

Análisis de regularidad para el reconocimiento de telómeros en *Candida parapsilosis mitochondrion* y el cromosoma XVI de *Saccharomyces cerevisiae*

Fecha de recepción: 06/12/2010

Fecha de aceptación: 04/03/2011

Tatiana Láscaris-Comneno Slepuin¹

Alejandro Ugalde León²

Yuri Morales López³

Palabras clave

Índice de máxima regularidad, telómeros.

Resumen

Desde hace varias décadas se sabe que los *telómeros*, o extremos eucarióticos de los cromosomas, tienen propiedades únicas. Las terminaciones naturales de cromosomas normales (no rotos) son estables y no muestran tendencia a fusionarse con otras terminaciones (nativas o rotas). En cambio, los nuevos extremos de cromosomas rotos son “pegajosos” y tienden a fusionarse entre sí, lo que indica que los telómeros deben tener estructuras especiales diferentes de los extremos producidos por rompimiento de cromosomas.

En el presente trabajo se profundiza el estudio sobre la aplicabilidad del índice de máxima regularidad ($i_{\max,r}$) para la detección de telómeros en secuencias de ADN. El $i_{\max,r}$ se aplica específicamente al estudio de los telómeros del cromosoma XVI de

Saccharomyces cerevisiae y el genoma completo de la *Candida parapsilosis mitochondrion*. En este caso, el telómero fue identificado exactamente en la región en que se encuentra documentado.

En el caso del cromosoma XVI las regiones de mayor nivel de regularidad, para longitudes de ventana de hasta aproximadamente 700 pares de bases, quedaron todas comprendidas en las regiones teloméricas. Los resultados obtenidos en los casos estudiados sugieren que el $i_{\max,r}$, conjuntamente con otras características genéticas, es una herramienta eficaz para la detección de telómeros.

Key words

Index of maximum regularity, telomeres.

Abstract

It has been known for several decades that the *telomeres* or eukaryotic extremes of the chromosomes have unique properties. The

1. Escuela de Matemática, Universidad Nacional, Heredia, Costa Rica.
Correo electrónico: tlascaris@una.ac.cr
2. Escuela de Matemática, Universidad Nacional, Heredia, Costa Rica.
Correo electrónico: augald@una.ac.cr
3. Escuela de Matemática, Universidad Nacional, Heredia, Costa Rica.
Correo electrónico: ymorales@una.ac.cr

natural endings of normal cromosomes (not broken) are stable and show no tendency to fuse together with other extremes (native or broken). Instead, the new extremes of broken cromosomes are “sticky” and tend to fuse between them, what indicates that the telomeres must have special structures which are different from the extremes produced when a cromosome breaks. In this analysis we do some additional work studying the applicability of the index of maximum regularity for the detection of telomeres in several DNA sequences.

The $i_{\max,r}$ is applied specifically to the study of the telomeres of the cromosome XVI of *Saccharomyces cerevisiae* and the complete genome of the *Candida parapsilosis mitochondrion*. In this case, the telomere was precisely identified in the region indicated by the scientific literature. In the cromosome XVI, when using windows up to approximately 700bp, all the regions with higher regularity levels were completely contained in the telomeric regions. The results obtained in these cases suggest that the $i_{\max,r}$, together with other genetic characteristics, is an effective tool for telomere detection.

Introducción

Diversos genomas han sido estudiados (Ugalde, Morales y Láscaris-Comneno, 2010a, 2010b; Morales, Ugalde y Láscaris-Comneno, 2010) para determinar la aplicabilidad del *índice de máxima regularidad* ($i_{\max,r}$) para la detección de regularidades en el ADN (Láscaris-Comneno, Skliar y Medina, 1999). En particular, este índice ha sido aplicado, con resultados positivos a la fecha, en la detección de telómeros, lo cual se evidencia en el estudio de la secuencia Z46902, de 8956 pares de bases (Morales et al, 2010).

En este caso, el estudio de la regularidad en esta secuencia permitió identificar el telómero (C1-3)A, rasgo documentado en la base de datos para esta secuencia

por el *National Center for Biotechnology Information* (NCBI), y correspondiente a la región en la cual el $i_{\max,r}$ alcanza su máximo valor en todo este genoma.

En este estudio se presentan los resultados obtenidos en la detección de telómeros en las secuencias NC005253 y NC001148 de *Candida parapsilosis mitochondrion* y *Saccharomyces cerevisiae*, respectivamente.

Telómeros

Los telómeros constituyen estructuras especializadas que forman los extremos de los cromosomas eucariotes y parecen estar implicados en numerosas funciones celulares, especialmente las relacionadas con el control de la duración de la vida de diferentes estirpes celulares. La función de los telómeros es proporcionar estabilidad al cromosoma, haciendo que sus extremos no interaccionen con los extremos de otros cromosomas. Tanto su ADN como las proteínas que los constituyen presentan características singulares que los diferencian del resto de los cromosomas.

Estas estructuras se replican durante el ciclo celular gracias a la acción de enzimas llamadas telomerasas, formadas por proteínas y ARN, y presentan un mecanismo peculiar. Recientemente se ha demostrado su relación con algunas enfermedades, especialmente con el cáncer (Hernández, 1999).

El descubrimiento de la singular estructura nucleoprotéica de los telómeros y su conservación filogenética estructural y funcional demuestran el carácter esencial de esas estructuras para la vida. Elizabeth Blackburn, Carol Greider y Jack Szostak fueron ganadores del Premio Nobel en Fisiología y Medicina en el 2009 gracias al descubrimiento de cómo los cromosomas son protegidos por los telómeros y la enzima telomerasa.

Entre los 6000 y 10 000 pares de bases de cada terminal de un cromosoma, es

posible diferenciar secuencias de ADN telomérico y secuencias asociadas a telómeros. Las secuencias teloméricas consisten de aproximadamente 250 hasta 1500 repeticiones cortas G-ricas *en tandem* (aproximadamente 9 mil pares de bases). Están altamente conservadas a lo largo de distintas especies, lo que refleja la función esencial que desempeñan en el mantenimiento de la integridad de los cromosomas (Passarge, 2001).

En algunos organismos, la longitud promedio de los telómeros responde a cambios genéticos o nutricionales. Las secuencias asociadas al telómero también son repetitivas y se encuentran tanto adyacentes al telómero como dentro del mismo, y contienen secuencias autónomamente replicativas (ARS).

Cromosoma XVI de la levadura *Saccharomyces cerevisiae*

Bradnam, Seoighe, Sharp y Wolfe (1999) señalan que la ventaja de estudiar los cromosomas de *Saccharomyces cerevisiae* como modelo de organismo unicelular es inferir patrones y modelos sobre otros organismos multicelulares.

Aunado a esto, otros grupos de especialistas se han dedicado a descifrar, cromosoma por cromosoma, la estructura de este organismo. En el Cromosoma XVI, trabajos como los de Purnelle, Coster y Goffeau (1996) fueron las bases de la identificación de proteínas y reguladores antes no conocidos.

Estudios recientes (Dolinski y Botstein, 2005) señalan que una década de trabajo luego de la secuencialización completa de este organismo apenas permite vislumbrar la gran cantidad de trabajo por realizar.

Mitocondrias de la *Candida parapsilosis*

Los organismos eucarióticos contienen información genética esencial separada del ADN nuclear en genomas

extracromosómicos llamados mitocondrias. Las mitocondrias de todos los eucariotas, los cloroplastos de plantas verdes y algas contienen moléculas circulares de ADN (mtDNA). Cada célula eucariota contiene entre 10^3 y 10^4 copias. La mitocondria y los cloroplastos son los sitios donde se realizan los procesos esenciales de distribución de energía y de fotosíntesis (Passarge, 2001).

Cada mitocondria está rodeada por dos membranas altamente especializadas, la interna y la externa. Se conoce un número de enfermedades en humanos originadas por mutaciones y supresiones en el mtDNA. Homologías en secuencias indican relaciones evolutivas. En particular, ha sido demostrada la transferencia evolutiva de segmentos de ADN de cloroplastos a mitocondria, y de cloroplastos al ADN nuclear de organismos eucarióticos.

Índice de máxima regularidad y metodología

En trabajos previos se ha explicado detalladamente el proceso de definición del $i_{\max,r}$ (Ugalde, et. al, 2010b). En síntesis, una cadena de ADN se considera como una sucesión de bases nitrogenadas. Para la secuencia de ADN de longitud I_c en estudio, se consideran las $I_c - 1$ secuencias que se generan al efectuar corrimientos, una base a la vez, de modo tal que la última base en la original pase a ser la primera de la segunda secuencia, y así sucesivamente.

La secuencia original, junto con cada una de las $I_c - 1$ generadas, se puede considerar como un conjunto de pares ordenados que pertenecen al conjunto $B \times B$, donde $B = \{A, G, T, C\}$. Adicionalmente, se repite este procedimiento considerando la secuencia original conjuntamente con la que se genera al invertir su orden, y se efectúa posteriormente corrimientos sucesivos desde $d = 1$ hasta $d = I_c - 1$.

Denotando por $n_{A,A,d}$, $n_{A,G,d}$, $n_{A,C,d}$, $n_{A,T,d}$, $n_{G,A,d}$, $n_{G,A,d}$, $n_{G,G,d}$, $n_{G,C,d}$, $n_{G,T,d}$, $n_{C,A,d}$, $n_{C,G,d}$

$n_{C.C.d}$, $n_{C.T.d}$, $n_{T.A.d}$, $n_{T.C.d}$, $n_{T.G.d}$ y $n_{T.T.d}$ la cantidad de pares ordenados efectivamente contados; el $i_{\max,r}$ se define como:

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

donde

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

Con el algoritmo descrito, se analiza el nivel de regularidad de las secuencias seleccionadas mediante el cálculo de los valores de regularidad. Esta información es representada gráficamente y se comparan los resultados obtenidos con los rasgos documentados en las bases de datos del NCBI que presentan una longitud mayor a 20 pares de bases.

Análisis de resultados

En esta etapa del estudio, el $i_{\max,r}$ fue computado para el genoma completo de la *Candida parapsilosis* *mitochondrion* (código de referencia NCBI: NC_005253) y para secciones de la secuencia del cromosoma XVI de la levadura *Saccharomyces cerevisiae* (código de referencia NCBI: NC_001148).

A continuación, se presentan los resultados más relevantes del estudio de los telómeros y las regiones teloméricas asociadas en ambos genomas indicados.

Secuencia NC_005253 (32745 pares de bases): “*DEFINITION Candida parapsilosis mitochondrion, complete genome*”.

La secuencia completa del ADN mitocondrial de la levadura patógena *Candida parapsilosis* fue determinada por Nosek et al (2004). El genoma mitocondrial está representado por moléculas de ADN lineales que concluyen con repeticiones de una unidad de 738 pares de bases.

El número de repeticiones varía, lo cual genera así una población de moléculas lineales de ADN que son heterogéneas en tamaño. La longitud de las moléculas más cortas es 30 922 pares de bases, en tanto que las moléculas más largas tienen terminales expandidas con matrices de repeticiones (738 pares de bases).

Los rasgos relativos a regiones teloméricas documentados en la base de datos son los siguientes:

repeat region “left telomeric tandem array”, ubicado entre las bases 1 y 1348.

repeat region “left subterminal repeat”, ubicado entre las bases 1349 y 1902.

repeat region “right subtelomeric repeat”, ubicado entre las bases 30844 y 31397.

repeat region “right telomeric tandem array”, ubicado entre las bases 31398. y 32745.

Resultados observados

Las regiones teloméricas de las terminales izquierda y derecha tienen cada una 1347 bases. Se analizaron los valores del $i_{\max,r}$ con un tamaño de ventana $V = 1347$, el cual coincide con la longitud de los rasgos 1 y 4, con el fin de determinar el grado de regularidad que presentan ambos telómeros.

Los resultados obtenidos se presentan en el gráfico 1.

El análisis del genoma completo de la *Candida parapsilosis mitochondrion* indica que el $i_{\max,r}$ alcanza sus valores más elevados en las regiones o subsecuencias que corresponden a los dos telómeros. Es decir, el $i_{\max,r}$ claramente identifica las regularidades teloméricas izquierda y derecha, y las señala como las regiones de este genoma con niveles más elevados de regularidad. El detalle se muestra en el cuadro 1.

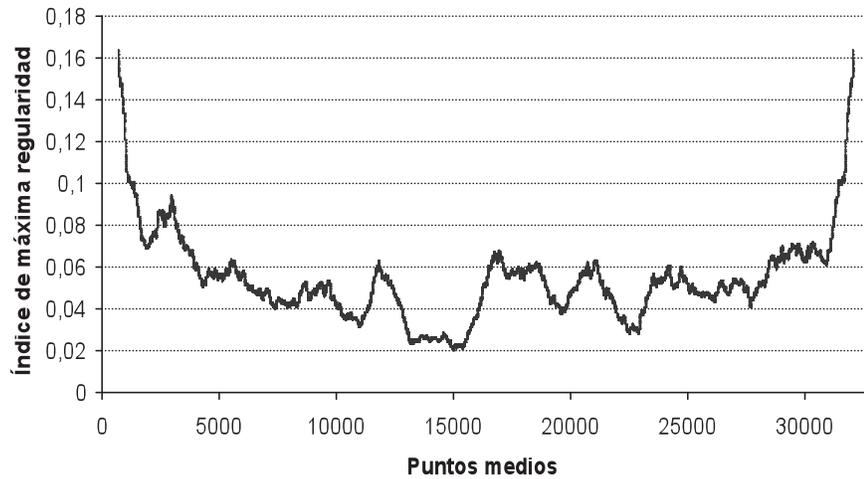


Gráfico 1. $i_{\max,r}$ de *Candida parapsilosis* mitochondrion, para $V=1347$

Nota: Los valores mayores del $i_{\max,r}$ son 0,16375959 y se alcanzan exactamente en las dos regiones telómericas documentadas, en las subsecuencias de puntos medios 675 y 32071.

Note que la aplicación del $i_{\max,r}$ ha permitido detectar que ambos telómeros son complementarios, dato al que no se hace referencia en la documentación del genoma en estudio.

Se analizó también el comportamiento de los valores del $i_{\max,r}$ en este genoma para un tamaño de ventana $V = 739$, que, tal como se menciona anteriormente, constituye una de las unidades de repetición de los extremos de las moléculas del genoma mitocondrial.

Los resultados obtenidos se presentan en el gráfico 2.

Los resultados generados se presentan en el cuadro 2.

En este caso, la aplicación del $i_{\max,r}$ detecta como las regiones con los valores

máximos regularidad a dos regiones de los extremos teloméricos del genoma de la *candida parapsilosis* mitochondrion, que, de nuevo, tienen los mismos valores de regularidad y son subsecuencias complementarias.

Secuencia NC_001148 (948062 pares de bases): “DEFINITION *Saccharomyces cerevisiae* chromosome XVI, complete sequence”

Bussey et al (1997) identificaron la secuencia nucleótida de los 948 061 pares de bases del cromosoma XVI, con lo que se completa así la secuencia del genoma de la levadura. La arquitectura de este cromosoma parece ser similar a la que presentan los cromosomas grandes de la levadura y presenta grandes coincidencias con cromosomas de otras levaduras. El

Cuadro 1.

Punto medio	Subsecuencia	$i_{\max,r}$	Composición
675	[1, 1349] ₁₃₄₇	0,16375959	(458A, 685T, 72G, 132C)
32071	[31397, 32745] ₁₃₄₇	0,16375959	(685A, 458T, 132G, 72C)

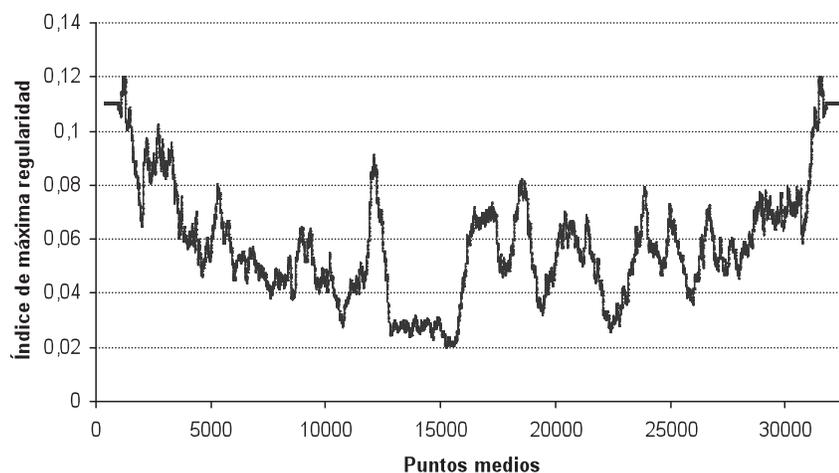


Gráfico 2. $i_{\max,r}$ de *Candida parapsilosis* mitochondrion, para $V=739$

Nota: Los valores mayores del $i_{\max,r}$ son 0,12003482 y se alcanzan en las subsecuencias de puntos medios 1255, 1256, 31490 y 31491.

Cuadro 2.

Punto medio	Subsecuencia	$i_{\max,r}$	composición
1255	[886, 1624] ₇₃₉	0,12003482	(266A, 375T, 37G, 61C)
1256	[887, 1625] ₇₃₉	0,12003482	(266A, 375T, 37G, 61C)
31490	[31121, 31859] ₇₃₉	0,12003482	(375A, 266T, 61G, 37C)
31491	[31122, 31860] ₇₃₉	0,12003482	(375A, 266T, 61G, 37C)

cromosoma XVI contiene potencialmente 487 codificadores de proteínas, 17 genes tRNA y dos pequeños genes de ARN nuclear; 27% de los genes tienen similitudes significativas con productos de genes humanos y 48% no tiene función biológica conocida.

De acuerdo con la documentación del NCBI, el cromosoma XVI de *Saccharomyces cerevisiae* tiene regiones teloméricas en sus terminales derecha e izquierda. La región telomérica de la terminal izquierda está constituida por 7223 pares de bases y está compuesta por las demás regiones teloméricas documentadas de dicha terminal. Similarmente, la región telomérica de la terminal derecha del cromosoma XVI, constituida por 5671

pares de bases, está compuesta por dos elementos teloméricos.

A continuación, se analizarán, por separado, los valores del $i_{\max,r}$ en los rasgos teloméricos de las terminales izquierda y derecha del cromosoma XVI.

Terminal izquierda del cromosoma XVI de *Saccharomyces cerevisiae*

Este análisis se realizó considerando la subsecuencia constituida por los primeros 14 000 pares de bases de la secuencia completa del cromosoma XVI.

Los rasgos relativos a regiones teloméricas documentados en la base de datos son los siguientes:

1. Repeat region “Telomeric region”, señalado en la base de datos como ubicado entre las bases 1 y 7223.
2. Repeat region “Telomeric long Y’ element”, ubicado entre las bases 1 y 6524.
3. Repeat region “Telomeric X element combinatorial repeat region”, ubicado entre las bases 6525 y 6777.
4. Repeat region “Telomeric X element Core sequence”, ubicado entre las bases 6778 y 7223.

Se calcularon los valores del $i_{\max,r}$ para tamaños de ventana de 85, 125, 161, 191, 253, 301, 325, 447, 651, 1301, 2601, 3901, 6525 y 7223 pares de bases, algunas de las cuales corresponden a longitudes de regiones teloméricas contenidas en el rasgo “Telomeric region”, que, en la base de datos, se encuentra entre las bases 1 y 7223. A continuación, se presentan los resultados obtenidos para los casos $V=253$ y $V=447$.

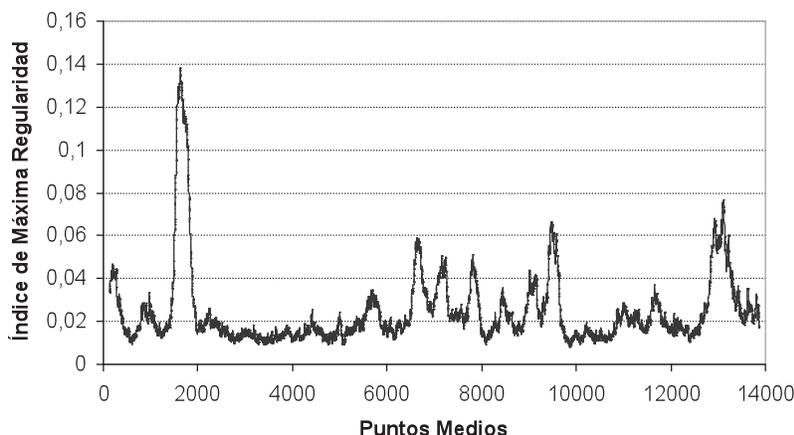


Gráfico 3. $i_{\max,r}$ de las regiones teloméricas de la terminal izquierda, para $V=253$

Nota: El mayor valor de $i_{\max,r}$ es 0,13808787 y se alcanza en la subsecuencia de punto medio 1628 de longitud 253.

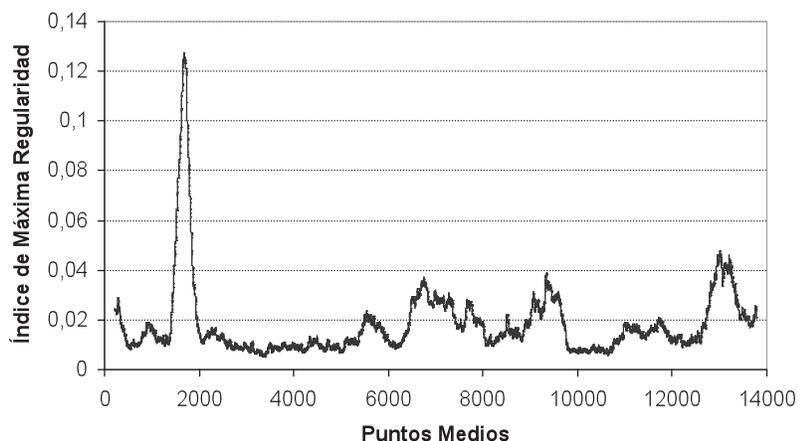


Gráfico 4. $i_{\max,r}$ de las regiones teloméricas de la terminal izquierda, para $V=447$

Nota: El mayor valor de $i_{\max,r}$ es 0,1273127 y se alcanza en la subsecuencia de punto medio 1679 de longitud 447.

Terminal derecha del cromosoma XVI de *Saccharomyces cerevisiae*

Este análisis se realizó tomando en consideración la subsecuencia constituida por los últimos 14 000 pares de bases de la secuencia completa del cromosoma XVI.

Los rasgos relativos a regiones teloméricas documentados en la base de datos son los siguientes⁴:

1. *Repeat region "Telomeric region"*, ubicado entre las bases 942392 y 948062.
2. *Repeat region "Telomeric X element Core sequence"*, ubicado entre las bases 942392 y 942767.
3. *Repeat region "Telomeric short Y' element"*, ubicado entre las bases 942806 y 948062.

Se calcularon los valores de $i_{\max,r}$ para tamaños de ventana de 85, 125, 161, 191, 253, 447, 1301, 2601 y 3901 pares de bases. A continuación, se presentan los resultados obtenidos con $V=253$ y $V=447$.

En el gráfico 5 se observa el $i_{\max,r}$ de la sección derecha de la secuencia completa del cromosoma XVI de *Saccharomyces cerevisiae*, analizado en ventanas de tamaño de 253 pares de bases.

Análogamente, en el gráfico 6 se presentan los valores del $i_{\max,r}$ y se analiza la sección derecha con ventanas de tamaño de 253 pares de bases.

En el estudio del cromosoma XVI en sus dos extremos, se puede observar que las zonas estudiadas que presentan mayor nivel de regularidad están completamente contenidas en las zonas teloméricas documentadas. Esto ocurre para valores de ventana de hasta aproximadamente 700 pares de bases. Para tamaños de ventana más grandes, los valores máximos del $i_{\max,r}$ se alejan de las regiones teloméricas.

Conclusiones

En ambos genomas estudiados, el $i_{\max,r}$ detectó los telómeros en los extremos de los cromosomas, en donde este índice alcanza los valores máximos.

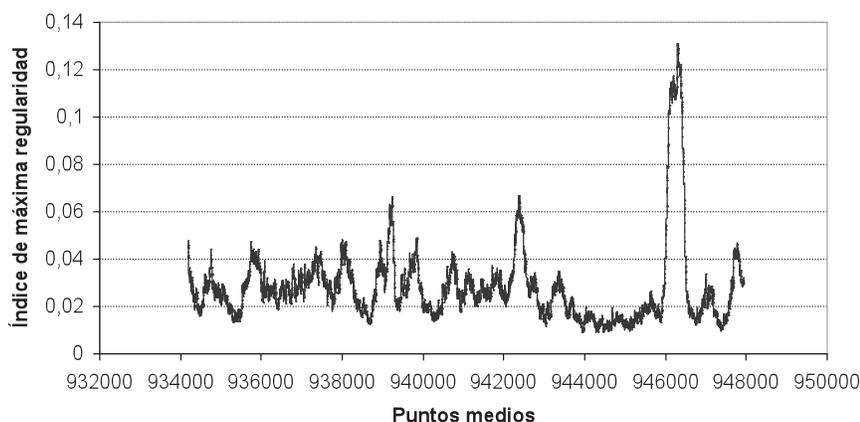


Gráfico 5. $i_{\max,r}$ de las regiones teloméricas del terminal derecha, para $V=253$

Nota: El mayor valor de $i_{\max,r}$ es 0,13082223 y se alcanza en la subsecuencia de punto medio 946303 de longitud 253.

4. Nótese que el rasgo 1 contiene a los rasgos 2 y 3.

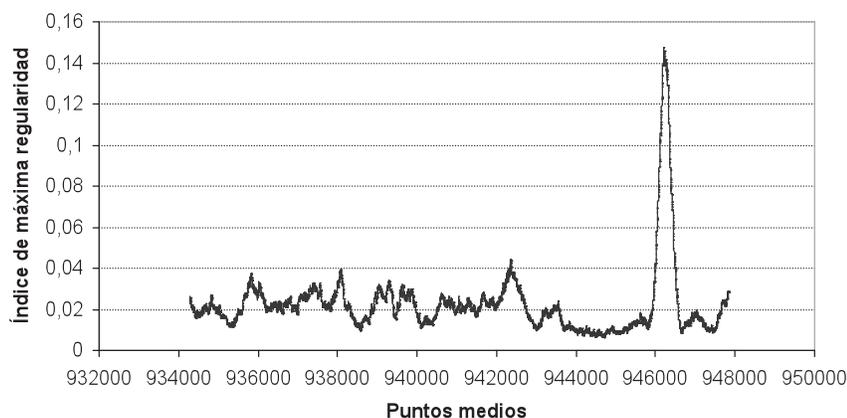


Gráfico 6. $i_{\max,r}$ de las regiones teloméricas de la terminal derecha, para $V=447$

Nota: El mayor valor de $i_{\max,r}$ es 0,14750587 y se alcanza en las subsecuencias de puntos medios ubicados entre 946205 y 946209, de longitud 447.

La aplicación del $i_{\max,r}$ permitió detectar subsecuencias complementarias, respecto a su composición, con niveles máximos de regularidad.

La información obtenida a la fecha sugiere que el $i_{\max,r}$ es una herramienta efectiva para la detección de telómeros. Las regiones de los genomas estudiados, en las cuales se alcanzan los mayores valores del $i_{\max,r}$, corresponden con las subsecuencias en las cuales han sido documentados los telómeros.

Esto permite considerar como hipótesis de trabajo que la aplicación del $i_{\max,r}$, conjuntamente con otras características, permitirá detectar telómeros aún no documentados en la bibliografía científica.

Bibliografía

Bradnam, K., Seoghe, C., Sharp, P., & Wolfe, K. (1999) *G+C Content Variation Along and Among Saccharomyces cerevisiae Chromosomes*. *Mol. Biol. Evol.* 16(5), 666–675.

Bussey, H., Storms, R., Ahmed, A., Albermann, K., Allen, E., Ansorge, W., Araujo, R., Aparicio, A., Barrell, B., Badcock, K., Benes, V., Botstein, D., Bowman, S., Bruckner, M., Carpenter, J., Cherry, J.M., Chung, E.,

Churcher, C., Coster, F., Davis, K., Davis, R.W., Dietrich, F.S., Delius, H., DiPaolo, T., Hani, J. (1997). *The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI*. *Nature* 387, 103-105.

Dolinski, K. & Botstein, D. (2005). *Changing perspectives in yeast research nearly a decade after the genome sequence*. *Genome Res.* 15, 1611-1619. Access at doi:10.1101/gr.3727505.

Hernández, R. (1999) *Telómeros y Telomerasas*. *Revista Cubana de Investigación Biomédica*. 18(2), 121-129.

Láscaris-Comneno, T., Skliar, O. y 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*, Concepción, Chile: (pp. 81-97).

Morales, Y., Ugalde, A. y Láscaris-Comneno, T. (2010) *Análisis de regularidad de genomas para detección de telómeros y secuencias autónomamente replicativas*. *Revista UNICIENCIA*. Manuscrito aceptado para publicación.

National Center for Biotechnology Information NCBI. *Taxonomy Browser Tool* (Gene Bank Database). Consultado el 20 de 07 del 2009 desde <http://www.ncbi.nlm.nih.gov/>

- Nosek, J., Novotna, M., Hlavatovicova, Z., Ussery, D.W., Fajkus, J., & Tomaska, L. (2004). *Complete DNA sequence of the linear mitochondrial genome of the pathogenic yeast Candida parapsilosis*. *Molecular Genetics Genomics*, 272, 173-180.
- Passarge, E. (2001). *Color Atlas of Genetics*. Germany: Georg Thieme Verlag.
- Purnelle, B., Coster, F., & Goffeau, A. (1996). *The sequence of 55 kb on the left arm of yeast chromosome XVI identifies a small nuclear RNA, a new putative protein kinase and two new putative regulators*. *Yeast*, 12. Recuperado de: <http://www.biomedsearch.com/nih/sequence-55-kb-left-arm/8948103.html>
- Ugalde, A., Morales, Y., & Láscaris-Comneno, T. (2010a). *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*
- Ugalde, A., Morales, Y., & Láscaris-Comneno, T. (2010b). *Mathematics Applied to the Detection of Genetic Regularities in the Yeast Yarrowia lipolytica*. *Revista de Matemática: Teoría y Aplicaciones*. Manuscrito remitido para publicación,