

# ANÁLISIS DE REGULARIDAD DE GENOMAS PARA DETECCIÓN DE TELÓMEROS Y SECUENCIAS AUTÓNOMAMENTE REPLICATIVAS

*Yuri Morales López*

Escuela de Matemática,  
Universidad Nacional, Heredia, Costa Rica.  
e-mail: ymorales@una.ac.cr

*Alejandro Ugalde León*

Escuela de Matemática,  
Universidad Nacional, Heredia, Costa Rica.  
e-mail: augald@una.ac.cr

*Tatiana Láscaris-Comneno Slepuiñ*

Escuela de Matemática,  
Universidad Nacional, Heredia, Costa Rica.  
e-mail: tlascaris@una.ac.cr

## RESUMEN

Hasta el momento, las levaduras han proporcionado mucha de la información referente a los orígenes eucarióticos de la replicación. La levadura *Yarrowia lipolytica* ha sido analizada con el fin de generar información que contribuya a la detección de regularidades de interés genético (Ugalde, Morales y Láscaris-Comneno, 2010) mediante la aplicación del índice de máxima regularidad  $i_{max,r}$  (Láscaris-Comneno, Skliar y Medina, 1999). En este trabajo, el  $i_{max,r}$  se aplica, en particular, al análisis de la secuencia correspondiente al cromosoma IX *cosmid 8224* y telómero derecho de la *Saccharomyces cerevisiae*, lo cual permite detectar el telómero (C1-3)A documentado en la base de datos especializada NCBI; asimismo identificar zonas completamente contenidas en este rasgo en las cuales el  $i_{max,r}$  alcanza su valor máximo de toda la secuencia. El análisis de genomas de la *Yarrowia lipolytica* por tamaños de ventana menores o iguales que la longitud de sus secuencias replicativas indicó, en todos los casos estudiados, que la región para la cual se obtiene el mayor  $i_{max,r}$  se encuentra totalmente contenida en la correspondiente secuencia autorreplicativa.

**Palabras clave:** índice de máxima regularidad, replicación, telómero, *Yarrowia lipolytica*, *Saccharomyces cerevisiae*

## ABSTRACT:

So far yeast has provided much of the information about eukaryotic origins of replication. The yeast *Yarrowia lipolytica* is analyzed with the purpose of generating information that will contribute to the detection of regularities of possible genetic interest (Ugalde, Morales & Láscaris-Comneno, 2010) by means of the application of the index of maximum regularity  $i_{max,r}$  (Láscaris-Comneno, Skliar & Medina, 1999). The  $i_{max,r}$  was used particularly to analyze the sequence that corresponds to the chromosome IX *cosmid 8224* and right telomere of *Saccharomyces cerevisiae*, which allowed to detect the telomere (C1-3)A documented in the specialized data base NCBI; and also to identify regions completely contained in the telomere in which the  $i_{max,r}$  attains its maximum value for all the sequence. The analysis of genomes of the *Yarrowia lipolytica* considering subsequences of a length smaller or equal to the length of its corresponding autonomously replicating sequences indicated in all the studied cases that the region with  $i_{max,r}$  the highest is totally contained in the corresponding replicative sequence.

**Keywords:** maximum regularity index, replication, telomere, *Yarrowia lipolytica*, *Saccharomyces cerevisiae*.

## INTRODUCCIÓN

En la actualidad, el conocimiento relacionado con biología molecular, ingeniería genética y biotecnología aumentó vertiginosamente. En gran medida, este crecimiento se ha debido tanto a la investigación de laboratorio como a su integración con otras disciplinas.

La Matemática aplicada no solo ofrece a este cuerpo de conocimiento la posibilidad de describir y analizar información, sino también plantea nuevas formas y estrategias para comprender gradualmente los fenómenos de interés.

Un ejemplo de este tipo de trabajo se desarrolla en el Proyecto de Investigación de Detección de Regularidades en el ADN, de la Escuela de Matemática de la Universidad Nacional (UNA), donde se han definido e implementado algoritmos para interpretar información del ADN.

La hipótesis principal de este proyecto consiste en que el grado de regularidad de las secuencias de ADN puede ser útil, conjuntamente con otras características de dichas secuencias, como centrómeros, telómeros, regiones autónomamente replicativas, entre otras, de particular relevancia genética.

En la siguiente sección se muestra el análisis efectuado a cadenas de ADN de diversos organismos.

## ORGANISMOS DE INTERÉS

### *Saccharomyces cerevisiae*

*Saccharomyces cerevisiae*<sup>1</sup> es una levadura unicelular eucariota, es decir, que posee una membrana que encierra el material genético. Según Goh y Kilmartin (1993), trabajar con el *Saccharomyces cerevisiae* ofrece muchas ventajas y, además, resulta útil para obtener información sobre los diferentes niveles de

funcionamiento de un organismo; de tal manera que es posible examinar la estructura y dinámica de la función celular y del organismo, más que las características aisladas de cada una de las partes de una célula u organismo. (Folch et al., 2004).

En primer lugar, el *Saccharomyces cerevisiae* es un modelo bien conocido en la actividad humana (Liti et al., 2006) relacionado a la "atenuación adecuada del mosto, estabilidad celular, sedimentación (permite la separación de la levaduras y contribuye a la clarificación de la cerveza), producción de alcohol, eficiencia fermentativa, y la reducción de compuestos indeseables durante la fase ulterior de maduración" (Abreu et al., 2008); en segundo lugar, ya se conoce la secuencialización completa de su código genético (Goffeau et al., 1996). En particular, la secuencia Z46902 posee 8956 pares de bases y se ubica en el cromosoma IX del *S. cerevisiae*.

### *Yarrowia Lipolytica*

La *Yarrowia lipolytica*<sup>2</sup> es un microorganismo con capacidad de producir sustancias metabólicas involucradas en el proceso de crecimiento, desarrollo y reproducción de organismos. Algunos trabajos (Casarégola et al., 1997) señalan que, inicialmente, esta levadura fue difícil de estudiar por una serie de características inusuales y la complejidad para construir su mapa genético. Superadas estas dificultades, *Yarrowia Lipolytica* se convirtió en punto de referencia para conocer otras múltiples características inusuales en otros organismos.

Antes de conocer su dinámica reproductiva, este organismo fue clasificado en otro grupo (*Deuteromycetous group*). Inclusive, se asemejaba mucho al *Saccharomyces cerevisiae* pero conforme se estudió más a fondo, se determinaron múltiples y significativas diferencias (Bankar et al., 2009).

<sup>1</sup> **Línea taxonómica:** cellular organisms; Eukaryota; Fungi/Metazoa group; Fungi; Dikarya; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces. **Nombre común:** levadura de panadería. **Sinónimos:** oviformis, italicus, capensis, uvarum var. Melibiosus (NCBI, 2009).

<sup>2</sup> **Línea taxonómica:** organismo celular; Eukaryota; Fungi/Metazoa group; Fungi; Dikarya; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Dipodascaceae; Yarrowia. Cepa natural conocida: Yarrowia lipolytica CLIB122 (NCBI, 2009).

*Yarrowia lipolytica* posee seis cromosomas clasificados del A al F y 7299 genes identificados. Las secuencias estudiadas en el presente trabajo se encuentran en el cromosoma A: la M91601, que posee 2309 pares de bases (bp); la M91600 compuesta por 1305 bp. Ambas están localizadas en la región centroamericana CEN de dicho cromosoma (Fournier et al., 1993).

Bankar et al. (2009) señalan que dentro de las utilidades más relevantes se encuentran:

- Bio-Remediación de suelos contaminados con petróleo.
- Bio-Remediación de ambientes acuáticos.
- Tratamiento o degradación de los desechos.
- Remoción de metales pesados para bio-remediación.
- Bio-transformaciones de productos de importancia económica.

**ÍNDICE DE MÁXIMA REGULARIDAD**

En este apartado se describe el índice de máxima regularidad ( $i_{max,r}$ ) propuesto por Láscaris-Commno, et al. (1999).

El  $i_{max,r}$  de una cadena de ADN es un número tal que  $0 \leq i_{max,r} \leq 1$ . Cuanto más cerca esté el  $i_{max,r}$  de 1, más regular será la cadena analizada.

Para el cálculo de  $i_{max,r}$  de una cadena de ADN, debe conocerse su longitud  $l_c$ , esto es, el número de bases que la componen.

En una cadena de ADN, que se puede representar como una sucesión de símbolos  $i_1 i_2 i_3 \dots i_n$ , donde los  $i$  (para  $1 \leq j \leq n$ ) son cualquiera de las bases nitrogenadas Adenina (A), Guanina (G), Citocina (C) y Timina (T). Para cada una de estas secuencias, se considera otra obtenida al efectuar un corrimiento o desfase de una letra en la secuencia original. Es decir, la última letra de la original pasa a ser la primera de la segunda secuencia y, a partir de ahí, el resto son las mismas letras de la original solo que corridas una posición a la derecha. Por ejemplo, obsérvese la siguiente cadena, para la cual  $l_c = 15$ :

A T C G G T T C T C T A T T G

Entonces, se obtiene la otra cadena bajo la original:

A T C G G T T C T C T A T T G  
T C G G T T C T C T A T T G A

La segunda secuencia se obtiene de la original efectuando un desfase de una letra ( $d = 1$ ).

Ambas secuencias se pueden considerar como un conjunto de pares ordenados (15 en este caso), tales que, la primera componente de cada uno es un elemento de la secuencia original y, la segunda, un elemento de la secuencia obtenida mediante el desfase. Los 15 pares del ejemplo anterior son los siguientes:

(A,T), (T,C), (C,G), (G,G), (G,T), (T,T), (T,C), (C,T), (T,C), (C,T), (T,A), (A,T), (T,T), (T,G) y (G,A).

Cada una de estos pares ordenados pertenece al conjunto  $BxB$ , donde  $B = \{A, G, T, C\}$ .

También, es posible definir una nueva secuencia, a partir de la original, efectuando el desfase de dos símbolos ( $d = 2$ ); otra más con el desfase de tres símbolos ( $d = 3$ ) y así sucesivamente hasta el desfase de catorce símbolos, pues un desfase de 15 símbolos (en este caso en particular) produce de nuevo la secuencia original.

Además, si se invierte el orden de las bases de la secuencia original, se obtiene una nueva secuencia que, junto con la original, forma el siguiente par de secuencias:

A T C G G T T C T C T A T T G  
G T T A T C T C T T G G C T A

Se pueden obtener de nuevo 14 pares de bases al ejecutar los desfasajes desde  $d = 1$  a  $d = 14$  con la secuencia invertida.

En general, para una secuencia de longitud  $l_c$ , se obtienen  $2 l_c - 1$  pares de secuencias,

de las cuales  $l_c - 1$  corresponden a los obtenidos con la secuencia dada y las  $l_c - 1$  secuencias generadas, a partir de ella, al efectuar los desfasajes sucesivos desde  $d = 1$  hasta  $d = l_c - 1$ . Los otros  $l_c$  pares de cadenas corresponden a los obtenidos con la secuencia dada y las  $l_c$  secuencias generadas, a partir de la original, al invertir el orden de las bases y, posteriormente, efectuar los desfasajes sucesivos desde  $d = 1$  hasta  $l_c - 1$ .

En este trabajo se parte de las siguientes suposiciones:

1. La probabilidad de que una base, elegida al azar, de una secuencia de bases sea cualquiera de las cuatro consideradas (A, G, C y T) es la misma, o sea,  $\frac{1}{4}$
- La probabilidad de que un par ordenado, elegido al azar, de cualquiera de los  $2l_c - 1$  pares de secuencias resultantes del proceso descrito anteriormente, sea cualquiera de las 16 posibles, es la misma, es decir,  $\frac{1}{16}$ .

Para cada uno de los pares de secuencias, correspondientes a un desfasaje de  $d$  símbolos, se denotan  $n_{A,A,d}$ ,  $n_{A,G,d}$ ,  $n_{A,C,d}$ ,  $n_{A,T,d}$ ,  $n_{G,A,d}$ ,  $n_{G,G,d}$ ,  $n_{G,C,d}$ ,  $n_{G,T,d}$ ,  $n_{C,A,d}$ ,  $n_{C,G,d}$ ,  $n_{C,C,d}$ ,  $n_{C,T,d}$ ,  $n_{T,A,d}$ ,  $n_{T,C,d}$ ,  $n_{T,G,d}$ , y  $n_{T,T,d}$  la cantidad de pares ordenados efectivamente contados. Por ello, se debe cumplir que

$$n_{A,A,d} + n_{A,G,d} + \dots + n_{T,T,d} = l_c$$

se denota con  $D_d$  la suma de los cuadrados de las diferencias entre el conteo obtenido y el valor esperado de cada par, correspondientes al primer conjunto de pares de secuencias para los diversos desfasajes posibles ( $d \in \{1, 2, \dots, l_c - 1\}$ ):

$$D_d = \left(n_{A,A,d} - \frac{l_c}{16}\right)^2 + \left(n_{A,G,d} - \frac{l_c}{16}\right)^2 + \dots + \left(n_{T,T,d} - \frac{l_c}{16}\right)^2$$

Además, se denota con  $D_d^*$  la suma de los cuadrados de las diferencias entre el conteo obtenido y el valor esperado de cada

par, correspondientes a los desfasajes ( $d \in \{1, 2, \dots, l_c - 1\}$ ), para el segundo conjunto de pares de bases<sup>3</sup>:

$$D_d^* = \left(n_{A,A,d} - \frac{l_c}{16}\right)^2 + \left(n_{A,G,d} - \frac{l_c}{16}\right)^2 + \dots + \left(n_{T,T,d} - \frac{l_c}{16}\right)^2$$

El,  $i_{max,r}$  para una secuencia de longitud  $l_c$ , se define por

$$i_{max,r} = \frac{\max\{D_1, \dots, D_{l_c-1}, D_0^*, \dots, D_{l_c-1}^*\}}{\left(l_c - \frac{l_c}{16}\right)^2 + 15\left(\frac{l_c}{16}\right)^2}$$

### METODOLOGÍA

Para la aplicación del algoritmo descrito en el apartado anterior, que calcula el  $i_{max,r}$  de una cadena de ADN, se utilizó dos software (Filetrunc y IMRProc) programados por el Yuri Morales de la Universidad Nacional y el César Arévalo de Arévalo Software y Multimedia S.A.; la plataforma utilizada fue Delphi Enterprise y la primera versión se lanzó en el segundo semestre del 2007.

A continuación, se describe el procedimiento utilizado:

1. Se eligieron varias cadenas de ADN contenidas en el genoma de la *Yarrowia lipolytica* y de la *Saccharomyces cerevisiae*, las cuales están en las bases de datos del National Center for Biotechnology Information (NCBI).
2. Para cada una de las cadenas, se analizó el nivel de regularidad aplicando el algoritmo descrito en subsecuencias cuyos puntos medios son consecutivos. Los tamaños de cada una de estas subsecuencias se denominan *ventanas* y se denotan V.
3. Se tabularon los valores del  $i_{max,r}$  para cada V y se construyeron gráficos.

<sup>3</sup> Formado por la secuencia invertida y sus  $l_c - 1$  desfasajes.

- Se compararon los resultados obtenidos con los rasgos documentados en las bases de datos del NCBI más largos que 20 bp.

El análisis efectuado se expone en el siguiente apartado.

### ANÁLISIS DE RESULTADOS

El  $i_{max,r}$  fue computado para secuencias de las levaduras *Saccharomyces cerevisiae* y *Yarrowia lipolytica*, que se encuentran documentados en la base de datos del NCBI, con los códigos de acceso Z46902 para la primera levadura y M91601 y M91600 para la segunda.

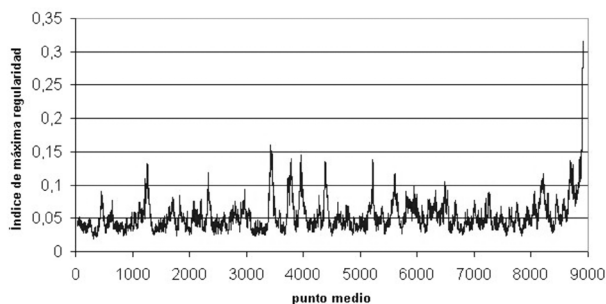
A continuación, se sintetizan los principales resultados del estudio de cada una de estas secuencias.

- Secuencia **Z46902** (8956 bp): Definición: “*Saccharomyces cerevisiae* chromosome IX cosmid 8224 and right telomere”.

Los rasgos documentados en la base de datos especializada considerados para el análisis son los siguientes:

- miscfeature: “similar to EMFUN:Scaarsa, L25318, *S. cerevisiae* autonomously replicating sequence (ARS120)”, para bases 7504 a 8415.
- misc feature: “85.8% identity with X-element in SCIXXY, M63935” para bases 8129 a 8681.
- misc feature: “telomere (C1-3)A sequence” para bases 8878 a 8956.
- El rasgo telomérico tiene una longitud documentada en NCBI de 79 bp. El estudio de la secuencia Z46902 mediante la aplicación del índice con  $V = 79$  genera los siguientes resultados:

**Figura 1.**  
 $i_{max,r}$  de la secuencia Z46902 para  $V = 79$



Como puede apreciarse, la aplicación del índice de máxima regularidad, con un tamaño de ventana igual a la longitud documentada del telómero (C1-3)A, genera un valor de  $i_{max,r}=0.3006249$  el cual difiere en 0.00538215 del mayor valor del índice en toda la secuencia, cuyo punto medio está a 2 bases del punto medio del rasgo documentado. Como puede observarse a continuación, el máximo valor del  $i_{max,r}=0.31600$  se alcanza en cinco subsecuencias consecutivas.

**Tabla 1**  
**Localización de los valores máximos del  $i_{max,r}$**

Punto medio	Subsecuencia	$i_{max,r}$
8916		0.3006249
8917	[8878, 8956] <sub>79</sub>	0.3006249
-telómero (C1-3)A-		
8911	[8872, 8950] <sub>79</sub>	0.31600705
8912	[8873, 8951] <sub>79</sub>	0.31600705
8913	[8874, 8952] <sub>79</sub>	0.31600705
8914	[8875, 8953] <sub>79</sub>	0.31600705
8915	[8876, 8954] <sub>79</sub>	0.31600705

Es el criterio de los autores que, con la aproximación descrita, puede considerarse que la aplicación del  $i_{max,r}$  ha permitido detectar este rasgo telomérico.

Así, el cálculo de los valores de  $i_{max,r}$  para tamaños de ventana  $V = 35, 45,$

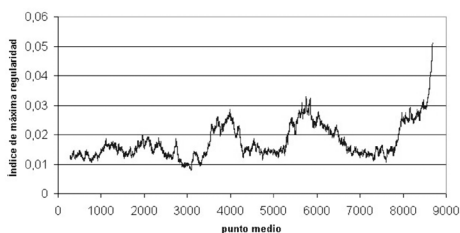
55, 65, 75 señaló, en todos estos casos, que las subsecuencias o regiones en las cuales se alcanza el valor más elevado del índice están completamente contenidas en el rasgo denominado “telómero (C1-3)A”.

Además, si se consideran para tamaños de ventana iguales a las longitudes de los otros dos rasgos documentados para esta secuencia, se obtiene que las regiones señaladas como las de máxima regularidad para los tamaños de ventana  $V = 553^4$  y  $V = 913^5$ , en todos los casos, contienen completamente al telómero documentado.

A manera de ejemplo, se ilustra en la siguiente figura el caso correspondiente a los resultados del análisis considerando  $V = 553$ :

**Figura 2**

$i_{max,r}$  de la secuencia Z46902 para  $V = 553$



2. Secuencia **M91601** (2309 bp): Definición: “Yarrowia lipolytica autonomously replicating sequence 68 and centromere”.

A partir del estudio de la relación entre la secuencia autónomamente replicativa y el centrómero (Fournier et al., 1993), y de la posición relativa entre ambos (Vernis et al., 1997 y Vernis et al., 1999) y se documentan en la base de datos los siguientes rasgos:

<sup>4</sup> La longitud de la región que presenta “85.8% identity with X-element in SC1XXY, M63935” es 553, con punto medio 8407.

<sup>5</sup> El rasgo “similar to EMFUN:Scaarsa, L25318, S. cerevisiae autonomously replicating sequence (ARS120)” es de longitud 912, con punto medio 7961 considerando  $V = 913$ .

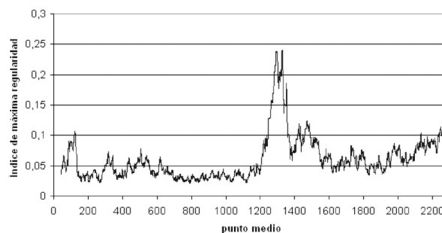
- misc feature ARS68, señalado en la base de datos entre las bases 734 y 1772
- misc feature CEN1-1 (segmento del centrómero 1), ubicado entre las bases 1545 y 1785

Se calcularon los valores de para los siguientes tamaños de ventana:  $V = 15, 25, 45, 65, 85, 105, 115, 251, 501$ . Para todos estos casos, la subsecuencia o región en la cual el  $i_{max,r}$  alcanza su valor más elevado está completamente contenida en la secuencia autónomamente replicativa ARS68.

Seguidamente se presentan, a manera de ejemplo, los datos obtenidos para  $V = 85$ :

**Figura 3**

$i_{max,r}$  de la secuencia M91601 para  $V = 85$



Nota: El mayor valor  $i_{max,r}=0.23271$  se alcanza en la subsecuencia de punto medio 1251 y longitud 85.

El grado de regularidad del rasgo ARS68 completo, considerando el punto medio 1254 y  $V = 1039$ , es  $i_{max,r} = 0.01899$ .

3. Secuencia **M91600** (1305 bp): Definición: “Yarrowia lipolytica centromere and autonomously replicating sequence 18”.

Las secuencias autónomamente replicativas ARS18 y ARS68, de M91600 y M91601, respectivamente, han sido comparadas entre sí y con los centrómeros (Fournier et al., 1993). A partir de este estudio, se documentan en la base de datos NCBI los siguientes rasgos para M91600:

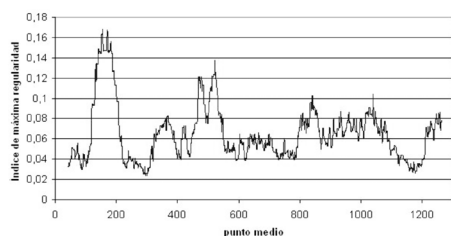


- misc feature ARS18 “autonomously replicative sequence 18” señalado en la base de datos como ubicado entre las bases 1 y 895.
- misc feature “centromere” ubicado entre las bases 283 y 1305.

Se calcularon los valores de  $i_{max,r}$  para los siguientes tamaños de ventana:  $V = 15, 25, 45, 65, 85, 105, 125, 441$ . En todos estos casos, la subsecuencia o región en la cual se alcanza el valor más elevado del índice está contenida en la secuencia autónomamente replicativa ARS18.

A manera de ejemplo, se registran los resultados para  $V = 85$ :

**Figura 4**  
 $i_{max,r}$  de la secuencia M91600 para  $V = 85$



Nota: El mayor valor  $i_{max,r} = 0.16763$  se alcanza en las subsecuencias de puntos medios 170 y 171, de longitud 85.

## CONCLUSIONES

En el análisis del índice de máxima regularidad de la secuencia Z46902, se detectaron dos posibles estrategias de acercamiento al rasgo de interés. Con un tamaño de ventana igual a la longitud del telómero, el  $i_{max,r}$  identificó este rasgo con valores más elevados. De igual manera, todas las regularidades obtenidas con ventanas menores a la longitud del telómero, alcanzaron sus valores máximos en subsecuencias completamente contenidas en este rasgo.

Para las secuencias M91600 y M91601, el  $i_{max,r}$  alcanza valores relativamente bajos. Sin embargo, todos los casos estudiados con tamaños de ventana mayores que 15 y menores que 501 en las cuales el  $i_{max,r}$  alcanza su

nivel más elevado, están contenidas en las secuencias autónomamente replicativas AR18 y ARS68. Ello sugiere que algunos rasgos de interés genético se constituyen por secuencias de menor longitud que presentan altos niveles de regularidad.

El trabajo prospectivo de esta investigación buscará profundizar en las estrategias de aplicación del  $i_{max,r}$  para la detección de regiones de interés genético.

## BIBLIOGRAFÍA

- Abreu, J., González, J., & Rodríguez, Y. (2008). Mejoramiento de la floculación de cepas de levadura cervecera. *Ciencia y Tecnología de los Alimentos*, 18: 22-30.
- Bankar, A., Kumar, A., & Zinjarde, S. (2009). Environmental and industrial applications of *Yarrowia lipolytica*. *Journal Applied Microbiology and Biotechnology*, 84: 847-865.
- Casarégola, S., Feynerol, C., Diez, M., Fournier, P. & Gaillardin, C. (1997). Genomic organization of the yeast *Yarrowia lipolytica*. *Chromosoma*, 106: 380-390
- Folch-Mallol, J., Garay-Arroyo, A., Lledías, F. & Covarrubias A. (2004). La respuesta a estrés en la levadura *Saccharomyces cerevisiae*. *Revista Latinoamericana de Microbiología*, 46: 24 – 46
- Fournier, P., Abbas, A., Chasles, M. Kudla, Ogrzydziak, D.M., Yaver, D., Xuan, J., Peito, A., Ribet, A.-M., Feynerol, C., He, y F. Gaillardin, C. (1993). Colocalization of centromeric and replicative functions on autonomously replicating sequences isolated from the yeast *Yarrowia lipolytica*. *Cell Biology*, 90: 4912-4916.
- Goffeau, A., Barrell, BG., Bussey, H., Davis, RW., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, JD., Jacq, C., Johnston, M., Louis, EJ., Mewes, HW., Murakami, Y., Philippsen, P., Tettelin, H. & Oliver, SG. (1996). Life with 6000 genes. *Science*, 274: 546-7
- Goh, P. & Kilmartin J. (1993). NDC10: A Gene Involved in Chromosome Segregation in *Saccharomyces cerevisiae*. *The Journal of Cell Biology*, 121: 503-512.
- Láscaris-Comneno, T., Skliar, y O. Medina, V. (1999). Determinación de valores del índice de máxima regularidad correspondientes a diversas secuencias de bases de ADN. Un nuevo método computacional en Genética. *Proceedings IX Congreso Internacional de BioMatemática*, pp. 81-97. Concepción, Chile.

- Liti, G., Barton, D. y Louis, E. (2006). Sequence Diversity, Reproductive Isolation and Species Concepts in *Saccharomyces*. *Genetics*, 174: 839-850
- National Center for Biotechnology Information NCBI, (2009). Taxonomy Browser Tool (Gene Bank Database). Consulta realizada el [20/07/2009] desde [<http://www.ncbi.nlm.nih.gov/>]
- Ugalde, A., Morales, Y. y Láscaris-Comneno, T. (2010). Mathematics Applied to the Detection of Genetic Regularities in the Yeast *Yarrowia lipolytica*. *Proceedings of the XVII International Symposium on Mathematical Methods Applied to the Sciences*. San José, Costa Rica.
- Vernis, L., Chasles, M., Pasero, P., Lepingle, A., Gaillardin, y C. Fournier, P. (1999). Short DNA Fragments without Sequence Similarity are Initiation Sites for Replication in the Chromosome of the Yeast *Yarrowia lipolytica*. *Mol. Cell Biol.* 10: 757-769.
- Vernis, L., Abbas, A., Chasles, M., Gaillardin, C., Brun, C., Huberman, & J. Fournier, P. (1997). An Origin of Replication and a Centromere are Both Needed to Establish a Replicative Plasmid in the Yeast *Yarrowia lipolytica*. *Molecular and Cellular Biology*, p.1995-2004.