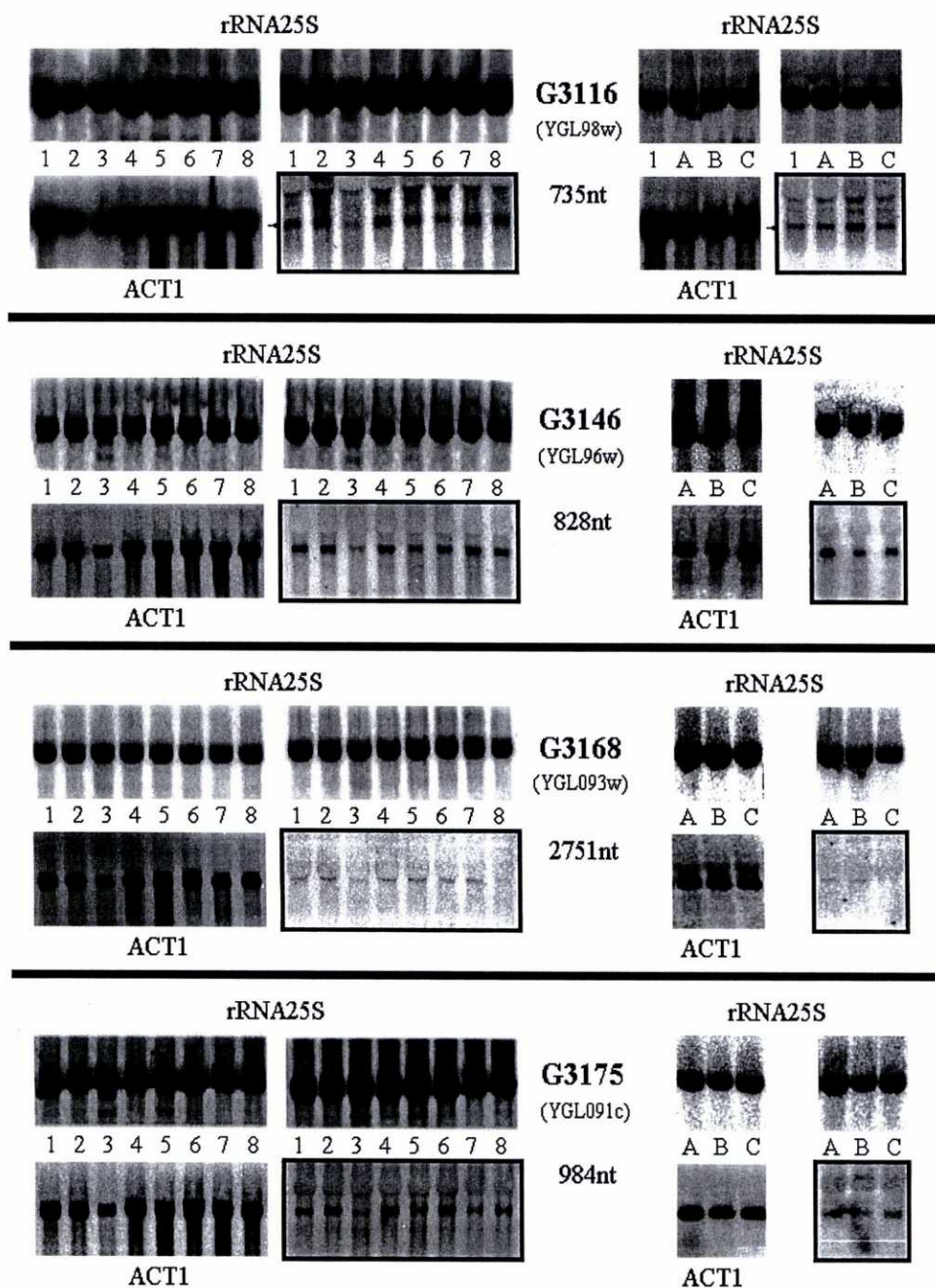


Figura 4.2.4. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida G3116, G3146, G3168 y G3175 del cromosoma VII de *S. cerevisiae*.

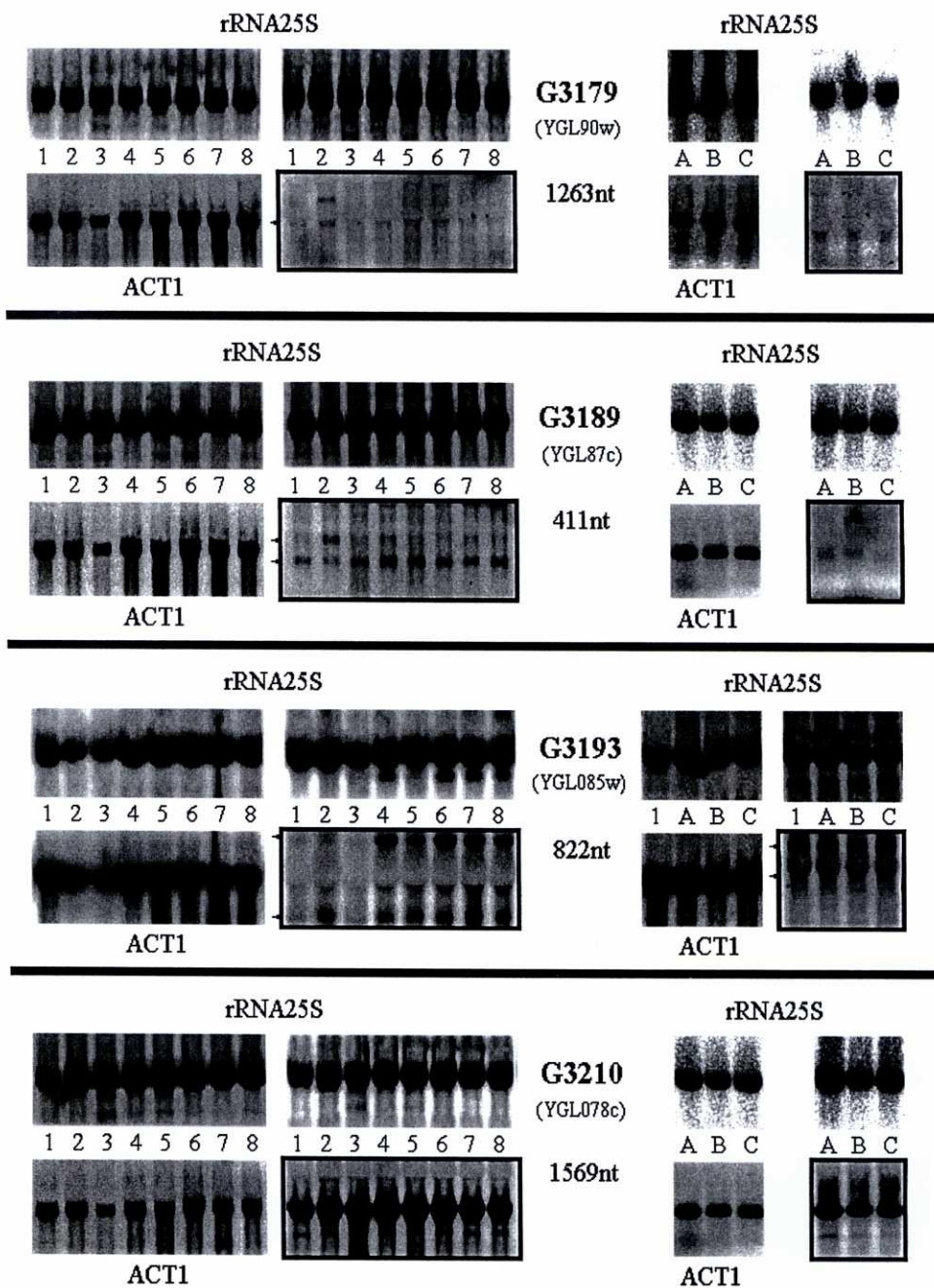


Figura 4.2.5. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida G3179, G3189, G3193 y G3210 del cromosoma VII de *S. cerevisiae*.

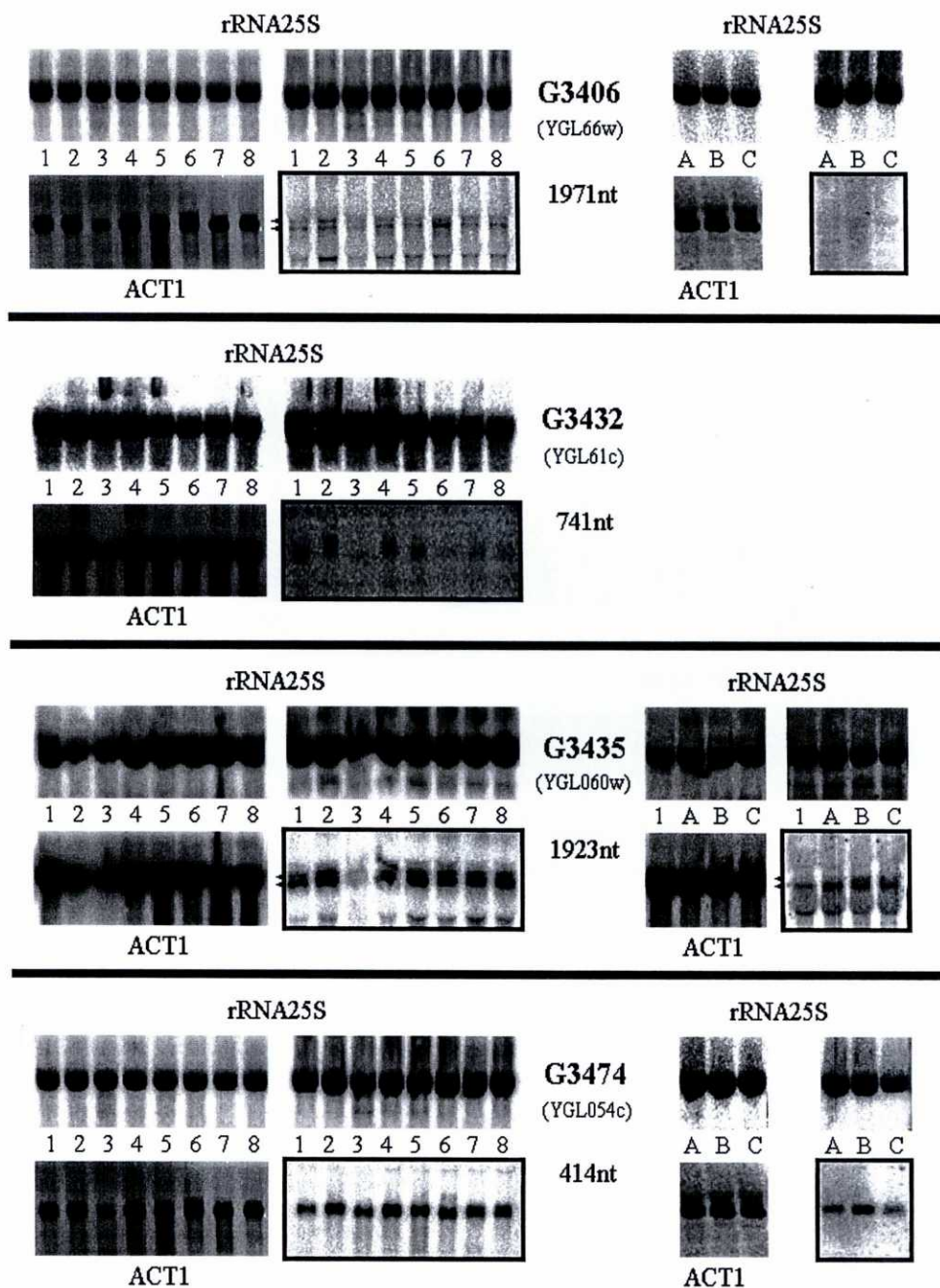


Figura 4.2.6. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida G3406, G3432, G3435 y G3474 del cromosoma VII de *S. cerevisiae*.

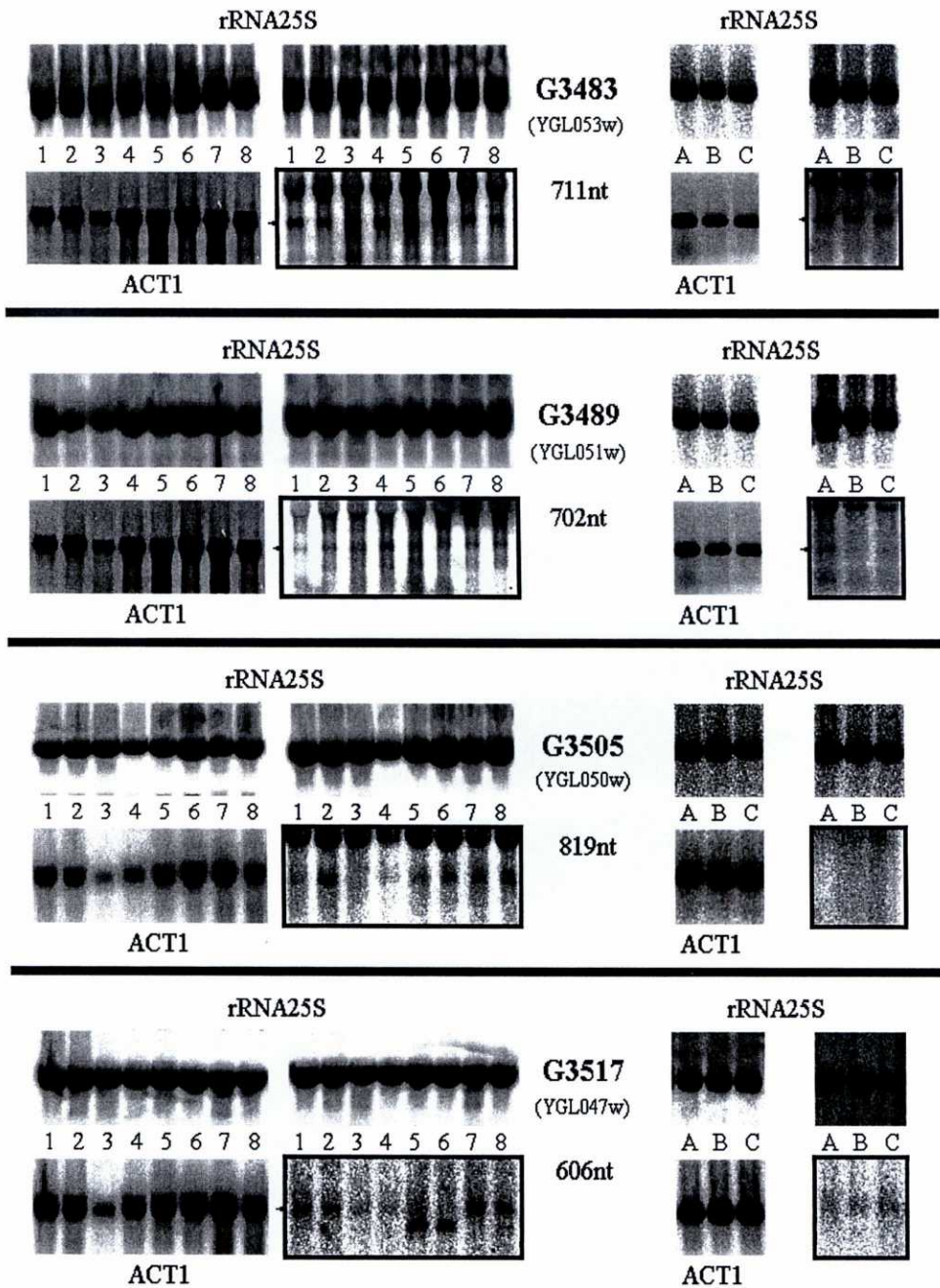


Figura 4.2.7. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida G3483, G3489, G3505 y G3517 del cromosoma VII de *S. cerevisiae*.

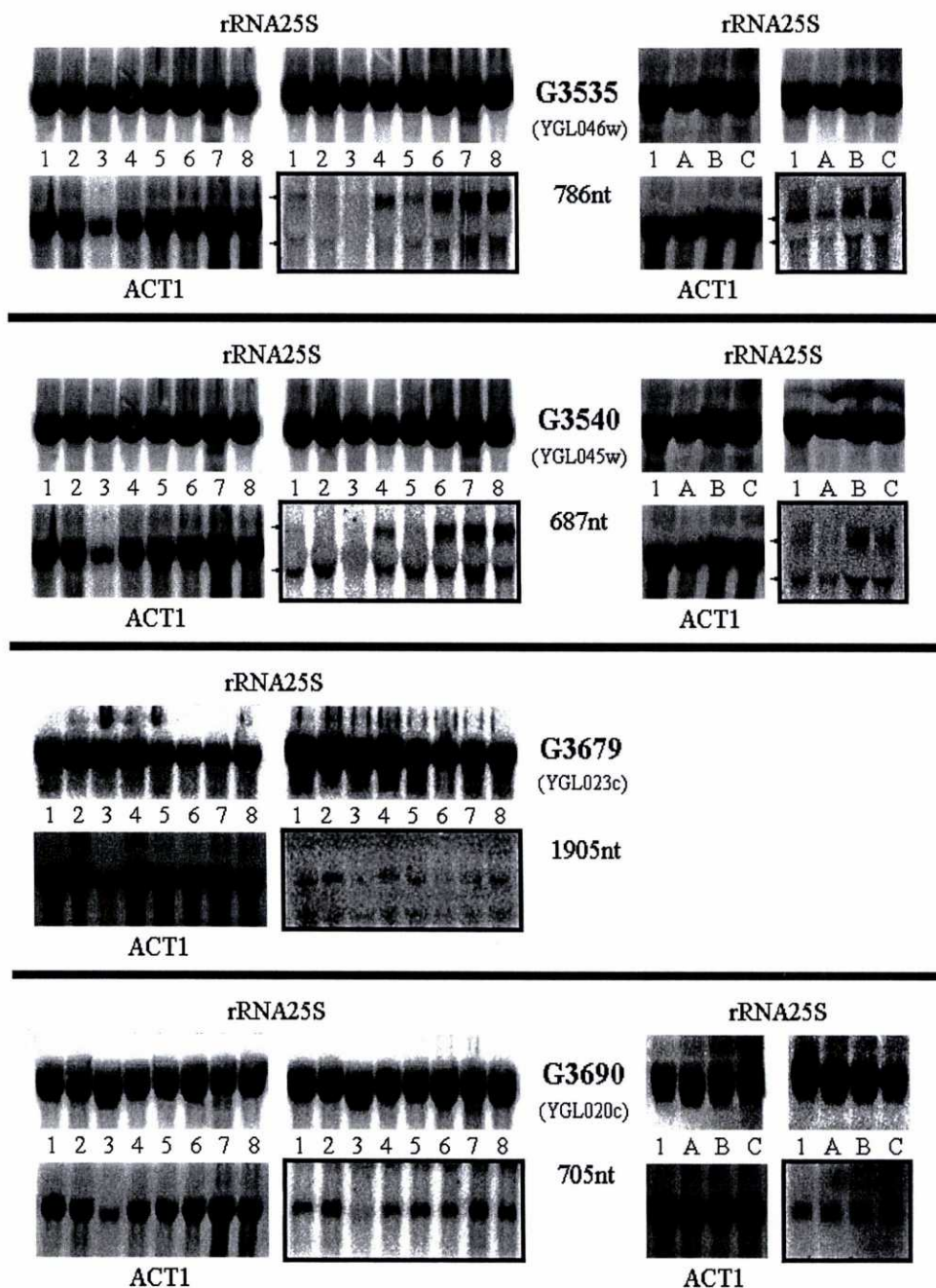


Figura 4.2.8. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida G3535, G3540, G3679 y G3690 del cromosoma VII de *S. cerevisiae*.

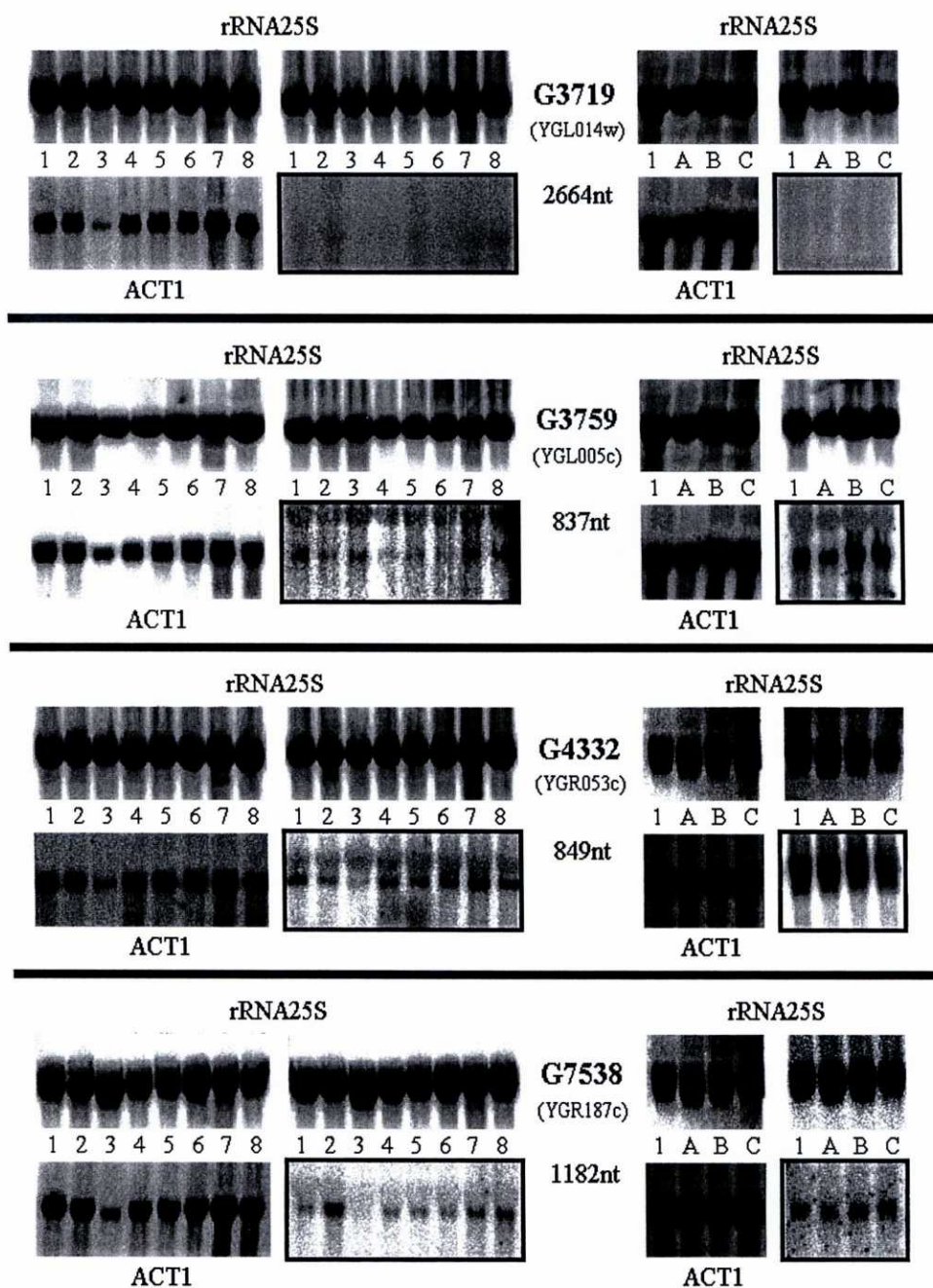


Figura 4.2.9. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida G3719, G3759, G4332 y G7538 del cromosoma VII de *S. cerevisiae*.

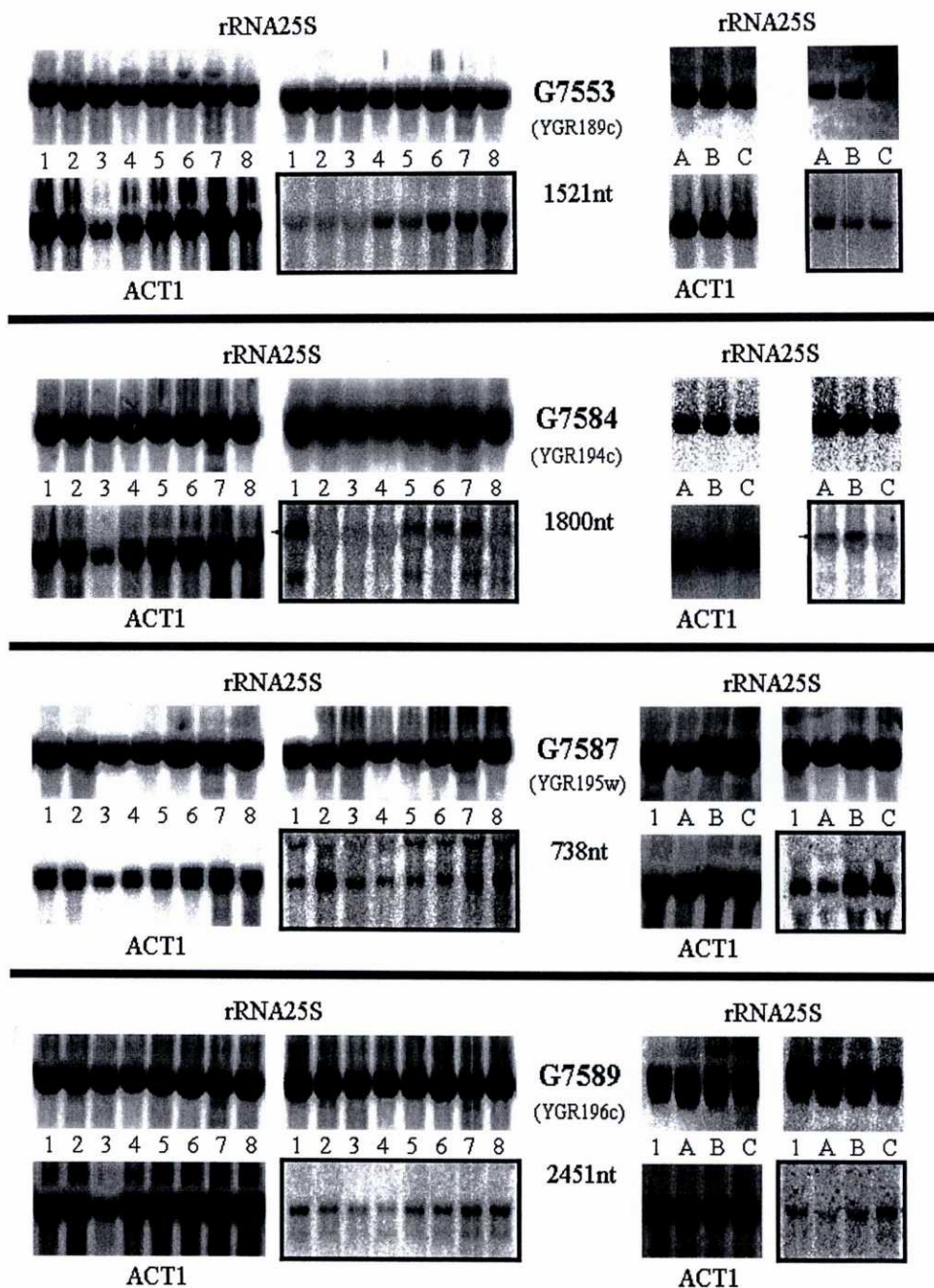


Figura 4.2.10. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida G7553, G7584, G7587 y G7589 del cromosoma VII de *S. cerevisiae*.

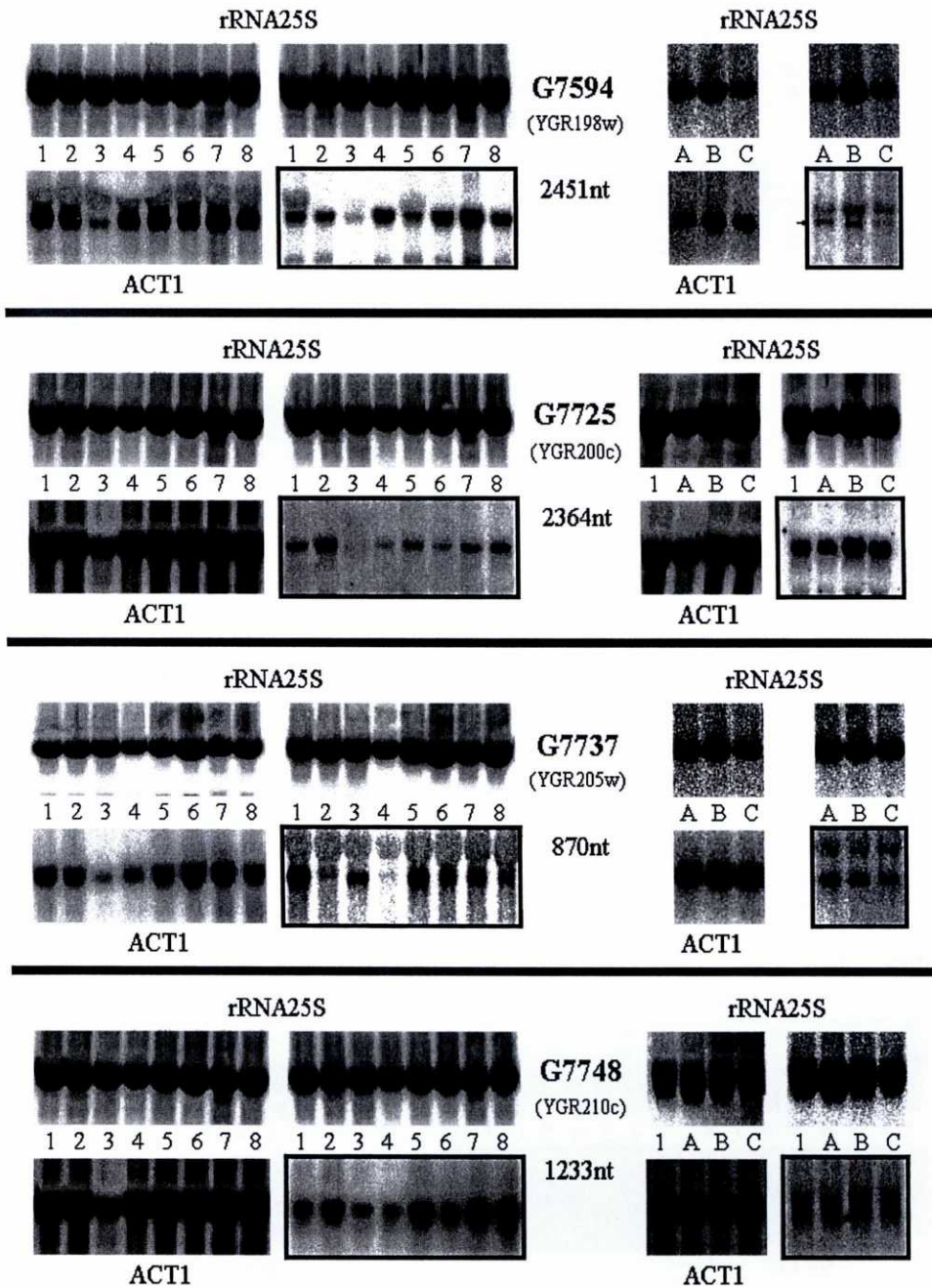


Figura 4.2.11. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida G7594, G7725, G7737 y G7748 del cromosoma VII de *S. cerevisiae*.

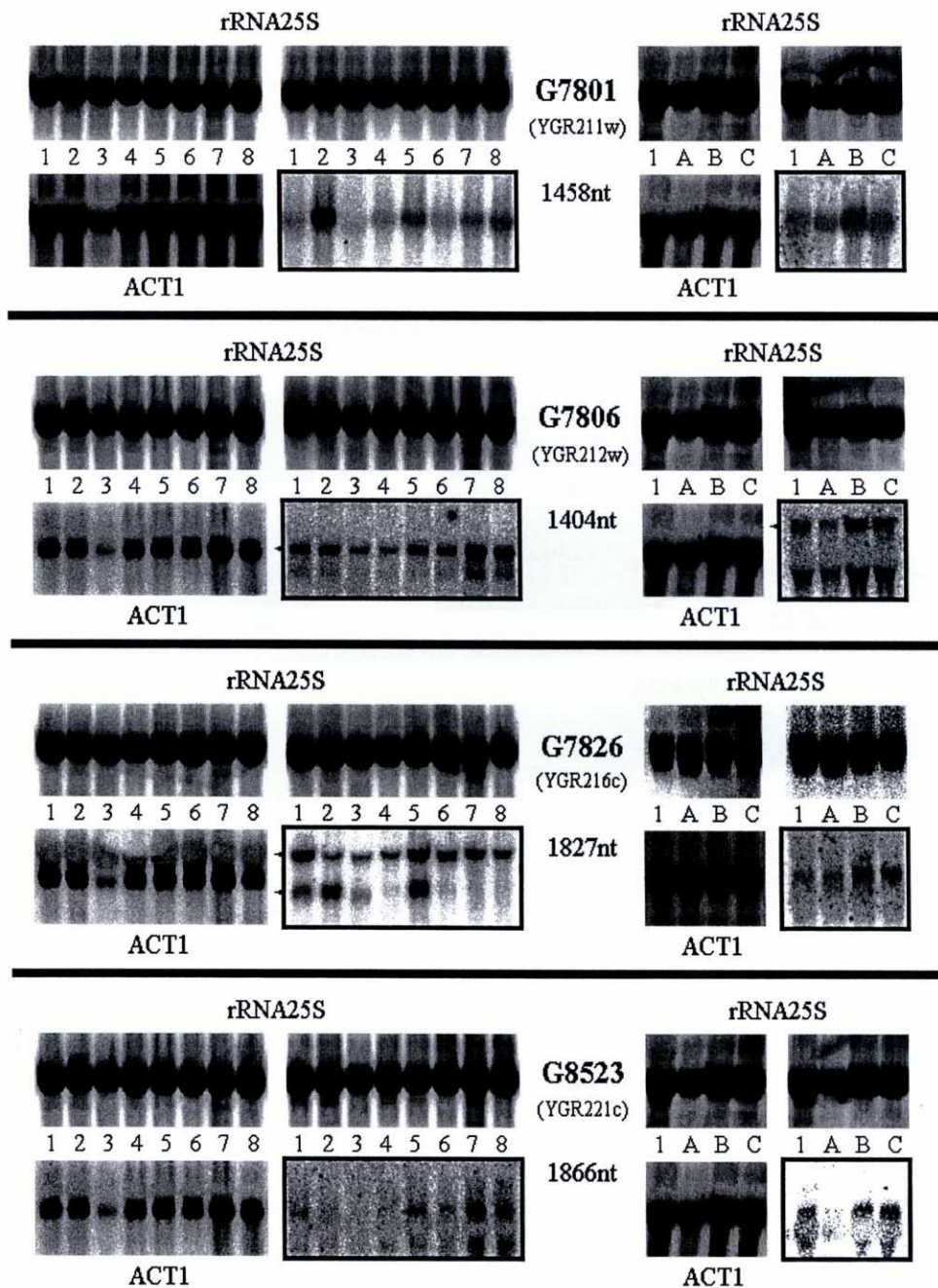


Figura 4.2.12. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida G7801, G7806, G7826 y G8523 del cromosoma VII de *S. cerevisiae*.

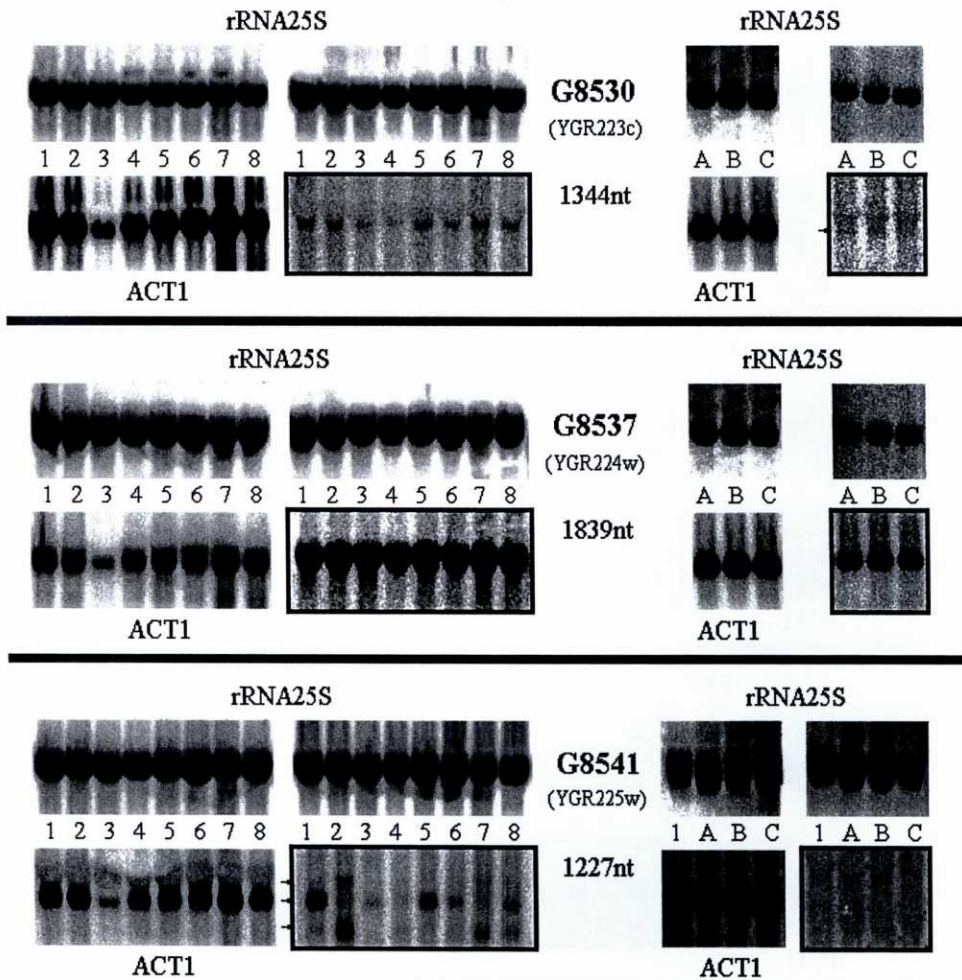


Figura 4.2.13. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida G8530, G8537 y G8541 del cromosoma VII de *S. cerevisiae*.

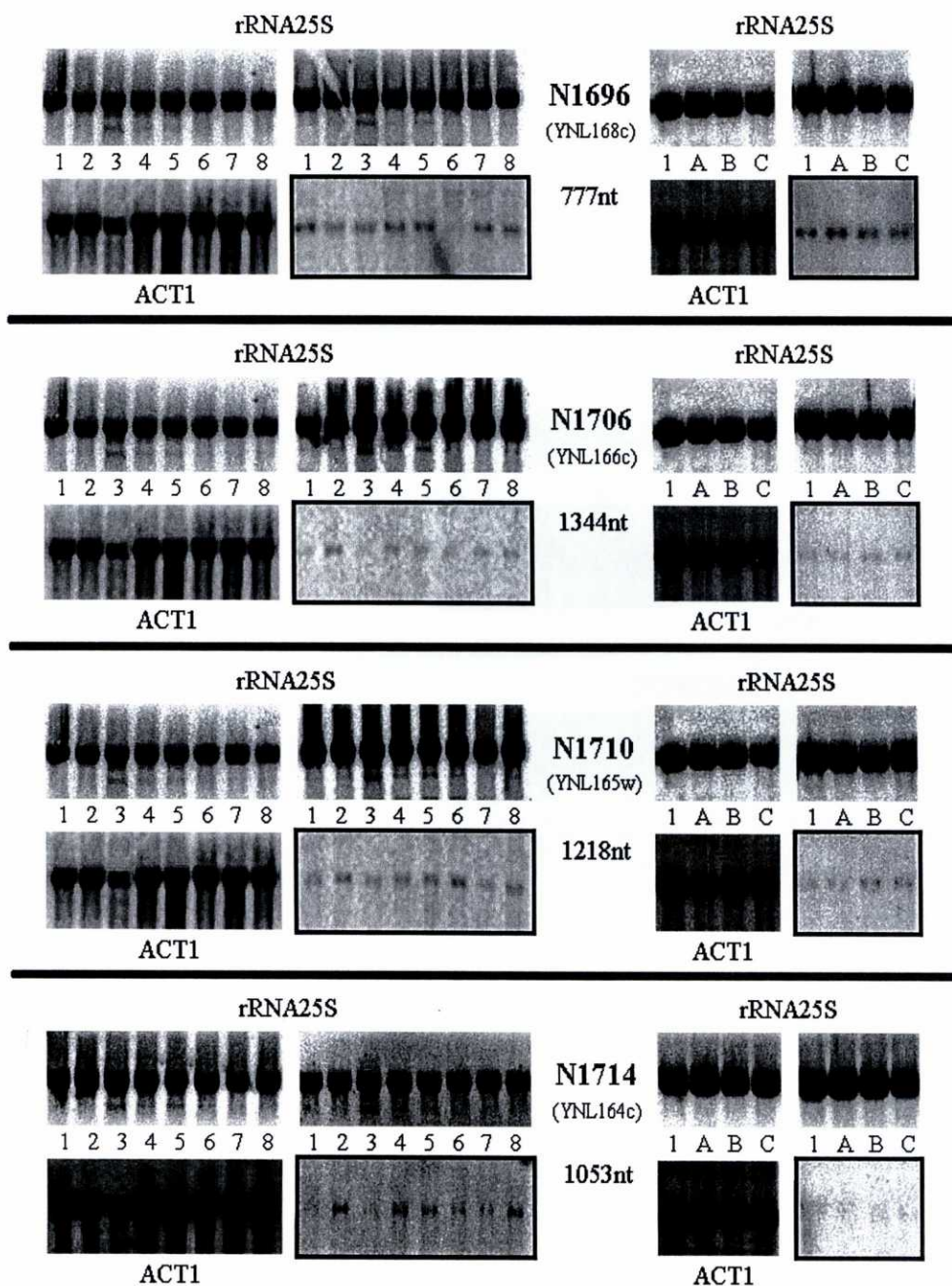


Figura 4.2.14. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida N1696, N1706, N1710 y N1714 del cromosoma XIV de *S. cerevisiae*.

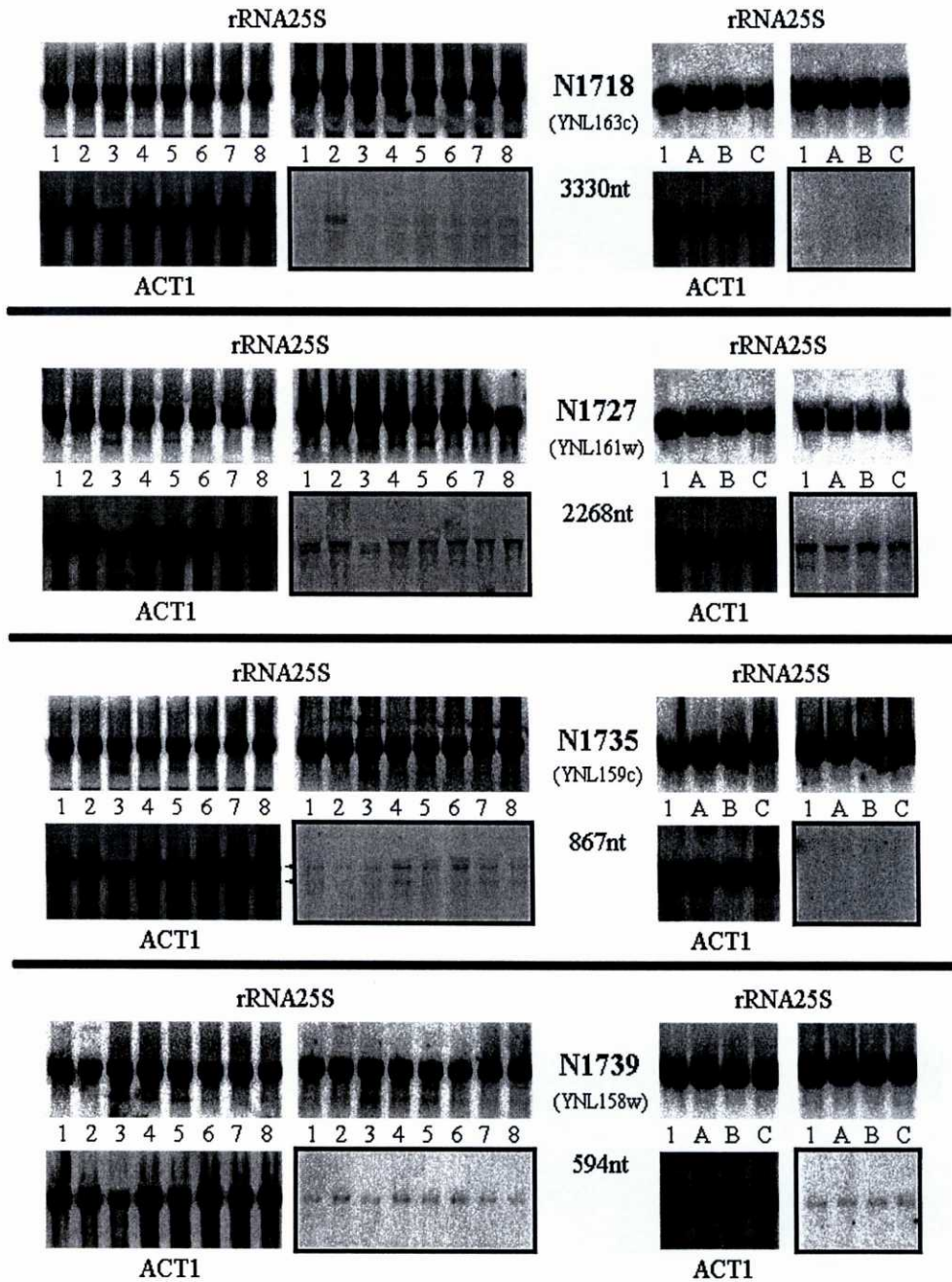


Figura 4.2.15. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida N1718, N1727, N1735 y N1739 del cromosoma XIV de *S. cerevisiae*.

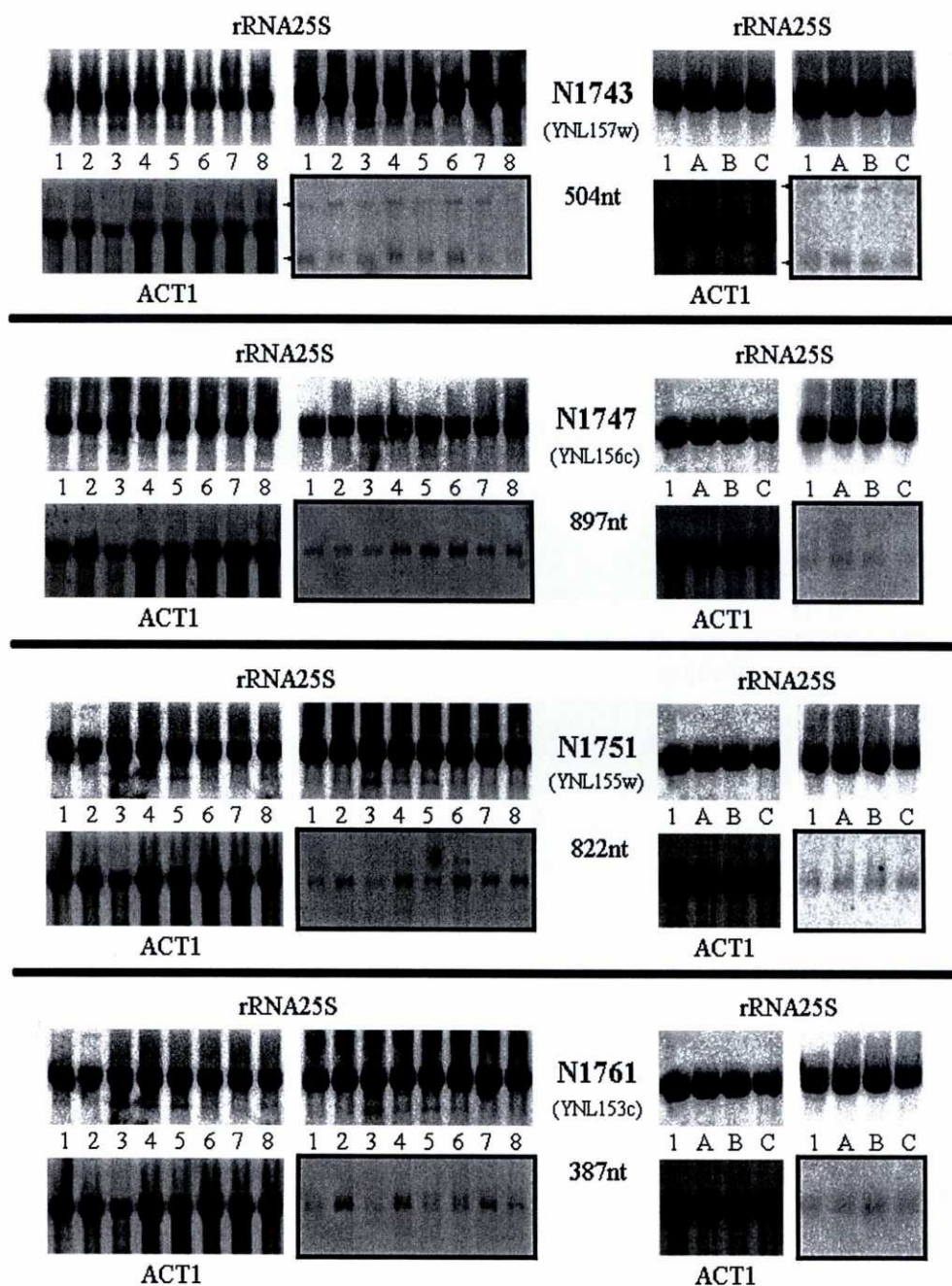


Figura 4.2.16. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida N1743, N1747, N1751 y N1761 del cromosoma XIV de *S. cerevisiae*.

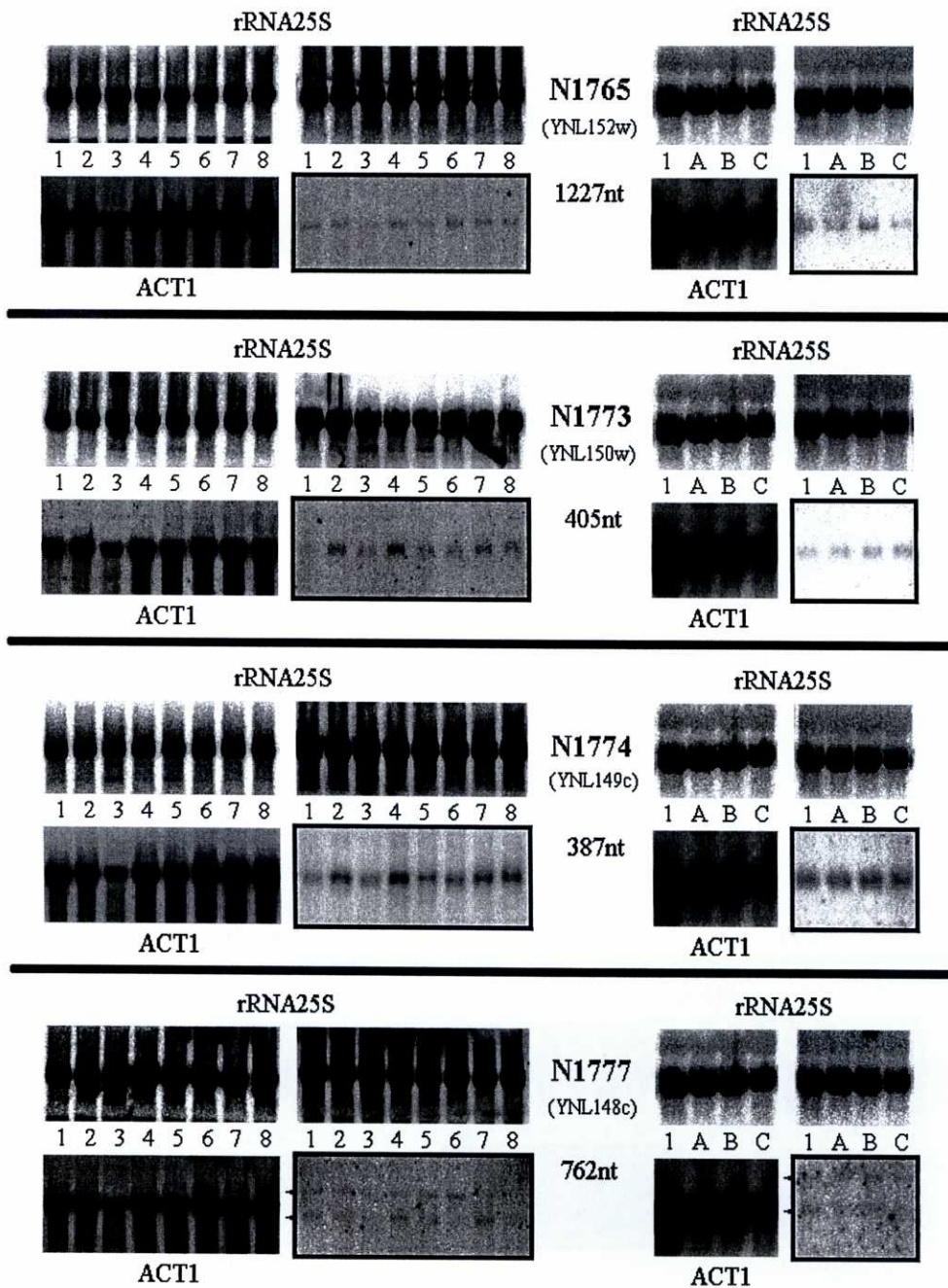


Figura 4.2.17. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida N1765, N1773, N1774 y N1777 del cromosoma XIV de *S. cerevisiae*.

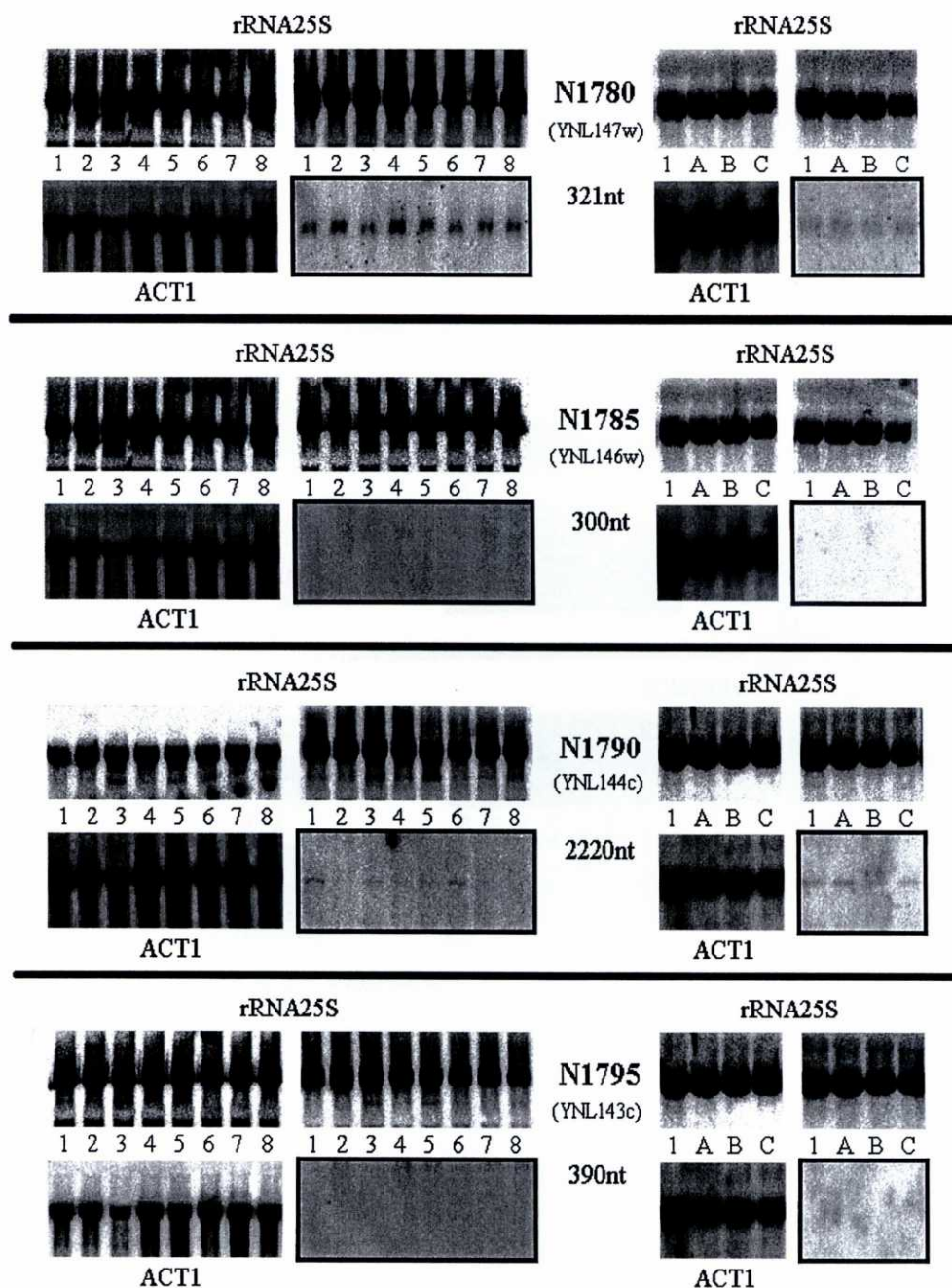


Figura 4.2.18. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida N1780, N1785, N1790 y N1795 del cromosoma XIV de *S. cerevisiae*.

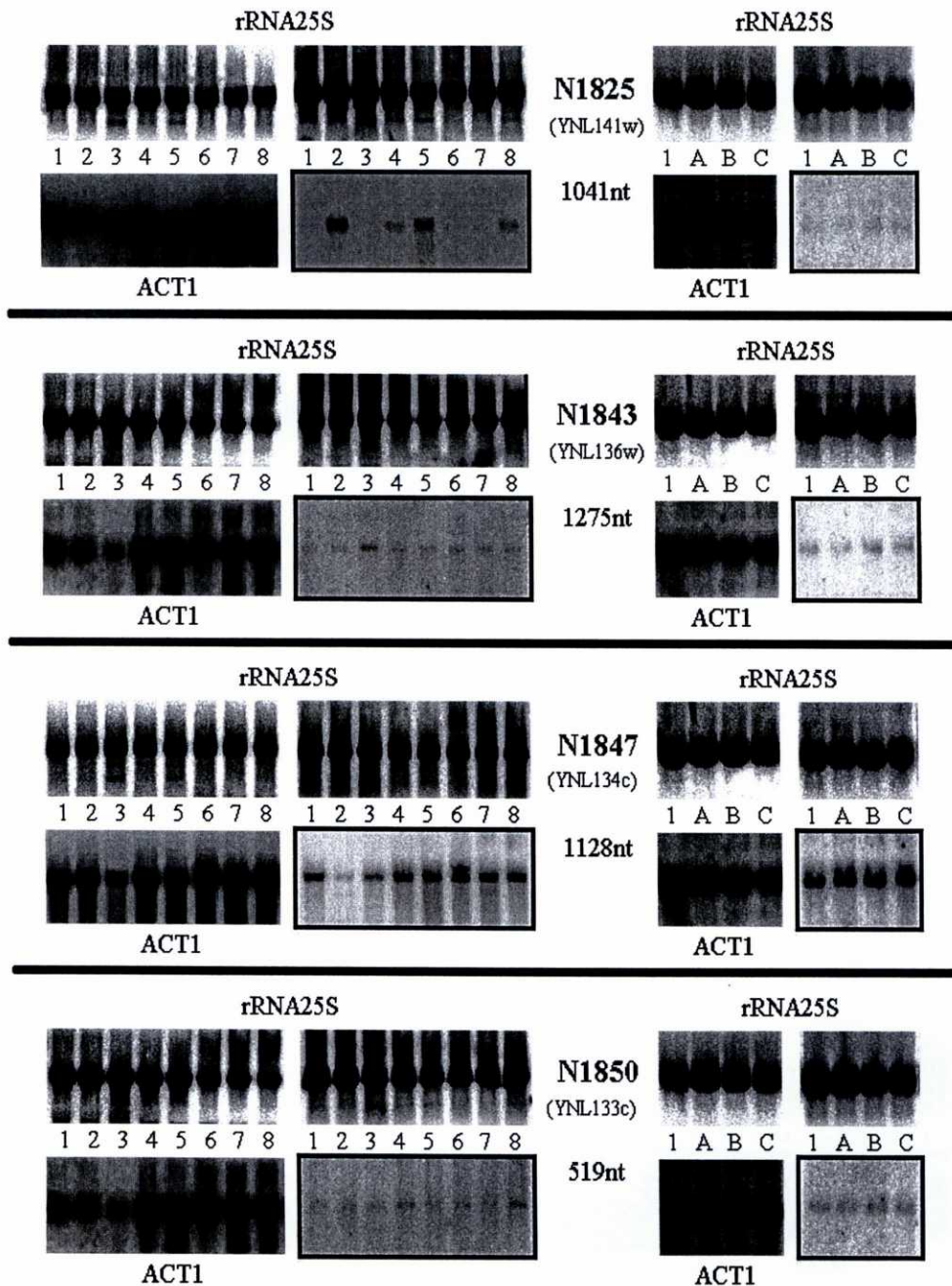


Figura 4.2.19. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida N1825, N1843, N1847 y N1850 del cromosoma XIV de *S. cerevisiae*.

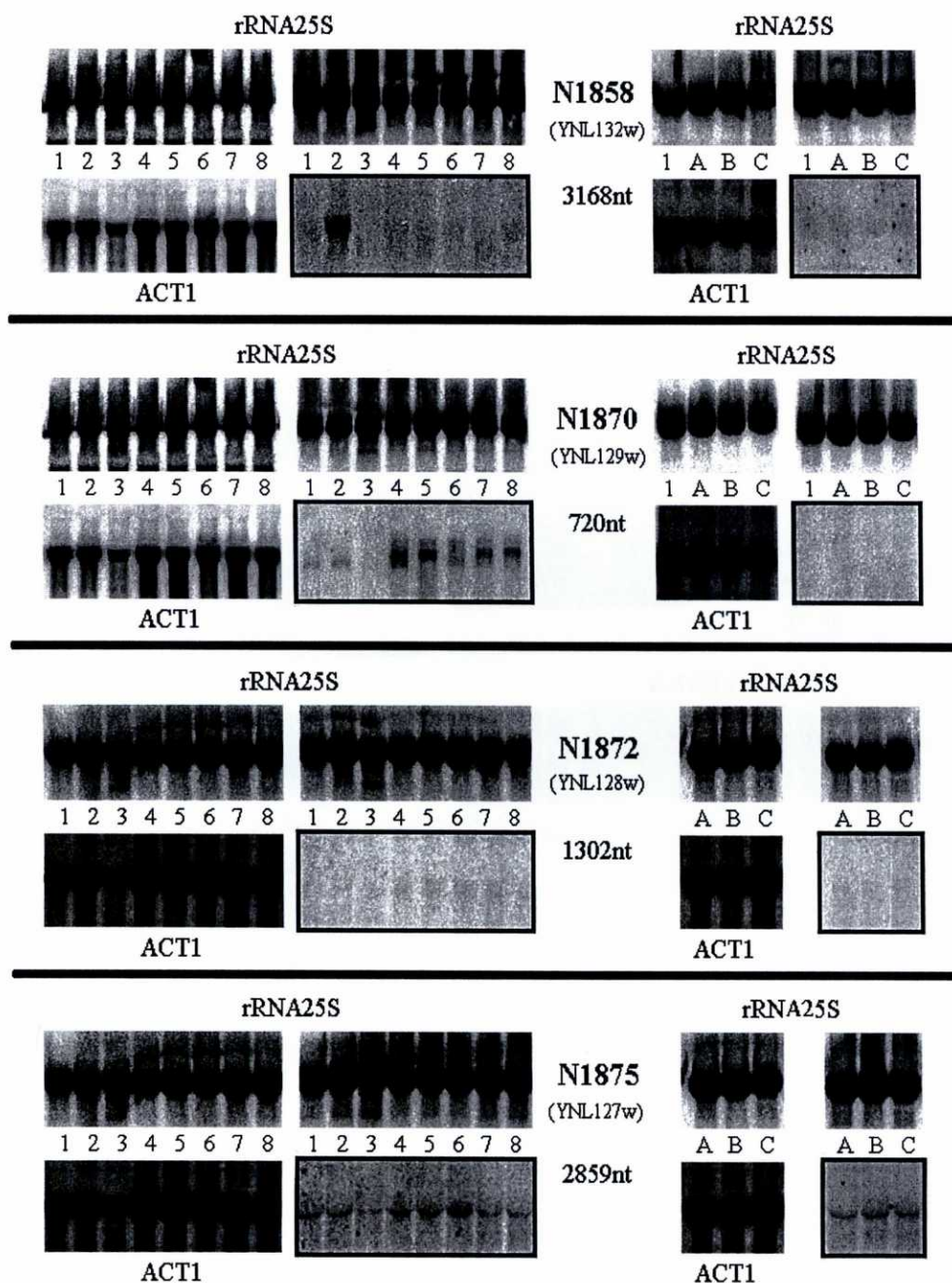


Figura 4.2.20. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida N1858, N1870, N1872 y N1875 del cromosoma XIV de *S. cerevisiae*.

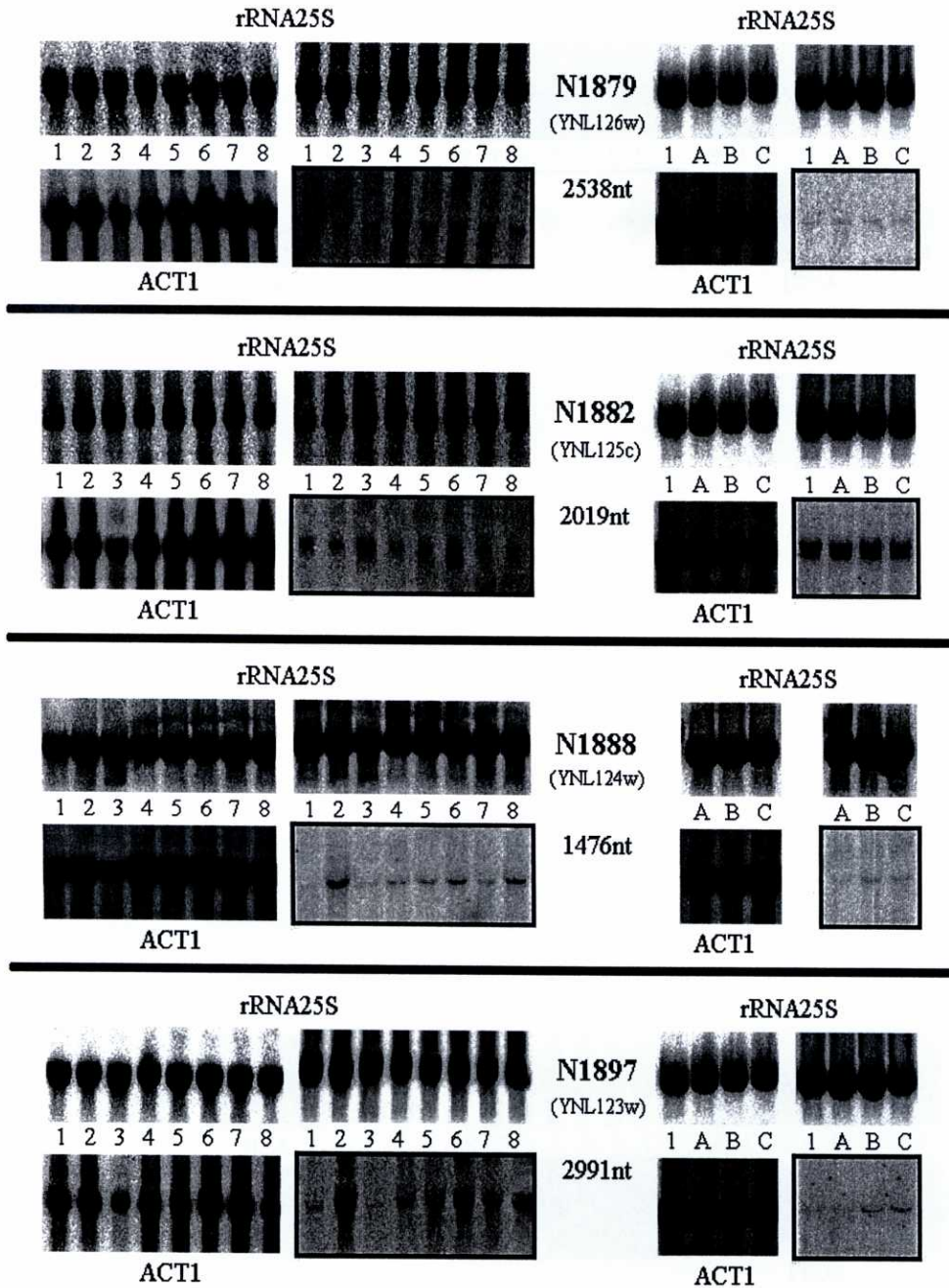


Figura 4.2.21. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida N1879, N1882, N1888 y N1897 del cromosoma XIV de *S. cerevisiae*.

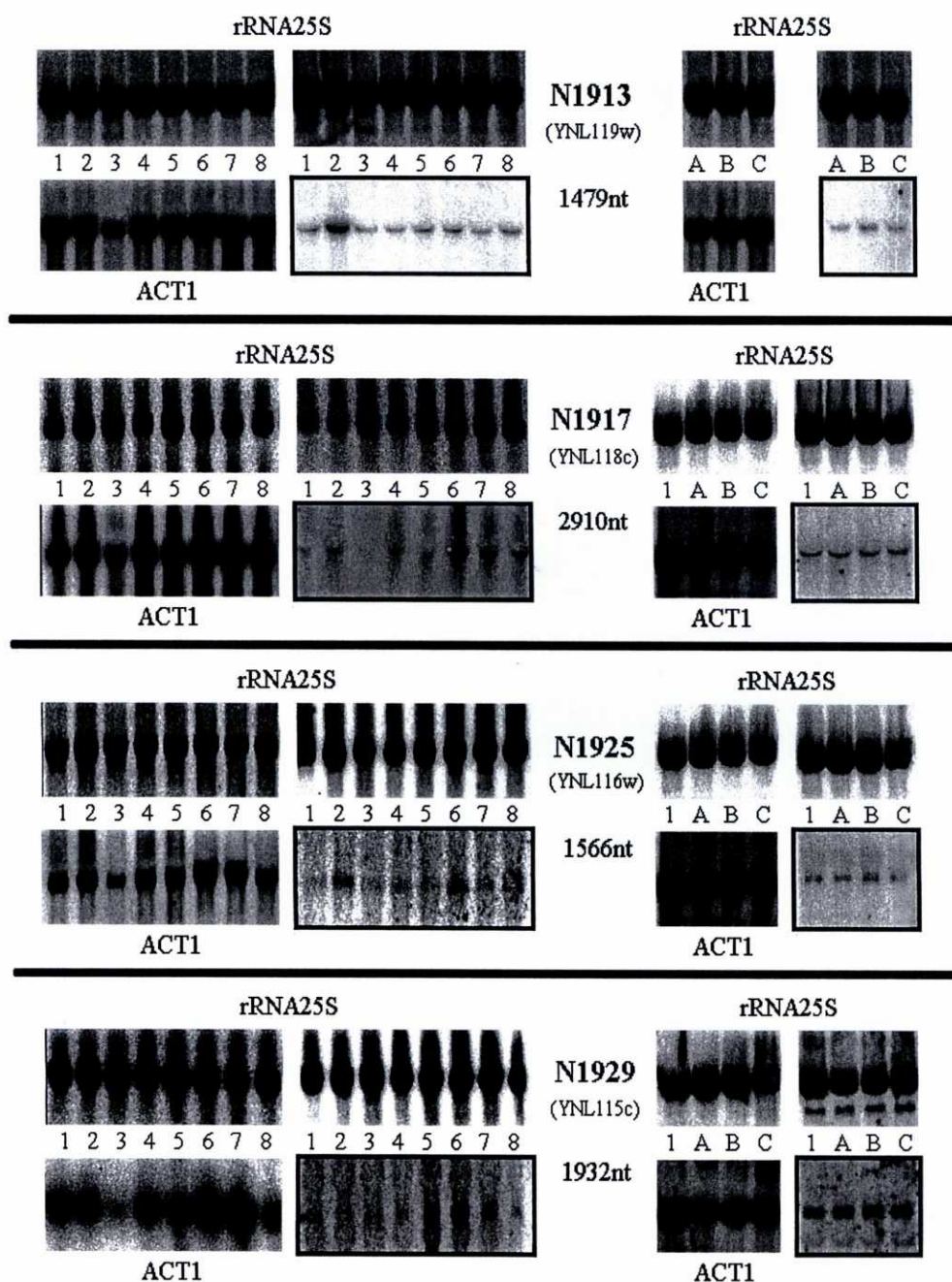


Figura 4.2.22. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida N1913, N1917, N1925 y N1929 del cromosoma XIV de *S. cerevisiae*.

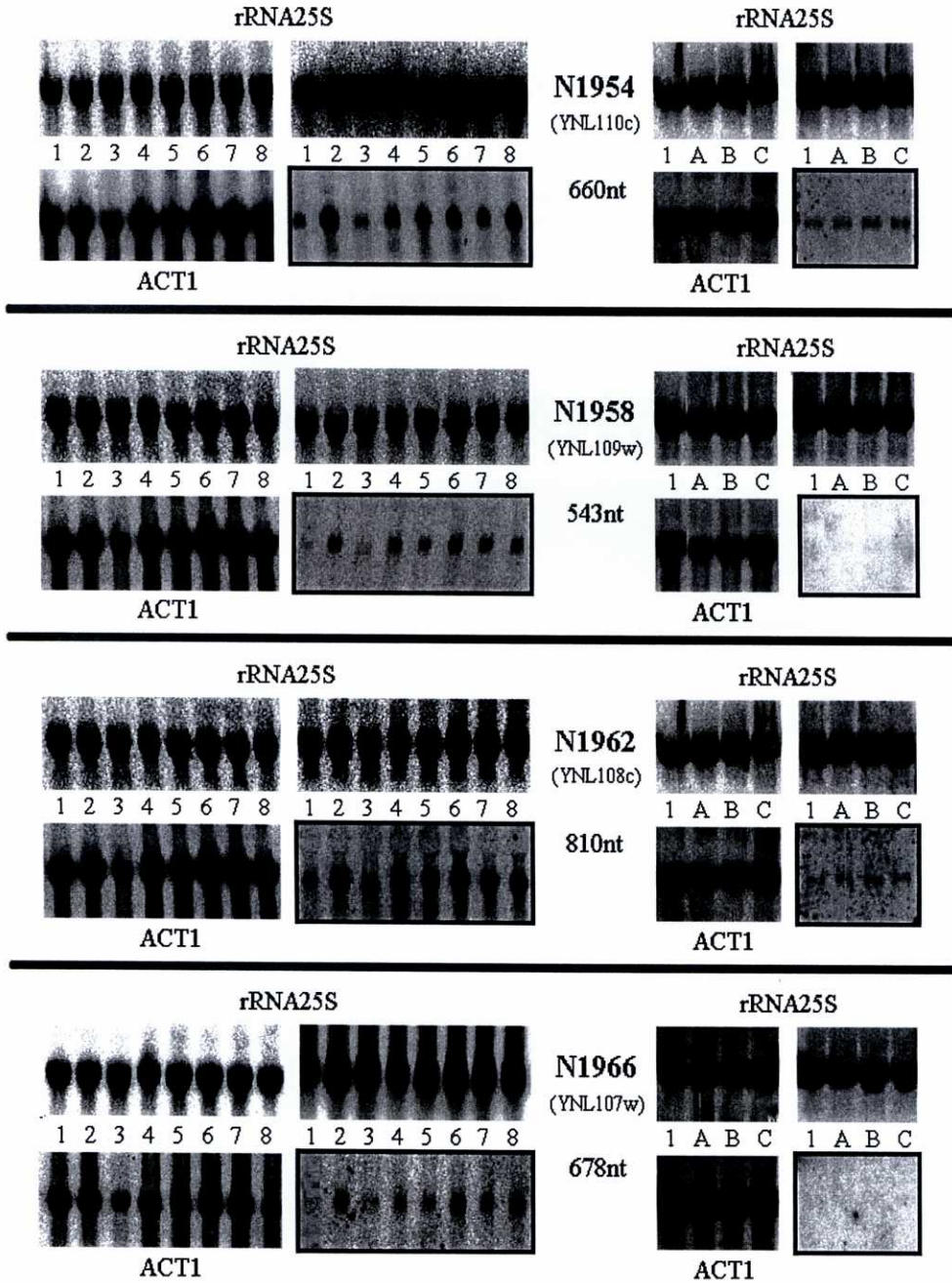


Figura 4.2.23. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida N1954, N1958, N1962 y N1966 del cromosoma XIV de *S. cerevisiae*.

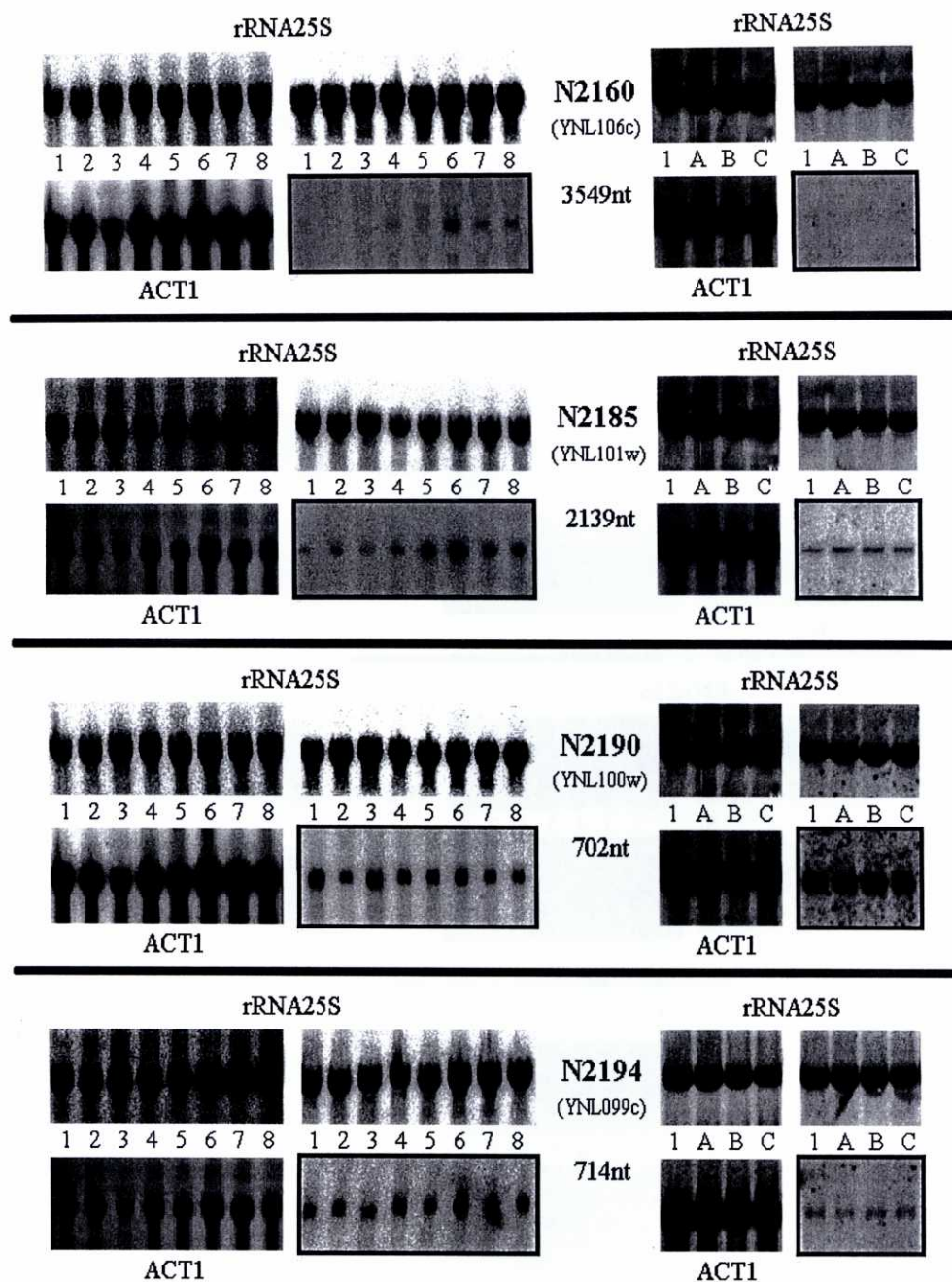


Figura 4.2.24. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida N2160, N2185, N2190 y N2194 del cromosoma XIV de *S. cerevisiae*.

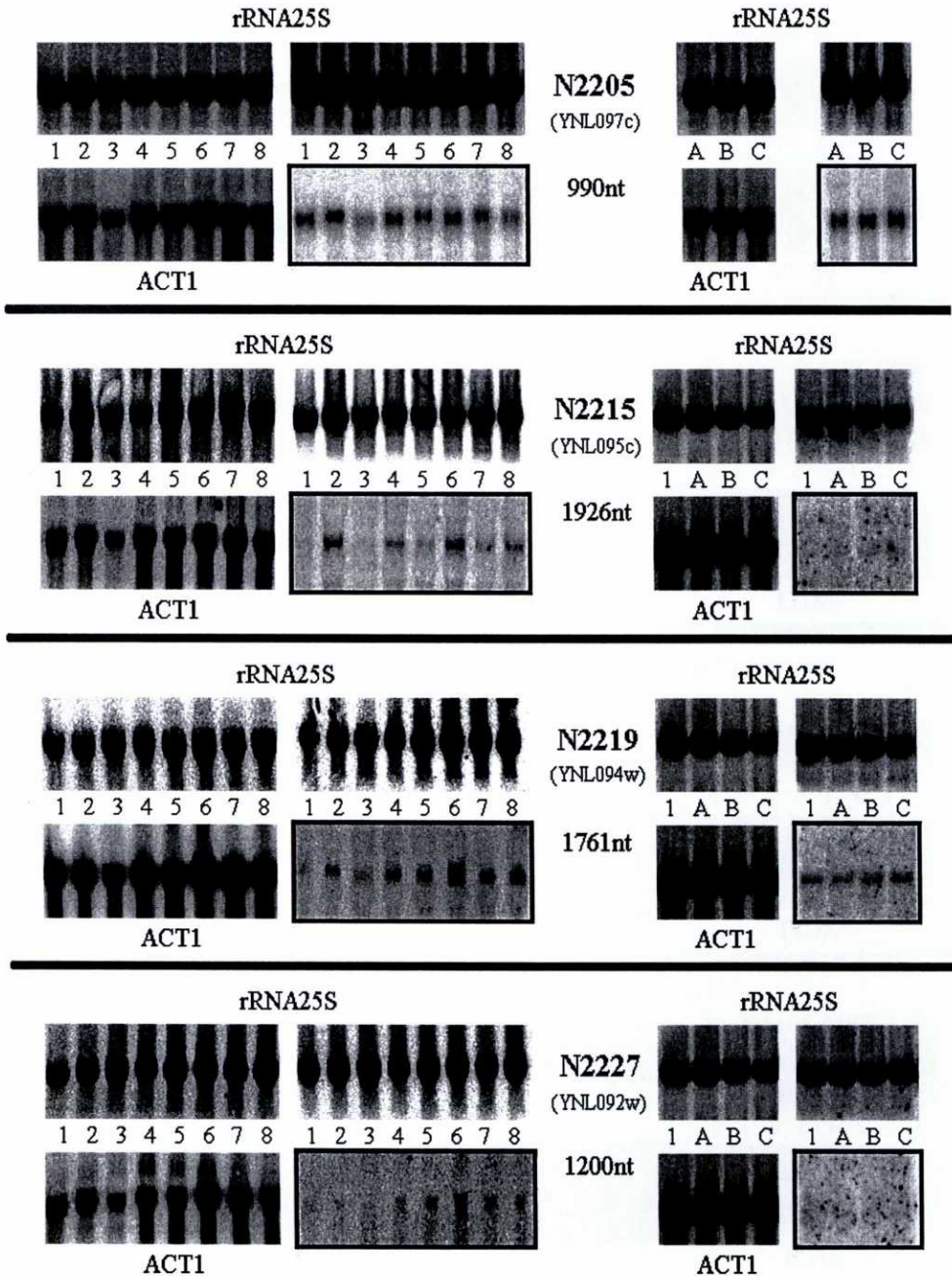


Figura 4.2.25. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida N2205, N2215, N2219 y N2227 del cromosoma XIV de *S. cerevisiae*.

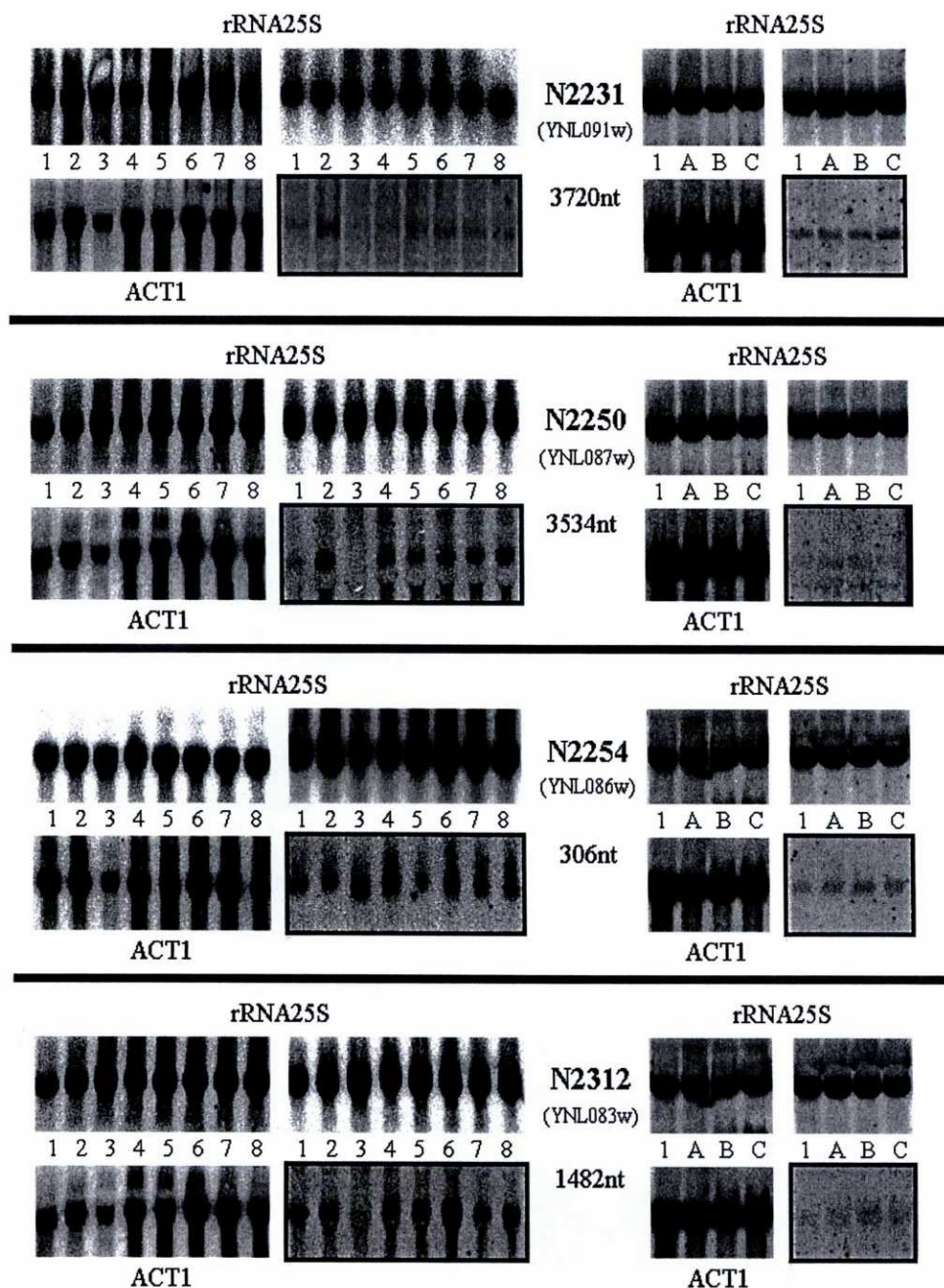


Figura 4.2.26. Hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo con las sondas correspondientes a las ORFs de función desconocida N2231, N2250, N2254 y N2312 del cromosoma XIV de *S. cerevisiae*.

Tabla 4.4. Valores de Nivel Relativo (N.R.) y Absoluto (N.A.) correspondientes a las hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo (1-YPGE, 2-YPD, 3-YPD/fase estacionaria, 4-YNBglu/privación de aminoácidos, 5-YNBglu/privación de nitrógeno, 6-YNBglu/choque osmótico, 7-YNBglu/23°C, YNBglu/36°C) con sondas específicas de 103 ORFs de función desconocida en los cromosomas VII y XIV de *S. cerevisiae*. (A.E.: Actividad Específica de las sondas en $\times 10^6$ de cpm/ μ g de DNA, ND: No Determinado, <1: No detectado, MT: Múltiples Transcritos).

ORF	A.E.	NIVEL RELATIVO (%)								N.A.
		1	2	3	4	5	6	7	8	
G2850	1411	<1	<1	<1	<1	<1	<1	<1	<1	<0,001
G2856	1068	100	68	32	33	38	48	37	38	0,20
G2861	1378	93	44	67	93	98	74	73	100	0,06
G2882	1279	10	11	9	70	21	100	72	50	0,16
G2889	1786	MT	MT	MT	MT	MT	MT	MT	MT	ND
G2913	1197	35	4	100	56	30	39	16	47	1,58
G2916	807	10	100	6	27	24	26	16	45	0,29
G3080	823	69	100	32	93	76	68	77	72	0,20
G3090	972	80	37	36	55	54	100	64	43	0,01
G3107	1167	45	100	27	95	76	36	57	51	0,04
G3110	1056	57	100	26	89	64	55	69	62	0,13
G3113	503	14	100	7	17	20	12	13	17	0,15
G3116	2402	96	100	50	83	86	52	78	75	0,02
G3146	1318	98	80	30	77	51	61	100	78	0,19
G3168	1267	91	93	60	71	100	68	78	70	0,02
G3175	1209	78	84	45	100	60	92	70	92	0,06
G3179	1245	51	54	40	47	71	87	45	100	0,02
G3189	1316	MT	MT	MT	MT	MT	MT	MT	MT	ND
G3193	2747	MT	MT	MT	MT	MT	MT	MT	MT	ND
G3210	1510	79	99	82	98	93	100	88	79	1,15
G3406	710	MT	MT	MT	MT	MT	MT	MT	MT	ND
G3432	593	100	87	42	83	100	56	81	45	0,02
G3435	2702	MT	MT	MT	MT	MT	MT	MT	MT	ND
G3474	871	44	92	51	100	75	68	54	55	0,41
G3483	1470	92	51	82	89	76	100	37	44	0,06
G3489	1303	72	100	77	92	92	68	86	79	0,02

Tabla 4.4. Valores de Nivel Relativo (N.R.) y Absoluto (N.A.) correspondientes a las hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo (1-YPGE, 2-YPD, 3-YPD/fase estacionaria, 4-YNBglu/privación de aminoácidos, 5-YNBglu/privación de nitrógeno, 6-YNBglu/choque osmótico, 7-YNBglu/23°C, YNBglu/36°C) con sondas específicas de 103 ORFs de función desconocida en los cromosomas VII y XIV de *S. cerevisiae*. (A.E.: Actividad Específica de las sondas en $\times 10^6$ de cpm/ μ g de DNA, ND: No Determinado, <1: No detectado, MT: Múltiples Transcritos). (Continuación)

ORF	A.E.	NIVEL RELATIVO (%)								N.A.
		1	2	3	4	5	6	7	8	
G3505	915	54	98	49	63	79	65	100	97	0,22
G3517	574	66	52	42	31	65	37	100	66	0,10
G3535	1483	MT	MT	MT	MT	MT	MT	MT	MT	ND
G3540	1358	MT	MT	MT	MT	MT	MT	MT	MT	ND
G3679	452	76	100	48	62	70	44	72	82	0,01
G3690	738	50	95	29	51	79	42	100	77	0,46
G3719	766	<1	<1	<1	<1	<1	<1	<1	<1	<0,001
G3759	1148	29	20	59	100	94	46	81	31	0,09
G4332	600	80	46	20	59	68	69	100	71	0,04
G7538	656	22	100	14	28	26	20	44	50	0,23
G7553	824	15	17	14	34	28	42	60	100	0,36
G7584	1272	100	18	23	20	44	33	64	32	0,05
G7587	1464	50	100	28	30	35	24	51	67	0,17
G7589	1427	67	58	40	29	67	51	90	100	0,03
G7594	625	28	26	10	54	21	46	100	36	0,42
G7725	1705	39	100	15	21	34	23	51	52	0,07
G7737	1300	100	15	26	12	52	25	35	16	0,74
G7748	1252	36	63	35	37	74	49	92	100	0,31
G7801	1414	30	100	18	34	53	26	51	62	0,09
G7806	890	58	46	33	27	57	48	100	77	0,32
G7826	836	MT	MT	MT	MT	MT	MT	MT	MT	ND
G8523	923	46	21	12	17	60	27	100	78	0,18
G8530	819	66	69	49	36	61	38	100	69	0,05
G8537	1082	71	87	100	96	93	70	81	82	0,50
G8541	1063	MT	MT	MT	MT	MT	MT	MT	MT	ND

Tabla 4.4. Valores de Nivel Relativo (N.R.) y Absoluto (N.A.) correspondientes a las hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo (1-YPGE, 2-YPD, 3-YPD/fase estacionaria, 4-YNBglu/privación de aminoácidos, 5-YNBglu/privación de nitrógeno, 6-YNBglu/choque osmótico, 7-YNBglu/23°C, YNBglu/36°C) con sondas específicas de 103 ORFs de función desconocida en los cromosomas VII y XIV de *S. cerevisiae*. (A.E.: Actividad Específica de las sondas en $\times 10^6$ de cpm/ μ g de DNA, ND: No Determinado, <1: No detectado, MT: Múltiples Transcritos). (Continuación)

ORF	A.E.	NIVEL RELATIVO (%)								N.A.
		1	2	3	4	5	6	7	8	
N1696	736	80	50	57	98	100	22	93	86	0,03
N1706	812	56	100	28	70	79	41	53	44	0,01
N1710	920	23	53	32	50	80	100	30	47	0,02
N1714	1441	36	100	30	72	72	39	39	75	0,01
N1718	1819	20	100	9	41	72	71	32	39	0,01
N1727	877	58	86	30	74	75	100	67	61	0,04
N1735	1827	MT	MT	MT	MT	MT	MT	MT	MT	ND
N1739	664	47	71	39	76	79	100	39	26	0,02
N1743	612	MT	MT	MT	MT	MT	MT	MT	MT	ND
N1747	1019	60	55	45	88	97	100	81	77	0,07
N1751	957	30	67	28	91	64	100	80	77	0,05
N1761	924	50	100	26	75	45	54	59	43	0,05
N1765	637	74	80	46	100	74	84	74	76	0,01
N1773	528	23	49	38	100	53	36	47	35	0,12
N1774	1156	11	53	29	100	57	47	67	45	0,13
N1777	1569	MT	MT	MT	MT	MT	MT	MT	MT	ND
N1780	1473	53	82	52	99	89	99	100	69	0,03
N1785	2140	<1	<1	<1	<1	<1	<1	<1	<1	<0,001
N1790	994	100	38	62	31	51	92	21	23	0,01
N1795	1408	<1	<1	<1	<1	<1	<1	<1	<1	<0,001
N1825	699	6	83	5	27	100	17	11	22	0,16
N1843	465	60	59	100	83	81	74	47	41	0,07
N1847	1313	24	5	21	74	100	84	83	42	0,29
N1850	518	22	34	69	100	88	57	33	48	0,13
N1858	732	9	100	18	32	26	15	11	14	0,05
N1870	560	10	20	5	64	100	35	55	59	0,08

Tabla 4.4. Valores de Nivel Relativo (N.R.) y Absoluto (N.A.) correspondientes a las hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo (1-YPGE, 2-YPD, 3-YPD/fase estacionaria, 4-YNBglu/privación de aminoácidos, 5-YNBglu/privación de nitrógeno, 6-YNBglu/choque osmótico, 7-YNBglu/23°C, YNBglu/36°C) con sondas específicas de 103 ORFs de función desconocida en los cromosomas VII y XIV de *S. cerevisiae*. (A.E.: Actividad Específica de las sondas en $\times 10^6$ de cpm/ μ g de DNA, ND: No Determinado, <1: No detectado, MT: Múltiples Transcritos). (Continuación)

ORF	A.E.	NIVEL RELATIVO (%)								N.A.
		1	2	3	4	5	6	7	8	
N1872	768	<1	<1	<1	<1	<1	<1	<1	<1	<0,001
N1875	868	92	76	36	78	71	100	64	58	0,01
N1879	1114	20	22	8	57	37	100	43	8	0,13
N1882	529	100	53	68	30	43	71	20	15	0,19
N1888	763	6	100	12	31	32	50	14	56	0,02
N1897	836	13	89	10	39	49	100	68	60	0,08
N1913	747	25	100	27	22	31	30	20	32	0,04
N1917	1226	93	96	44	81	61	100	64	56	0,06
N1925	580	45	97	45	65	31	79	48	100	0,12
N1929	692	34	35	27	23	85	100	40	44	0,25
N1954	734	20	100	13	34	37	31	19	45	0,47
N1958	751	16	96	25	99	54	100	80	63	0,04
N1962	666	18	98	35	100	57	97	67	51	0,17
N1966	1004	41	85	36	98	60	100	59	100	0,04
N2160	780	29	18	17	41	33	100	40	55	0,05
N2185	757	12	18	18	27	47	100	47	42	0,10
N2190	741	100	24	54	45	40	58	29	25	0,18
N2194	583	53	59	39	81	58	100	95	56	0,14
N2205	702	86	97	59	100	84	87	77	66	0,05
N2215	799	14	84	15	56	40	100	39	61	0,08
N2219	929	25	36	25	50	37	100	65	75	0,07
N2227	761	29	17	14	43	37	100	45	32	0,04
N2231	726	48	100	28	53	70	56	42	49	0,04
N2250	688	43	100	10	100	50	80	74	82	0,05
N2254	716	55	40	74	100	27	52	50	48	0,16
N2312	846	57	50	26	55	42	100	46	56	0,10

El análisis de *Northern* realizado con las 103 ORFs de los cromosomas VII y XIV de *S.cerevisiae*, permitió detectar la expresión de 86 ORFs (83,5%). Para cinco de ellas (4,9%), no fue posible visualizar una señal de hibridación (G2850, G3719, N1785, N1795 y N1872). Por otro lado, once ORFs (10.7%) presentaron dos transcritos de tamaño compatible con el de la secuencia de la ORF correspondiente (G2889, G3189, G3193, G3406, G3435, G3535, G3540, G7826, N1735, N1743 y N1777) no siendo posible descartar uno de los dos. Finalmente, una ORF (G8541) presento varios transcritos.

4.4. Niveles absolutos de expresión:

Los Niveles Absolutos de cada ORF (N.A.) se calcularon refiriéndolos al nivel de expresión de *ACT1* en la condición de cultivo número 2 tal como se describe en Materiales y Métodos.

4.4.1. Clasificación de las ORFs en función de su nivel absoluto:

Con los valores de nivel absoluto calculados se realizó una clasificación de las distintas ORFs en función de sus niveles de expresión (Tabla 4.5.):

- Alta expresión: $N.A. > \cong 0.5$
- Mediana expresión: $\cong 0.1 < N.A. < 0.5$
- Baja expresión: $N.A. < 0.1$
- No detectable: ausencia de señal.

Aquellas ORFs en las que no fue posible identificar un único transcrito de manera inequívoca se excluyeron de la cuantificación (MT).

Tabla 4.5. Niveles Absolutos (N.A.) y clasificación de las ORFs de los cromosomas VII y XIV de *S. cerevisiae* según su grupo de expresión (Exp.). MT: Múltiples transcritos; ND: No Detectado; A: Alta ; M: Moderada; B: Baja.

ORF	N.A.	Exp.	ORF	N.A.	Exp.	ORF	N.A.	Exp.	ORF	N.A.	Exp.
G2850	<0,001	--	G3505	0,22	M	N1696	0,03	B	N1872	<0,001	--
G2856	0,20	M	G3517	0,10	M	N1706	0,01	B	N1875	0,01	B
G2861	0,06	B	G3535	MT	--	N1710	0,02	B	N1879	0,13	M
G2882	0,16	M	G3540	MT	--	N1714	0,01	B	N1882	0,19	M
G2889	MT	--	G3679	0,01	B	N1718	0,01	B	N1888	0,02	B
G2913	1,58	A	G3690	0,46	M	N1727	0,04	B	N1897	0,08	B
G2916	0,29	M	G3719	<0,001	--	N1735	MT	--	N1913	0,04	B
G3080	0,20	M	G3759	0,09	B	N1739	0,02	B	N1917	0,06	B
G3090	0,01	B	G4332	0,04	B	N1743	MT	--	N1925	0,12	M
G3107	0,04	B	G7538	0,23	M	N1747	0,07	B	N1929	0,25	M
G3110	0,13	M	G7553	0,36	M	N1751	0,05	B	N1954	0,47	M
G3113	0,15	M	G7584	0,05	B	N1761	0,05	B	N1958	0,04	B
G3116	0,02	B	G7587	0,17	M	N1765	0,01	B	N1962	0,17	M
G3146	0,19	M	G7589	0,03	B	N1773	0,12	M	N1966	0,04	B
G3168	0,02	B	G7594	0,42	M	N1774	0,13	M	N2160	0,05	B
G3175	0,06	B	G7725	0,07	B	N1777	MT	--	N2185	0,10	M
G3179	0,02	B	G7737	0,74	A	N1780	0,03	B	N2190	0,18	M
G3189	MT	--	G7748	0,31	M	N1785	<0,001	--	N2194	0,14	M
G3193	MT	--	G7801	0,09	B	N1790	0,01	B	N2205	0,05	B
G3210	1,15	A	G7806	0,32	M	N1795	<0,001	--	N2215	0,08	B
G3406	MT	--	G7826	MT	--	N1825	0,16	M	N2219	0,07	B
G3432	0,02	B	G8523	0,18	M	N1843	0,07	B	N2227	0,04	B
G3435	MT	--	G8530	0,05	B	N1847	0,29	M	N2231	0,04	B
G3474	0,41	M	G8537	0,50	A	N1850	0,13	M	N2250	0,05	B
G3483	0,06	B	G8541	MT	--	N1858	0,05	B	N2254	0,16	M
G3489	0,02	B				N1870	0,08	B	N2312	0,10	M

Se puede observar que la mayor parte de las ORFs analizadas pertenecen a los grupos de mediana (33%) o baja expresión (46,6%). No se pudieron cuantificar 17 ORFs, sea porque no se visualizó ninguna señal de hibridación, sea porque se observaron varios transcritos posibles para la ORF estudiada.

La clasificación de las ORFs estudiadas en grupos de niveles de expresión nos permitió proceder a una evaluación de la metodología utilizada en la detección de los niveles de expresión absolutos comparando los datos obtenidos con otros aportados por diversos autores (Tabla 4.6.).

Tabla 4.6. Comparación de las proporciones de ORFs que pertenecen a los cuatro grupos de nivel absoluto de expresión (A: alto; M: moderado; B: bajo; N.C.: no cuantificado) obtenidos en diferentes laboratorios con distintos cromosomas (crom.) de *S. cerevisiae*

Crom.	Referencia	Técnica	NºORFs	NIVELES DE EXPRESIÓN (%)			
				A	M	B	N.C.
Todos	Velculescu <i>et al.</i> (1997)	<i>SAGE</i>	6121	12,9 ^(a)	28,9 ^(b)	34,4 ^(c)	23,8
III	Yoshikawa y Isono (1990)	<i>Northern blot</i>	160	1,9	75,0		23,1
VI	Naitou <i>et al.</i> (1997)	<i>Northern blot</i>	126	7,9	32,5	42,9	16,7
VIII	Hegemann y Johnston (comunicación personal)	Fusiones a <i>GFP</i>	127	1,0	10,0	64,0	25,0
XI	Richard <i>et al.</i> (1997)	<i>Northern blot/ Southern blot</i>	333	1,5	18,6	32,4	47,5
II	Consortio B2 EUROFANI I	<i>Northern blot</i>	141	4,2	29,8	49,6	16,4
IV			159	5,1	22,6	58,5	13,8
VII			132	8,3	31,0	33,3	27,3
X			118	2,5	14,4	45,8	37,3
XII			118	2,6	14,5	55,5	27,4
XIV			250	2,0	7,2	69,2	21,6
XV			82	6,1	24,4	51,2	18,3
VII + XIV	Este trabajo	<i>Northern blot</i>	103	3,9	33,0	46,6	16,5

(a) Con más de 100 copias de mRNA/célula, (b) Entre 10 y 100 copias de mRNA/célula, (c) Con menos de 10 copias de mRNA/célula.

Teniendo en cuenta la obligada variabilidad, debida, tanto al cromosoma estudiado como a las estrategias de estudio (originadas por la diferente sensibilidad de la técnica y los umbrales adoptados para diferenciar cada grupo de nivel de expresión), cuando se comparan los resultados obtenidos por diferentes técnicas en varios laboratorios podemos considerar que los resultados obtenidos en este trabajo son concordantes con los

encontrados por otros autores que han analizado otras ORFs de función desconocida en otros cromosomas de la levadura.

Es de destacar la incapacidad de cualquiera de los métodos utilizados para detectar la totalidad de las ORFs analizadas. Aún exceptuando el mapeo transcripcional del cromosoma XI (Richard *et al.*, 1997) que no fue capaz de detectar la expresión de casi la mitad de las ORFs pronosticadas por la secuenciación, los otros tres métodos no pudieron detectar la expresión de una proporción significativa de las ORFs estudiadas.

También es interesante observar la variabilidad de resultados obtenidos, usando los mismos criterios y la misma tecnología, en cuanto a la distribución de ORFs en los diferentes grupos de niveles absolutos de expresión en los cromosomas II, IV, VII, X, XII, XIV y XV, destacándose principalmente variabilidad en el porcentaje de ORFs que no pudieron ser cuantificadas en cada caso (del 13,8% al 37,3%). Todos estos cromosomas han sido analizados en el consorcio B2 de EUROFAN con el que colabora nuestro laboratorio. Si estos datos se confirman una vez analizados la totalidad de los genes de la levadura, estas diferencias podrían reflejar un aspecto funcional en lo que se refiere a la comprensión de los procesos evolutivos que llevaron a la organización actual del genoma de *S. cerevisiae*.

Contrastando los datos presentados en la tabla 4.5. con los que proporciona el estudio teórico de las secuencias nucleotídicas y relacionándolos con la información disponible en la literatura podemos sacar algunas conclusiones en cuanto a la distribución de las 103 ORFs analizadas en los cuatro grupos de niveles absolutos de expresión:

4.4.1.1. ORFs cuya expresión no fue detectada:

Cinco ORFs no presentaron ninguna señal de hibridación en los *Northern* realizados (N1785, N1795, G2850, G3719 y N1872); la incapacidad para detectar la expresión de estas ORFs puede ser debida a diversas circunstancias:

- a- La sensibilidad de la técnica es inferior al nivel de expresión. Esto es aplicable para aquellas ORFs que se expresan a niveles inferiores al 1% del nivel de *ACT1*.
- b- Entre las condiciones analizadas en este trabajo, no se encuentra(n) presente(s) aquella(s) en las que la ORF se expresa.
- c- O finalmente que la ORF aparece en la secuencia por casualidad y no se corresponde con un gen funcional.

En el caso de las ORFs G2850 y N1872, existen datos que indican que se trata de genes funcionales; G2850 codifica una proteína que presenta similitudes con la proteína hipotética C3H1.12C de *S. pombe* y que contiene en su secuencia un motivo característico de peptidasas señal tipo I. En el análisis funcional, realizado en nuestro laboratorio, se pudo comprobar que las colonias derivadas de una espora portadora de la ORF interrumpida crecen más lentamente y son de menor tamaño que las silvestres, lo que revela un fenotipo asociado a su función (Tizón *et al.*, 1998, enviado). Recientemente, N1872 ha sido identificada como el gen *TEP1* que codifica una tirosin-fosfatasa y que presenta similitudes con el gen supresor de tumores humanos *TEP1/MMAC/PTEN1* (Li *et al.*, 1997).

Aunque no existen datos sobre la funcionalidad de G3719 en *S.cerevisiae*, su homología con otras proteínas funcionales de levadura o de

Drosophila nos hace suponer que en este caso la ausencia de señal sea debida a los factores relacionados con los expuestos en a y/o b. G3719 presenta similitudes con la proteína Mtp5p; que se requiere para la redistribución de las proteínas Sir (*Silent mating type Information Regulation*) desde el telómero hasta el nucléolo (Kennedy *et al.*, 1997). También es similar a la proteína *pumilio*, de unión a RNA, y que interviene en el desarrollo embrionario de *D. melanogaster* (Forbes y Lehmann, 1998).

Respecto a N1785 y N1795, se trata de ORFs de 300 y 390 nucleótidos respectivamente y que por tanto se encontrarían cercanas al límite de tamaño establecido como mínimo para considerar una ORF no cuestionable. Al no existir datos adicionales sobre la posible funcionalidad de estas ORFs, no podemos descartar que se trate de ORFs fortuitas.

Teniendo en cuenta que este trabajo se llevó a cabo con RNA extraído de células haploides con el tipo de apareamiento α (*Mating-type α*) existía la posibilidad de que algunos de los genes no detectados fuesen α -específicos; por este motivo se decidió buscar en las zonas promotoras de estas ORFs (en una región de 800 pb 5' del ATG de inicio de cada ORF) la presencia de sitios de unión a la proteína Mat α 2p, responsable de la represión de la transcripción de los genes α -específicos (Wilson y Herskowitz, 1984). El resultado negativo de la búsqueda permite suponer que la ausencia de expresión no es atribuible a este motivo.

4.4.1.2. ORFs para las que fueron detectados varios transcritos:

12 ORFs dieron lugar a la detección de múltiples transcritos. Según los datos recogidos en la base de datos: G3535, G3540, N1735 y N1743 codificarían proteínas hipotéticas y G3189, G7826 y N1777 han sido

caracterizadas por otros laboratorios en el transcurso de este trabajo, como los genes *MMS2* (que codifica una protein-ligasa implicada en la resistencia del DNA a agentes dañinos), *GPII* (que se requiere en la síntesis de la N-acetilglucosaminil fosfatidilinositol) y *ALF1* (codificador de una tubulina alfa asociada al citoesqueleto) respectivamente. Cinco ORFs (G2888, G3193, G3406, G3435 y G8541) presentan similitudes con proteínas de levadura o de otros organismos. En este caso, la multiplicidad de los transcritos, debida a hibridaciones inespecíficas de la sonda con otras regiones del genoma de *S. cerevisiae*, puede explicarse como el resultado de la redundancia de la información genética observada en la levadura. Por ejemplo, es interesante observar que la ORF G3435 (1923 pb de longitud) presenta una gran similitud de secuencia con la ORF YBR216c (2022 pb), siendo sus tamaños también muy parecidos; las dos señales observadas en la figura 4.2.7. (de tamaños próximos) podrían ser el reflejo de la hibridación de la sonda de la ORF G3435 con los transcritos de dos ORFs resultantes de una duplicación evolutiva. También es posible que una ORF de lugar a la aparición de 2 ó varios transcritos de tamaño diferente, bien por existir varias señales de inicio o por existir varias señales de terminación (Freire Picos *et al.*, 1995).

4.4.1.3. ORFs de alta expresión:

Cuatro ORFs (3,9 %) han sido clasificadas como de alta expresión. La ORF G3210 ha sido caracterizada por otro laboratorio en el período de realización de este trabajo, como una probable helicasa ATP-dependiente que sería utilizada en el procesamiento del pre-RNA ribosómico (Weaver *et al.*, 1997); siendo el rRNA uno de los ácidos nucleicos más abundantes en la célula, el elevado nivel de expresión de este gen podría indicar que su producto es también abundante para facilitar dicho procesamiento.

Los elevados niveles de expresión obtenidos para las otras 3 ORFs no parecen sin embargo fácilmente interpretables, así por ejemplo G2913 no ha sido aún caracterizada y se considera como una proteína hipotética; a pesar de sus altos niveles de expresión, la delección del gen, realizada en nuestro laboratorio, no da lugar a un fenotipo relevante en las condiciones ensayadas (Tizón *et al.*, 1998 enviado). Respecto a las otras dos ORFs, no se tienen datos que puedan estar relacionados con su alto nivel de expresión; la ORF G7737 codifica una proteína hipotética y la ORF G8537 presenta una gran similitud con el gen *SGE1* que codifica una proteína de resistencia a fármacos (Goffeau *et al.*, 1997).

4.4.1.4. ORFs de expresión media:

Se encontraron 34 ORFs (33%) de expresión media; estos resultados incluyen a 12 ORFs consideradas hipotéticas porque no poseen ninguna homología con cualquiera de las secuencias presentes en las bases de datos (G3505, G3690, G7594, N1773, N1850, N1929, N2190 y N2254) o presentan similitudes significativas con otras ORFs de función desconocida en *S. cerevisiae* (G8523, N1925, N1962, N2194). El hecho de que se haya detectado su expresión indica que se trata de genes funcionales.

Para las otras 22 ORFs originalmente consideradas como desconocidas y pertenecientes a este grupo, existen otros datos adicionales que corroboran su funcionalidad (Ver Tabla I del apéndice).

4.4.1.5. ORFs de baja expresión:

48 ORFs presentaron un nivel de expresión absoluto inferior al 1 % del nivel de *ACT1*; de estas, 18 ORFs se consideraban proteínas hipotéticas

(G3116, G4332, G3107, G3483, G3489, G7801, N1706, N1714, N1739, N1751, N1765, N1888, N2227, N1710, N1747, N1790, N1913 y N2219). Cinco ORFs (G3175, N1761, N1917, N2160 y N2205) han sido caracterizadas como los genes *NBP35*, *YNB4*, *PSU1*, *PIE3* y *PHO23* respectivamente, y sus funciones no están completamente dilucidadas. Las 23 ORFs restantes presentan similitudes con proteínas de *S. cerevisiae* o de otros organismos. La información disponible no nos permite relacionar los niveles de expresión observados con la función hipotética de las ORFs estudiadas.

Si comparamos la distribución de expresión de las ORFs analizadas en este trabajo con otros estudios realizados en levaduras (Yoshikawa e Isono, 1990, Veculescu *et al.*, 1997; Richard *et al.*, 1997; DeRisi *et al.*, 1997; Wodicka K *et al.*, 1997; Naitou *et al.*, 1997; consorcio B2 de EUROFAN I) los datos son coincidentes e indican que prácticamente la mitad de los transcritos están presentes a bajos niveles.

Además, es interesante observar que según los datos obtenidos en nuestro estudio, la mayoría de estos transcritos se expresaron a niveles detectables en todas las condiciones probadas, a pesar de que, probablemente, no todas las ORFs implicadas codifiquen proteínas con funciones esenciales para el crecimiento. Se puede especular que durante la evolución de las células de levadura, resultó mas rentable tolerar una expresión global (aunque generalmente de bajo nivel) de funciones inapropiadas bajo una condición de crecimiento particular, que hacer frente a los requisitos energéticos de una regulación transcripcional total de los genes correspondientes.

Además de las condiciones generales que acabamos de comentar y que permiten interpretar el elevado porcentaje de genes de baja expresión, también es posible pensar que existen condiciones específicas en el fondo

genético de la línea FY73 utilizada que determinen bajos niveles de expresión de genes concretos.

Aunque la cepa FY73 se consideró silvestre, se trata de una línea haploide de tipo $\text{Mat}\alpha$. Se pensó que dentro de las ORFs que presentaron un bajo nivel de expresión, podía haber genes α -específicos parcialmente reprimidos en las células α utilizadas para hacer las extracciones de RNA. Se estudió la zona promotora (800pb) de estas ORFs con un bajo nivel absoluto y se destacaron nueve ORFs (G2861, G3090, G3116, G3179, G7589, G8530, N1761, N1958 y N2160) que presentaban sitios de unión para el factor represor de los genes α -específicos, $\text{Mat}\alpha 2\text{p}$ (Figura 4.3.).

ORF	N.A.	% H	SECUENCIA	Pos.	-800	-700	-600	-500	-400	-300	-200	-100	ATG
G2861	0,06	71,4	TATTTACACGGTTTT	-570			■						
G3090	0,01	71,4	TATTA CTGGTTTT	-180							■		
G3116	0,02	71,4	GATTTGTTTCGTTTT	-161							■		
G3179	0,02	75,0	TTATTATTATACAA	-185							■		
G7589	0,03	71,4	GATGCTACGTGGT	-755	■								
G8530	0,05	71,4	GCTTCTGGCTTGT	-142							■		
N1761	0,05	71,4	GAAAACAGGGTTTT	-431				■					
		71,4	GATCCTTGGTTTC	-290					X				
N1958	0,04	71,4	CCTGTATAGGTTTT	-551			■						
N2160	0,03	71,4	GATGTCGGTGATTC	-529			■						
Con.			GATGCTGGGTTTT										

Figura 4.3. Representación de las ORFs de baja expresión que presentaron en su zona promotora (800pb) secuencias con un porcentaje de homología (%H) superior a 70% con la secuencia consenso (Con.) del sitio de unión para el factor $\text{Mat}\alpha 2\text{p}$. Además de dar la posición (Pos.) exacta del sitio de unión se indica la zona donde está presente en el promotor de la ORF considerada. X: localización en el promotor del sitio de unión del factor $\text{Mcm}1\text{p}$.

Excepto para las ORF G7589, cuyo sitio de unión a $\text{Mat}\alpha 2\text{p}$ está más alejado, y N1761 que presentó dos sitios de unión, las demás ORFs parecen tener los sitios de unión a $\text{Mat}\alpha 2\text{p}$ localizados en dos regiones concretas

respecto del ATG de inicio (-600 a -500 para 3 ORFs y -200 a -100 para 4 ORFs). La ORF N1761 presenta una alta probabilidad de ser un gen *a*-específico ya que se demostró que Mat α 2p suele reprimir la expresión de los genes *a*-específicos cuando está asociado con Mcm1p (Johnson, 1992), una proteína que aumenta la afinidad de unión a DNA de Mat α 2p por lo menos 50 veces (Keleher *et al.*, 1988). El sitio de unión de Mcm1p en el promotor de los genes *a*-específicos suele estar localizado entre dos sitios de unión a Mat α 2p (Herskowitz *et al.*, 1992) lo que sucede con la ORF N1761, cuyo promotor contiene un sitio de unión a Mcm1p situado en la posición - 334.

Estos resultados obtenidos en el promotor sugieren que un porcentaje (7.8%) de ORFs de baja expresión pudiera incrementar su expresión en un fondo genético *a*, sin embargo sería necesario comprobar experimentalmente esta hipótesis.

4.4.2. Relación de los niveles absolutos de expresión con otros parámetros génicos y genómicos:

4.4.2.1. Comparación de los niveles de expresión con los índices de CAI:

El origen del índice de *CAI* (*Codon Adaptation Index* propuesto por Sharp y Li, 1987) está en la observación de que, pese que el código genético es degenerado, los codones sinónimos son raramente utilizados con la misma frecuencia. Los factores selectivos que controlan el uso de codones han sido más extensamente estudiados en organismos unicelulares procariotas (bacterias) y eucariotas (levaduras).

El índice de *CAI* cuantifica la frecuencia relativa del uso de un codón particular en un gen. Los valores que toma oscilan de 0 a 1 y los valores más

elevados indican mayor desviación en el uso de codones. De forma teórica, el *CAI* está relacionado con una mayor eficacia en la traducción de determinados mensajeros que contienen aquellos codones cuyos anticodones son más abundantes en la población de tRNAs.

Adicionalmente, se ha encontrado (Benetzen y Hall, 1982) que los genes de levadura que se expresan a un alto nivel utilizan casi exclusivamente 25 de los 61 codones posibles. Por el contrario, en genes que se expresan a niveles muy bajos, todos los codones son utilizados y el uso de codones es por lo general menos selectivo, es decir, hay una menor desviación (*bias*) en el uso aleatorio de codones.

Si se utilizan los valores de *CAI*, correspondientes a las ORFs analizadas, recopilados en la tabla 4.7. y recogidos en la base de datos de *MIPS* (*Martinsride Institute for Protein Sequences*) para compararlos con los valores de nivel absoluto obtenidos en este trabajo, se obtiene una representación gráfica que puede ser orientativa de la relación existente entre ambos (Figura 4.4.).

Tabla 4.7. Niveles absolutos (N.A.) y valores de índice de *CAI* de las ORFs analizadas en los cromosomas VII y XIV de *S. cerevisiae*

ORF	N.A.	CAI	ORF	N.A.	CAI	ORF	N.A.	CAI	ORF	N.A.	CAI
G2850	<0,001	0,13	G3505	0,22	0,12	N1696	0,03	0,17	N1872	<0,001	0,12
G2856	0,20	0,15	G3517	0,10	0,12	N1706	0,01	0,13	N1875	0,01	0,12
G2861	0,06	0,12	G3535	MT	0,10	N1710	0,02	0,10	N1879	0,13	0,11
G2882	0,16	0,22	G3540	MT	0,14	N1714	0,01	0,12	N1882	0,19	0,15
G2889	MT	0,11	G3679	0,01	0,10	N1718	0,01	0,15	N1888	0,02	0,15
G2913	1,58	0,15	G3690	0,46	0,16	N1727	0,04	0,16	N1897	0,08	0,18
G2916	0,29	0,21	G3719	<0,001	0,15	N1735	MT	0,10	N1913	0,04	0,15
G3080	0,20	0,22	G3759	0,09	0,12	N1739	0,02	0,12	N1917	0,06	0,13
G3090	0,01	0,12	G4332	0,04	0,13	N1743	MT	0,11	N1925	0,12	0,13
G3107	0,04	0,18	G7538	0,23	0,18	N1747	0,07	0,15	N1929	0,25	0,15
G3110	0,13	0,14	G7553	0,36	0,23	N1751	0,05	0,16	N1954	0,47	0,18
G3113	0,15	0,20	G7584	0,05	0,14	N1761	0,05	0,14	N1958	0,04	0,09
G3116	0,02	0,14	G7587	0,17	0,14	N1765	0,01	0,12	N1962	0,17	0,16
G3146	0,19	0,10	G7589	0,03	0,13	N1773	0,12	0,11	N1966	0,04	0,12
G3168	0,02	0,12	G7594	0,42	0,13	N1774	0,13	0,22	N2160	0,05	0,13
G3175	0,06	0,12	G7725	0,07	0,18	N1777	MT	0,08	N2185	0,10	0,12
G3179	0,02	0,13	G7737	0,74	0,17	N1780	0,03	0,20	N2190	0,18	0,11
G3189	MT	0,22	G7748	0,31	0,13	N1785	<0,001	0,13	N2194	0,14	0,15
G3193	MT	0,13	G7801	0,09	0,24	N1790	0,01	0,13	N2205	0,05	0,13
G3210	1,15	0,27	G7806	0,32	0,13	N1795	<0,001	0,13	N2215	0,08	0,13
G3406	MT	0,13	G7826	MT	0,12	N1825	0,16	0,25	N2219	0,07	0,16
G3432	0,02	0,13	G8523	0,18	0,11	N1843	0,07	0,13	N2227	0,04	0,11
G3435	MT	0,15	G8530	0,05	0,13	N1847	0,29	0,24	N2231	0,04	0,13
G3474	0,41	0,31	G8537	0,50	0,14	N1850	0,13	0,11	N2250	0,05	0,15
G3483	0,06	0,13	G8541	MT	0,14	N1858	0,05	0,24	N2254	0,16	0,10
G3489	0,02	0,10				N1870	0,08	0,13	N2312	0,10	0,13

Como hemos visto en el apartado anterior, la gran mayoría de las ORFs analizadas en este trabajo pueden clasificarse como de baja o moderada expresión. En La figura 4.4. podemos ver que los valores de *CAI* están dentro de un rango relativamente homogéneo, con un valor mínimo de 0,08 (N1777) y máximo de 0,31 (G3474), compatible con genes de expresión baja o moderada. Sin embargo algunas ORFs de alta expresión no muestran un valor de *CAI* elevado, lo que indica que, a pesar de tener unas tasas de transcripción altas, puede que su eficacia a nivel traduccional no sea óptima.

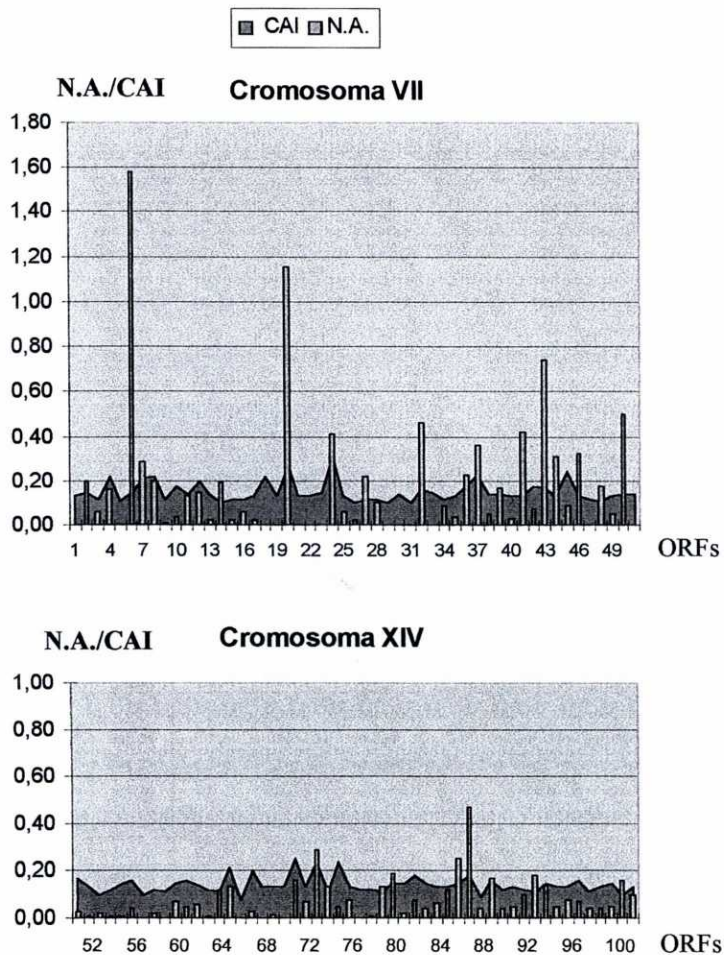


Figura 4.4. Comparación de los niveles absolutos (N.A.) de expresión de las 103 ORFs de función desconocida en los cromosomas VII y XIV de *S. cerevisiae* con sus respectivos índices de *CAI*. Las ORFs están numeradas siguiendo su orden creciente en la nomenclatura.

4.4.2.2. Comparación de los niveles de expresión con el tamaño de las ORFs:

Cuando, después del proyecto de secuenciación del genoma de levadura, se eligieron las ORFs de función desconocida que iban a ser analizadas en EUROFAN I, se descartaron todas las que poseían un tamaño

inferior a 300 nucleótidos ya que la probabilidad de que una ORF mayor de 100 codones ocurriera sólo por casualidad era muy baja (0,2%) (Barry *et al.*, 1996). ¿Se puede especular que una ORF pueda ver su expresión influenciada por la longitud de su secuencia nucleotídica?

Cuando correlacionamos el nivel absoluto de expresión de las ORFs analizadas con sus respectivos tamaños (Figura 4.5.) podemos ver que todas las ORFs con un tamaño superior a 2500 nucleótidos presentaron un bajo nivel de expresión, incluso dos de ellas, G2850 y G3719 de 4.209 y 2.664 nucleótidos respectivamente, no fueron detectadas. Por otro lado, las ORFs con un tamaño inferior a 2.000 nucleótidos se reparten de manera relativamente homogénea entre las ORFs de baja y moderada expresión y las 4 ORFs que se expresan a altos niveles no tienen un tamaño homogéneo aunque en todos los casos éste es inferior a 2.000 nucleótidos. Finalmente, cabe destacar que dos ORFs (N1785 y N1795) que poseen tamaños (300 y 390 nucleótidos respectivamente) cercanos al elegido para considerar una ORF fortuita no fueron detectadas en el análisis transcripcional.

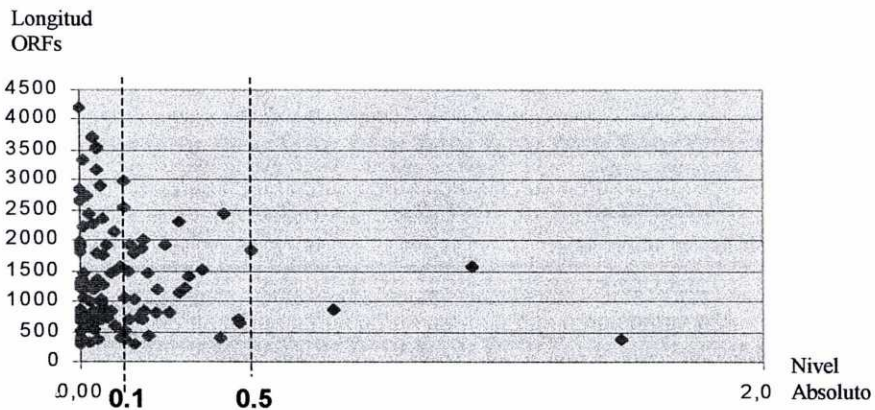


Figura 4.5. Gráfica de correlación entre los niveles absolutos de expresión y los tamaños (en nucleótidos) de las 103 ORFs de función desconocida en los cromosomas VII y XIV de *S. cerevisiae*.

4.4.2.3. Análisis de los niveles absolutos de expresión en relación con la posición en el cromosoma:

- Análisis realizado a partir de las 103 ORFs estudiadas en este trabajo:

Una representación esquemática de la distribución de los valores de nivel absoluto de expresión obtenidos para las ORFs analizadas a lo largo de las regiones estudiadas en este trabajo dentro de los cromosomas VII y XIV de *S. cerevisiae* se muestra en la figura 4.6. En un primer análisis se estudió la distribución de los niveles de expresión teniendo en cuenta si la ORF está situada en la hebra de Watson o de Crick. Los datos recopilados en la tabla 4.8. nos indican que la distribución de los niveles es muy homogénea, lo que revela un reparto aleatorio de los niveles de expresión a lo largo de las dos hebras de los cromosomas.

Tabla 4.8. Distribución de las ORFs en las hebras de Watson (W) y Crick (C) de los cromosomas (crom.) VII y XIV de *S. cerevisiae* en función de su nivel de expresión absoluto (N.A.).

Crom.	HEBRA	C	W	C+W
VII	Total ORFs	24	27	51
	Alta expresión	2 (8,3%)	2 (7,4%)	4 (7,8%)
	Mediana expresión	8 (33,3%)	10 (37,1%)	18 (35,3%)
	Baja expresión	11 (45,8%)	7 (25,9%)	18 (35,3%)
	No detectadas	1 (4,2%)	1 (3,7%)	2 (3,9%)
	Múltiples transcritos	3 (12,5%)	6 (22,2%)	9 (17,7%)
XIV	Total ORFs	22	30	52
	Alta expresión	--	--	--
	Mediana expresión	8 (36,4%)	8 (26,7%)	16 (30,8%)
	Baja expresión	11 (45,8%)	19 (63,3%)	30 (57,6%)
	No detectadas	1 (4,2%)	2 (6,7%)	3 (5,8%)
	Múltiples transcritos	2 (8,3%)	1 (3,3%)	3 (5,8%)

Si la distribución de los niveles de expresión absolutos de las distintas ORFs dentro de las dos hebras de un mismo cromosoma parece homogénea, la comparación entre los dos cromosomas estudiados muestra diferencias significativas. Así la región estudiada del cromosoma XIV no presentó ninguna ORF con un alto nivel de expresión pero sí una alta concentración de ORFs de baja expresión (Tabla 4.8.).

Un estudio de la distribución de los genes activos y de los niveles de expresión de los transcritos en el cromosoma VI de *S. cerevisiae*, mostró que el gen de alta expresión *ACT1* se situaba con otros genes de expresión comparable en la región central de los brazos cromosómicos, en oposición con los genes de menor expresión, situados todos a ambos lados de esta zona (Yoshikawa y Isono, 1991). En este trabajo, dos ORFs (G2913 y G3210) que presentaron niveles de expresión absoluta comparables con los de *ACT1* (1,58 y 1,15 respectivamente) también están situados en la zona central del brazo izquierdo del cromosoma VII. Sin embargo, otras dos ORFs consideradas como de alta expresión (G7737 y G8537), pero que mostraron niveles de expresión absoluta inferiores al de *ACT1* (0,74 y 0,50 respectivamente), están relativamente cerca de la región telomérica del brazo derecho del cromosoma VII. Finalmente en el cromosoma XIV, a pesar de haber estudiado una región céntrica de su brazo izquierdo, no apareció ninguna ORF de nivel de expresión absoluta comparable al de *ACT1*.

Otro aspecto que muestra la figura 4.6. es la ubicación de las ORFs que no pudieron ser cuantificadas al no presentar ninguna señal de hibridación. Se encuentran relativamente aisladas unas de otras y no demasiado lejos de otras ORFs cuyos transcritos fueron detectados (incluso con niveles absolutos de expresión altos y moderados), lo que indica que la ausencia de transcrito no es

atribuible en estos casos a un efecto posicional de la ORF considerada en una determinada región del cromosoma.

CROMOSOMA VII

CROMOSOMA XIV

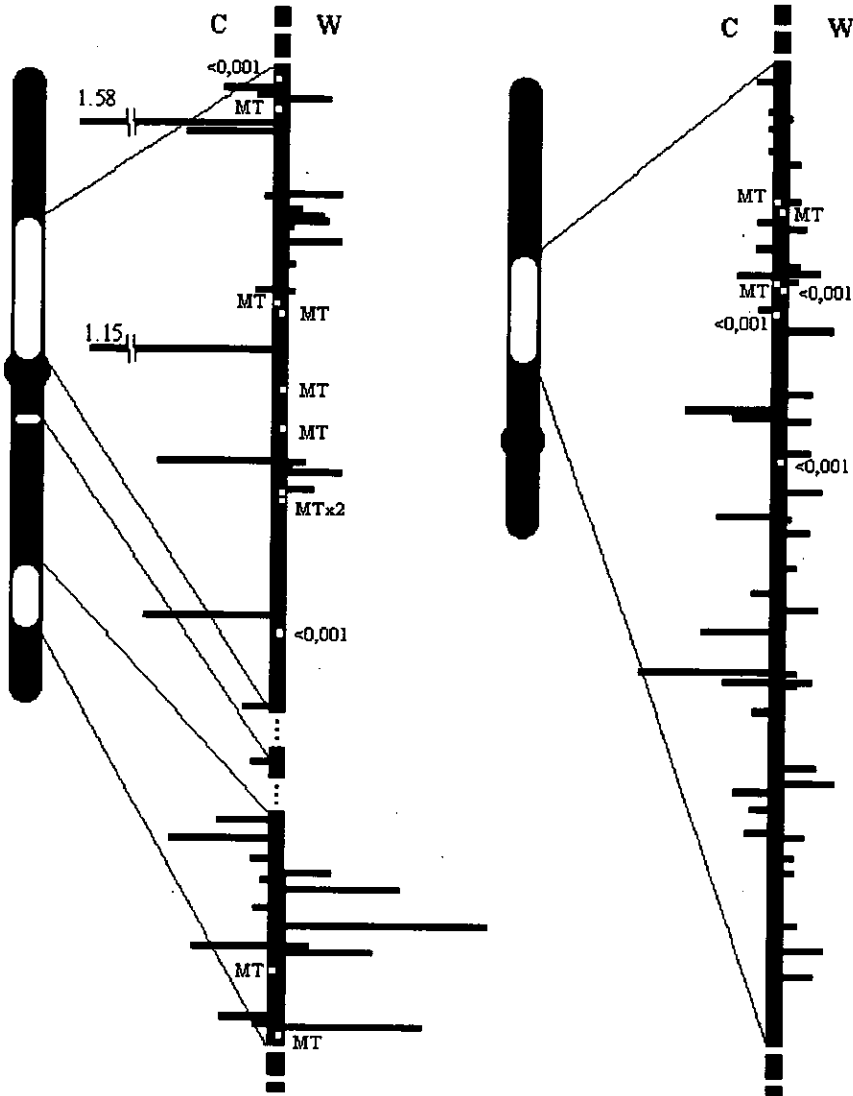


Figura 4.6. Representación esquemática del nivel absoluto de expresión de las 103 ORFs de función desconocida en los cromosomas VII y XIV de *S. cerevisiae*. (W: hebra de Watson, C: hebra de Crick). Los cuadros blancos dentro de las regiones estudiadas indican las ORFs que presentaron múltiples transcritos (M.T.) o ninguna señal de hibridación (0) para las cuales no fue posible calcular un valor de nivel absoluto.

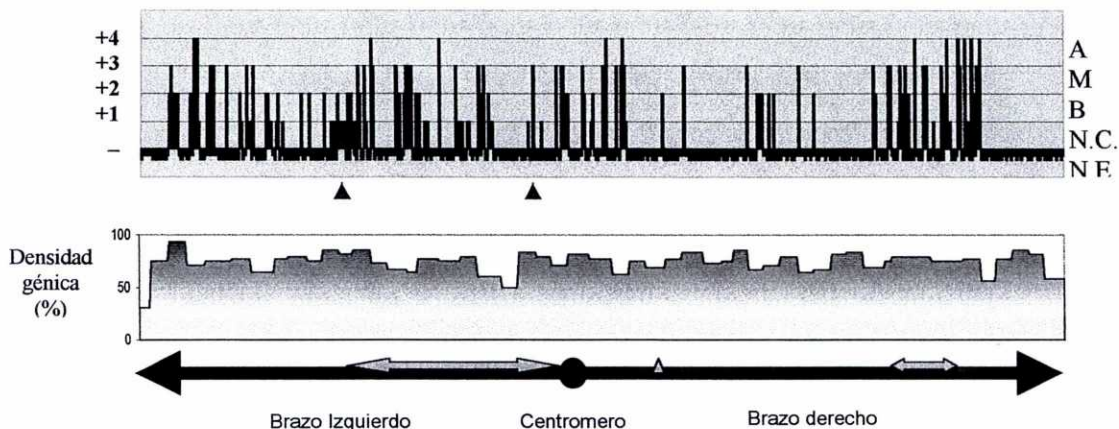
- Análisis realizado sobre las 132 y 250 ORFs estudiadas en los cromosomas VII y XIV respectivamente:

Dado que las regiones analizadas en este trabajo podían no resultar representativas de lo que sucede realmente en los dos cromosomas estudiados, ampliamos el análisis con los datos aportados por otros laboratorios que participan en el consorcio B2. Una confirmación de si existe correlación entre la localización de los diferentes tipos de ORFs y su nivel absoluto de expresión puede lograrse en un estudio a escala cromosómica. El análisis de la distribución de los valores de nivel absoluto resultantes de la totalidad del análisis realizado por el consorcio B2 (132 ORFs del cromosoma VII y 250 del cromosoma XIV) se muestra en la figura 4.7.

El análisis de los datos reflejados en la figura 4.7. conjuntamente con los de la tabla 4.8., permite confirmar la alta concentración de ORFs de baja expresión en el cromosoma XIV respecto al cromosoma VII (69,2% vs 33,3%). Por el contrario el cromosoma VII contiene un alto porcentaje de ORFs de alta y moderada expresión (39,3%) frente al cromosoma XIV (9,2%).

En lo que se refiere a la distribución de las ORFs a lo largo del cromosoma en función de su nivel absoluto, se puede observar que en los cromosomas VII y XIV, las ORFs de alta expresión tienden a distribuirse tanto en las regiones centrales de los brazos cromosómicos como en las regiones subtelo méricas y pericentroméricas; por tanto las observaciones de Naitou *et al.* (1997) en el cromosoma VI no parecen extrapolables a la totalidad de los cromosomas, hecho que coincide con los datos aportados por el reciente estudio de Vesulescu *et al.* (1997).

CROMOSOMA VII



CROMOSOMA XIV

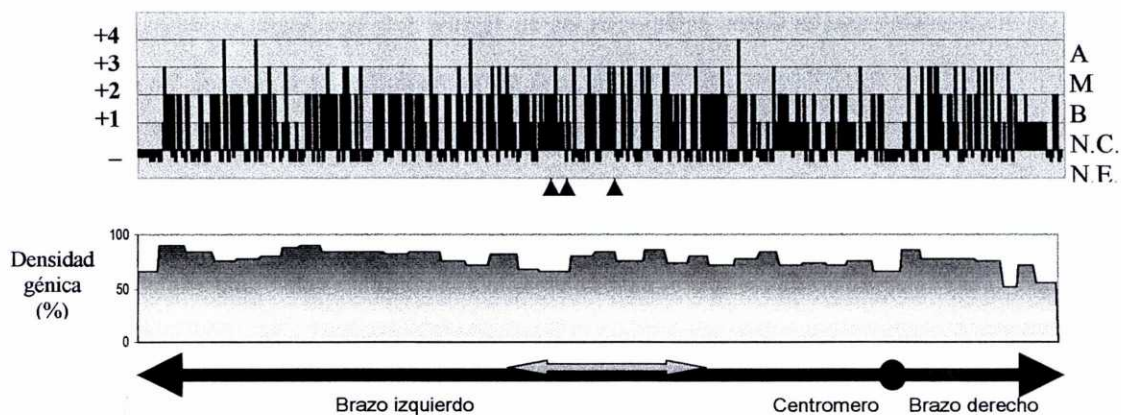


Figura 4.7. Comparación de la densidad génica con la distribución de los valores de nivel absoluto [A: alto (con un valor relativo de +4); M: moderada (+3); B: baja (+2); N.C.: no calculado; ORFs no detectadas y con múltiples transcritos (+1); N.E.: no estudiados (-)] a lo largo de los cromosomas VII y XIV de *S. cerevisiae*. Se señalan con flechas de doble cabeza las regiones estudiadas en este trabajo y con puntas de flecha negra la situación de las ORFs que no presentaron ningún transcrito en este trabajo.

4.4.2.4. Expresión y densidad génica:

Los niveles de expresión pueden estar relacionados no sólo con la localización en el cromosoma sino también con las características particulares de una determinada región, como por ejemplo la densidad génica.

Cuando analizamos la localización de las ORFs no detectadas en este trabajo y la comparamos con la densidad génica a lo largo del cromosoma, podemos ver (Figura 4.7.) que, para el cromosoma XIV dos de estas ORFs (N1785 y N1795) se localizan en una región de baja densidad génica media (65%) mientras que la ORF N1872 y las dos ORFs no detectadas del cromosoma VII están situadas en una zona de alta densidad génica (superior al 80%).

Tabla 4.9. Distribución de las ORFs de función desconocida no detectadas en los cromosomas VII y XIV de *S. cerevisiae* en función de la densidad génica. P representa el porcentaje de la secuencia total de los cromosomas VII y XIV con una densidad génica determinada y D.A. la distribución aleatoria de las ORFs no detectadas esperada en función de P.

Densidad génica	ORFs no detectadas	P.	D.A.
>80%	38 (55,0%)	34,4%	23
80%-70%	23 (33,4%)	45,2%	31
<70%	8 (11,6%)	20,4%	15

Sin embargo, si analizamos la totalidad de las ORFs no detectadas en los cromosomas VII y XIV por todos los laboratorios del consorcio B2, podemos observar que un porcentaje elevado (55%) está situado en zonas de alta densidad génica (>80%) mientras que sólo un 11% está situado en regiones de baja densidad génica (<70%). Esta distribución se aleja de la distribución aleatoria esperada en función del porcentaje de secuencia génica total correspondiente a las citadas densidades génicas (Tabla 4.9.).

4.5. Niveles relativos de expresión:

Los niveles relativos de mRNA en cada condición de cultivo nos permiten comparar las señales de hibridación obtenidas en las distintas situaciones de crecimiento. Están representados por el porcentaje relativo a la señal máxima obtenida en una de las ocho condiciones del proyecto EUROFAN.

4.5.1. Niveles de expresión en las distintas condiciones de cultivo:

La representación gráfica de los valores obtenidos para las 103 ORFs de los cromosomas VII y XIV de *S. cerevisiae* nos permite estudiar el comportamiento general de la transcripción de estas ORFs en las 8 condiciones estudiadas (Figura 4.8.). Se puede así destacar la reducción de la expresión génica durante la fase estacionaria para numerosas ORFs. También hay que señalar el elevado porcentaje de ORFs del cromosoma XIV que responden al choque osmótico.

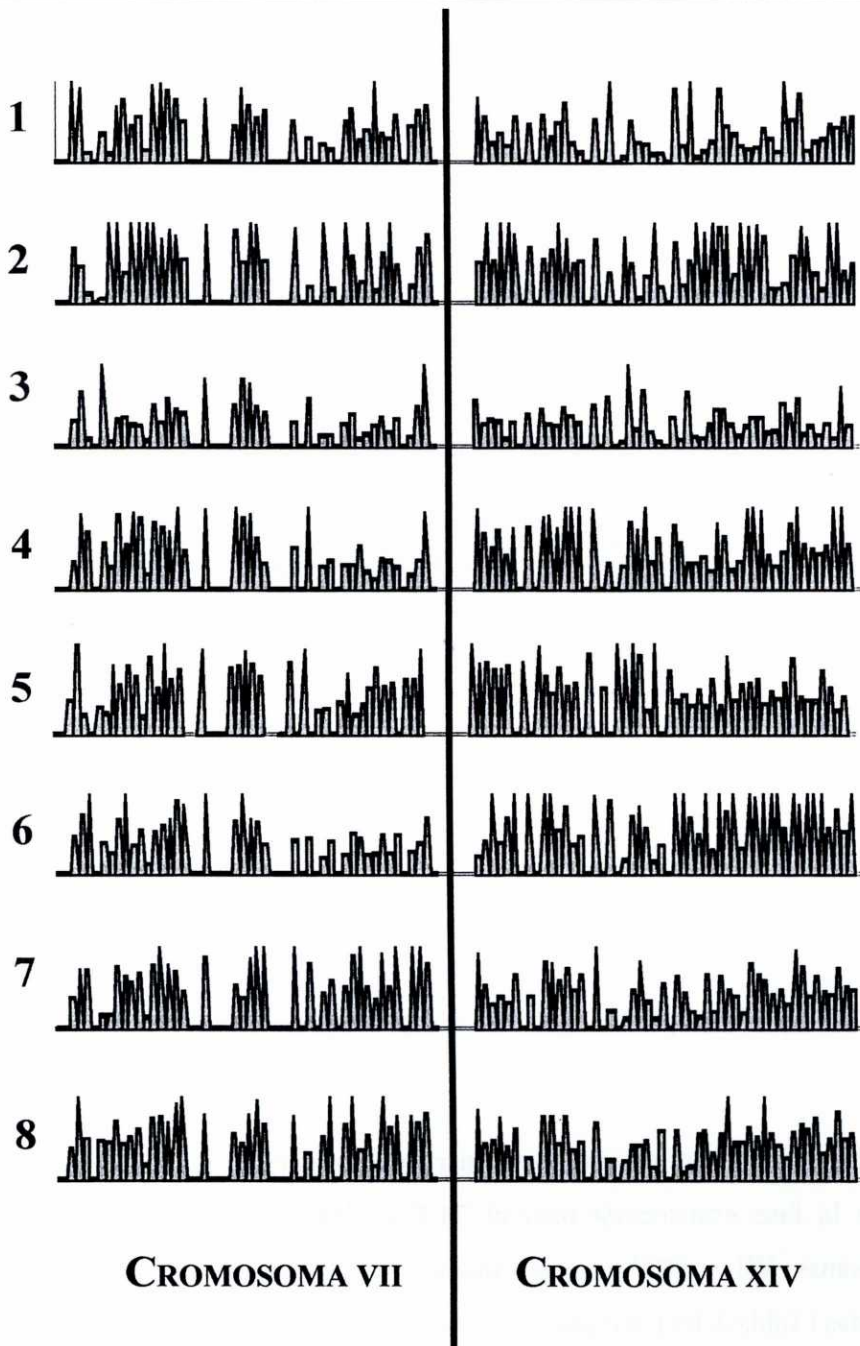


Figura 4.8. Representación de los niveles relativos de las 103 ORFs de función desconocida de los cromosomas VII y XIV de *S. cerevisiae* (1-YPGE, 2-YPD, 3-YPD/fase estacionaria, 4-YNBglu/privación de aminoácidos, 5-YNBglu/privación de nitrógeno, 6-YNBglu/choque osmótico, 7-YNBglu/23°C, 8-YNBglu/36°C).

Un análisis mas detallado nos permite afinar estas observaciones y comparar los resultados obtenidos para cada cromosoma. Cuando calculamos el porcentaje de las ORFs cuantificadas que se expresan por encima del 50 % de su nivel absoluto en cada condición de cultivo, obtenemos la gráfica de la Figura 4.9.



Figura 4.9. Proporción de las ORFs cuantificadas que se expresaron como mínimo a una tasa del 50% de su nivel absoluto en las distintas condiciones de cultivo (1-YPGE, 2-YPD, 3-YPD/fase estacionaria, 4-YNBglu/privación de aminoácidos, 5-YNBglu/privación de nitrógeno, 6-YNBglu/choque osmótico, 7-YNBglu/23°C, 8-YNBglu/36°C).

Esta figura nos permite confirmar la disminución de la transcripción durante la fase estacionaria para el 74,7% y 80,4% de las ORFs de los cromosomas VII y XIV respectivamente. En el resto de las condiciones ensayadas (Tabla 4.10.), en general el porcentaje de ORFs que superan el 50% de su expresión absoluta superan el 60 % de la totalidad de las ORFs estudiadas, excepto para el crecimiento en glicerol-etanol dónde son aproximadamente la mitad.

Tabla 4.10. Proporción de las ORFs cuantificadas que se expresaron como mínimo a una tasa del 50% de su nivel absoluto en los cromosomas (Crom.) VII y XIV en las distintas condiciones (Cond.) de cultivo (1-YPGE, 2-YPD, 3-YPD/fase estacionaria, 4-YNBglu/privación de aminoácidos, 5-YNBglu/privación de nitrógeno, 6-YNBglu/choque osmótico, 7-YNBglu/23°C, 8-YNBglu/36°C).

Cond.	Crom. VII (%)	Crom. XIV (%)	Total (%)
1	65,8	36,9	51,4
2	68,4	67,3	67,9
3	26,3	19,6	23,0
4	55,3	69,6	62,5
5	71,1	63,0	67,1
6	47,3	78,3	62,8
7	78,9	47,8	63,4
8	71,1	50,0	60,6

Cuando comparamos la regulación de las ORFs en los dos cromosomas, podemos ver que no existen discrepancias en el porcentaje de ORFs que se expresan cercanas a su nivel absoluto en YPD, sin embargo en las otras condiciones de crecimiento se observan diferencias, principalmente en YPGE, durante el choque osmótico o a temperaturas extremas.

4.5.2. Clasificación de las ORFs en función de su regulación:

Las 8 condiciones de cultivo elegidas en el consorcio B2 de EUROFAN permitieron observar 8 tipos distintos de respuesta de las 103 ORFs analizadas en cuanto a su regulación transcripcional:

- Expresión constitutiva.
- Activación en glucosa (G+) si 1-YPGE < 2-YPD.
- Represión en glucosa (G-) si 1-YPGE > 2-YPD .
- Activación en nitrógeno (N+) si 4-YNBglu > 5-Privación de nitrógeno.
- Represión en nitrógeno (N-) si 4-YNBglu < 5-Privación de nitrógeno.
- Activación en fase estacionaria (S+) si 3-YPDestacionario > 1-YPGE o 2-YPD.
- Activación durante el choque osmótico (O+) si 6-choque osmótico > 4-YNBglu.
- Activación por choque térmico (H+) si 8-YNBglu 37°C > 7-YNBglu 23°C.

Si la variación de expresión entre las condiciones consideradas anteriormente podía cuantificarse al menos por un factor 2, se consideraba que una ORF estaba regulada. De esta forma pudimos proceder a una clasificación de las ORFs en función de su regulación (Tabla 4.11.).

Numerosas ORFs mostraron estar reguladas ya que sólo el 37,2 % de las ORFs cuantificadas mostraron una expresión constitutiva. Sin embargo hay que tener en cuenta que el umbral fijado para considerar una ORF como regulada es relativamente bajo y que muchas respuestas fisiológicas están mediadas por variaciones transcripcionales mayores. La mayoría de estas ORFs reguladas (45,9%) lo hicieron en respuesta sólo a una de las diferentes condiciones de cultivo, mientras que 14 de ellas (16,1%) lo hicieron frente a dos o tres.

Tabla 4.11. Clasificación de las 103 ORFs analizadas en los cromosomas VII y XIV de *S. cerevisiae* en función de su regulación (G+/-: activación o represión en glucosa; N+/-: activación o represión en nitrógeno; S+: activación durante la fase estacionaria; O+: activación por choque osmótico; H+: activación por choque térmico). NC: No Cuantificada; C: Constitutiva.

ORF	Regulación	ORF	Regulación	ORF	Regulación	ORF	Regulación
G2850	NC	G3505	C	N1696	C	N1872	NC
G2856	C	G3517	N-	N1706	C	N1875	C
G2861	G-	G3535	NC	N1710	G+O+	N1879	C
G2882	N+	G3540	NC	N1714	G+	N1882	O+
G2889	NC	G3679	C	N1718	G+	N1888	G+H+
G2913	G-S+H+	G3690	C	N1727	C	N1897	G+O+
G2916	G+H+	G3719	NC	N1735	NC	N1913	G+
G3080	C	G3759	S+	N1739	C	N1917	C
G3090	G-	G4332	C	N1743	NC	N1925	G+N+H+
G3107	G+	G7538	G+	N1747	C	N1929	N-O+
G3110	C	G7553	C	N1751	G+	N1954	G+H+
G3113	G+	G7584	G-N-	N1761	G+	N1958	G+
G3116	C	G7587	G+	N1765	C	N1962	G+
G3146	C	G7589	N-	N1773	G+	N1966	G+
G3168	C	G7594	N+	N1774	G+	N2160	O+
G3175	C	G7725	G+	N1777	NC	N2185	O+
G3179	H+	G7737	G-N-O+	N1780	C	N2190	G-
G3189	NC	G7748	N-	N1785	NC	N2194	C
G3193	NC	G7801	G+	N1790	G-O+	N2205	C
G3210	C	G7806	N-	N1795	NC	N2215	G+
G3406	NC	G7826	NC	N1825	G+N-H+	N2219	O+
G3432	C	G8523	G-N-	N1843	C	N2227	O+
G3435	NC	G8530	C	N1847	G-	N2231	G+
G3474	G+	G8537	C	N1850	S+	N2250	G-N-
G3483	C	G8541	NC	N1858	G+	N2254	N+
G3489	C			N1870	G+	N2312	C

La expresión de numerosas ORFs parece estar influenciada por la presencia de glucosa (44,2 %); el nivel de expresión del 32,6 % de las ORFs aumentó después del cambio de las células a un medio con glucosa. Sólo el 10.8% sufrió una represión por glucosa (Tabla 4.12.).

El cambio de condiciones de cultivo desde el medio YNBglu a la privación de fuentes de nitrógeno mostró que el 5,8 % de las ORFs eran activadas por fuente de nitrógeno mientras que el 10,5 % eran reprimidas (Tabla 4.12.).

Como se comentó anteriormente, la mayoría de las ORFs ve su expresión reducirse durante la fase estacionaria y por ese motivo no se incluye este tipo de respuesta como una regulación específica. Por el contrario, sí resulta significativo, desde el punto de vista de la regulación, que el 3,6% de las ORFs aumenten su expresión en fase estacionaria.

Varias ORFs (19,8 %) respondieron a condiciones de estrés; el 11,6 % fueron inducidas durante el choque osmótico; de las 10 ORFs inducidas, 9 son del cromosoma XIV. Los niveles del 8,1% de los transcritos aumentaron en respuesta al choque térmico (Tabla 4.12.).

Tabla 4.12. Distribución de las 103 ORFs analizadas en los cromosomas VII y XIV de *S. cerevisiae* en función de su regulación (G+/-: activación o represión en glucosa; N+/-: activación o represión en nitrógeno; S+: activación durante la fase estacionaria; O+: activación por choque osmótico; H+: activación por choque térmico). NC: No Cuantificada; C: Constitutiva.

	CROMOSOMA VII		CROMOSOMA XIV	
	Este trabajo	Total (132 ORFs)	Este trabajo	Total (250 ORFs)
C	18 (45%)	47,4%	14 (30,4%)	33,5%
G	+ 8 (20%)	16,8%	21 (46,7%)	23%
	- 6 (15%)	15,8%	3 (6,6%)	12%
N	+ 2 (5%)	12,6%	3 (6,6%)	10%
	- 7 (17,5%)	7,4%	2 (4,4%)	11%
S+	2 (5%)	9,5%	1 (2,2%)	3%
O+	1 (2,5%)	4,2%	9 (20%)	13%
H+	3 (7,5%)	3,1%	4 (8,9%)	11%
NC	11	37	6	53

Cuando comparamos los resultados obtenidos en este trabajo con el conjunto total de ORFs analizadas por el consorcio B2 en cada cromosoma, podemos observar que en algunos casos aparecen diferencias en cuanto al porcentaje de ORFs con una respuesta regulada determinada. En la figura 4.10. se representa el tipo de respuesta observado para una ORF en función de su situación en el cromosoma.

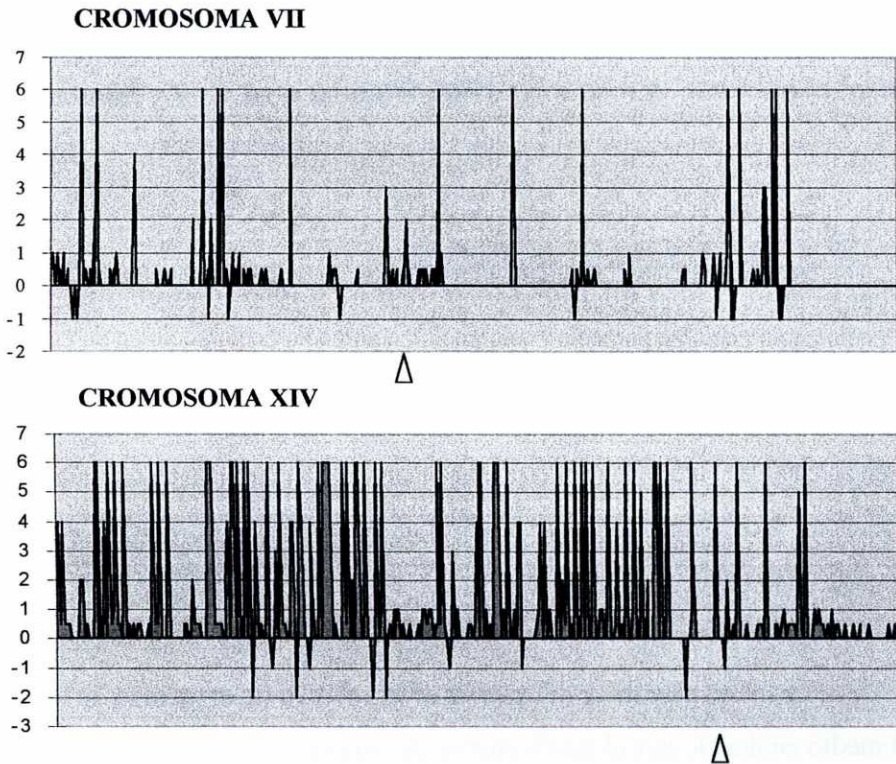


Figura 4.10. Distribución de las 132 y 250 ORFs analizadas por el consorcio B2, respectivamente en los cromosomas VII y XIV de *S. Cerevisiae* en función de su regulación. Las diferentes respuestas observadas fueron relacionadas con valores arbitrarios: Constitutiva: +0.5; G+: +1; G-: -1; N+: +2 N-: -2; S+: +3; O+: +4; H+: +5 y +6 para las ORFs que presentaron mas de una respuesta. Se localiza con una punta de flecha a los centrómeros.

La distribución de la regulación de las ORFs en función de su localización indica que no parece haber agrupaciones o “clusters” de ORFs

con el mismo patrón de regulación, sin embargo hay que tener en cuenta que en muchos casos las ORFs estudiadas no son adyacentes al encontrarse separadas por genes previamente caracterizados. Pese a que el análisis detallado de la expresión de las 250 ORFs de función desconocida en el cromosoma XIV mostró una concentración de ORFs que estaban reprimidas durante la fase estacionaria en la región subtelomérica del brazo izquierdo (Planta *et al.*, enviado), otros estudios de expresión a gran escala muestran que en general, las ORFs co-reguladas están dispersas en el genoma de levadura (Richard *et al.*, 1997; DeRisi *et al.*, 1997; Wodicka *et al.*, 1997; Naitou *et al.*, 1997).

4.5.3. Cuantificación del factor de regulación:

Los cambios de expresión de las ORFs en respuesta a las diferentes condiciones de cultivo son, no sólo cualitativamente, sino también cuantitativamente variables. La intensidad de estos cambios puede estimarse mediante un factor que refleja la intensidad de la inducción o represión que afecta a la ORF regulada.

Este factor se calcula dividiendo el nivel relativo de expresión de la ORF en el medio analizado por el nivel relativo de expresión de la ORF en el medio control (Tabla 4.13.).

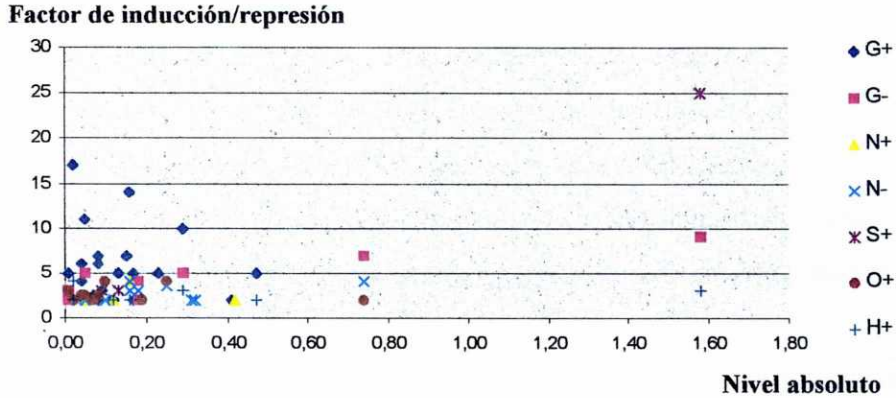
Tabla 4.13. Valores del factor de inducción o represión obtenidos para las ORFs analizadas en los cromosomas VII y XIV de *S. cerevisiae* que presentaron algún tipo de regulación. G+/-: activación o represión en glucosa; N+/-: activación o represión en nitrógeno; S+: activación durante la fase estacionaria; O+: activación por choque osmótico; H+: activación por choque térmico; N.A.: Nivel absoluto.

CROMOSOMA VII								CROMOSOMA XIV									
ORF	G		N		S	O	H	N.A.	ORF	G		N		S	O	H	N.A.
	+	-	+	-	+	+	+			+	-	+	-	+	+	+	
G2861		2						0,06	N1710	2					2		0,02
G2882			3					0,16	N1714	3							0,01
G2913		9			25		3	1,58	N1718	5							0,01
G2916	10						3	0,29	N1751	2							0,05
G3090		2						0,01	N1761	2							0,05
G3107	2							0,04	N1773	2							0,12
G3113	7							0,15	N1774	5							0,13
G3179							2	0,02	N1790		3				3		0,01
G3474	2							0,41	N1825	14			4			2	0,16
G3517				2				0,10	N1847		5						0,29
G3759					3			0,09	N1850					3			0,13
G7538	5							0,23	N1858	11							0,05
G7584		5		2				0,05	N1870	2							0,08
G7587	2							0,17	N1882						2		0,19
G7589				2				0,03	N1888	17						4	0,02
G7594			2					0,42	N1897	7					3		0,08
G7725	3							0,07	N1913	4							0,04
G7737		7		4			2	0,74	N1925	2		2				2	0,12
G7748				2				0,31	N1929				4		4		0,25
G7801	3							0,09	N1954	5						2	0,47
G7806				2				0,32	N1958	6							0,04
G8523		2		3				0,18	N1962	5							0,17
									N1966	2							0,04
									N2160						3		0,05
									N2185						4		0,10
									N2190		4						0,18
									N2215	6							0,08
									N2219						2		0,07
									N2227						3		0,04
									N2231	2							0,04
									N2250	3		2					0,05
									N2254			4					0,16

La mayoría de los cambios de expresión observados fueron relativamente poco intensos; el 62,5 % de las respuestas representan un factor de inducción/represión de apenas 2 a 3, y sólo 10 ORFs, de las 60 que dieron algún tipo de respuesta, mostraron cambios importantes en su expresión (superiores a un factor de 5).

Como se ha comentado anteriormente, los niveles absolutos de expresión de las ORFs analizadas son relativamente pequeños. Un análisis más detallado muestra que el 53,7 % de las ORFs que presentaron algún tipo de respuesta fueron clasificadas como de baja expresión ($N.A. < 0,1$) (Figura 4.11.A). El significado de pequeños cambios en los niveles relativos de transcritos poco abundantes es cuestionable, mientras que cambios análogos en la expresión de ORFs con altos o moderados niveles absolutos pueden resultar más notables. Una alternativa para diferenciar estas dos situaciones, consiste en multiplicar el factor de inducción/represión observado para una ORF por su nivel absoluto. Teniendo en cuenta que el umbral elegido para diferenciar las ORFs de bajo nivel absoluto fue de 0,1 y que la mayoría de los factores de regulación observados eran 2 o 3, se puede tomar el valor 0.3 como umbral mínimo para seleccionar las ORFs reguladas (Figura 4.11.B).

A



B

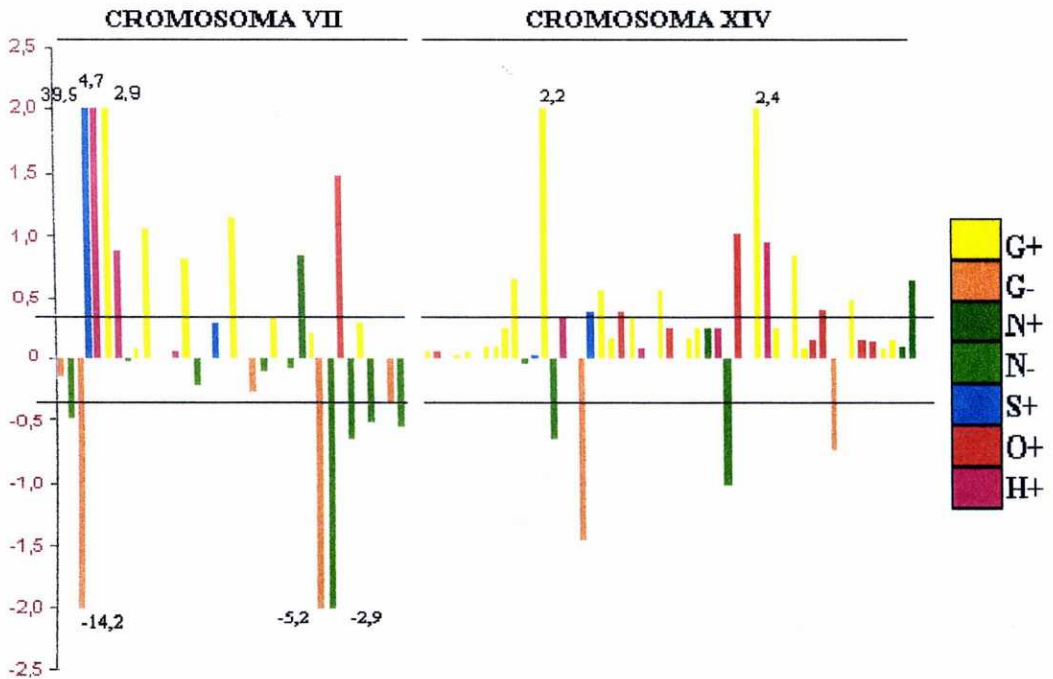


Figura 4.11. A: Correlación entre los valores de nivel absoluto y el factor de regulación observado. **B:** Representación del perfil de las diferentes respuestas observadas después de haber multiplicado el nivel absoluto por el factor de inducción/represión de cada ORF que presentaron algún tipo de regulación. G+/-: activación o represión en glucosa; N+/-: activación o represión en nitrógeno; S+: activación durante la fase estacionaria; O+: activación por choque osmótico; H+: activación por choque térmico. Los valores obtenidos que salen del cuadro de la gráfica se indican.

En la figura 4.11.B se puede ver como, tras introducir este criterio, un gran número de respuestas se queda por debajo del límite elegido. Cuando volvemos a calcular el número de las diferentes respuestas obtenidas después de haber descartado las que poseían una “regulación factorizada” inferior a 0,3, aparece una mayor homogeneidad entre los datos de las ORFs analizadas en este trabajo en los dos cromosomas (Tabla 4.14.).

Tabla 4.14. Proporciones de las distintas respuestas observadas antes (A.F.: Antes de la factorización) y después (D.F.: Después de la factorización) de haber descartado las que se consideraban cuestionables al haber presentado una “regulación factorizada” inferior al umbral de 0,3. (G+/-: activación o represión en glucosa; N+/-: activación o represión en nitrógeno; S+: activación durante la fase estacionaria; O+: activación por choque osmótico; H+: activación por choque térmico).

Respuestas		CROMOSOMA VII (% respuestas)		CROMOSOMA XIV (% respuestas)	
		A.F.	D.P.	A.F.	D.P.
G	+	21,0	13,2	46,7	17,8
	-	15,8	7,9	6,6	4,4
N	+	5,2	2,6	6,6	2,2
	-	18,4	13,2	4,4	4,4
S	+	5,3	2,6	2,2	2,2
O	+	2,6	2,6	20,0	6,7
H	+	7,9	5,3	8,9	4,4

4.5.4. Regulación y función:

El principal objetivo del análisis transcripcional llevado a cabo en este trabajo como parte del consorcio B2 de EUROFAN consiste en obtener información sobre la posible función de las ORFs analizadas.

Al principio de este estudio, las funciones de la totalidad de las ORFs que iban a ser analizadas se desconocían. Sin embargo, paralelamente al análisis transcripcional llevado a cabo, otros grupos colaboradores en EUROFAN (*e.g.* consorcio B0) generaron información funcional para algunas de las ORFs desconocidas.

De este modo, los datos derivados de nuestro trabajo pueden relacionarse con datos funcionales originados por la observación de los fenotipos provocados por la disrupción de la ORF considerada y con otras funciones descritas en la bibliografía y deducidas por diversas metodologías. Sin embargo, los datos funcionales disponibles son todavía muy escasos (apéndice, Tabla A1) y por tanto, sólo parcialmente aplicables a la discusión de los datos obtenidos en este trabajo y a su relevancia en la posible asignación de funciones.

Las ORFs de expresión constitutiva encontradas en este trabajo resultaron ser abundantes (37,2%). A medida que se han ido conociendo otros datos funcionales para estas ORFs, se ha podido comprobar que algunas de ellas son esenciales para la viabilidad celular o para asegurar unas tasas de crecimiento normales. Su expresión constitutiva estaría por tanto garantizando la presencia de la proteína en diferentes condiciones y asegurando su función esencial para la viabilidad celular. Podemos citar en este grupo, por ejemplo a tres ORFs cuya delección resulta letal y de las que se dispone de datos adicionales sobre su función: G3080 identificada como el gen que codifica la proteína Mlc1p con sitios de unión a calcio, G3175 es el gen *NBP35* que codifica una proteína de unión a nucleótidos y está relacionada con la familia de las ATPasas bacterianas (Vitale *et al.*, 1996) y N1879 ha sido identificada

como el gen *SPC98* que codifica un componente del cuerpo polar del huso mitótico (Geissler *et al.*, 1996).

Otras ORFs cuya delección causa letalidad, N1954, N1888 y N1925 resultaron presentar algún tipo de regulación relacionado con el choque térmico, lo que indica que las proteínas que codifican, no sólo son esenciales para la adaptación a una situación de estrés, sino que desempeñan papeles importantes en las condiciones de crecimiento normal de la célula.

La presencia de glucosa en el medio induce la expresión de una serie de genes de levadura relacionados con el transporte de monosacáridos o su utilización. Entre las ORFs analizadas en este trabajo y que presentaron una clara inducción por glucosa, no se encuentra ninguna que por los datos de su secuencia o estructura pueda relacionarse con enzimas implicadas en el transporte y/o metabolismo de azúcares. Sin embargo se incluyen ORFs relacionadas con otras funciones. Por ejemplo, G3113 posee sitios de unión a ATP/GTP y podría formar parte de rutas de transmisión de señales. N2215 presenta similitudes con la proteína Ecm3p implicada en la biogénesis de la pared celular, asociada a procesos de división y crecimiento celular especialmente activos en medios ricos en nutrientes. G7538 (*HGHI*) está relacionada con isozimas que participan en la biosíntesis de esteroides.

Los genes reprimidos por glucosa se relacionan con funciones que son necesarias para la respiración celular: enzimas del ciclo de Krebs y genes que codifican para proteínas mitocondriales. En este trabajo sólo se han identificado dos ORFs reprimidas por glucosa, N1847 y N2190 para las que no existen otros datos funcionales adicionales que nos permitan interpretar la respuesta observada.

Respecto a las ORFs inducidas por la presencia de fuente nitrogenada, se han caracterizado tres. Para dos de ellas, G7594 y N2254, no existen datos funcionales procedentes de otros análisis, pero en el caso de la ORF G2882 se ha podido demostrar que las líneas que contienen el gen interrumpido son incapaces de crecer sobre medios carentes de aminoácidos y que el gen codifica para una enzima implicada en la biosíntesis de metionina (Tizón *et al.* 1996; Tizón *et al.* 1998, enviado).

Las dos ORFs que son inducidas por choque osmótico, N1882 y N2185, presentan además otra característica común, la presencia de 10 segmentos transmembrana, lo que indica que puede tratarse de genes que codifican para transportadores de membrana. Este hecho tiene una confirmación adicional en el caso de la ORF N1882 que ha sido caracterizado como el gen *ESBP6* (De Antoni *et al.*, 1997) homólogo a un transportador desconocido ligado al cromosoma X humano.

Otro caso, en el que el análisis de otros laboratorios revela una coincidencia con la función aportada por la regulación encontrada en este trabajo, es el de la ORF N1825 que se induce en ausencia de nitrógeno y que ha sido caracterizada como el gen *AAH1* que codifica para la enzima adenosina desaminasa que participa en la degradación de las bases nitrogenadas del grupo de las purinas.

A medida que se vayan caracterizando y haciendo públicas otras funciones para estas ORFs, será posible encontrar una justificación más precisa sobre el tipo de regulación encontrada. De momento los resultados obtenidos permiten dirigir algunas de estas ORFs hacia grupos de investigación especializados e interesados en su análisis.

4.6. Búsqueda de sitios de unión a factores transcripcionales en el promotor de las ORFs analizadas:

La regulación transcripcional está mediada por proteínas reguladoras que reconocen secuencias situadas en el promotor. Con el objetivo de obtener información adicional sobre los mecanismos de regulación existentes en las ORFs analizadas, nos planteamos un estudio de la secuencia de sus promotores.

La búsqueda de sitios de reconocimiento para reguladores transcripcionales se llevó a cabo mediante dos programas, *MatInspector 2.1* (Quandt *et al.*, 1995) y *Pattern Search 1.1* (GBF-Braunschweig), asociados a la base de datos TRANSFAC – *The transcription factor database* (Wingender *et al.*, 1997) de acceso “<http://transfac.gbf.de>” en una secuencia de 800 pb en posición 5’ del ATG de inicio de cada ORF analizada.

En un primer paso se recopilaron todos los factores que habían sido encontrados por estos dos programas. Teniendo en cuenta que *MatInspector 2.1* y *Pattern Search 1.1* pueden encontrar los sitios de unión, respectivamente para 12 y 75 factores transcripcionales diferentes en *S. cerevisiae* (ver Materiales y Métodos), para simplificar el estudio de los promotores se decidió considerar únicamente los factores directamente relacionados con una de las respuestas observadas en el análisis experimental.

Esta selección nos permitió restringir el análisis y considerar sólo los sitios de unión para ocho factores de regulación cuya función había sido ampliamente estudiada (Tabla 4.15.).

Tabla 4.15. Recopilación de las secuencias consenso para los factores de regulación transcripcional en levadura utilizados en el análisis de los promotores de las 103 ORFs desconocidas de *S. cerevisiae*. (en las secuencias; R = A o G, W = A o T, N = A, T, C o G, S = C o G, Y = C o T, M = A o C y a = A, G o C). ND: No Determinado; F.E.: Fase Estacionaria; E. O. y T.: Estrés Osmótico y térmico.

REGULACIÓN	FACTOR ASOCIADO	Secuencia consenso de unión	Referencia
MATING	Mato2p Mcm1p	CATGTAATT CTTCCTAATTAGGAAG	Johnson y Herskowitz (1985) Kéléher <i>et al</i> (1988)
GLUCOSA	Mig1p Gcr1p	SYGGGG + zona rica en A-T flanqueante GMWTCCW	Lundin <i>et al.</i> , 1994 Svetlov y Cooper, 1995
PRIVACIÓN DE AMINOÁCIDOS	Gcn4p	ATGASTCAT	Svetlov y Cooper, 1995
REGULACIÓN POR NITRÓGENO	Gln3p Dal80p Gat1p Deh1p	GATAA (<i>UAS</i> _{NTR}) GATAA (<i>URS</i> _{GATA}) GATAA GATAA	Blinder y Magasanik, 1995 Cunningham y Cooper, 1993 Coffman <i>et al.</i> , 1996 Coffman <i>et al.</i> , 1997
F. E. + E. O. y T.	Msn2p Msn4p	Sitio <i>STRE</i> AGGGG	Schüller <i>et al.</i> , 1994 Martinez-Pastor <i>et al.</i> , 1996
CHOQUE TÉRMICO	Hsp	(NTTCNNGAAN) ₂ > (NGAANNITCN) ₂	Svetlov y Cooper, 1995

Los sitios *STRE* han sido intensamente estudiados y son conocidos por estar relacionados con respuestas a distintos tipos de estrés (ver apartados 1.4.4. a 1.4.6.).

Los sitios de unión para los factores GATA, no aparecieron recopilados en la base de datos de TRANSFAC. Sin embargo se decidió incluirlos por su relación con la regulación por nitrógeno. Están formados al menos por dos secuencias (T/A)GAT(T/A)(A/G) próximas (separadas por un máximo de 40 pb). La búsqueda de estos sitios fue realizada mediante alineamiento del promotor con la secuencia consenso.

La búsqueda correspondiente a los sitios de unión al factor Mat α 2p tenía como objetivo relacionar la represión que este factor provoca sobre los genes a-específicos con las ORFs analizadas en este trabajo que no habían podido ser detectadas o que presentaban un nivel absoluto de expresión bajo (ver apartados 4.3.1.1. y 4.3.1.4.). Para el resto de los factores estudiados no se tuvieron en cuenta, ni las ORFs que no pudieron ser detectadas ni las que presentaron múltiples transcritos ya que, para estas ORFs era imposible relacionar una respuesta con un eventual sitio de unión a un factor en su promotor.

Para poder rechazar sitios de unión poco significativos se procedió a una modelización individual de estos sitios para cada factor. Para ello se recopilamos de la bibliografía los genes para los cuales ha sido localizada y/o demostrada la unión de dicho factor en sus respectivos promotores. También se determinaron las localizaciones de estos sitios en los promotores de los genes que están regulados por alguno de los factores seleccionados pero sobre los que se carece de referencia bibliográfica sobre su situación. Esta información permite definir una zona preferencial del promotor donde estarían situados los sitios de unión a un factor determinado en la mayoría de los genes regulados por este factor. El umbral para delimitar esta zona fue establecido de forma que incluyese el 70% de los genes que poseían estos sitios (Tabla 4.16.).

Tabla 4.16. Recopilación de las localizaciones de los sitios de unión para los factores de regulación transcripcional de genes funcionalmente conocidos de *S. cerevisiae*. El área sombreada señala la zona preferencial elegida para la unión del factor considerado. E.C.: Estimador de confianza.

FACTOR	GEN	REFERENCIA	E.C.	POSICIÓN	800	700	600	500	400	300	200	100	A	T	G	
					[Gráfico de barras de unión con sombreado preferencial]											
MIG1	FPS1	Klein <i>et al.</i> , 1996	2	-227												
	GAL1	Nehlin <i>et al.</i> , 1991	2	-205												
	GAL4		2	-78												
	HXT2	Ozcan y Johnston, 1996	3	-778, -676, -631, -432												
	SNF3	Walsh <i>et al.</i> , 1996	3	-326, -250												
	SUC2	Lundin <i>et al.</i> , 1994	2	-119, -66												
GCR1	ENO2	Wilett <i>et al.</i> , 1993	2	-420												
	ENO1	Huie <i>et al.</i> , 1992	2	-491												
	TPI1		2	-377, -366												
	PGK1		2	-459												
	ADH1		2	-694												
	PYK1		2	-649												
GCN4	TRP4		Moesch <i>et al.</i> , 1990	2	-252, -140											
	ADE4	Schmidheini <i>et al.</i> , 1990	1	-358, -313, -276												
	ARO7		2	-503												
	HIS4		2	-259, -203, -178, -151												
	ILV1	Arndt y Fink, 1986	2	-128, -51, -18												
	ILV2		2	-359												
	HIS3		2	-337, -287, -254, -209, -126												
			Hope y Struhl, 1985	2	-102											
	LEU4	Beltzer <i>et al.</i> , 1986	4	-103, -320, -364, -427, -717												
ARG4	Beacham <i>et al.</i> , 1984	4	-181, -221, -336, -779													
FACTORES GATA	GLN1	Minehart y Magasanik, 1992	1	-659 / 6+5												
	GDH2	Miller y Magasanik, 1991	1	-411 / 4+6												
	DAL5	Rai <i>et al.</i> , 1989	1	-170 / 9+7												
	DAL7	Yoo y Cooper, 1989	1	-166 / 8+5												
	DEH1	Coffman <i>et al.</i> , 1997	4	-141 / 8+4												
	DAL80		2	-307 / 9+4												
	GAP1		2	-343 / 11+7												
	UGA4		2	-196, -389, -437 / 6+3												
HSF	SSA3	Silar <i>et al.</i> , 1991	4	-236												
	HSP83	Wiederrecht <i>et al.</i> , 1987	2	-211												
	CUP1	Yang <i>et al.</i> , 1991	2	-175												
		Huibregtse <i>et al.</i> , 1989	2	-146												
	HSP104	Sanchez y Lindquist, 1990	4	-297												
	HSP90	McDaniel <i>et al.</i> , 1989	4	-230, -112												
	SSA1	Elwood y Craig, 1984	4	-353												
	SSA2		4	-226												
	ENO1		4	-536												

Tabla 4.16. Recopilación de las localizaciones de los sitios de unión para los factores de regulación transcripcional de genes funcionalmente conocidos de *S. cerevisiae*. El área sombreada señala la zona preferencial elegida para la unión del factor considerado. E.C.: Estimador de confianza (Continuación).

FACTOR	GEN	REFERENCIA	E.C.	POSICIÓN	800	700	600	500	400	300	200	100	A T G
STRE	<i>SOD2</i>	Flattery O'Brien <i>et al.</i> , 1997	4	-178, -138									
	<i>TPS2</i>	Mager y De Kruijff, 1995	1	-523,-490,-441,-421,-308									
	<i>DDR2</i>		1	-472-248-203-175									
	<i>CCT1</i>		1	-330,-345,-100									
	<i>SSA1</i>		4	-211,-160									
	<i>ENA1</i>		4	-651									
	<i>HSP26</i>		4	-659, -484, -466, -328									
	<i>HSP104</i>		4	-252, -220, -172									
	<i>UBI4</i>		4	-655, -252									
	<i>HAL1</i>		4	-399									

Cada gen está acompañado de un estimador de la confianza que puede presentar en cuanto a la validez funcional del sitio de unión para un factor transcripcional determinado; éste toma los valores siguientes:

- 1- Para los sitios de unión a un factor confirmados funcionalmente (principalmente por *Foot-printing*).
- 2- Para los sitios que unen factores transcripcionales en experimentos de unión DNA-proteína (*Binding*).
- 3- Para los genes cuya respuesta fisiológica varía cuando se les muta el sitio de unión al factor transcripcional implicado en dicha respuesta.
- 4- Para los genes que muestran, a la vez, una respuesta fisiológica determinada y secuencias consenso para la unión de un factor transcripcional relacionado con esa respuesta.

Para los sitios de unión a los factores GATA, se indican en la tabla la localización del principio de la doble secuencia (T/A)GAT(T/A)(T/G), así

como el número de secuencias estabilizadoras de la unión de estos factores y que están formadas tanto por la secuencia (T/A)GAT(T/A)(T/G) como por (T/A)GAT respectivamente.

Para analizar los promotores de las ORFs que presentaron algún tipo de regulación en este trabajo, se estableció un valor “estimador” de la calidad de cada sitio potencial, resultado del producto entre la valoración de la idoneidad de su localización en el promotor y de su homología con una hipotética secuencia consenso de unión al factor. La calificación o valoración de estos sitios potenciales se realizó mediante una factorización basada en los criterios que se recogen en la tabla 4.17.

Tabla 4.17. Método de valoración de los sitios de unión a factores transcripcionales encontrados en los promotores de las ORFs analizadas.

	FACTOR		
	1,00	0.75	0.50
Ubicación dentro del promotor	Coincide dentro de la ventana delimitada para los genes del grupo modelo	Está fuera de la ventana pero dentro de un margen de +/- 50pb	Está alejada en más de 50 pb de la ventana.
Homología con la secuencia consenso	Homología > 90%	90% > Homología > 80%	Homología < 80%

En el caso de los factores GATA se valoraron también las secuencias esparcidas en el promotor y que estabilizan la unión del factor a su sitio; para cada secuencia (T/A)GAT(T/A)(A/G) encontrada se le fijó un factor arbitrario de 0,2, mientras que para las secuencias (T/A)GAT, de menor “peso”, de 0,1.

Después de haber hecho los cálculos, se desestimaron los sitios potenciales cuyo producto entre homología y localización era inferior a un umbral correspondiente a un sitio teórico fiable para cada factor considerado y que se estimó, teniendo en cuenta las referencias a genes conocidos, con el criterio siguiente:

- Para Gcr1p, Mig1p y Hsfp se valoró en 0,75, correspondiendo a un sitio dentro de su ventana y con una homología superior al 80% o a un sitio cercano de esta ventana pero con una homología superior al 90%.
- Para Gcn4p y los elementos *STRE* el umbral considerado fue de 1,5; éste equivale a por lo menos dos sitios de unión a Gcn4p o a dos elementos *STRE* dentro de su ventana y con una homología superior al 80% o a dos sitios cercanos a esta ventana pero con una homología superior al 90%.
- Finalmente, para los factores GATA, se consideraron fiables los sitios cuyo producto había sido superior a 2,0; equivalente a una doble secuencia (T/A)GAT(T/A)(A/G) en su región preferencial y a por lo menos 5 dispersas en todo el promotor.

Estos umbrales nos permitieron seleccionar para las 86 ORFs de los cromosomas VII y XIV que habían podido ser cuantificadas en este trabajo, sitios de unión potenciales para los diferentes factores transcripcionales estudiados (Tabla 4.18).

Para valorar la calidad de los sitios seleccionados, se efectuó un análisis comparativo de la existencia de sitios, en los promotores de las ORFs que habían dado una respuesta concorde con la regulación atribuible al factor, antes y después de haber rechazado los que no se consideraron adecuados por no superar el umbral asignado a cada factor. Este análisis se hizo considerando las 250 ORFs analizadas por el consorcio B2 en el cromosoma XIV de *S. cerevisiae*, ya que la muestra era mas representativa que si nos limitáramos a las 86 ORFs analizadas en este trabajo (Tabla 4.19.).

Tabla 4.19. Distribución de los sitios de unión a factores reguladores en los promotores de las 250 ORFs del cromosoma XIV de *S. cerevisiae*. G+/-: activación/represión en glucosa; A+: Activación durante la privación de aminoácidos; N+/-: activación/represión en nitrógeno; S+: activación durante la fase estacionaria; O+: activación por choque osmótico; H+: activación por choque térmico; % R+: % de ORFs que presentaron la respuesta relacionada con el factor considerado.

Factor / Respuesta	Nº ORFs que presentan sitios	%R+	Nº ORFs que presentan sitios probables	% R+
Gcr1p/G+	140	27,8	80	22,5
Mig1p/G-	60	10,0	16	25,0
Gcn4p/A+	156	10,3	70	11,4
GATA/N+, N-	199	26,8	57	33,3
Hsfp/H+	74	17,6	44	6,8
STRE/S+, O+, H+	86	29,1	20	20,0

Los resultados obtenidos en el análisis realizado indican que sólo de un 10 a un 30 % de las ORFs que presentan sitios de unión a factores de regulación específicos, muestran después una regulación concordante al variar las condiciones del medio.

Cuando se consideran únicamente aquellos sitios que cumplen las restricciones impuestas respecto a su número y localización en el promotor, el porcentaje de concordancia con el tipo de regulación observado mejora sólo muy levemente y en el caso de algunos factores como Gcr1p, Hsfp o los elementos *STRE*, incluso disminuye. Este resultado puede estar motivado por diversos hechos. Por una parte, se sabe que hay una distribución aleatoria de secuencias coincidentes con sitios de reconocimiento y que contribuiría a la aparición de falsos positivos (Schuldiner *et al.*, 1997). Por otra parte, la estructura de los promotores es muy compleja y la regulación final no es dependiente de un único factor sino de interacciones sinérgicas o antagónicas de varios sucesos reguladores.

En los últimos años se han realizado varias aproximaciones para determinar la función de nuevos genes a partir de secuencias promotoras conocidas por su participación en procesos de regulación. En este sentido cabe destacar el trabajo realizado por Fondrat y Kalogeropoulos en el cromosoma III (Fondrat y Kalogeropoulos, 1996) o el más reciente de Schuldiner *et al.*, sobre la totalidad del genoma (Schuldiner *et al.*, 1997). Al comparar sus resultados con los encontrados en nuestro trabajo, sólo podemos hacer referencia a las conclusiones sobre *GCN4* ya que fue el único analizado por Schuldiner *et al.*, y el único en común respecto al trabajo de Fondrat y Kalogeropoulos.

Estos últimos autores, utilizando condiciones altamente restrictivas para la localización de sitios *GCN4*, identificaron 7 sitios en las regiones correspondientes a ORFs de función desconocida del cromosoma III y que asciende a un número de 76. De estas siete, tres fueron descartadas por coincidir dentro de las regiones codificadoras y otras tres por quedar fuera del

intervalo previsto de la región promotora. El único sitio no descartado en dicho estudio estaba localizado en la posición -145 de la ORF YCL009c que muestra homologías con la acetolactato sintetasa de *E. coli* (gen *ILVH*) implicada en el metabolismo de aminoácidos. A pesar de que estos autores admiten que un solo sitio de unión para Gcn4p puede resultar funcional, sus condiciones respecto a la coincidencia y estructura interna del sitio son mucho más restrictivas que las planteadas en nuestro trabajo, como queda reflejado por el hecho que hemos llegado a detectar 70 ORFs con sitios potenciales de unión a Gcn4p en un total de 250 ORFs analizadas en el cromosoma XIV. Esto puede explicar el mayor número de falsos positivos que hemos hallado al comparar estas ORFs con su regulación.

Sin embargo, si las condiciones utilizadas para considerar un sitio como significativo son muy restrictivas se corre el riesgo de no tomar en cuenta algunas ORFs que si están reguladas. Por ejemplo en el trabajo realizado por Sculdiner *et al.* (1998), sólo señalan como posiblemente reguladas por *GCN4* dos ORFs del cromosoma VII (YGL060w e YGL109w) y una ORF del cromosoma XIV (YNL091w). Otras ORFs descartadas en el citado estudio muestran sin embargo una activación de su expresión en ausencia de aminoácidos: G2861 (YGL128c), N1773 (YNL150w), N1847 (YNL134c) y N2160 (YNL106c) y fueron consideradas como potencialmente reguladas por *GCN4* con los criterios menos restrictivos de nuestro estudio. Incluso con nuestros criterios, han quedado fuera algunas ORFs que hemos podido comprobar participan en el metabolismo de aminoácidos y están reguladas por la concentración de estos en el medio; es el caso de G2882 (YNL125w) que codifica para una metilen-tetrahidrofolato reductasa que participa en la biosíntesis de metionina y contiene dos sitios de unión a Gcn4p en las posiciones -255 y 386 de su promotor (Tizón, Tesis de Doctorado, 1997) pero

cuya homología y posición no superaron los criterios de selección predeterminados.

Resulta evidente que el análisis “*in silico*” de promotores puede resultar una herramienta eficaz para la asignación de funciones a ORFs desconocidas, sin embargo será necesario automatizar y complicar los algoritmos de búsqueda a fin de que tengan en cuenta, no sólo la homología de la secuencia respecto a un consenso y su posición en la secuencia lineal del promotor, sino también su localización en la estructura tridimensional del promotor, su interferencia con sitios de posicionamiento de nucleosomas o interrelaciones multi-factoriales.

4.7. Regulación de la expresión génica en mutantes *hap1*, *rox1* y *rox6*:

Dado el interés de nuestro laboratorio en el estudio de la expresión de genes aerobios e hipóxicos en la levadura *S. cerevisiae*, decidimos aprovechar las circunstancias de este estudio sistemático de ORFs de función desconocida para localizar algunos genes implicados en estas vías. Contabamos con la disponibilidad de cepas mutadas en diversos genes reguladores.

Rencientemente se ha demostrado (Bécam *et al.*, 1998) que una mutación natural afectaba el gen regulador *HAP1* en algunas cepas de *S. cerevisiae* utilizadas para el análisis funcional. El análisis de su secuencia nucleotídica mostró que esta mutación consiste en la inserción de un elemento Ty modificado en la región 3' de *HAP1* en la cepa FY1679 que deriva de la cepa FY73 utilizada en este trabajo. Esta inserción podría afectar la transcripción de *HAP1* debido a la formación de un transcrito híbrido inestable, pero también a la actividad de la proteína resultante al modificar la región

C-terminal de Hap1p dónde se cree está el dominio de activación (Bécam *et al.*, 1998). Con el objetivo de comprobar la idoneidad para esta búsqueda de genes del circuito aeróbico e hipóxico tanto de las tres cepas mutadas utilizadas como de la cepa FY73 que nos sirve de control, se hibridaron los RNAs extraídos en las distintas condiciones ensayadas en este trabajo con sondas específicas de los genes *HAP1*, *CYC1* y *ROX1* (Figura 4.12.).

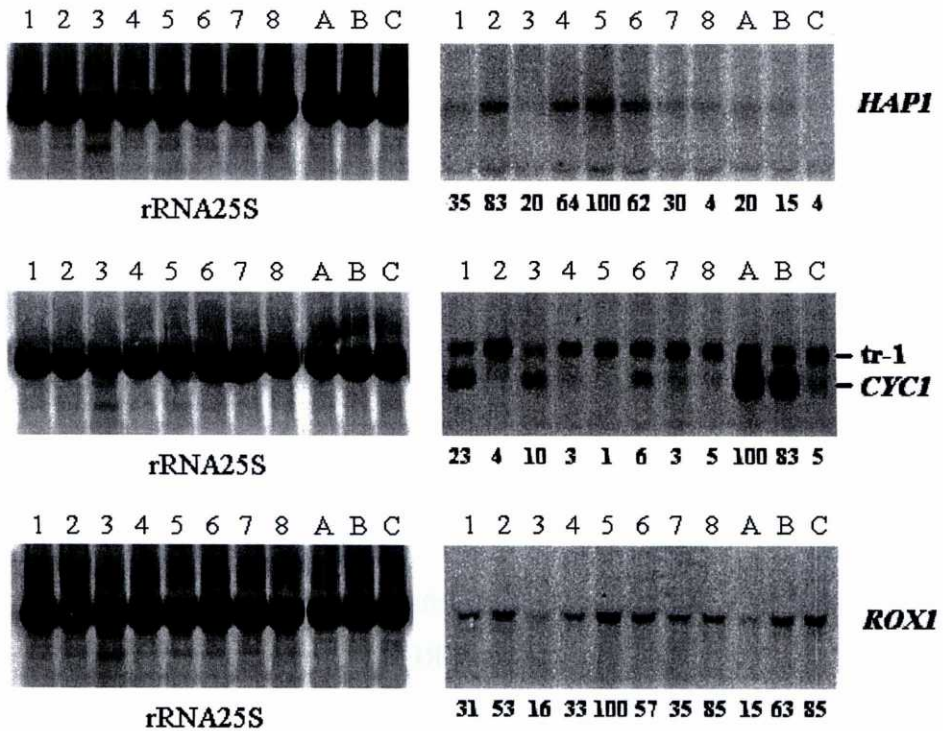


Figura 4.12. Valores de Nivel relativo (N.R.) correspondientes a las hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo (1-YPGE, 2-YPD, 3-YPD/fase estacionaria, 4-YNBglu/privación de aminoácidos, 5-YNBglu/privación de nitrógeno, 6-YNBglu/choque osmótico, 7-YNBglu/23°C, 8-YNBglu/36°C A-YPGE/*Sc.rox1*, B-YPGE/*Sc.rox6*, C-YPGE/*Sc.hap1*) con sondas correspondientes a los genes *HAP1*, *CYC1* y *ROX1*.

En esta figura podemos ver que, como se esperaba, los genes *HAP1* y *ROX1* no se expresan en las cepas mutadas para estos genes (condiciones C y A respectivamente). Sin embargo y contrariamente a las observaciones de Bécam *et al.* (1998), en la cepa FY73 el gen *HAP1* se transcribe en todas las condiciones estudiadas (1 a 8) sin mostrar un transcrito de tamaño superior (la inserción del transposon en la secuencia codificadora supone un incremento en el tamaño de 32 pb); por otro lado, la inducción de *CYC1* en condiciones aeróbicas (condición 1) muestra que tampoco parece haber sido anulada su actividad reguladora a pesar de que esta inducción sea notablemente inferior a las observadas en las condiciones A y B donde *HAP1* es completamente funcional, ya que en estas líneas no hay inserción de transposón.

Podemos concluir que la mutación provocada por la inserción de un transposón en *HAP1* y observada por Bécam *et al.* (1998) pudo haber sido posterior a la distribución de las cepas utilizadas en el análisis funcional y aún en el caso de encontrarse presente, sólo afecta al papel funcional de Hap1p en un pequeño grado que nos permite utilizarla como referencia comparativa en el análisis de la expresión de las ORFs en la cepa donde *HAP1* está completamente anulado. Cabe también señalar que la mutación *hap1* utilizada (condición C) no afecta a la expresión de *ROX1*.

La mutación *hap1* causa una disminución de la expresión de diversos genes relacionados con la respiración como *CYC1*, ya que *HAP1* codifica un activador de genes aeróbicos. Esperábamos encontrar en un fondo genético *hap1* una disminución de la expresión de genes que pudiesen estar relacionados con las funciones respiratorias. De las 103 ORFs, tres de ellas, N1747, N1765 y N2250 presentan este tipo de respuesta (Tabla 4.19.) y su función podría relacionarse con genes necesarios para la adaptación de la

célula a condiciones aerobias. En los tres casos se trata de proteínas de función desconocida que no han sido caracterizadas previamente por lo que su estudio en relación con los procesos de aerobiosis/anaerobiosis podría resultar interesante para establecer su función.

Las mutaciones *rox1* y *rox6* han sido caracterizadas porque están asociadas a una sobre-expresión de genes anaerobios o hipóxicos en condiciones aerobias. *Rox1p* es un represor de genes hipóxicos en condiciones aerobias (Lowry y Zitomer, 1984) y la mutación *rox6* es complementada por el gen *SRB10* que codifica un componente del par ciclina-quinasa responsable de la fosforilación del extremo carboxilo-terminal de la RNA polimerasa II (Cadahia, Tesis de licenciatura, 1996). Esperábamos que aquellas ORFs que estuviesen directamente o indirectamente implicadas en las vías metabólicas en condiciones anaerobias (ausencia de oxígeno) y/o hipóxicas (bajos niveles de oxígeno) viesen su expresión incrementada en uno o ambos de estos fondos genéticos. Como se puede observar en la tabla 4.19., solamente dos ORFs, G3168 y N1929, presentan una activación de su expresión tanto en los mutantes *rox1* como *rox6*. También en este caso se trata de proteínas para las que no existía ni una previa comprobación sobre su funcionalidad ni una posible hipótesis sobre su función, por lo que un futuro análisis en esta dirección de trabajo puede resultar interesante.

Son mas numerosas, ocho, las ORFs cuya expresión se ve incrementada sólo en el mutante *rox6* (G3116, G3690, G3759, G7587, G7725, G7806, N1710 y N1888). Esto no indica necesariamente que todas estas ORFs estén implicadas en las condiciones de crecimiento anaerobio o hipóxico dado el efecto pleiotrópico de la mutación *rox6*. *SRB10* (=UME5/ARE1/SSN3/GIG2) ha sido clonado por diferentes laboratorios partiendo de fenotipos diversos

(Surosky *et al.*, 1994; Wahi y Johnson, 1995; Kuchin *et al.*, 1995; Liao *et al.*, 1995; Balciunas y Ronne, 1995). El efecto pleiotrópico asociado a la mutación *rox6* puede ser interpretado teniendo en cuenta que una serie de factores transcripcionales relacionados con diferentes funciones de la célula pueden transmitir su efecto a la RNA polimerasa II a través de la proteína Sbr10p directamente o por fosforilación de otro componente del complejo mediador. Si Sbr10p participa en la transmisión de señales represoras, estas 8 ORFs podrían estar sometidas a la acción de reguladores represores específicos no relacionados que ven interrumpida la transmisión de la señal por la mutación de la quinasa Sbr10p. Aunque también es posible pensar que Sbr10p pueda ser un transmisor de señales reguladoras positivas, resulta significativo el hecho de que sólo una ORF, G7553, vea disminuida su expresión en el mutante.

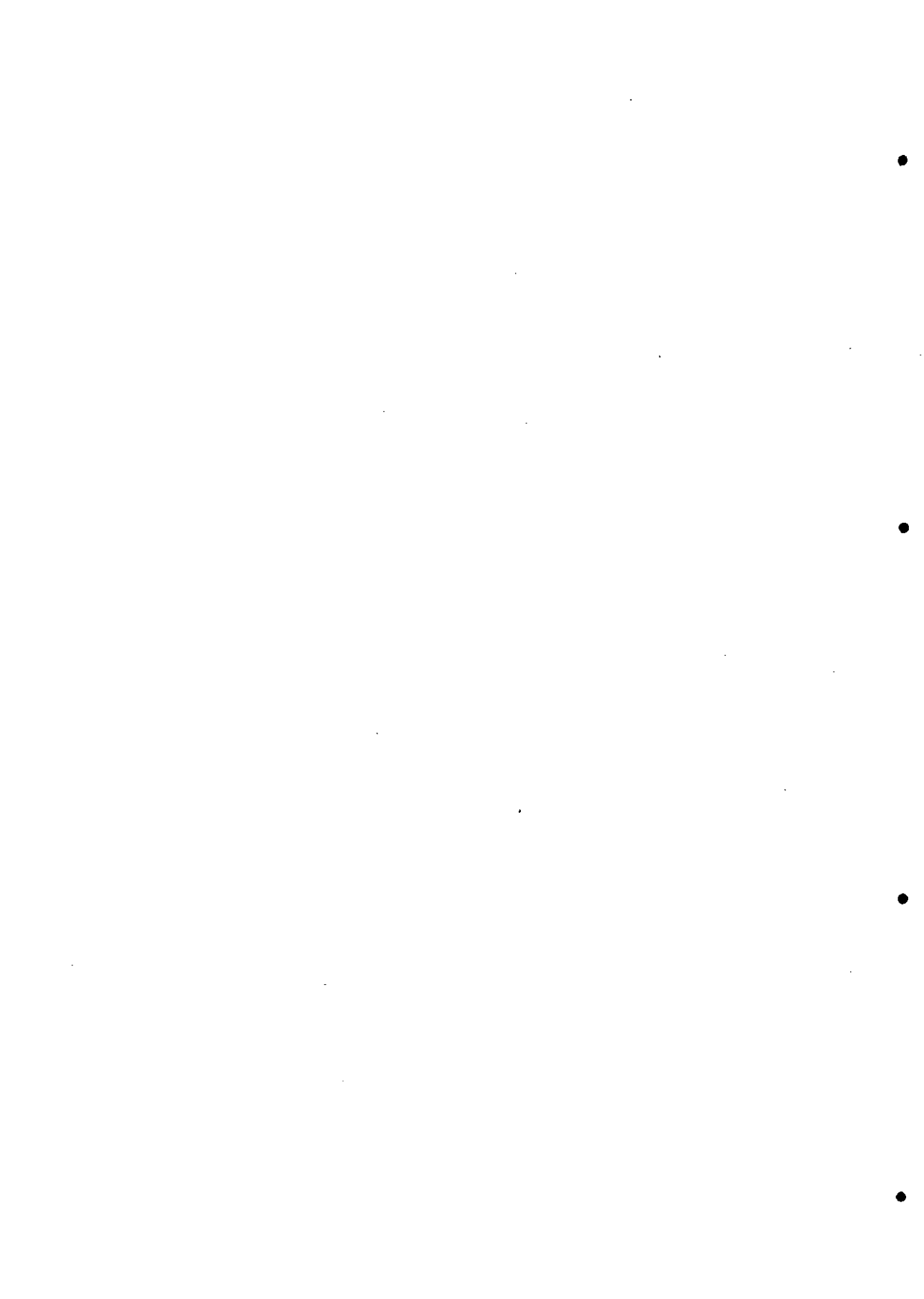
Tabla 4.20. Valores de Nivel relativo (N.R.) correspondientes a las hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo (1-YPGE, A-YPGE/*Sc.rox1*, B-YPGE/*Sc.rox6*, C-YPGE/*Sc.hap1*) con sondas específicas de 51 ORFs de función desconocida en el cromosoma VII de *S. cerevisiae*. (<1: no detectado, MT: Múltiples transcritos).

ORF	N.R. (%)				ORF	N.R. (%)			
	1	A	B	C		1	A	B	C
G2850	<1	<1	<1	<1	G3505	<1	<1	<1	<1
G2856	56	52	100	40	G3517	72	71	100	81
G2861	58	51	100	71	G3535	MT	MT	MT	MT
G2882	46	78	79	100	G3540	MT	MT	MT	MT
G2889	MT	MT	MT	MT	G3679	<1	<1	<1	<1
G2913	72	66	74	100	G3690	36	43	73	100
G2916	53	39	60	100	G3719	<1	<1	<1	<1
G3080	94	69	100	92	G3759	36	39	80	100
G3090	<1	<1	<1	<1	G4332	89	82	82	100
G3107	78	100	76	81	G7538	100	53	72	88
G3110	84	66	100	59	G7553	100	51	49	64
G3113	91	96	66	100	G7584	75	100	83	86
G3116	46	59	100	60	G7587	46	36	100	74
G3146	74	100	41	55	G7589	67	36	60	100
G3168	37	92	100	63	G7594	100	63	67	74
G3175	67	76	100	72	G7725	50	50	100	65
G3179	58	88	100	70	G7737	100	89	98	96
G3189	MT	MT	MT	MT	G7748	61	61	87	100
G3193	MT	MT	MT	MT	G7801	60	75	100	78
G3210	100	84	86	77	G7806	31	38	100	71
G3406	MT	MT	MT	MT	G7826	MT	MT	MT	MT
G3432	<1	<1	<1	<1	G8523	58	46	81	100
G3435	MT	MT	MT	MT	G8530	78	100	90	89
G3474	65	65	100	100	G8537	80	100	99	92
G3483	100	81	99	96	G8541	MT	MT	MT	MT
G3489	100	89	90	96					

Tabla 4.20. Valores de Nivel relativo (N.R.) correspondientes a las hibridaciones de los RNAs extraídos en diferentes condiciones de cultivo (I-YPGE, A-YPGE/*Sc.rox1*, B-YPGE/*Sc.rox6*, C-YPGE/*Sc.hap1*) con sondas específicas de 52 ORFs de función desconocida en el cromosoma XIV de *S. cerevisiae*. (<1: no detectado, MT: Múltiples transcritos) (continuación).

ORF	N.R. (%)				ORF	N.R. (%)			
	I	A	B	C		I	A	B	C
N1696	52	67	82	100	N1872	<1	<1	<1	<1
N1706	74	56	72	100	N1875	76	64	89	100
N1710	35	68	100	81	N1879	81	56	76	100
N1714	<1	<1	<1	<1	N1882	72	66	100	65
N1718	<1	<1	<1	<1	N1888	39	48	100	66
N1727	83	57	100	100	N1897	30	31	59	100
N1735	<1	<1	<1	<1	N1913	92	65	100	80
N1739	68	66	100	96	N1917	90	85	87	100
N1743	MT	MT	MT	MT	N1925	82	75	88	100
N1747	82	100	57	38	N1929	42	100	100	74
N1751	56	64	88	100	N1954	58	64	78	100
N1761	72	63	100	92	N1958	37	41	60	100
N1765	100	65	78	44	N1962	60	68	100	63
N1773	44	46	61	100	N1966	<1	<1	<1	<1
N1774	60	61	73	100	N2160	<1	<1	<1	<1
N1777	MT	MT	MT	MT	N2185	53	63	74	100
N1780	66	75	62	100	N2190	97	89	100	100
N1785	<1	<1	<1	<1	N2194	100	58	83	87
N1790	100	100	91	73	N2205	43	53	76	100
N1795	<1	<1	<1	<1	N2215	<1	<1	<1	<1
N1825	44	38	63	100	N2219	73	60	91	100
N1843	81	60	81	100	N2227	<1	<1	<1	<1
N1847	51	56	70	100	N2231	78	68	78	100
N1850	68	66	100	89	N2250	79	90	100	54
N1858	100	48	68	72	N2254	63	74	100	90
N1870	<1	<1	<1	<1	N2312	59	66	100	90

5. CONCLUSIONES



Las conclusiones obtenidas al término de este trabajo se exponen a continuación:

- 1- De las 103 ORFs analizadas, ha sido posible caracterizar un transcrito específico para 86, doce mostraron múltiples transcritos y cinco de ellas quedaron por debajo del límite de detección de la técnica. Entre estas 5 últimas, se encuentran tres, G2850, G3719 y N1872, cuya función génica ha quedado demostrada en otros análisis. Este resultado corrobora que un 98 % o más de las ORFs determinadas por análisis "*in silico*" se corresponden con genes funcionales.
- 2- Los niveles absolutos de expresión indican que un 3,9 % de las ORFs son de alta expresión, 33 % de expresión media y 46,6 % de baja expresión.

Los bajos niveles de CAI calculados para estas ORFs se corresponden con el predominio de las ORFs de baja y mediana expresión.

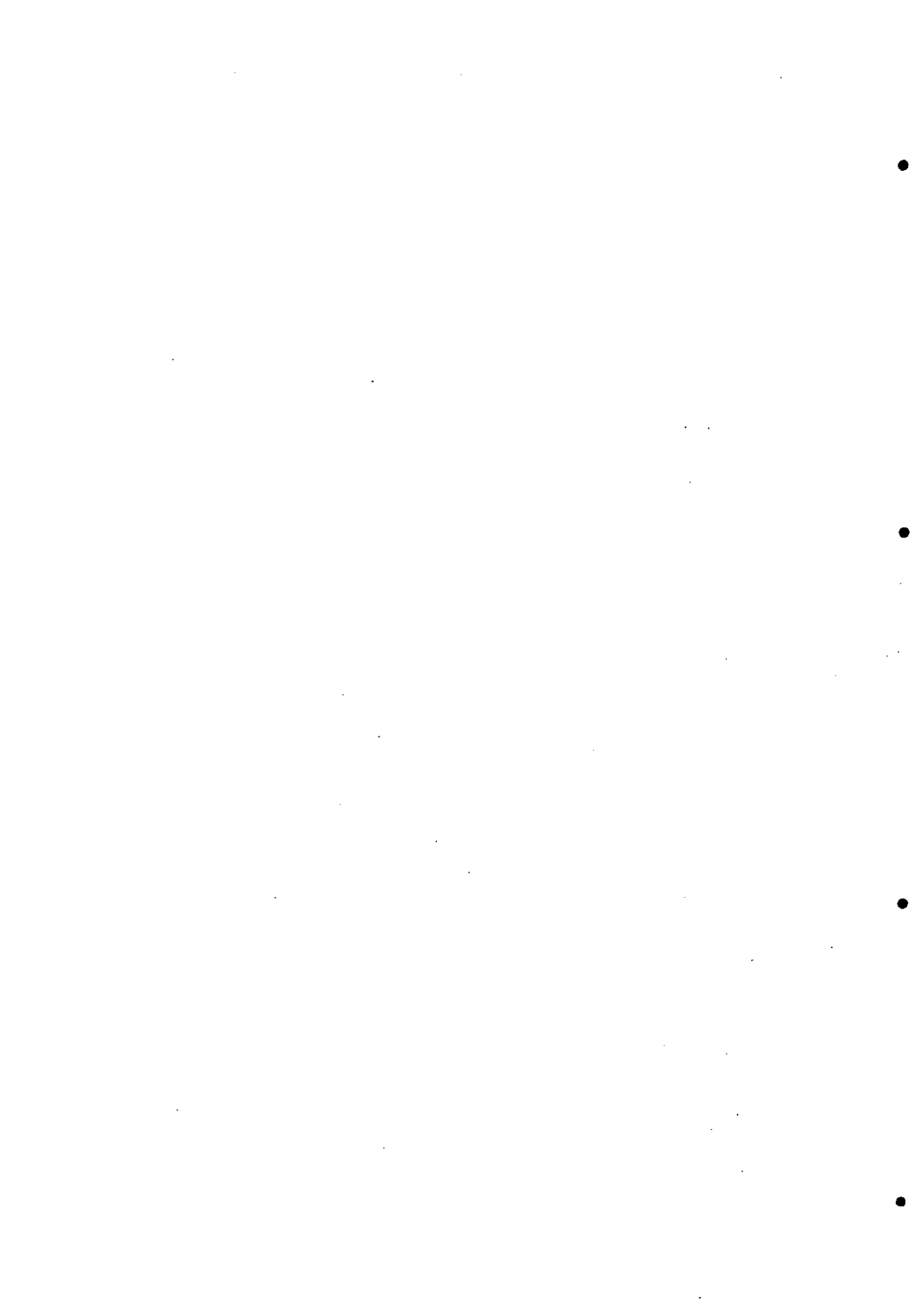
Los niveles absolutos de expresión son independientes de la hebra analizada y de la posición en el cromosoma, aunque se observa un predominio de ORFs no detectadas en las regiones de elevada densidad génica cuando se analizan todos los datos disponibles del cromosoma XIV.

- 3- Respecto a la regulación a nivel transcripcional, el 45 % de las ORFs analizadas en el cromosoma VII y el 30 % de las del cromosoma XIV resultaron ser constitutivas. La regulación observada para el resto de las ORFs se distribuyó entre los diferentes tipos de regulación (inducción/represión en glucosa, inducción/represión en nitrógeno, inducción en la fase estacionaria, durante los choques osmótico y térmico)

siendo el grupo mayoritario las que presentan inducción en glucosa. Un porcentaje pequeño, 16 %, presentó respuestas a varios tipos de regulación. En general la respuesta reguladora fue pequeña con un factor de inducción ó represión cercano a 2 y sólo un 18,5 % mostró respuestas cuantificables por un factor superior a 5. Respecto a la regulación en mutantes relacionados con la adaptación a condiciones aerobias ó anaerobias, el porcentaje de ORFs implicadas es pequeño, 6 %, lo que coincide con la extensión de este regulón en levaduras.

- 4- La búsqueda de sitios de unión a factores reguladores de la transcripción mediante análisis "*in silico*" del promotor no resulta adecuada para la asignación de funciones cuando se utiliza como herramienta única porque si se utilizan criterios muy restrictivos se descartan procesos funcionales, y si se hacen menos restrictivos aparecen un elevado número de falsos positivos. Sin embargo, utilizada conjuntamente con el análisis transcripcional resulta útil para la asignación de funciones.

6. BIBLIOGRAFÍA



- Adams, B.G. (1972). Induction of galactokinase in *Saccharomyces cerevisiae*: kinetics of induction and glucose effects. *J. Bacteriol.*, **111**: 308-315.
- Arndt, K. & Fink, G.R. (1986). GCN4 protein, a positive transcription factor in yeast, binds general control promoters at all 5' TGACTC 3' sequences. *Proc. Natl. Acad. Sci. USA*, **83**: 8516-8520.
- Arroyo, J., García-Gonzalez, M., García-Saez, M.I., Sanchez, M & Nombela, C. (1997). DNA sequence analysis of a 23,002 bp DNA fragment of the right arm of *Saccharomyces cerevisiae* chromosome VII. *Yeast*, **13**: 357-363.
- Baker, H.V. (1986). Glycolytic gene expression in *Saccharomyces cerevisiae*: Nucleotide sequence of *GCR1*, null mutants, and evidence for expression. *Mol. Cell. Biol.*, **6**: 3774-3784.
- Balasubramanian *et al.*, (1993). The Rox1 repressor of the *Saccharomyces cerevisiae* hypoxic genes is a specific DNA-binding protein with a high-mobility-group motif. *Mol. Cell. Biol.*, **13**: 6071-6078.
- Balciunas, D. & Ronne, H. (1995). Three subunits of the RNA polymerase II mediator complex are involved in glucose repression. *Nucleic Acids Res.*, **21**: 4421-4425.
- Barry, C., Fichant, G., Kalogeropoulos, A. & Quentin, Y. (1996). A computer filtering method to drive out tiny genes from the yeast genome. *Yeast*, **12**: 1163-1178.
- Beacham, I.R., Schweitzer, B.W., Warrick, H.M. & Carbon, J. (1984). The nucleotide sequence of the yeast *ARG4* gene. *Gene*, **29**: 271-279.
- Becam, A.-M, Verdière, J., & Herbert C.J. (1998). A "natural" mutation in some *Saccharomyces cerevisiae* strains used for functional analysis affects the complex regulatory gene *HAPI* (*CYP1*). EUROFAN'98: The third meeting. Noordwijkerhout, The Netherlands. May 16th-19th, 1998. Biotech programme.
- Beltzer, J.P., Chang, L.-F.L., Hinkannen, A.E. & Kohlhaw, G.B. (1986). Structure of yeast *LEU4*. The 5' flanking region contains features that predict two modes of control and two productive translation starts. *J. Biol. Chem.*, **261**: 5160-5167.
- Benetzen, J.L. & Hall, B.D. (1982). Codon selection in yeast. *J. Biol. Chem.*, **257**: 3026-3031.
- Bettany, A.J., Moore, P.A., Cafferkey R., Bell L.D., Goodey, A.R., Carter, B.L. & Brown, A.J. (1989). 5'-secondary structure formation, in contrast to a short string of non-preferred codons, inhibits the translation of the pyruvate kinase mRNA in yeast. *Yeast*, **5**: 187-198.
- Bisson, L.F. (1988). High-affinity glucose transport in *Saccharomyces cerevisiae* is under general glucose repression control. *J. Bacteriol.*, **170**: 4838-4845.
- Blinder, D., Coshi Gano, P. & Magasanik, B. (1996). Interaction of the GATA factor Gln3p with the nitrogen regulator Ure2p in *Saccharomyces cerevisiae*. *J. Bacteriol.*, **178**: 4734-4736.

- Blomberg, A. & Adler, L. (1989). Roles of glycerol and glycerol-3-phosphate dehydrogenase (NAD⁺) in acquired osmotolerance of *Saccharomyces cerevisiae*. *J. Bacteriol.*, **171**: 1087-1092.
- Boguslawski, G. (1992). PBS2, a yeast gene encoding a putative protein kinase, interacts with the RAS2 pathway and affects osmotic sensitivity of *Saccharomyces cerevisiae*. *J. Gen. Microbiol.*, **138**: 2425-2432.
- Bolivar, F. & Backman, K. (1979). Plasmids of *Escherichia coli* as cloning vectors. *Methods Enzymol.*, **68**: 245-267.
- Brewster, J.L., De Valoir, T., Dwier, N.D., Winter, E. & Gustin, M.C. (1993). An osmosensing signal transduction pathway in yeast. *Science*, **259**: 1760-1763.
- Brown, A.J.P. (1995). Preparation of total RNA. In: *Methods in Yeast Molecular Biology* (Evans, I., ed.) pp 269-276. Humana Press, NY, USA.
- Burlini, N., Morandi, S., Pellegrini, R., Tortora, P. & Guerritore A. (1989). Studies on the degradative mechanism of phosphoenolpyruvate carboxykinase from the yeast *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta*, **1014**: 153-161.
- Cadahía, J.L. (1996). Clonación del gen *ROX6* de *Saccharomyces cerevisiae*. Tesis de licenciatura, Universidad de La Coruña.
- Capieaux, E., Vignais M.-L., Sentenac, A. & Goffeau, A. (1989). The yeast H⁺-Atpase gene is controlled by the promoter binding factor *TUF*. *J. Biol. Chem.*, **264**: 7437-7446.
- Carlson, M. & Botstein, D. (1982). Two differentially regulated mRNAs with different 5' ends encode secreted with intracellular forms of yeast invertase. *Cell*, **28**: 145-154.
- Church, G.M. & Gilbert, W. (1984). Genomic sequencing. *Proc. Natl. Acad. Sci. USA*, **81**: 1991-1995.
- Coffman, J.A., Rai, R., Cunningham, T., Svetlov, V. & Cooper, T.G. (1996). Gat1p, a GATA family protein whose production is sensitive to nitrogen catabolite repression, participates in transcriptional activation of nitrogen-catabolic genes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **16**: 847-858.
- Coffman, J.A., Rai, R., Loprete, D.M., Cunningham, T., Svetlov, V. & Cooper, T.G. (1997). Cross regulation of four GATA factors that control nitrogen catabolic gene expression in *Saccharomyces cerevisiae*. *J. Bacteriol.*, **179**: 3416-3429.
- Cohen, R., Holland, J.R., Yokoi, T. & Holland, M.J. (1986). Identification of a regulatory region that mediates glucose-dependant induction of the *Saccharomyces cerevisiae* enolase gene *ENO2*. *Mol. Cell. Biol.*, **6**: 2287-2297.
- Crauwels, M., Winderickx, J., De Winde, J.H. & Thevelein, J. (1997). Identification of genes with nutrient-controlled expression by PCR-mapping in the yeast *Saccharomyces cerevisiae*. *Yeast*, **13**: 973-984.

- Dang, V.-D., Bohn, C., Bolotin-Fukuhara, M. & Daighan-Fornier, B. (1996). The CCAAT box-binding factor stimulates ammonium assimilation in *Saccharomyces cerevisiae*, defining a new cross-pathway regulation between nitrogen and carbon metabolisms. *J. Bacteriol.*, **178**: 1842-1849.
- DeAntoni, A., D'Angelo, M., Dal Pero, F., Sartorello, F., Pandolfo, D., Pallaricini, A., Lanfranchi, G. & Valle, G. (1996). The DNA sequence of cosmid 14-13b from chromosome XIV of *Saccharomyces cerevisiae* reveals an unusually high number of overlapping open reading frames. *Yeast*, **13**: 261-266.
- Decottignies, A. & Goffeau, A. (1997). Complete inventory of the yeast ABC proteins. *Nature Genet.*, **15**: 137-145.
- Delbruck, S. & Ernst, J.F. (1993). Morphogenesis-independent regulation of actin transcript levels in the pathogenic yeast *Candida albicans*. *Mol. Microbiol.*, **10**: 859-866.
- Denis, C.L., Ciriacy M. & Young, E.T. (1981). A positive regulatory gene is required for accumulation of the functional messenger RNA for the glucose-repressible alcohol dehydrogenase from *Saccharomyces cerevisiae*. *J. Mol. Biol.*, **148**: 355-368.
- Denis, C.L., Ferguson, J. & Young, E.T. (1983). mRNA levels for the fermentative alcohol dehydrogenase of *Saccharomyces cerevisiae* decrease upon growth on a nonfermentable carbon source. *J. Biol. Chem.*, **258**: 1165-1171.
- DeRisi, J.L. Iyer, V.R. & Brown, P.O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, **278**: 680-686.
- Dubois, E. & Messenguy, F. (1997). Integration of the multiple controls regulating the expression of the arginase gene *CAR1* of *Saccharomyces cerevisiae* in response to different nitrogen signals: role of *Gln3p*, *ArgRp-Mcm1p*, and *Ume6p*. *Mol. Gen. Genet.*, **253**: 568-580.
- Dujon B. (1996). The yeast genome project: what did we learn? *Trends Genet.*, **12**: 263-270.
- Dujon, B., Alexandraki, D., Andre, B., Ansorge, W., Baladron, V., Ballesta, J.P., Banrevi, A., Bolle, P.A., Bolotin-Fukuhara, M., Bossier, P., *et al.* (1994). Complete DNA sequence of yeast chromosome XI. *Nature*, **369**: 371-378.
- Einerhand, A.W., Voorn-Brouwer, T.M., Erdmann, R., Kunau, W.H. & Tabak, H.F. (1991). Regulation of transcription of the gene coding for peroxisomal 3-oxoacyl-CoA thiolase of *Saccharomyces cerevisiae*. *Eur. J. Biochem.*, **200**: 113-122.
- Ellwood, M.S. & Craig, E.A. (1984). Differential regulation of the 70K heat shock gene and related genes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **4**: 1454-1459.
- Federoff, H.J., Eccleshall, T.R. & Marmur, J. (1983). Carbon catabolite repression of maltase synthesis in *Saccharomyces carlsbergensis*. *J. Bacteriol.*, **156**: 301-307.

- Feinberg, A.P. & Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.*, **132**: 6-13.
- Fink, G.R. (1987). Pseudogenes in yeast ? *Cell*, **49**: 5-6.
- Flattery-O'Brien, J.A., Grant, C.M. & Dawes, I.W. (1997). Stationary-phase regulation of the *Saccharomyces cerevisiae* *SOD2* gene is dependant on additive effects of *HAP2/3/4/5* and *STRE* binding elements. *Mol. Microbiol.*, **23**: 303-312.
- Fodor, S.P.A., Rava, R.P., Huang, X.C., Pease, A.C., Holmes, C.P. & Adams, C.L. (1993). Multiplexed biochemical assays with biological chips. *Nature*, **364**: 555-556.
- Fondrat, C. & Kalogeropoulos, A. (1996). Approaching the function of new genes by detection of their potential upstream activation sequences in *Saccharomyces cerevisiae*: application to chromosome III. *CABIOS*, **12**: 363-374.
- Forbes, A. & Lehmann, R. (1998). Nanos and pumilio have critical roles in the development and function of *Drosophila* germline stem cells. *Development*, **125**: 679-690.
- Foury, F. (1997). Human genetic diseases: a cross-talk between man and yeast. *Gene*, **195**: 1-10.
- Freire-Picos, M.A., Hollenberg, C.P., Breuning, K.D. & Cerdán, M.E. (1995). Regulation of cytochrome *c* expression in the aerobic respiratory yeast *Kluyveromyces lactis*. *FEBS letters*, **360**: 39-42.
- Fuge, E.K., Braun, E.L. & Werner-Washburne, M. (1994). Protein synthesis in long-term stationary-phase cultures of *Saccharomyces cerevisiae*. *J. Bacteriol.*, **176**: 5802-5813.
- Galibert, F., Alexandraki, D., Baur, A., *et al.* (1996). Complete nucleotide sequence of *Saccharomyces cerevisiae* chromosome X. *EMBO J.*, **15**: 2031-2049.
- Gancedo, C. & Schwerzmann, N. (1976). Inactivation by glucose of phosphoenolpyruvate carboxikinase from *Saccharomyces cerevisiae*. *Arch. Microbiol.*, **109**: 221-225.
- Gancedo, J.M. (1992). Catabolite repression in yeast. *Eur. J. Biochem.*, **206**: 297-313.
- Gee, S., Krauss, S.W., Miller, E., Aoyagi, K., Arenas, J. & Conboy, J.G. (1997). Cloning of mDEAH9, a putative RNA helicase and mammalian homologue of *Saccharomyces cerevisiae* splicing factor Prp43. *Proc. Natl. Acad. Sci. USA*, **94**: 11803-11807.
- Geissler, S., Pereira, G., Spang, A., Knop, M., Soues S., Kilmartin, J. & Schiebel, E. (1996). The spindle pole body component Spc98p interacts with the gamma-tubulin-like Tub4p of *Saccharomyces cerevisiae* at the sites of microtubule attachment. *EMBO J.* **15**: 3899-3911.
- Goffeau, A. & Vassaroti, A. (1991). The European project for sequencing the yeast genome. *Res. Microbiol.*, **142**: 901-903.

- Goffeau, A., Park, J., Paulsen, I.T., Jonniaux, J.L., Dinh, T., Mordant & P., Saier, M.H. Jr (1997). Multidrug-resistant transport proteins in yeast: complete inventory and phylogenetic characterization of yeast open reading frames with the major facilitator superfamily. *Yeast*, **13**: 43-54.
- Greer, H., Penn, M., Hauge, B., & Galgoci, B. (1982). Control of aminoacid biosynthesis in yeast. *Rec. Adv. Yeast Mol. Biol.*, **1**: 122-142.
- Ha, N., Hellauer, K. & Turcotte, B. (1996). Mutations in target DNA elements of yeast *HAPI* modulates its transcriptional activity without affecting DNA binding. *Nucleic Acids Res.*, **24**: 1453-1459.
- Haarasilta, S. & Oura, E. (1975). On the activity and regulation of anaplerotic and gluconeogenetic enzymes during the growth process of baker's yeast. The biphasic growth. *Eur. J. Biochem.*, **52**: 1-7.
- Herruer, M.H., Mager, W.H., Doorenbosch, T.M., Wessels, P.L., Wassenaar, T.M. & Planta R.J. (1989). The extended promoter of the gene encoding ribosomal protein S33 in yeast consists of multiple protein binding elements. *Nucleic Acids Res.*, **17**: 7427-7439
- Herskowitz, I., Rine, J. & Strathern, J. (1992). Mating-type determination mating-type interconversion in *Saccharomyces cerevisiae*. In: *The molecular and cellular biology of the yeast Saccharomyces: Gene expression.* (Jones, E.W., Pringle, J.R. & Broach, J.R. ed.) Vol. **2**, pp 591-595. Cold spring harbor Laboratory press NY, USA.
- Hoheisel, J. (1997). Oligomer chip technology. *Trends Biotech.*, **15**: 465-469.
- Hope, I.A. & Struhl, K. (1985). GCN4 protein, synthesized *in vitro*, binds HIS3 regulatory sequences: implications for general control of amino acid biosynthetic genes in yeast. *Cell*, **43**: 177-188.
- Huibregtse, J.M., Engelke, D.R. & Thiele, D.J. (1989). Copper-induced binding of cellular factors to yeast metallothionein upstream activation sequences. *Proc. Natl. Acad. Sci. USA*, **86**: 65-69.
- Huie, M.A., Scott, E.W., Drazinic, C.M., Lopez, M.C., Hornstra, I.K., Yang, T.P. & Baker, H.V. (1992). Characterization of the DNA-binding activity of GCR1: *in vivo* evidence for two GCR1-binding sites in the upstream activating sequence of TPI of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **12**: 2690-2700.
- Igual, J.C., Matallana, E., Gonzalez-Bosch, C., Franco, L. & Perez-Ortin, J.E. (1991). A new glucose-repressible gene identified from the analysis of chromatin structure in deletion mutants of yeast *SUC2* locus. *Yeast*, **7**: 379-389.
- Johnson, A.D. (1992). A combinatorial regulatory circuit in budding yeast. In *transcription regulation* (de S. McKnight and K. Yamamoto). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York. pp. 975-1006.

- Johnston, M. (1996). Genome sequencing: the complete code for a eukariotic cell. *Curr. Biol.*, **6**: 500-503.
- Johnston, M. & Carlson, M. (1992). Regulation of carbon and phosphate utilization In: *The molecular and cellular biology of the yeast Saccharomyces: Gene expression*. (Jones, E.W., Pringle, J.R. & Broach, J.R. eds.) Vol. 2, pp 238-240. Cold spring harbor Laboratory press NY, USA.
- Johnston, M., Andrews, S., Brinkman, R., Cooper, J., Ding, H., Dover, J., Du, Z., Favello, A., Fulton, L., Gattung, S. *et al.* (1994). Complete nucleotide sequence of *Saccharomyces cerevisiae* chromosome VIII. *Science*, **265**: 2077-2082.
- Kaback, D.B., Angerer, L. & Davidson, D. (1979). Improved methods for the formation and stabilisation of R-loops. *Nucleic Acids Res.* **6**: 2499-2517.
- Kafatos, F.C., Jones, C.W. & Stratiadis, E.F. (1979). Determination of nucleic acid sequence, homologies and relative concentrations by a dot hybridization procedure. *Nucleic Acids Res.*, **7**: 1541-1552.
- Kalogeropoulos, A. (1995). Automatic intron detection in nuclear DNA sequence of *Saccharomyces cerevisiae*. *Yeast*, **11**: 555-565.
- Keleher, C.A., Goutte, C. & Johnson, A.D. (1988). The yeast cell-type-specific repressor alpha 2 acts cooperatively with a non-cell-type-specific protein. *Cell*, **53**: 927-936.
- Keng, T. (1992). *HAP1* and *ROX1* form a regulatory pathway in the repression of *HEM13* transcription in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **12**: 2616-2623.
- Kennedy, B.K., Gotta, M., Sinclair, D.A., Mills, K., McNabb, D.S., Murthy, M., Pak, S.M., Laroche, T., Gasser, S.M. & Guarente, L. (1997). Redistribution of silencing proteins from telomeres to the nucleolus is associated with extension of life span in *S. cerevisiae*. *Cell*, **89**: 381-391.
- Klein, P., Kanemisa, M. & Delisi, C. (1985). The detection and classification of membrane-spanning proteins. *Biochem. Biophys. Acta*, **815**: 468-476.
- Klein, C.J.L.; Olsson, L.; Ronnow, B.; Mikkelsen, J.D.; Nielsen, J. (1996). Alleviation of glucose repression of maltose metabolism by *mig1* disruption in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.*, **62**: 4441-4449.
- Kraakman, L.S., Griffioen, G., Zerp, S., Groeneveld, P., Thevelein, J.M., Mager, W.H. & Planta, R.J. (1993). Growth-related expression of ribosomal protein genes in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.*, **239**: 196-204.
- Kuchin, S., Yeghiayan, P. & Carlson, M. (1995). Cyclin-dependant protein kinase and cyclin homologs *SSN3* and *SSN8* contribute to transcriptional control in yeast. *Proc. Natl. Acad. Sci. USA*, **92**: 4006-4010.

- Lagunas, R. (1986). Misconceptions about the energy metabolism of *Saccharomyces cerevisiae*. *Yeast*, **12**: 221-228.
- Larsson, K., Ansell, R., Erikson, P. & Adler, L. (1993). A gene encoding sn-glycerol 3-phosphate dehydrogenase (NAD⁺) complements an osmosensitive mutant of *Saccharomyces cerevisiae*. *Mol. Microbiol.*, **10**: 1101-1111.
- Lauer, G.D., Roberts, T.M. & Klotz, L.C. (1977). Determination of the nuclear DNA content of *Saccharomyces cerevisiae* and implication for the organization of DNA in yeast chromosomes. *J. Mol. Biol.*, **114**: 507-526.
- Lee, F.-J.S., Lin, L.-W. & Smith, J.A. (1990). A glucose-repressible gene encodes acetyl-CoA hydrolase from *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **265**: 7413-7418.
- Leidich, S.D. & Orlean, P. (1996). Gpi1, a *Saccharomyces cerevisiae* protein that participates in the first step in glycosylphosphatidylinositol anchor synthesis. *J. Biol. Chem.*, **271**: 27829-27837.
- Li, L., Ernsting, B.R., Wishart, M.J., Lohse, D.L. & Dixon, J.E. (1997). A family of putative tumor suppressors is structurally and functionally conserved in humans and yeast. *J. Biol. Chem.*, **272**: 29403-29406.
- Liao, S.-M., Zhang, J., Jeffery, D.A., Koleseske, A.J., Thompson, C.M., Chao, D.M., Viljoen, M., Van Vuuren, H.J.K. & Young, R.A. (1995). A kinase-cyclin pair in the RNA polymerase II holoenzyme. *Nature*, **374**: 193-196.
- Lillehaug, J.R. & Kleppe, K. (1975). Effects of salts and polyamines on T4 polynucleotide kinase. *Biochemistry*, **14**: 1225-1230.
- Lindquist, S. (1981). Regulation of protein synthesis during heat shock. *Nature*, **293**: 311-314.
- Lindquist, S. & Craig, E.A. (1988). The heat-shock proteins. *Annu. Rev. Genet.*, **22**: 631-677.
- Lloyd, A.T. & Sharp, P.M. (1992). Codons: A microcomputer program for codon usage analysis. *J. Heredity*, **83**: 239-240.
- Lobo, Z., & Maitra, P.K. (1977). Genetics of yeast hexokinase. *Genetics*, **86**: 727.
- Lombardia, L. (1996). Análisis de la expresión de 52 ORFs de función desconocida en *Saccharomyces cerevisiae*. Tesis de licenciatura. Universidad de la Coruña.
- Lombardo, A., Carine, K. & Scheffler, I.E. (1990). Cloning and characterization of the iron-sulfur subunit gene of succinate dehydrogenase from *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **265**: 10419-10423.
- Lowry, C.V. & Zitomer, R.S. (1984). Oxygen regulation of aerobic and anaerobic genes mediated by common factor in yeast. *Proc. Natl. Acad. Sci. USA*, **81**: 6129-6133.

- Lundin, M., Nehlin, J.D. & Ronne, H. (1994). Importance of a flanking AT-rich region in target site recognition by the GC-box-binding zinc finger protein *MIG1*. *Mol. Cell. Biol.*, **14**: 1979-1985.
- Magasanik, B. (1992). Regulation of nitrogen utilization in: The molecular and cellular biology of the yeast *Saccharomyces*: Gene expression. (Jones, E.W., Pringle, J.R. & Broach, J.R. eds.) Vol. 2, pp 283-317. Cold spring harbor Laboratory press, NY, USA.
- Mager, W.H. & Moradas Ferreira, P. (1993). Stress response of yeast. *Biochem. J.*, **290**: 1-13.
- Mager, W.H. & De Kruijff, A.J. (1995). Stress-induced transcriptional activation. *Microbiol. Rev.*, **59**: 506-531.
- Mager, W.H., Maurer, K. & Piper, P.W. (1998). Identifying stress genes. In "Methods in Microbiology: Yeast Gene Analysis" (Brown, A.J.P. & Tuite, M.F., eds.), pp 337-349. Academic Press, London, UK.
- Mahler, M.R., Jaynes, P.K., McDonough, J.P. & Hanson, D.K. (1981). Current topics in cellular regulation. (Horeker, B.L. & Stadtman, E.R. eds.), vol. **18**, pp 454-474.
- Marchler, G., Schüller, C., Adam, G. & Ruis, H. (1993). A *Saccharomyces cerevisiae* UAS element controlled by protein kinase A activates transcription in response to a variety of stress conditions. *EMBO J.*, **12**: 1997-2003.
- Martinez-Pastor, M.T., Marchler, G., Schuller, C., Marchler-Bauer, A., Ruis, H. & Estruch F (1996). The *Saccharomyces cerevisiae* zinc finger proteins Msn2p and Msn4p are required for transcriptional induction through the stress response element (STRE). *EMBO. J.*, **15**: 2227-2235.
- Marzluf, G.A. (1997). Genetic regulation of nitrogen metabolism in the fungi. *Microbiol. Mol. Biol. Rev.*, **61**: 17-32.
- McDaniel, D., Caplan, A.J., Lee, M.-S., Adams, C.C., Fishel, B.R., Gross, D.S. & Garrard, W.T. (1989). Basal-level expression of the yeast *HSP82* gene requires a heat shock regulatory element. *Mol. Cell. Biol.*, **9**: 4789-4798.
- Mercado, J.J., Smith, R., Sogliocco, F.A., Brown, A.J.P. & Gancedo J.M. (1994). The levels of yeast gluconeogenic mRNAs respond to environmental factors. *Eur. J. Biochem.*, **224**: 473-481.
- Messing, J. (1979). A multipurpose cloning system based on stringle-stranded DNA bacteriophage M13. *Recomb. DNA Tech. Bull.*, **2**: 43.
- Mewes, H.W., Albermann, K., Heumann, K., Liebls, S. & Pfeiffer, F. (1997a). MIPS: A database for protein sequence, homology data and yeast genome information. *Nucleic Acids Res.*, **25**: 28-30.

- Mewes, H.W., Albermann, K., Bähr, M., Frishman, D., Gleissner, A., Hani, J., Heumann, K., Maier, L.A., Oliver, S.G., Pfeiffer, F. & Zollner, A. (1997b). Overview of the yeast genome. *Nature*, **387**: 7-65.
- Miller, M.J., Xuong, N.-H. & Geiduschek, E.P. (1982). Quantitative analysis of the heat shock response of *Saccharomyces cerevisiae*. *J. Bacteriol.*, **151**: 311-327.
- Miller, S.M. & Magasanik, B. (1991). Role of the complex upstream region of the GDH2 gene in nitrogen regulation of the NAD-linked glutamate dehydrogenase in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **11**: 6229-6247.
- Minehart, P.L. & Magasanik, B. (1992). Sequence of the *GLN1* gene of *Saccharomyces cerevisiae*: Role of the upstream region in regulation of glutamine synthetase expression. *J. Bacteriol.*, **174**: 1828-1836.
- Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M. & Tollervey, D. (1997). The exosome: A conserved eukaryotic RNA processing complex containing multiple 3'→5' Exoribonucleases. *Cell*, **91**: 457-466.
- Moehle, C.M. & Jones, E.W. (1990). Consequences of growth media, gene copy number and regulatory mutations on the expression of the *PRB1* gene of *Saccharomyces cerevisiae*. *Genetics*, **124**: 39-55.
- Moehle, C.M. & Hinnebush, A.G. (1991). Association of RAP1 binding sites with stringent control of ribosomal protein gene transcription in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **11**: 2723-2735.
- Moesch, H.U., Graf, R., Schmidheini, T. & Braus, G. (1990). Three GCN4 responsive elements act synergistically as upstream and as TATA-like elements in the yeast TRP4 promoter. *EMBO J.*, **9**: 2951-2957.
- Mortimer, R.K. & Schild, D. (1989). Genetic map of *Saccharomyces cerevisiae*. Edition 10. *Yeast*, **5**: 321-403.
- Mullis, K.B. & Faloona, F.A. (1987). Specific synthesis of DNA *in vitro* via a polymerase-catalysed chain reaction. In: *Methods in Enzymology* (Wu, R. ed.), Vol. **155**, pp 335-350. Academic press, San Diego, CA, USA.
- Naitou, M., Hagiwara, H., Hanaoka, F., Eki, T. & Murakami, Y. (1997). Expression profiles of transcripts from 126 open reading frames in the entire chromosome VI of *Saccharomyces cerevisiae* by systematic northern analyses. *Yeast*, **13**: 1275-1290.
- Nakai, K. & Kanehisa, M. (1992). A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics*, **14**: 897-911.
- Nasr, F., Becam, A.M. & Herbert, C.J. (1996). The sequence of 36.8 Kb from the left arm of chromosome XIV reveals 24 complete open reading frames: 18 correspond to new genes, one of which encodes a protein similar to the human myotonic dystrophy kinase. *Yeast*, **12**: 169-175.

- Nehlin, J.O., Carlberg, M. & Ronne, H. (1991). Control of yeast *GAL* genes by *MIG1* repressor: A transcriptional cascade in the glucose response. *EMBO J.*, **10**: 3373-3377.
- Nelissen, B., De Wachter, R. & Goffeau A. (1997). Classification of all putative permeases and other membrane plurispansers of the major facilitator superfamily encoded by the complete genome of *Saccharomyces cerevisiae*. *FEMS Microbiology Reviews*, **21**: 113-134.
- Niedenthal, R.K., Riles, L., Johnston, M. & Hegemann, J.H. (1996). Green fluorescent protein as a marker for gene expression and subcellular localization in budding yeast. *Yeast*, **12**: 773-786.
- Niederberger, P., Miozzor, G. & Heutter, R. (1981). Biological role of the general control of aminoacid biosynthesis in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **1**: 584-593.
- Nieto-Sotelo, J., Wiederrecht, G., Okuda, A. & Parker, C.S. (1990). The yeast heat shock transcription factor contains a transcriptional activation domain whose activity is repressed under nonshock conditions. *Cell*, **62**: 807-817.
- Nomura, M., Gourse, R. & Baughman, G. (1984). Regulation of the synthesis of ribosomes and ribosomal components. *Ann. Rev. Biochem.*, **53**: 75-117.
- Oliver, S. (1996). A network approach to the systematic analysis of yeast gene function. *Trends Genet.*, **12**: 241-242.
- Oliver, S.G., Van Der Aart, O.J., Agostoni-Carbone, M.L., Aigle, M., Alberghina, L., Alexandraki, D., Antoine, G., Anwar, R., Ballesta, J.P., Benit, P. *et al.* (1992). The complete DNA sequence of yeast chromosome III. *Nature*, **357**: 38-46.
- Orlean, P., Leidish, S.D., Drapp, D.A., Colussi, P. & Braz, J. (1994). Isolation of temperature-sensitive yeast GPI-anchoring mutants. *Med. Biol. Res.*, **27**: 145-150.
- Ozcan, S., Johnston, M. (1996). Two different repressors collaborate to restrict expression of the yeast glucose transporter genes *HXT2* and *HXT4* to low levels of glucose. *Mol. Cell. Biol.*, **16**: 5536-5545.
- Paoluzi, S., Minenkova, O. & Castagnoli, L. (1997). The genes encoding the transcription factor ytaf(ii)60, the g4p1 protein and a putative glucose-transporter protein are contained in a 12.3 Kb DNA fragment on the left arm of *Saccharomyces cerevisiae* chromosome VII. *Yeast*, **13**: 85-91
- Park, H.O. & Craig, E.A. (1989). Positive and negative regulation of basal expression of a yeast *HSP70* gene. *Mol. Cell. Biol.*, **9**: 2025-2033.
- Penn, M.D., Galcoci, B. & Greer, H. (1983). Identification of AAS genes and their regulatory role in general control of aminoacids biosynthesis in yeast. *Procl. Natl. Acad. Sci. USA*, **80**: 2704-2708.
- Perlman, D. & Halvorson, H.O. (1981). Distinct repressible mRNAs for cytoplasmic and secreted yeast invertase are encoded by a single gene. *Cell*, **25**: 525-536.

- Persson, B. & Argos, P. (1994). Prediction of transmembrane segments in proteins utilising multiple sequence alignments. *J. Mol. Biol.*, **237**: 182-192.
- Pfeifer, K. (1988). Coordinate regulation of cytochrome *c* genes in *Saccharomyces cerevisiae* by heme and *HAPI*. PhD Thesis, MIT, Cambridge, USA.
- Philippsen, P. *et al.* (1997). The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XIV and its evolutionary implications. *Nature*, **387**: 93-98.
- Piper, P.W. (1993). Molecular events associated with acquisition of heat tolerance by the yeast *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.*, **11**: 339-355.
- Piper, P., Curran, B., Davies, M., Lockheart, A. & Reid, G. (1986). Transcription of the phosphoglycerate kinase gene of *Saccharomyces cerevisiae* increases when fermentation cultures are stressed by heat shock. *Eur. J. Biochem.*, **161**: 525-531.
- Planta, R.J., Brown, A.J., Cadahia, J.L., Cerdan, M.E., DeJonge, M., Gent, M.E., Hayes, A., Kolen, C.P.A.M., Lombardia, L.J., Sefton, M., Oliver, S.G., Thevelein, J., Tournu, H., Van Delft, Y.J., Verbart, D.J. & Winderickx, J. (1998). Transcript analysis of 250 novel yeast genes from chromosome XIV. *Yeast*. Enviado.
- Polakis, E.S., Bartley, W. & Meek, G.A. (1964). Changes in the structure and enzyme activity of *Saccharomyces cerevisiae* in response to changes in the environment. *Biochem. J.*, **90**: 369-374.
- Polakis, E.S. & Bartley, W. (1966). Changes in dry weight, protein, deoxyribonucleic acid, ribonucleic acid and reserve and structural carbohydrate during the aerobic growth cycle of yeast. *Biochem. J.*, **98**: 883-887.
- Praekelt, U.M. & Meacock, P.A. (1990). HSP12, a new small heat shock gene of *Saccharomyces cerevisiae*: analysis of structure, regulation and function. *Mol. Gen. Genet.*, **223**: 97-106.
- Pringle, J.R. & Hartwell, L.H. (1981) In: *The molecular biology of the yeast Saccharomyces: life, cycle and inheritance* (Strathern, S.N., Jones, E.W. & Broach, J.R. eds.) pp 97-142. Cold Spring Harbor Laboratory Press, NY, USA.
- Quandt, K., Frech, K., Karas, H., Wingender, E. & Werner, T. (1995). MatInd and MatInspector- New fast and versatile tools for detection of consensus matches in nucleotides sequence data. *Nucleic Acids Res.*, **23**: 4878-4884.
- Rai, R., Genbauffe, F.S., Sumrada, R.A. & Cooper, T.G. (1989). Identification of sequences responsible for transcriptional activation of the allantoin permease gene in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **9**: 602-608.

Repetto, B. & Tzagoloff, A. (1990). Structure and regulation of *KGD2*, the structural gene for yeast dihydrolipoyl transsuccinylase. *Mol. Cell. Biol.*, **10**: 4221-4232.

Richard, G.F., Fairhead, C. & Dujon, B. (1997). Complete transcriptional map of yeast chromosome XI in different life conditions. *J. Mol. Biol.*, **268**: 303-321.

Rodríguez-Belmonte, E., Rodríguez-Torres, A., Tizón, B., Cadahía, J.L., González-Siso, M.I., Ramil, E., Becerra, M., González-Domínguez, M. & Cerdán, E. (1996). Sequence analysis of a 10Kb DNA fragment from yeast chromosome VII reveals a novel member of the DnaJ family. *Yeast*, **12**: 145-148.

Ronne, H. (1995). Glucose repression in fungi. *Trends Genet.*, **11**: 12-17

Roseblum-Vos, L.S. (1988). Isolation and characterization of mutations wich cause overexpression of the *CYC7* gene in *Saccharomyces cerevisiae*. PhD thesis. State University of New York at Albany. USA.

Roth, S., Neuman-Silberberg, F.S., Barcelo, G. & Schupbach, T. (1995). Cornichon and the EGF receptor signaling process are necessary for both anterior-posterior and dorsal-ventral pathern formation in *Drosophila*. *Cell*, **16**: 967-978.

Rothstein, R.J. & Sherman, F. (1980). Genes affecting the expression of cytochrome *c* in yeast: Genetic mapping and genetic interactions. *Genetics*, **94**: 871-889.

Ruis, H. & Schüller, C. (1995). Stress signaling in yeast. *Bioessays*, **17**: 959-965.

Salmon, J.-M., Piñon, R. & Gancedo, C. (1989). Isolation and characterization of mutants of *Saccharomyces cerevisiae* able to sporulate in the presence of glucose. *J. Gen. Microbiol.*, **135**: 203.

Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989). *Molecular cloning: a laboratory manual* (Ford, N., Nolan, C. & Fergusson, M. eds.). Cold Spring Harbor Laboratory Press, NY, USA.

Sanchez, Y. & Lindquist, S. (1990). *HSP104* required for induced thermotolerance. *Science*, **248**: 1112-1115.

Schmidheini, T., Mõsh, R. Graf, R. & Braus, G.H. (1990). A *GCN4* protein recognition element is not sufficient for *GCN4*-dependent regulation of transcription in the *ARO7* promoter of *Saccharomyces cerevisiae*. *Mol. Gen. Genet.*, **224**: 57-64.

Schul, W., Groenhout, B., Koberna, K., Takagaki, Y., Jenny, A., Manders, E.M., Raska, I., Van Driel, R., De Jong, L. (1996). The RNA 3' cleavage factors CstF 64 kDa and CPSF 100 kDa are concentrated in nuclear domains closely associated with coiled bodies and newly synthesized RNA. *EMBO J.*, **15**: 2883-2892.

Schudiner, O., Yanover, C. & Benvenisty, N. (1997). Computer analysis of the entire budding yeast genome for putative targets of the *GCN4* transcription factor. *Curr. Genet.* **33**: 16-20.

- Schüller, H.-J. & Entian, K.-D. (1987). Isolation and expression analysis of two yeast regulatory genes involved in the derepression of glucose-repressible enzymes. *Mol. Gen. Genet.*, **209**: 366-373.
- Schüller, C., Brewster, J.L., Alexander, M.R., Gustin, M.C. & Ruis, H. (1994). The HOG pathway controls osmotic regulation of transcription via the stress response element (STRE) of the *Saccharomyces cerevisiae* *CTT1* gene. *EMBO J.*, **13**: 4382-4389.
- Sharp, P.M. & Li, W.-H. (1987). The codon Adaptation Index. A measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **15**: 1281-1295.
- Shena, M., Shalon, D., Davis, R.W. & Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complimentary DNA microarray. *Science*, **270**: 467-470.
- Silar, P., Butler, G. & Thiele, D.J. (1991). Heat shock transcription factor activates transcription of the yeast metallothionein gene. *Mol. Cell. Biol.*, **11**: 1232-1238.
- Smart, W.C., Coffman, J.A. & Cooper, T.G. (1996). Combinatorial regulation of the *Saccharomyces cerevisiae* *CAR1* (arginase) promoter in response to multiple environmental signals. *Mol. Cell. Biol.*, **16**: 5876-5887.
- Sogin, S.J. & Saunders, C.A. (1980). Fluctuation in polyadenylate size and content in exponential- and stationary-phase cells of *Saccharomyces cerevisiae*. *J. Bacteriol.*, **144**: 74-81.
- Sorger, P.K. (1991). Heat shock factor and the heat shock response. *Cell*, **65**: 363-366.
- Sprague, G.F. & Cronan Jr., J.E. (1977). Isolation and characterization of *Saccharomyces cerevisiae* mutants defective in glycerol catabolism. *J. Bacteriol.*, **129**: 1335-1342.
- Stanbrough, M. & Magasanik, B. (1996). Two transcription factors, Gln3p and Nillp, use the same GATAAG sites to activate the expression of *GAP1* of *Saccharomyces cerevisiae*. *J. Bacteriol.*, **178**: 2465-2468.
- Stanbrough, M., Rowen, D.W. & Magasanik, B. (1996). Role of the GATA factors Gln3p and Nit2p of *Saccharomyces cerevisiae* in the expression of the nitrogen-regulated genes. *Proc. Natl. Acad. Sci. USA*, **92**: 9450-9454.
- Surosky, R.T.; Strich, R. & Exposito, R.E. (1994). The yeast *UME5* gene regulates the stability of meiotic mRNAs in response to glucose. *Mol. Cell. Biol.*, **14**: 3446-3458.
- Svetlov, V.V. & Cooper, T.G. (1995). Review: Compilation and Characteristics of dedicated transcription factors in *Saccharomyces cerevisiae*. *Yeast*, **11**: 1439-1484.
- Termier, M. & Kalogeropoulos, A. (1996). Discrimination between fortuitous and biologically constrained open reading frames in DNA sequences of *Saccharomyces cerevisiae*. *Yeast*, **12**: 369-384.
- Tettelin, H. *et al.* (1997). The nucleotide sequence of *Saccharomyces cerevisiae* chromosome VII. *Nature*, **387**: 81-84.

- Tizón, B. (1997). Secuenciación de un fragmento de DNA del cromosoma VII de *Saccharomyces cerevisiae* y análisis funcional de las ORFs YGL131c, YGL129c, YGL125w, YGL124c, e YGL121c. Tesis de doctorado. Universidad de La Coruña.
- Tizón, B., Rodríguez-Torres, A., Rodríguez-Belmonte, E., Cadahía, J.L. & Cerdán, E. (1996). Identification of a putative methylenetetrahydrofolate reductase by sequence analysis of a 6.8Kb DNA fragment of yeast chromosome VII. *Yeast*, **12**: 1047-1051.
- Treitel, M.A. & Carlson, M. (1995). Repression by *SSN6-TUP1* is directed by *MIG1*, a repressor/activator protein. *Proc. Natl. Acad. Sci. USA*, **92**: 3132-3136.
- Trumbly, R.J. (1986). Isolation of *Saccharomyces cerevisiae* mutants constitutive for invertase synthesis. *J. Bacteriol.*, **166**: 1123-1127.
- Trumbly, R.J. (1992). Glucose repression in the yeast *Saccharomyces cerevisiae*. *Mol. Microbiol.*, **6**: 15-21.
- Tugendreich, S., Basset Jr., D.E., Mc Kusick, V.A., Boguski, M.S. & Hieter, P. (1994). Genes conserved in yeast and humans. *Human molecular genetics*, **3**: 1509-1517.
- Tzamarías, D. & Struhl, K. (1995). Functional dissection of the yeast *Cyc8-Tup1* transcriptional corepressor complex. *Nature*, **369**: 758-761.
- Valdés-Hevia, M.D., De La Guerra, R. & Gancedo, C. (1989). Isolation and characterization of the gene encoding phosphoenolpyruvate carboxykinase from *Saccharomyces cerevisiae*. *FEBS Lett.*, **258**: 313-316.
- Varela, J.C.S., Praekelt, U.M., Meacock, P.A., Planta, R.J. & Mager, W.H. (1995) The *Saccharomyces cerevisiae* HSP12 gene is activated by the high-osmolarity glycerol pathway and negatively regulated by protein kinase A. *Molecular and Cellular Biology*, **15**: 6232-6245.
- Velculescu, V.E., Zhang, L., Vogelstein, B. & Kinzler, W.K. (1995). Serial analysis of gene expression. *Science*, **270**: 484-487.
- Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett Jr., D.E., Hieter, P., Vogelstein, B. & Kinzler, W.K. (1997). Characterization of the yeast transcriptome. *Cell*, **88**: 243-251.
- Vitale, G., Fabre, E. & Hurt, E.C. (1996). *NBP35* encodes an essential and evolutionary conserved protein in *Saccharomyces cerevisiae* with homology to a superfamily of bacterial ATPases. *Gene*, **178**: 97-106.
- Wahi, M. & Johnson, A.D. (1995). Identification of genes required for *alfa2* repression in *Saccharomyces cerevisiae*. *Genetics*, **140**: 79-90.
- Walsh, M.C., Scholte, M., Valkier, J., Smits, H.P. & Vandam, K. (1996). Glucose sensing and signaling properties in *Saccharomyces cerevisiae* require the presence of at least 2 members of the glucose-transporter family. *J. Bacteriol.*, **178**: 2593-2597.

- Weaver, P.L., Sun, C. & Chang, T.H. (1997). Dbp3p, a putative RNA helicase in *Saccharomyces cerevisiae*, is required for efficient pre-ribosomal-RNA processing predominantly at site a(3). *Mol. Cell. Biol.*, **17**: 1354-1365.
- Werner-Washburne, M., Braun, E., Johnston, G.C. & Singer, R.A. (1993). Stationary phase in the yeast *Saccharomyces cerevisiae*. *Microbiol. Rev.*, **57**: 383-401.
- Werner-Washburne, M., Braun, E.L., Crawford, M.E. & Peck, V.M. (1996). Stationary phase in *Saccharomyces cerevisiae*. *Mol. Microbiol.*, **19**: 1159-1166.
- Wiederrecht, G., Shuey, D.J., Kibbe, W.A. & Parker, C.S. (1987). The *Saccharomyces* and *Drosophila* heat shock transcription factors are identical in size and DNA binding properties. *Cell*, **48**: 507-515.
- Willett, C.E., Gelfman, C.M. & Holland, M.J. (1993). A complex regulatory element from the yeast gene *ENO2* modulates *GCR1*-dependent transcriptional activation. *Mol. Cell. Biol.*, **13**: 2623-2633.
- Wilson, K.L. & Herskowitz, I. (1984). Negative regulation of *STE6* gene expression by the $\alpha 2$ product of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **4**: 2420-2427.
- Wingender, E., Kel, A.E., Kel, O.V., Karas, M., Heimeyert, T., Dietze, P., Knueppel, R., Romaschenko, A.G. & Kolchanov, N.A. (1997). TRANSFAC, TRRD, and COMPEL: Towards a federated database system on transcriptional regulation. *Nucl. Acid. Res.*, **25**: 265-268.
- Winston, F., Dollard, C. & Ricupero-Hovasse, S.L. (1995). Construction of a set of convenient *Saccharomyces cerevisiae* strains that are isogenic to S288C. *Yeast*, **11**: 53-55.
- Wodicka, K., Dong, H., Mittmann, M., Ho, M. & Lockheart, D.J. (1997). Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nature Biotech.*, **15**: 1359-1367.
- Wolfe, K. & Shields, D. (1997). Molecular evidence for an ancient duplication of the entire *Saccharomyces cerevisiae* genome. *Nature*, **387**: 708-713.
- Woods, R.A., Roberts, D.G., Stein, D.S. & Filpula, D. (1984). Adenine phosphoribosyltransferase mutants in *Saccharomyces cerevisiae*. *Gen. Microbiol.*, **184**: 2629-2657.
- Wright, R.M. & Poyton, R.O. (1990). Release of two *Saccharomyces cerevisiae* cytochrome genes, *COX6* and *CYC1*, from glucose repression requires the *SNF1* and *SSN6* gene products. *Mol. Cell. Biol.*, **10**: 1297.
- Xing, Y., Fikes, J.D. & Guarente, L. (1993). Mutations in yeast *HAP2/HAP3* define a hybrid CCAAT box binding domain. *EMBO J.*, **12**: 4647-4655.
- Yang, S.S., Wei, Y.L., Sung, T.F. (1991). Chemical analysis and amino acid content of temperature-sensitive mutants of *Saccharomyces cerevisiae*. *Chung Hua Min Kuo Wei Sheng Wu Chi Mien I Hsueh Tsa Chih* **24**: 311-320.

- Yin, Z., Smith, R.J. & Brown A.J.P. (1996). Multiple signalling pathways trigger the exquisite sensitivity of yeast gluconeogenic mRNAs to glucose. *Mol. Microbiol.*, **20**: 751-764.
- Yoo, H.S. & Cooper, T.G. (1989). The DAL7 promoter consists of multiple elements that cooperatively mediate regulation of the gene's expression. *Mol. Cell. Biol.*, **9**: 3231-3243
- Yoshikawa, A. & Isono, K. (1990). Chromosome III of *Saccharomyces cerevisiae*: an ordered clone bank, a detailed restriction map and analysis of transcripts suggest the presence of 160 genes. *Yeast*, **6**: 383-401.
- Yoshikawa, A. & Isono, K. (1991). Construction of an ordered clone bank and systematic analysis of the whole transcripts of chromosome VI of *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **9**: 1189-1195.
- Zamaroczy, M. & Bernardi, G. (1986). The primary structure of the mitochondrial genome of *Saccharomyces cerevisiae* - a review. *Gene*, **47**: 155-177.
- Zhang, L., Hach, A. & Wang, C. (1998). Molecular mechanism governing heme signaling in yeast: a higher-order complex mediates heme regulation of the transcriptional activator *HAP1*. *Mol. Cell. Biol.* **18**: 3819-3828.
- Zitomer, R.S. & Lowry, C.V. (1992). Regulation of gene expression by oxygen in *Saccharomyces cerevisiae*. *Microbiol. Rev.*, **56**: 1-11.
- Zitomer, R.S., Limbach, M.P., Rodríguez-Torres, A.M., Balasubramanian, B., Deckert, J. & Snow, P.M. (1997). Approaches to the study of *ROX1* repression of the hypoxic genes in the yeast *Saccharomyces cerevisiae*. *Methods: A companion to Methods Enzymol.*, **11**: 279-288.

7. APÉNDICE

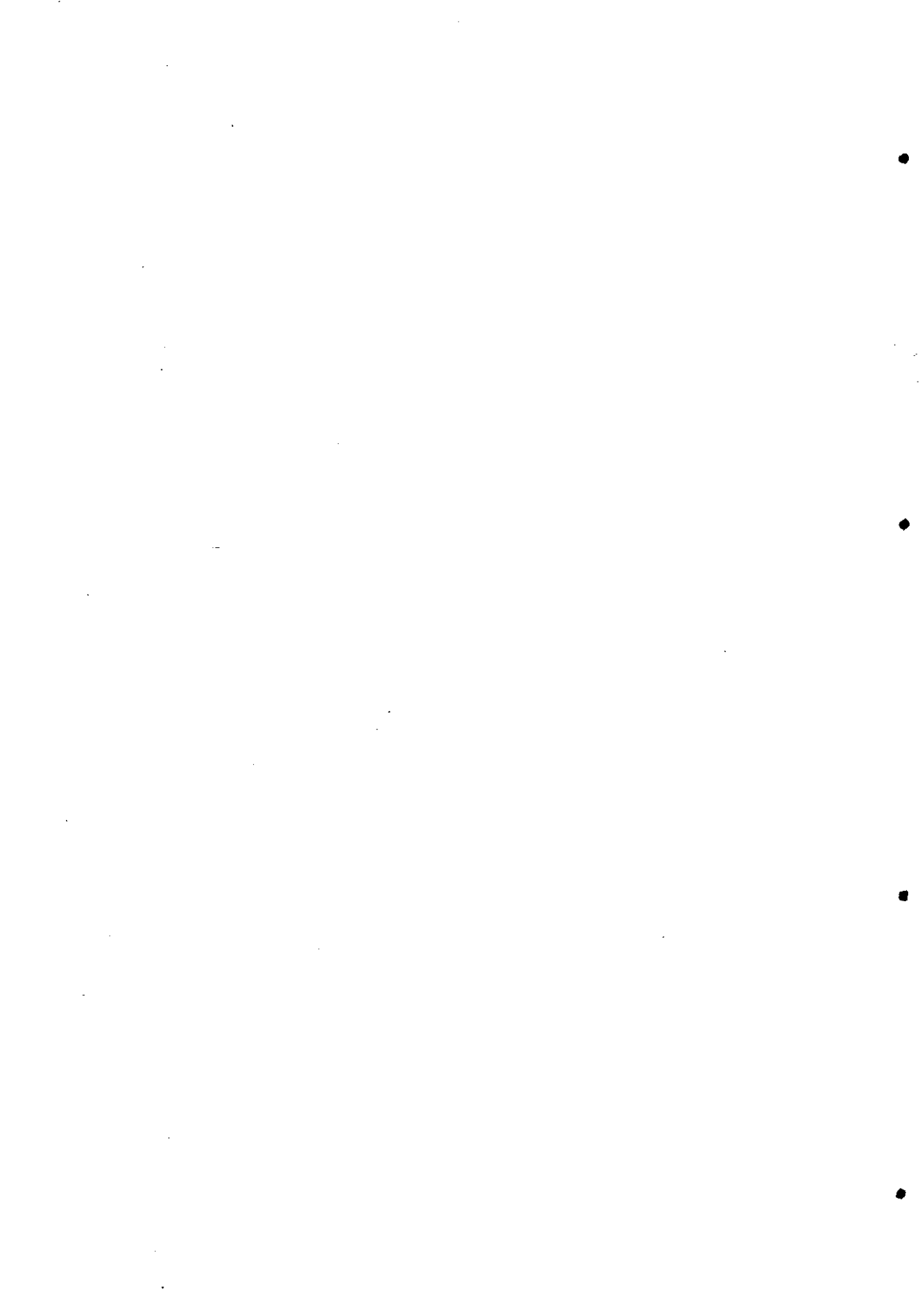


Tabla Apéndice. Descripción de las funciones potenciales y/o similitudes con otras proteínas de las 51 y 52 ORFs de los cromosomas VII y XIV de *Saccharomyces cerevisiae* analizadas en este trabajo. Se indica la nomenclatura estándar correspondiente a cada ORF. (Datos obtenidos de MIPS).

NOMENCLATURA			CROMOSOMA VII	
Trabajo	Estandar	GEN	POSIBLE FUNCIÓN	
G2850	YGL131c		Codificaría una proteína que presenta similitudes con la proteína hipotética C3H1.12C de <i>Schizosaccharomyces pombe</i> y que contiene en su secuencia un motivo característico de peptidasas señal tipo I. En el análisis funcional, las colonias derivadas de una espora portadora de la ORF interrumpida crecieron más lentamente y fueron de menor tamaño que las salvajes.	
G2856	YGL129c		El análisis funcional mostró que podría estar relacionada con la función mitocondrial ya que las líneas que contienen esta ORF interrumpida fueron incapaces de crecer sobre medios con fuentes de carbono no fermentables en todas las temperaturas de cultivo ensayadas. Además la secuencia aminoácida contiene dos motivos característicos, uno de ellos típico de transportadores mitocondriales y el otro de proteínas de unión a ATP/GTP.	
G2861	YGL128c		Su producto presenta un alto grado de homología en su región N-terminal con las proteínas DnaJ, de choque térmico, de levaduras y de <i>E. coli</i> (Rodríguez-Belmonte <i>et al.</i> , 1996). El análisis fenotípico reveló que las líneas con la ORF interrumpida son incapaces de crecer o crecen muy lentamente a 15°C y 37°C respectivamente, además se observaron colonias pequeñas con una morfología celular inusual.	
G2882	YGL125w	<i>MET11</i>	El análisis fenotípico mostró que las líneas que contienen el gen interrumpido son incapaces de crecer sobre medios sin aminoácidos y que la adición de metionina al medio de cultivo restaura el crecimiento de la línea delecionada. También se observaron grandes vacuolas y condensaciones bi o multipolares de material denso en las células al microscopio óptico. Se sabe ahora que es el gen <i>MET11</i> que codifica la metilentetrahidrofolato reductasa, implicada en la biosíntesis de la metionina (Tizón <i>et al.</i> , 1998).	
G2889	YGL124c		Muestra homología con una ORF que codifica una proteína hipotética de <i>Schizosaccharomyces pombe</i> . Su análisis funcional reveló un defecto de esporulación cuando era delecionada	
G2913	YGL121c		Proteína hipotética	
G2916	YGL120c	<i>PRP43</i>	Codifica un factor de corte-empalme (splicing) de los pre-mRNA de la familia de las cajas DEAH, unas helicasas implicadas en el corte empalme del RNA (Gee <i>et al.</i> , 1997).	
G3080	YGL106w	<i>MLC1</i>	El producto que codifica contiene 4 sitios de unión a calcio y por lo tanto presenta grandes similitudes con las calmodulinas, proteínas reguladoras de numerosas actividades enzimáticas en las células eucariotas. El análisis funcional mostró que su disrupción era letal para la célula.	

Tabla Apéndice. Descripción de las funciones potenciales y/o similitudes con otras proteínas de las 51 y 52 ORFs de los cromosomas VII y XIV de *Saccharomyces cerevisiae* analizadas en este trabajo. Se indica la nomenclatura estándar correspondiente a cada ORF. (Datos obtenidos de MIPS) (continuación).

NOMENCLATURA		CROMOSOMA VII	
Trabajo	Estandar	GEN	POSIBLE FUNCIÓN
G3090	YGL104c		Se predicen 11 hélices transmembrana para su producto que es parecido a las proteínas de transporte de la glucosa (Paoluzi et al., 1997).
G3107	YGL101w		El producto que codifica presenta una gran similitud con la proteína hipotética codificada por la ORF YBR242w. El análisis de tétradas mostró que dos esporas crecían correctamente mientras que dos revelaron un fenotipo complejo con una segregación anormal de los marcadores auxotróficos.
G3110	YGL100w	SEHI	Codifica una proteína formadora de un poro nuclear. Su disrupción es viable, sin embargo los mutantes nulos crecen lentamente, especialmente a bajas temperaturas.
G3113	YGL099w		Al poseer un motivo típico de los sitios de unión a ATP/GTP, el producto de esta ORF es similar al de <i>MMR1</i> que codifica una posible proteína humana de unión a GTP. El análisis funcional mostró que la interrupción de la ORF provoca un defecto en el apareamiento y en la sensibilidad a la temperatura, observándose células de forma poliédrica.
G3116	YGL098w		Proteína hipotética
G3146	YGL096w		Homología con una proteína que interviene en la homeostasis del cobre por represión de las proteínas de importación y que permite el crecimiento en un medio con altos niveles de cobre, especialmente cuando la fuente de carbono es lactato. G3146 posee además un homeodominio característico de los factores de transcripción.
G3168	YGL093w		El producto que codifica es ligeramente parecido a una proteína de <i>M. genitalium</i> y posee una hélice transmembrana.
G3175	YGL091c	NBP35	Codifica una proteína de unión a nucleótidos (con un motivo de unión a ATP/GTP, en el cual un cambio de nucleótidos conduce a una proteína no funcional) que presenta una homología con una superfamilia de ATPasas bacterianas (Vitale et al., 1996). La disrupción del gen es letal para la célula.
G3179	YGL090w		El producto de esta ORF mostró semejanzas con la proteína p100 de la enfermedad de Lyme.
G3189	YGL087c	MMS2	Codificaría una protein-ligasa implicada en la resistencia del DNA a agentes dañinos
G3193	YGL085w		Pequeña homología con una nucleasa de <i>Staphylococcus aureus</i> . El análisis fenotípico reveló pequeñas colonias blancas de forma regular; su crecimiento es reducido en medio mínimo.

Tabla Apéndice. Descripción de las funciones potenciales y/o similitudes con otras proteínas de las 51 y 52 ORFs de los cromosomas VII y XIV de *Saccharomyces cerevisiae* analizadas en este trabajo. Se indica la nomenclatura estándar correspondiente a cada ORF. (Datos obtenidos de MIPS) (continuación).

NOMENCLATURA		CROMOSOMA VII	
Trabajo	Estandar	GEN	POSIBLE FUNCIÓN
G3210	YGL078c	<i>DBP3</i>	RNA-helicasa necesaria en el procesamiento del pre-rRNA (Weaver <i>et al.</i> , 1997). La disrupción de este gen es viable pero el mutante deletionado crece lentamente. Al ser también deficiente en sub-unidad ribosomal 60S se retrasa en la formación de rRNA 25S. Por otro lado, el análisis fenotípico mostró diferencias entre las líneas α y α .
G3406	YGL066w		Similitud con el factor de unión a cajas G en <i>Dicystostelium discoideum</i>
G3432	YGL061c		Proteína hipotética.
G3435	YGL060w		Tiene una gran similitud con la ORF YBR216c.
G3474	YGL054c		Su producto es similar a la proteína cni (cornichon) que interviene en la señalización dorso-ventral de las líneas germinales de <i>D. Melanogaster</i> (Roth <i>et al.</i> , 1995). Se predicen 3 hélices transmembrana en su secuencia aminoacídica.
G3483	YGL053w		Es muy similar a las ORFs YAR028w, YAR033w, YCR007c y G3489 analizadas en este trabajo. La proteína que codifica tiene dos hélices transmembrana potenciales pero aun no fue incluida en ninguna categoría funcional.
G3489	YGL051w		Muy similar a YAR033w y contiene dos dominios transmembrana potenciales.
G3505	YGL050w		Proteína hipotética.
G3517	YGL047w		Su producto es similar a una proteína específica del inicio de la gametocitogénesis en <i>Plasmodium falciparum</i> .
G3535	YGL046w		Proteína hipotética.
G3540	YGL045w		Proteína hipotética.
G3679	YGL023c		La proteína que codifica presenta similitudes con Vsp27p y posee motivos de sitios de unión a ATP/GTP y de sitios de unión a hemo típico de la familia del citocromo c.
G3690	YGL020c		Codifica una proteína desconocida para la que fueron pronosticados tres segmentos transmembrana y una localización en la membrana del retículo endoplasmático.
G3719	YGL014w		Presenta similitudes con la proteína Mtp5p; que se requiere para la re-distribución de las proteínas Sir (Silent mating type Information Regulation) en el nucléolo de las células viejas (Kennedy <i>et al.</i> , 1997), para el crecimiento a altas temperaturas y para la recuperación de la detención del crecimiento provocada por el factor alpha. También es similar a la proteína pumilio, de unión a RNA, y que interviene en el desarrollo embrionario de la de <i>D. Melanogaster</i> (Forbes y Lehmann, 1998).

Tabla Apéndice. Descripción de las funciones potenciales y/o similitudes con otras proteínas de las 51 y 52 ORFs de los cromosomas VII y XIV de *Saccharomyces cerevisiae* analizadas en este trabajo. Se indica la nomenclatura estándar correspondiente a cada ORF. (Datos obtenidos de MIPS) (continuación).

NOMENCLATURA		CROMOSOMA VII	
Trabajo	Estandar	GEN	POSIBLE FUNCIÓN
G3759	YGL005c		Su producto es ligeramente similar a la proteína Eg5 relacionada con una quinesina de <i>Xenopus</i> ; son unas proteínas implicadas en el transporte de organelos, en la mitosis y en la meiosis.
G4332	YGR053c		Proteína hipotética.
G7538	YGR187c	HGHI	Codifica un producto que presenta alguna similitud con las proteínas humanas Hmg1p y Hmg2p (isozimas de la HMG-CoA reductasa) (Arroyo <i>et al.</i> , 1997).
G7553	YGR189c		Codifica una proteína similar a Utr2p que a su vez se parece a una glucanasa de <i>Bacillus</i> .
G7584	YGR194c		Su producto es similar a los miembros de una familia que incluye a la fucosinasa, la glucoquinasa y la gliceroquinasa.
G7587	YGR195w	ECM20 RRP41	Codifica un componente del complejo exoribonucleasa 3'-5' implicada en el procesamiento del extremo 3' del rRNA 5.8S (Mitchell <i>et al.</i> , 1997).
G7589	YGR196c		Su producto presenta una ligera similitud con la proteína ácida repetitiva arp I del protozoario ciliado <i>Tetrahymena</i> .
G7594	YGR198w		Proteína hipotética.
G7725	YGR200c		La proteína que codifica posee dos hélices transmembrana potenciales y presenta una ligera similitud con una proteína reguladora del secuestro del nucleótido guanina.
G7737	YGR205w		Su producto es similar a una proteína de transporte de la putrescina en <i>E. coli</i> .
G7748	YGR210c		Codifica una proteína similar a una proteína de unión a GTP en <i>Methanococcus jannaschii</i> .
G7801	YGR211w		Su producto presenta semejanzas con una proteína hipotética del organismo procarionta <i>Methanococcus jannaschii</i> .
G7806	YGR212w		Su producto es algo parecido a una proteína hipotética de <i>S. pombe</i> .
G7826	YGR216c	GPII	Se requiere para la síntesis de la N-acetilglucosaminil fosfatidilinositol. Se mostró que la disrupción del gen era viable pero que la delección Δ gpi1 provocaba un fenotipo sensible a la temperatura y el anclaje de la GPI (glucosa-6-fosfato isomerasa) en la membrana celular (Orlean <i>et al.</i> , 1994; Leidich y Orlean, 1996).
G8523	YGR221c		La proteína que codifica se parece a Yhr149p.
G8530	YGR223c		Su producto presenta similitudes con la proteína hipotética codificada por YFR021w. El promotor de esta ORF tiene un sitio potencial de unión al factor transcripcional Mat α 2p, y G8530 podía haber sido, por lo tanto, un gen específico reprimido por este factor; sin embargo el análisis funcional mostró que la interrupción de la ORF no provoca diferencias entre las líneas α y a, no observándose tampoco un defecto en el apareamiento.

Tabla Apéndice. Descripción de las funciones potenciales y/o similitudes con otras proteínas de las 51 y 52 ORFs de los cromosomas VII y XIV de *Saccharomyces cerevisiae* analizadas en este trabajo. Se indica la nomenclatura estándar correspondiente a cada ORF. (Datos obtenidos de MIPS) (continuación).

NOMENCLATURA		CROMOSOMA VII + XIV	
Trabajo	Estandar	GEN	POSIBLE FUNCIÓN
G8537	YGR224w		Su producto es similar a la proteína de resistencia a fármacos codificada por el gen <i>SGE1</i> , a la bomba de toxinas de <i>Cochitobolus carbonum</i> y a la proteína emrB de resistencia a fármacos de <i>E. coli</i> . Además como posee 12 hélices transmembrana potenciales, podría pertenecer a la familia de proteínas transportadoras que intervienen en la resistencia a múltiples fármacos (Goffeau <i>et al.</i> , 1997; Nelissen <i>et al.</i> , 1997). El análisis funcional indicó que las células con esta ORF delecionada crecen mas lentamente a 15°C.
G8541	YGR225w		Codificaría una proteína con similitudes con las que intervienen en el ciclo de división celular humano (p55CDC) y de <i>Candida elegans</i> (Cdc20p).
N1696	YNL168c		Su producto tiene alguna similitud con proteínas en <i>C. Elegans</i> (ZK688.3 de función desconocida) y <i>E. Coli</i> . (hpcEp con actividades isomerasa y decarboxilasa).
N1706	YNL166c		Proteína hipotética.
N1710	YNL165w		Es similar a las ORFs YOR385w e YMR316w.
N1714	YNL164c		Proteína hipotética.
N1718	YNL163c		Gran similitud de su producto con los factores de alargamiento de la traducción en mamíferos eEF-2, Eft1p y Eft2p.
N1727	YNL161w		Codificaría un producto de gran similitud con las protein-quinasas cot-1 de <i>Neurospora crassa</i> y MDK de la distrofia miotónica humana (Nasr <i>et al.</i> , 1996).
N1735	YNL159c		Proteína hipotética.
N1739	YNL158w		Proteína hipotética.
N1743	YNL157w		Proteína hipotética.
N1747	YNL156c		Su producto es similar al de la ORF YHR133c.
N1751	YNL155w		Proteína hipotética.
N1761	YNL153c	PFD4	Codifica la sub-unidad 4 de la prefoldina, supuesto homólogo de la chaperona bovina del mismo nombre. Interviene también en la biogénesis de la actina y de la tubulina. El análisis funcional mostró que el mutante nulo provoca un crecimiento sensible al frío (desorganizándose los microtubulos a 14°C) y al estrés osmótico. Sin embargo, las células mutadas permanecen viables tras 20 horas a la temperatura restrictiva.

Tabla Apéndice. Descripción de las funciones potenciales y/o similitudes con otras proteínas de las 51 y 52 ORFs de los cromosomas VII y XIV de *Saccharomyces cerevisiae* analizadas en este trabajo. Se indica la nomenclatura estándar correspondiente a cada ORF. (Datos obtenidos de MIPS) (continuación).

NOMENCLATURA		CROMOSOMA XIV	
Trabajo	Estandar	GEN	POSIBLE FUNCIÓN
N1765	YNL152w		Proteína hipotética.
N1773	YNL150w		Proteína hipotética.
N1774	YNL149c		Su producto presenta una ligera similitud con una proteína hipotética de <i>S. pombe</i> .
N1777	YNL148c	<i>ALFI</i>	Codificaría una tubulina alfa asociada al citoesqueleto la cual actuaría sobre un componente del complemento, el factor B.
N1780	YNL147w		Codificaría una proteína similar a varias snRNPs (small nuclear RiboNucleoProteins) de humanos, alfalfa, ratón y rata.
N1785	YNL146w		Proteína hipotética.
N1790	YNL144c		La proteína que codifica es similar a la de YHR131c.
N1795	YNL143c		Proteína hipotética.
N1825	YNL141w	<i>AAHI</i>	Codifica la adenosina deaminasa (o adenina aminohidrolasa). Se mostró que su actividad enzimática no es inducida por la adenina y que se reduce 7 veces cuando las células crecen en un medio que contiene prolina en vez de amonio como única fuente de carbono. El análisis funcional mostró que la mutación <i>aaH1</i> afecta la represión de la adenina, mientras que la doble mutación <i>aaH1apt1</i> anula la represión de la adenina (Woods <i>et al.</i> , 1984).
N1843	YNL136w		El producto que codifica sería similar a una proteína (triplet M) que participa en la formación de los filamentos intermedios en las neuronas.
N1847	YNL134c		El producto que codifica es similar al del gen <i>toxD</i> de <i>C. carbonum</i> .
N1850	YNL133c		Proteína hipotética.
N1858	YNL132w		La proteína que codifica es similar al receptor esteroídico del anteridiol de <i>Achlya ambisexualis</i> .
N1870	YNL129w		Su producto presenta una ligera similitud con la uridin-quinasa udk de <i>Mycoplasma pneumoniae</i> .
N1872	YNL128w	<i>TEPI</i>	Codifica una tirosin-fosfatasa y que presenta similitudes con el gen supresor de tumores humanos <i>TEPI/MMAC/P TEN1</i> (Li <i>et al.</i> , 1997).
N1875	YNL127w		Su producto es similar a una proteína hipotética de <i>C. elegans</i> y posee alguna homología con Fus2p, una proteína que interviene en la fusión celular durante el apareamiento. Se pronosticó para el producto de N1875 una hélice transmembrana que se puede relacionar con su posible función en la superficie celular.

Tabla Apéndice. Descripción de las funciones potenciales y/o similitudes con otras proteínas de las 51 y 52 ORFs de los cromosomas VII y XIV de *Saccharomyces cerevisiae* analizadas en este trabajo. Se indica la nomenclatura estándar correspondiente a cada ORF. (Datos obtenidos de MIPS) (continuación).

NOMENCLATURA		CROMOSOMA XIV	
Trabajo	Estandar	GEN	POSIBLE FUNCIÓN
N1879	YNL126w	SPC98	Codifica un componente del cuerpo polar del huso mitótico; interacciona con una alfa-tubulina (producto del gen <i>TUB4</i>) para formar un complejo que estaría implicado en la organización microtubular del cuerpo polar (Geisler <i>et al.</i> , 1996). El análisis funcional mostró que su disrupción era letal y que la sobre-expresión de <i>SPC98</i> resulta tóxica, parándose la división de las células con gemas grandes, estructuras microtubulares defectuosas y núcleos sin dividirse.
N1882	YNL125c	ESBP6	Similar a la ORF YKL221w, y el producto de su transcripción es homólogo de un transportador ligado al cromosoma X humano, además, teniendo en cuenta la respuesta observada y el hecho que se le pronosticaron 10 segmentos transmembrana en su secuencia aminoácida, se puede apoyar su pertenencia al grupo funcional de las proteínas facilitadoras del transporte.
N1888	YNL124w		Su producto es similar a una proteína hipotética de <i>S. pombe</i> . Esta proteína es esencial ya que la delección de la ORF provoca un fenotipo letal para la célula.
N1897	YNL123w		Su producto presenta similitudes con una serin-proteasa de <i>Campylobacter jejuni</i> .
N1913	YNL119w		Su producto presenta alguna semejanza con una proteína hipotética de <i>M. jannaschii</i> .
N1917	YNL118c	PSU1	Codifica una proteína supresora de la deficiencia respiratoria en mutantes <i>pet</i> (el fenotipo <i>petite</i> corresponde a una incapacidad para las células de levadura de crecer en fuentes de carbono no fermentables).
N1925	YNL116w		Su producto presenta similitudes con una proteína "dedo de zinc" nuclear de <i>Gallus gallus</i> . El análisis funcional reveló que la delección de la ORF era letal para la célula.
N1929	YNL115c		Codifica para una proteína hipotética que posee 4 segmentos transmembrana potenciales
N1954	YNL110c		Su producto muestra semejanzas con una proteína de unión a RNA de la mosca de la fruta. El análisis fenotípico mostró que la delección de esta ORF es letal para la célula.
N1958	YNL109w		El producto que codifica posee tres hélices transmembrana y se parece levemente a una oxidasa del citocromo c.
N1962	YNL108c		Presenta una gran homología con el de la ORF YOR110w.
N1966	YNL107w		Su producto presentó una identidad del 36% en un solapamiento de 94 aminoácidos con la proteína humana AF-9 implicada en translocaciones cromosómicas en leucemias agudas.

Tabla Apéndice. Descripción de las funciones potenciales y/o similitudes con otras proteínas de las 51 y 52 ORFs de los cromosomas VII y XIV de *Saccharomyces cerevisiae* analizadas en este trabajo. Se indica la nomenclatura estándar correspondiente a cada ORF. (Datos obtenidos de MIPS) (continuación).

NOMENCLATURA			CROMOSOMA XIV
Trabajo	Estandar	GEN	POSIBLE FUNCIÓN
N2160	YNL106c	PIE3	Codifica una fosfatasa que cataliza la hidrólisis del fosfatidilinositol-fosfato. El análisis funcional reveló que la disrupción del gen es viable pero que el mutante nulo presenta vacuolas anormales.
N2185	YNL101w		Se pronosticaron 10 hélices transmembrana para su producto, el cual presenta similitudes con miembros (Ybr089p e Yer064p) de una probable familia de transportadores.
N2190	YNL100w		Proteína hipotética.
N2194	YNL099c		Su producto es similar al de las ORFs YNL032w, YNL056w e YDR067c; el hecho de que las dos primeras también estén situadas en el cromosoma XIV podría representar un ejemplo de duplicación génica.
N2205	YNL097c	PHO23	Su producto está implicado en la regulación transcripcional de <i>PHO5</i> (el mutante nulo provoca la expresión constitutiva de <i>PHO5</i>).
N2215	YNL095c		Es muy parecida a la ORF YOR092w y presenta similitudes con <i>Ecm3p</i> , implicada en la biogénesis de la pared celular así como en la arquitectura de la célula. Se le pronosticaron 8 segmentos transmembrana en su secuencia aminoácídica.
N2219	YNL094w		Su producto es similar a una proteína hipotética de <i>S. pompe</i> .
N2227	YNL092w		Su producto es similar a una proteína hipotética de <i>C. elegans</i> .
N2231	YNL091w		Similitud con una proteína participante en el transporte de las proteínas intracelulares del retículo endoplasmático al aparato de Golgi y, por su capacidad en formar estructuras filamentosas en la célula, también podría intervenir en la organización celular.
N2250	YNL087w		Similitud con una proteína que posee dos hélices transmembrana potenciales y presenta la firma característica de un sitio de unión de la fosfopantetina, un factor de crecimiento en <i>Lactobacillus bulgaricus</i> y un intermediario de la biosíntesis del Coenzima A en el hígado de los mamíferos y de otros organismos.
N2254	YNL086w		Proteína hipotética.
N2312	YNL083w		Codificaría una proteína de la familia de los transportadores mitocondriales MCF (Mitochondrial Carrier Family).

Universidad de La Coruña.
Facultad de Ciencias.
Departamento de Biología Celular y Molecular.
Área de Bioquímica y Biología Molecular.
La Coruña, Junio de 1998.

**Análisis de la transcripción
de 103 ORFs localizadas en los cromosomas VII y XIV de
Saccharomyces cerevisiae.**

Memoria presentada por **Luis José Lombardia Ferreira**
Para aspirar al grado de Doctor en Biología.

Figura portada: Representación esquemática del nivel absoluto de expresión de las 103 ORFs de función desconocida estudiadas en los cromosomas VII y XIV de *Saccharomyces cerevisiae*.

UNIVERSIDADE DA CORUÑA
Servicio de Bibliotecas



1700757440

1998

TD

FACULTAD DE CIENCIAS
DEPARTAMENTO DE
BIOLOGÍA CELULAR Y MOLECULAR
AREA DE BIOQUÍMICA
Y BIOLOGÍA MOLECULAR



UNIVERSIDADE DA CORUÑA