

## Resumen

Con base en los análisis del espectro de potencias para secuencias de bases (nucleótidos) de varios genes, estudiamos correlaciones de largo alcance dentro del total de secuencias de bases que son expresadas como  $1/f^\alpha$ . El comportamiento del exponente para la secuencia de bases acumulada, así como para periodicidades de corto alcance, están dentro tanto del espectro de potencias como del promedio para las bases individuales. Observamos de tales análisis que podemos extraer diversa información independientemente de aquella proveniente de las secuencias de pares de bases obtenidas de los codones. También hemos llevado a cabo análisis comparativos usando la función de información mutua para una muestra de datos y para datos aleatorios, y hemos exhibido diferencias distintivas, particularmente en la región con una menor distancia entre bases.

## 1 Introducción

Dentro del proyecto del genoma humano la principal atención se ha enfocado en la determinación de todas las secuencias de ADN del genoma humano. En la actualidad cuando se ha logrado esta meta, la prioridad está dándose a los análisis de la función de los genomas, llamado análisis post-genoma. Sin embargo la técnica de información-matemática para tales análisis todavía no se establece y su desarrollo urgente está esperando.

Como uno de rasgos importantes de las secuencias de ADN en los eucariontes, se puede señalar la existencia de la repetición periódica de secuencias. Tal repetición puede existir en el rango corto así como en el largo (por ejemplo la secuencia Alu). Este tipo de periodicidades podría ser explicable en términos de la duplicación de genes defendida por Ohno (1). De acuerdo con este mecanismo se piensa que los genes que parecían ser generados recientemente en los procesos de evolución no eran nuevos en un verdadero significado, sino que se generaron de la tediosa copia de los genes que ya habían existido y habían jugado un papel modulador, particularmente en la fase inicial de evolución. De hecho se sabe que los genes están constituidos por genes particulares con longitud individual más corta que puede tener funciones particulares.

Sin embargo el significado evolutivo del por qué existen repeticiones de secuencias tales como la Alu todavía no se entiende bien. Mediante la generación de datos simbólicos

unidimensionales que consisten en 0 y 1 dentro del armazón del modelo del autómatas celular que contiene las operaciones análogas para el proceso de repetición en la expansión de genes así como para el proceso de mutación Li pudo mostrar la existencia de correlaciones de largo alcance dentro de estas (2) (3). De la investigación con este modelo se puede anticipar no sólo las periodicidades sino también las correlaciones de largo alcance que podrían existir en las secuencias de ADN reales. Li y Kaneko examinaron la función de correlación de secuencias específicas de ADN (4). Aunque ellos no pudieron deducir resultados suficientemente claros debido a que las secuencias de ADN no son lo suficientemente largas para satisfacer estadísticamente el espectro de potencias obtenido por estos, se sugirió la forma de una correlación de largo alcance:  $1/f^\alpha$ . Subsecuentemente cuando el espectro de potencias se despliega en un gráfico con escala logaritmo-logaritmo, el exponente  $\alpha$  mostrara la caída de la línea recta,  $\alpha$  lo llamaremos simplemente como pendiente.

Tiempo después las correlaciones de largo alcance dentro de las secuencias de ADN también eran confirmadas por Peng y cols.(5) y Voss (6) mediante otra aproximación. Es más, Takushi y cols. han examinado la relación entre la pendiente  $\alpha$  y el contenido promedio del intrón para el gen de supresor tumoral y ha sostenido la existencia de una posible correlación entre el comportamiento de la fase transitoria de  $\alpha$  para una secuencia de bases acumulada y las probabilidades de generación de cáncer debido a la anormalidad del gene supresor tumoral (7).

En este artículo analizamos la presencia de las periodicidades y las correlaciones de largo alcance de varias secuencias de ADN mediante la evaluación de los datos de espectros de potencias. Primero procedemos a convertir base en un dato numérico para después calcular su transformada de Fourier. Examinamos rasgos estructurales de las secuencias de ADN para varios genes (VHL del Homo Sapiens y el homólogo del gen humano NF(1).

Organizamos este artículo como sigue. En la sección 2 describimos el método del cálculo. La sección 3 se dedica al análisis de los datos de las secuencias de bases de varios genes desde diferentes puntos de vista en base al análisis del espectro de potencias, i.e. correlaciones de largo alcance dentro del total de secuencias de bases, los valores de las pendientes  $\alpha$  en dependencia de la longitud de secuencias de bases acumuladas

y periodicidades de corta distancia. En la sección 4 también realizamos los análisis usando la función de información mutua. Comparamos la función de información mutua para los datos de la muestra y los datos aleatorios y señalamos la diferencia clara entre ellos. En la sección 5 resumimos los resultados obtenidos en este artículo. En el Apéndice demostramos que el cálculo numérico directo de tres métodos de conversión de las secuencias de bases en secuencias numéricas no producen ninguna diferencia esencial en el comportamiento de los espectros de potencia resultantes.

## 2 Método de cálculo.

En esta Sección describimos el método del cálculo de los espectros de potencia de secuencias de ADN del cual podemos examinar la periodicidad y las correlaciones (como la correlación de largo alcance así como la correlación de corto alcance). Primero se convierte la secuencia de ADN, la cual consiste en cuatro tipos de bases (A, T, G, C) en un código numérico. Entonces es posible evaluar los espectros de potencia y examinar varios rasgos estructurales ocultos en la secuencia de bases.

El espectro de potencia puede obtenerse mediante Fourier transformando la función de autocorrelación sobre los datos numéricos a través del teorema de Wiener-Khinchin. Denotemos una secuencia de bases como  $X(n)$ , entonces la función de autocorrelación  $X_b(t)$  de la base  $\beta$  ( $\beta$ : A, T, G, C) es evaluada como:

$$X_b(t) = \langle X_b(n)X_b(n+t) \rangle_b \quad (\text{Ec. 1})$$

Esta función toma el valor siguiente: Si en una secuencia de bases  $X(n)$ , dos bases que tienen una distancia  $\tau$  pueden ser:

$$\begin{aligned} X_b(t) &= 1, \text{ si dos bases son idénticas.} \\ X_b(t) &= 0, \text{ si dos bases son diferentes.} \end{aligned} \quad (\text{Ec.2})$$

Se obtiene el espectro de potencia  $S_b(f)$  tomando la transformada de Fourier de  $X_b(t)$  como sigue:

$$S_b(f) = \frac{1}{N} \left| \sum_{n=0}^{N-1} X_b(n) \exp(-2\pi i f n) \right|^2 \quad (\text{Ec. 3})$$

donde  $N$  es el número de bases, y  $f^k$  es la frecuencia ( $\sim 1/\text{número de bases}$ ) definida como:

$$f^k = \frac{k}{N}; k=1,2,\dots, \frac{N-1}{2}; \quad (\text{Ec. 4})$$

De (4) tenemos  $f^k \leq 0.5$ . Si tratamos una secuencia de bases dada de un ciclo como un todo, entonces el espectro de potencia  $S_b(f)$  parece ser simétrico con respecto al límite en  $f^k=0.5$  y entonces  $S_b(0.5\sim 1.0)$  es idénticamente igual a  $S_b(0\sim 0.5)$ .

El espectro de potencia promedio de las cuatro bases:  $S_\sigma(f)$  es calculado de la fórmula (3) como sigue:

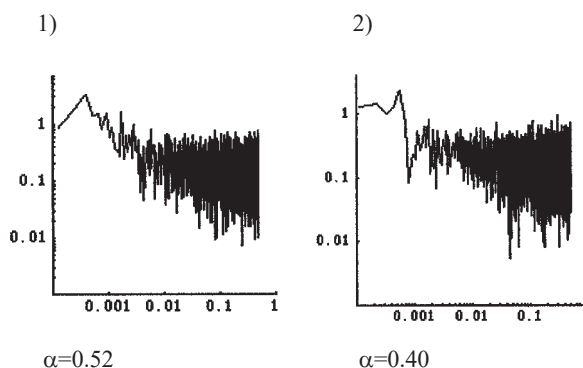
$$S_\sigma(f) = \frac{1}{4} \sum_{b:A,G,C,T} S_b(f) \quad (\text{Ec. 5})$$

Si se grafica el espectro de potencia  $S_\sigma(f)$  contra  $f$  en una escala logaritmo-logaritmo puede ser aproximado por una línea recta en una región de bajas frecuencias como  $S_\sigma(f) \sim 1/f^\alpha$ , lo que significa que hay una correlación de largo alcance. La pendiente puede ser determinada por el método de mínimos cuadrados.

## 3 Análisis de datos

### 3.1 Correlaciones de largo alcance en los espectros de potencia de varios genes

Utilizamos como secuencias de bases las registradas en la Base de datos de ADN de Japón (DDBJ) (<http://www.ddbj.nig.ac.jp/Welcome-j.html>). Presentamos los gráficos de espectros de potencia de varios genes en Fig.1, donde los valores de la pendiente  $\alpha$  también se dan. Se pueden observar las correlaciones de largo alcance en cada gráfico de 1) 3) 5) y 6). En 2) no se observa una correlación excepto en la región de baja frecuencia. Contrariamente en cada gráfica de 4) y 5) se puede observar un pico en un región de frecuencia aunque en otras regiones los gráficos exhiben comportamientos cercanos a los del ruido blanco. Estos picos corresponden a la periodicidad debido a una repetición de tres bases. Para examinar la presencia de las correlaciones de largo alcance por el método presente necesitamos tener 1000 bases por lo menos. En el caso de 2) que consiste sólo en intrones, no pudimos observar ninguna periodicidad. El patrón característico del espectro aparece en la mayoría de los gráficos como los unos vistos en los casos de 1) y 3).



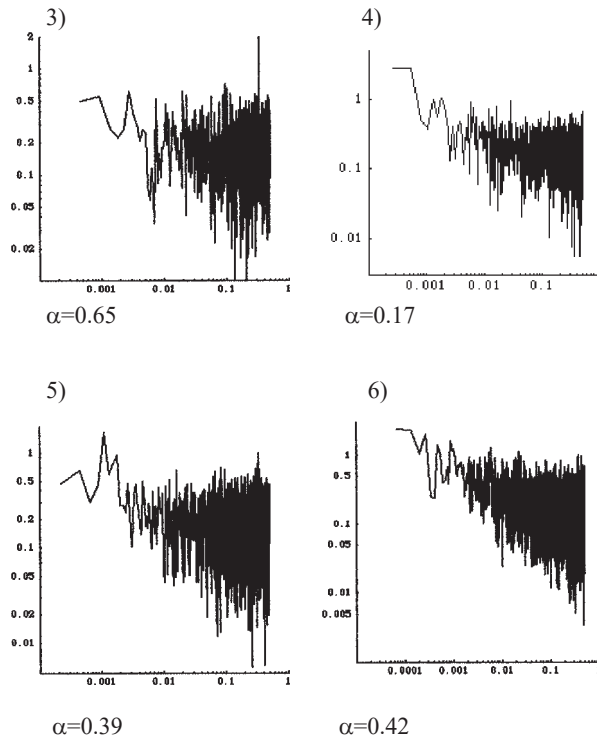


Fig.1 Espectros de potencia de secuencias de bases para los varios genes. El eje vertical corresponde a la densidad del espectro de potencia mientras que el eje horizontal para  $f_k = k / N$  (siendo el número total de la secuencia de bases,  $k = 1, 2, \dots, (N-1) / 2$ ) que satisface la condición  $f_k \leq 0.5$ . Los siguientes genes se han usado para dibujar los gráficos:

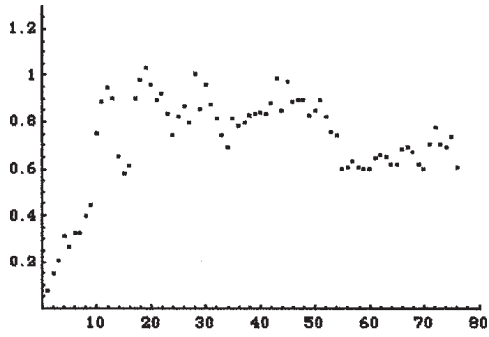
- 1) Gen homólogo humano NF1. NF1 (homo): Homólogo del gen represor de neurofibromatosis. El número de bases es 7675.
- 2) intrón 5 del gen histamina N-metilo transferasa del Homo sapiens (HNMT). Este es un intrón que forma parte del gen que transcribe el fermento que reestructurar el radical N metil de histamina. El número de bases es 8554.
- 3) El gen humano BRCA1 cds parcial el cual es una parte de gen del represor de la familia de los cánceres del pecho humanos. El número de bases es 3798.
- 4) Gen de la proteína p56 relacionado con la 2'-5' oligoadenilato sintetiza del Homo sapiens 2'-5'. Es un gen que transcribe proteína que relaciona con la síntesis del fermento del ácido 2'-5' oligoadenilil de Homo Sapiens. El número de bases es 2155.
- 5) ARNm completo cds de la proteína relacionada p130. El número de bases es 4513.
- 6) el gen supresor tumoral de von Hippel-Lindau del Homo Sapiens. El número de bases es 14543.

### 3.2 Valores de la pendiente en dependencia de la longitud de las secuencias de bases acumuladas.

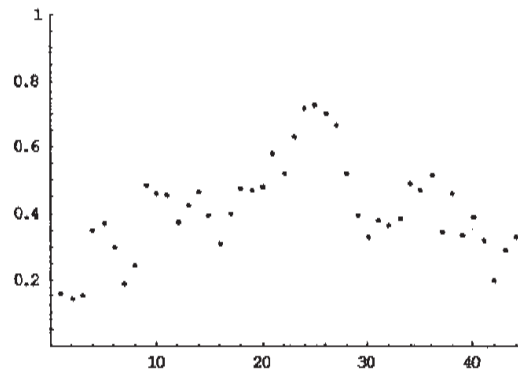
En general, el valor de la pendiente  $\alpha$  varía dependiendo del número de bases acumuladas. Evaluamos los espectros

de potencia aumentando el número de bases desde la base número 1 por 100, 200, 300 y entonces la pendiente  $\alpha$  para cada espectro de potencia. En la Fig.2 graficamos  $\alpha$  contra el número de bases acumulado. Particularmente, observamos en el gráfico del gen supresor tumoral p51 un cambio claro que se parece a una transición de fase de primero-orden alrededor de la base 450 (vea también (7)). Observamos cambios similares alrededor de la base 1400 en HUMCFVII y de alrededor de la base 1000 en el ARNm del APC de ratón. La gráfica respectiva de los genes analizados tienen varios contornos peculiares, en cada una, y esto hace difícil averiguar algún rasgo común. Por ejemplo en el gráfico de HUMCFVII (gráfico 8) de la Fig.2 la pendiente aumenta junto con un incremento del número de bases manteniendo un valor constante alto ( $\alpha \sim 1.3$ ) después de alrededor de la base 4000. En otros genes también podemos observar rasgos como aumento o disminución de la pendiente después de alcanzar el umbral con un cierto número de bases acumulado. De tales gráficas uno puede suponer que el valor a cambiar no simplemente depende del número de bases sino de la estructura de las secuencias de bases. Esto puede verse claramente en el gráfico 9) de t-s p16 MTS1/CDKN2 en la Fig.2 donde  $\alpha$  toma el valor más alto ( $\alpha \sim 0.7$ ) en la vecindad de la base 4100 pero disminuye casi hasta cero al final de la región.

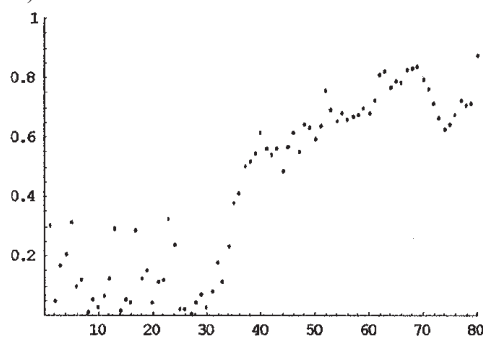
1)



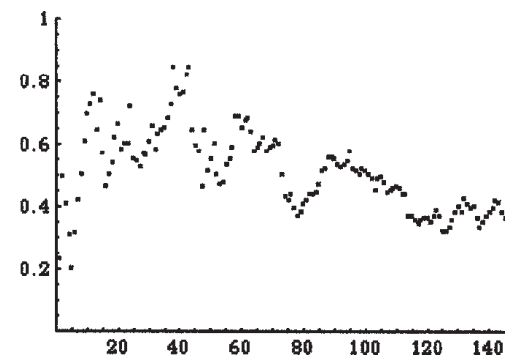
6)



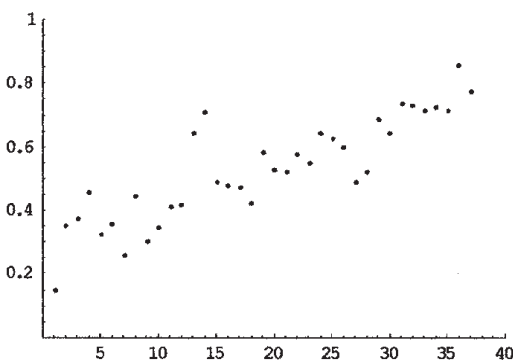
2)



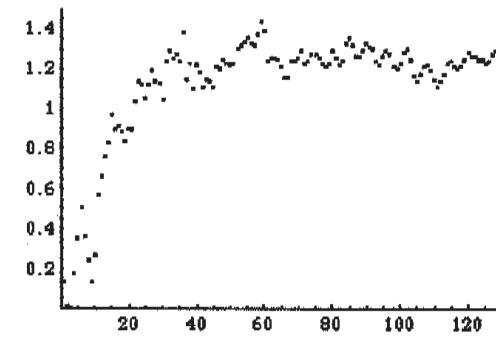
7)



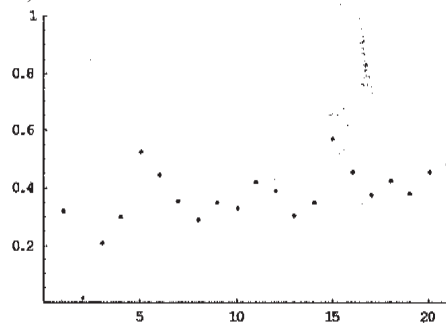
3)



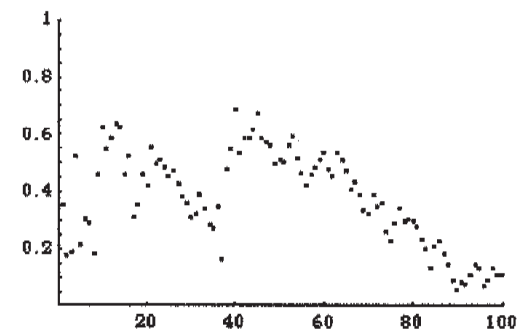
8)



4)



9)



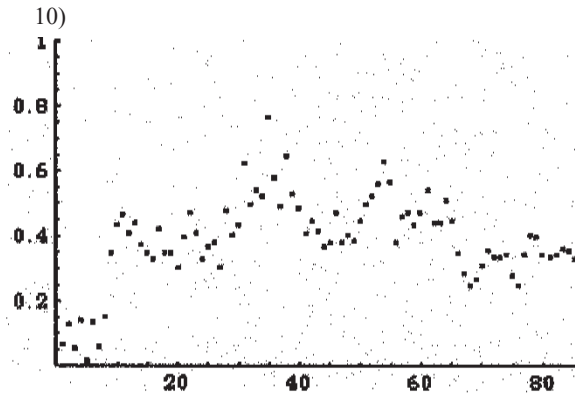


Fig.2. El valor de la pendiente para el acumulado de secuencias de bases. El eje vertical corresponde al valor de la pendiente y el eje horizontal al número de bases/100. Los gráficos del 1)-6) corresponden a los genes de la Fig.1 mientras que los gráficos de 7) 10) a los genes siguientes. 7) El gen supresor tumoral p51. El número de bases es 2816. 8) HUMCFVII (el factor de coagulación sanguíneo VII de humano). El número de bases es 12850. 9) el t-s p16 MTS1/CDKN2. El número de bases es 10020. 10) ARNm de APC de ratón (proteína de poliposis adenomatosa coli). El número de bases es 8505.

Entonces surge una pregunta: ¿cuál es el significado de las estructuras en las secuencias de bases? Basado en la comparación de las gráficas en la Fig.2 suponemos que la proporción satisfecha entre exones y/o intrones así como su arreglo son los factores cruciales que causan el cambio de la pendiente  $\alpha$ . Para verificar tal hipótesis examinamos cómo están colocados los exones en el rango donde  $\alpha$  se mantuvo alto (entre las bases 3000-12850) en la gráfica de HUMCFVII (el número total de secuencia de bases es 12850). Los resultados se resumen en tabla 1.

exon 1	exon 2	exon 3
522 ... 585	1654 ... 1719	4294 ... 4454
exon 4	exon 5	exon 6
6383 ... 6407	6478 ... 6591	8307 ... 8447
exon 7	exon 8	exon 9
9419 ... 9528	10124 ... 10247	11064 ... 11659

TABLA 1. ARREGLO DE PARTE DEL EXON EN HUMCFVII

De esta tabla confirmamos que los exones e intrones se repiten alternadamente en el rango donde  $\alpha$  se mantiene alto

de la gráfica 8.

### 3.3 Periodicidad para abreviar el rango.

En esta Sección examinamos las periodicidades en la secuencia de bases de HUMCFVII (12850 bases) dentro de la región de corto alcance. Para este propósito dividimos la secuencia de bases en un intervalo corto (100 bases) y evaluamos el espectro de potencia en cada intervalo para verificar en que región y en que tamaño hay repeticiones de secuencia de bases. Como se muestra en la Fig.3 hay picos claros, aparecen con período aproximado entre la base 2000 y 3000 de la secuencia de bases.

En la Fig.3 el tamaño de la repetición de la secuencia de bases es  $100/\alpha$  (valor del eje horizontal en el pico) mientras el eje vertical representa la densidad del espectro de potencia que muestra cuántas secuencias de bases de un tamaño repetido son incluidas. Note que esta figura no es trazada en una escala logaritmo-logaritmo. Como resultado de lo anterior para un intrón, aparece un pico en  $6n$  ( $n=1,2,3,\dots$ ) en el eje horizontal el tamaño de la repetición es  $100/6 \approx 17$  bases. De hecho observamos que un conjunto de 17 bases (CCCGGGGCG-TGGGTGT) aparece repetidamente en el rango considerado. En este rango también igualmente observamos frecuentemente secuencias como (CCAGGAGTCGTGGGTGT) donde algunos caracteres difieren de los regulares. De igual modo pudimos ver claramente la apariencia de las repeticiones con  $100/3 \approx 33$  bases en los análisis de corto alcance dentro del rango de entre 4800~5400 bases (parte del intrón) para los espectros de potencia del oncogen T24 de la vejiga humana.

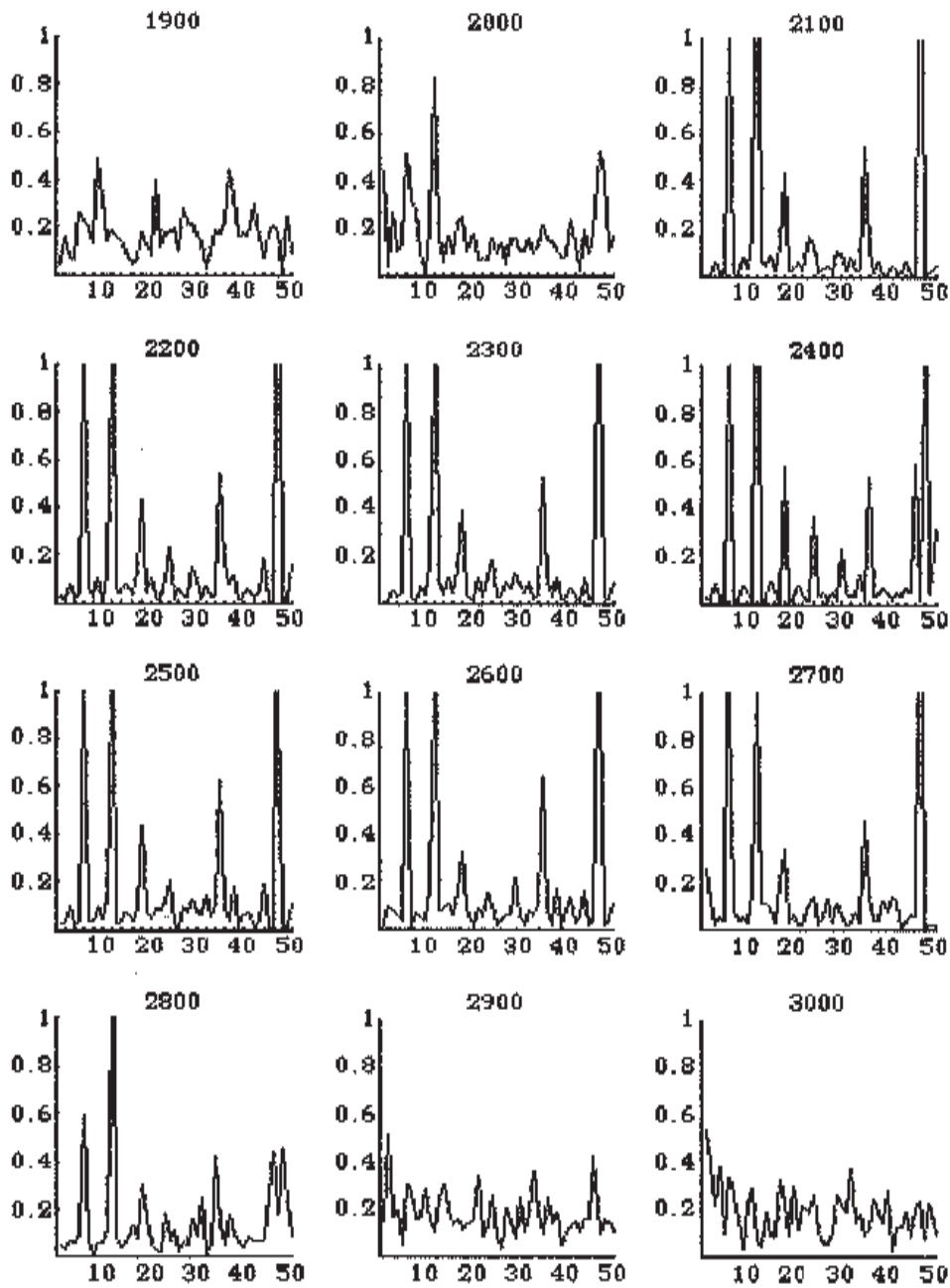
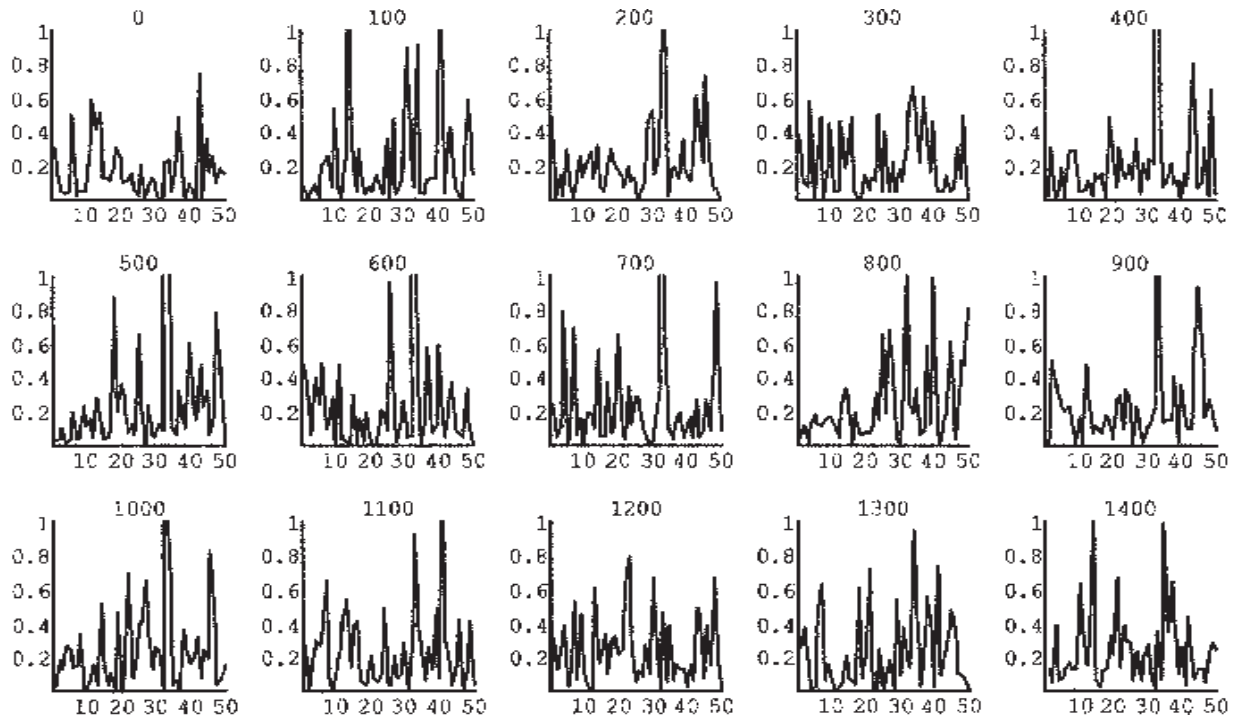
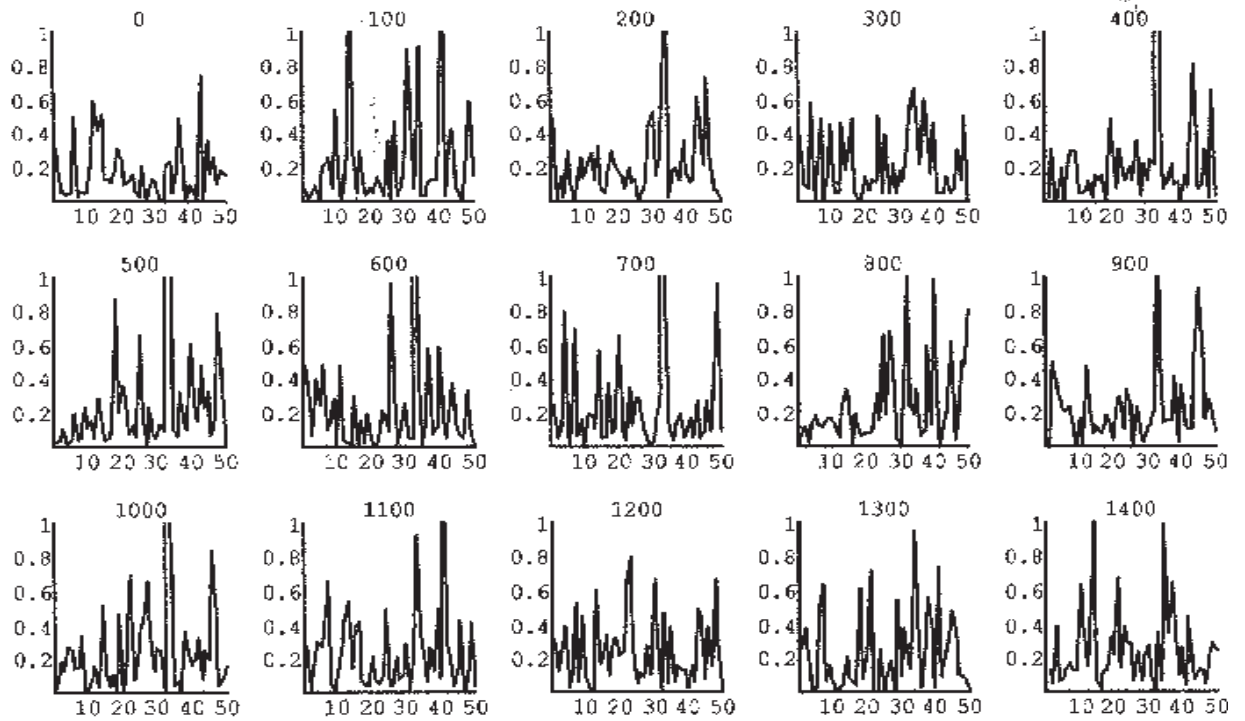


Fig.3. Periodicidad de corto alcance en las secuencias de bases del HUMCFVII. El tamaño de la repetición de la secuencia de bases es 100 / (un valor del eje horizontal en el pico) en tanto que el eje vertical representa la densidad de espectro de potencia.

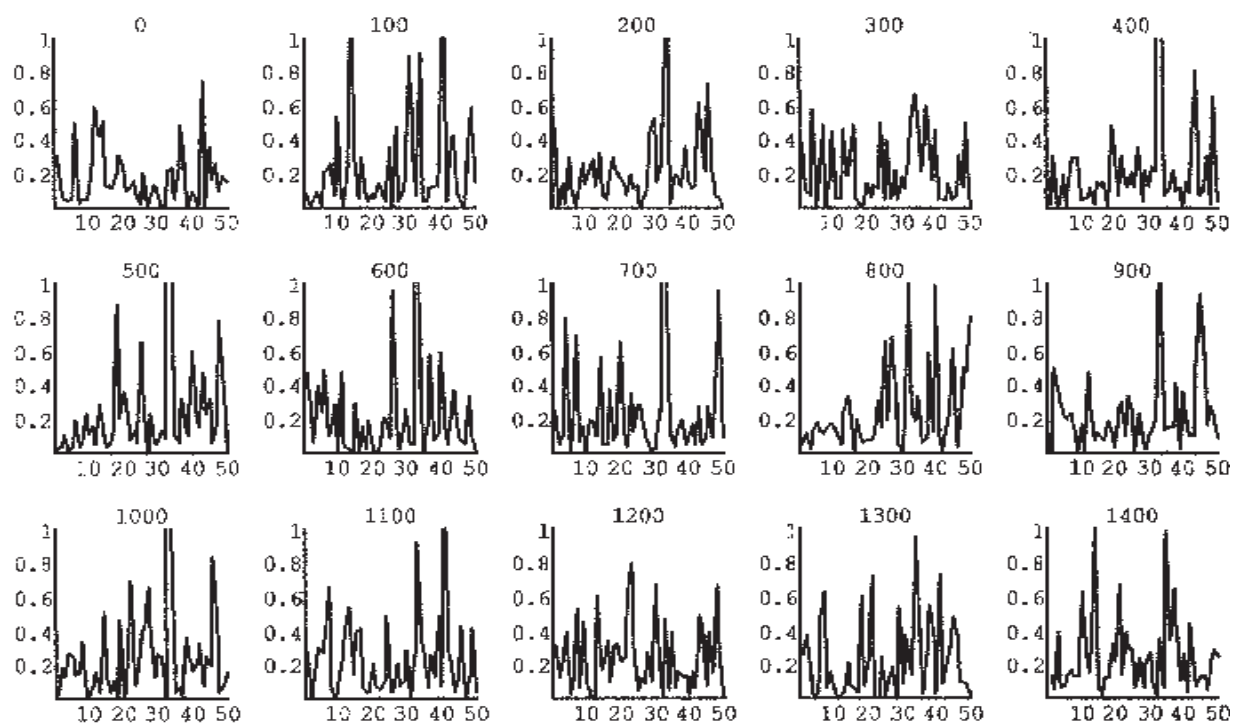
A



C



G



T

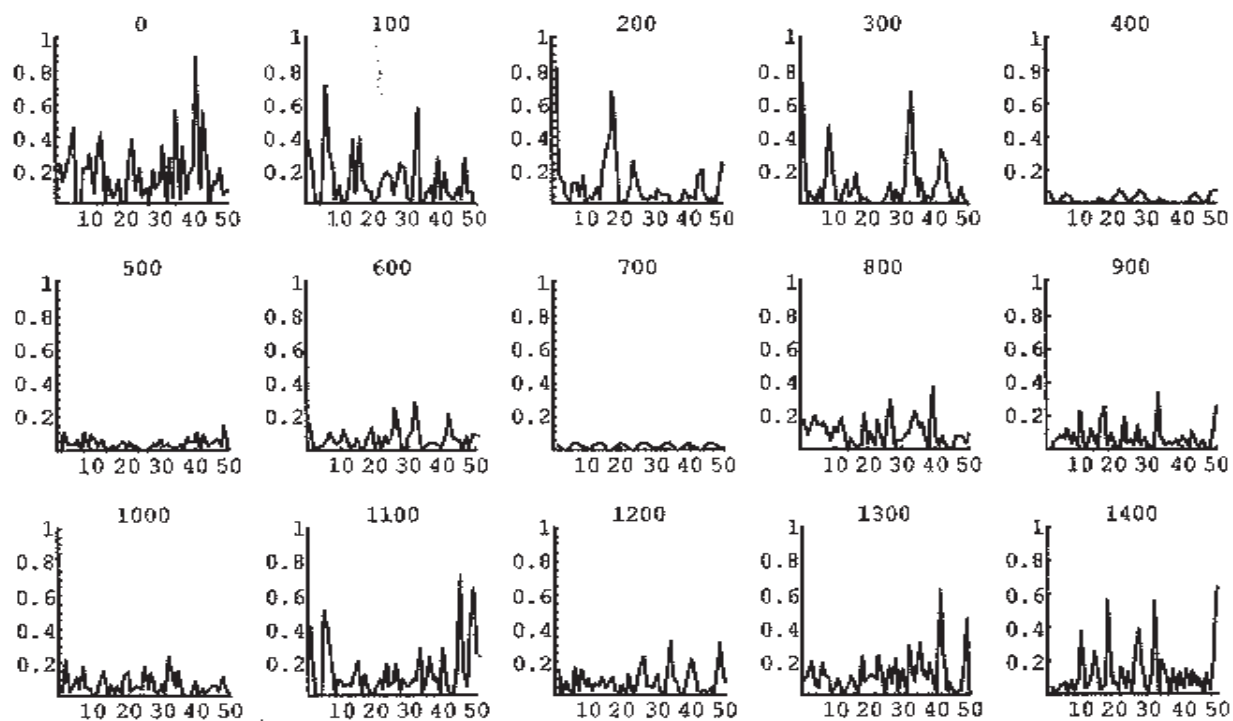


Fig.4. La Periodicidad del espectro de potencia para cada base del gen AF237379 del neurofilamento del *Xenopus laevis* (NF-M2) cds completo (cuenta: 716A, 246C, 631G, 228 T).

Hasta ahora hemos analizado el espectro de potencia de las secuencias de bases en genes. Claro que también se pueden ana-

lizar las correlaciones de corto alcance (periodicidades) dentro del respectivo espectro de potencia de cada base (G, T, C, A). De los ejemplos siguientes de análisis para el exón AF237379 que es parte del gen que codifica para el neurofilamento del *Xenopus laevis* (NF-M2) el cds completo (cuenta: 716 A 246C 631G 228 T) se puede ver que la periodicidad difiere dependiendo fuertemente de GTC y A como se muestra en Fig.4.

En el gráfica para G se puede observar aspectos de los picos claramente entre la base 100/3~33. Ésta es la reflexión de la repetición de tres bases (codón) conteniendo G. En la gráfica A la repetición de tres bases que contienen A también aparece aunque no tan claro como el de G. Comparado con estos casos las repeticiones de tres bases conteniendo C y T parecen ser menos claros.

#### 4. Funciones de la información mutual de las secuencias de bases del gen.

En esta Sección analizamos las secuencias de bases del gen usando la función de información mutual. Esta cantidad la cual se define mediante la siguiente fórmula basada en el entropía de Shannon nos permite examinar las correlaciones de las secuencias en detalle.

$$M(d) \equiv \sum_{a,b=A,G,C,T} P_{a,b}(d) \log_2 \left( \frac{P_{a,b}(d)}{P_a P_b} \right)$$

donde  $P_a$  y  $P_b$  denota la probabilidad de encontrar el símbolo a y b en la secuencia de bases respectivamente mientras  $P_{a,b}(d)$  denota la probabilidad combinada de encontrar a y b a una distancia d. Cada uno de  $P_a$ ,  $P_b$  y  $P_{a,b}(d)$  es calculado según las estadísticas a lo largo de la secuencia. Usamos como la muestra el gen supresor tumoral de von el Hippel-Lindau del homo sapiens. Para efecto de comparación propusimos una secuencia aleatoria que posee la misma longitud y la proporción de bases contenida tal como en los datos de la muestra. La función de información mutual calculada para este datos compuestos también se muestra en la Fig 5.

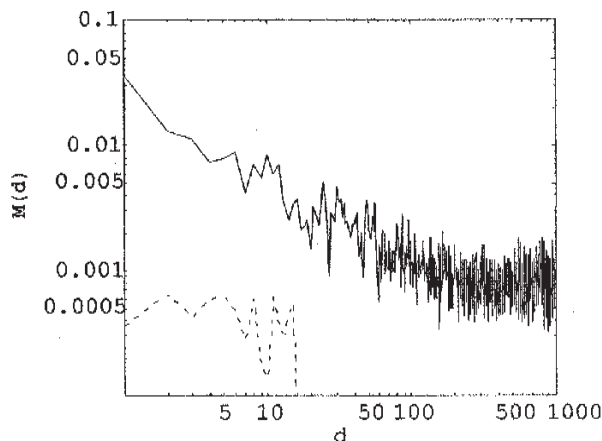


Fig.5. La función de información mutual del gen supresor tumoral de von Hippel-Lindau del homo sapiens (el número de bases es 14543). El eje vertical corresponde a la función de información mutual  $M(d)$ , mientras el eje horizontal a la distancia  $d$  entre dos bases. La línea sólida representa  $M(d)$  de la muestra mientras que la línea punteada representa los datos aleatorios

En la Fig.5 observamos una clara diferencia entre las funciones de información mutual de la muestra  $M(d)$  y de los datos aleatorios. La diferencia es particularmente notoria a una pequeña distancia d (típicamente  $d=1\sim 100$ ). La función de información mutual para los datos aleatorios son casi un valor constante (fluctuando). Por el contrario  $M(d)$  de los datos de la muestra disminuye gradualmente con d. Esto sugiere que allí existen algunas estructuras en la secuencia de bases de acuerdo con los resultados obtenidos en Sección 3. De esta manera la función de información mutual puede proporcionar la información detallada adicional sobre las secuencias de bases.

#### 5. Resumen

En el presente artículo hemos estudiado varias correlaciones en las secuencias de bases de genes en base al análisis del espectro de potencia como las correlaciones de largo alcance, los valores de la pendiente en la dependencia con la longitud de secuencias de bases acumulado y la región con periodicidades de corto alcance. En lo subsiguiente resumimos los resultados. Primero estudiamos la existencia de correlaciones de largo alcance analizando los espectros de potencia de secuencias de la bases totales así como de la parte del exón e intrón separadamente para los datos de muestras de ocho genes supresores tumorales. Los análisis sugieren que las correlaciones de largo, tienden a manifestarse en los espectros de potencia de secuencias de la base totales en lugar de aquellos de la parte del exón y el intrón de forma separada. Aunque aún es demasiado prematuro para esbozar conclusiones definitivas con el análisis de este número limitado de datos. De los análisis sobre los valores de la pendiente  $\alpha$  en dependencia con la longitud de secuencias de bases acumulada podremos confirmar la existencia de un cambio rápido en el valor de la pendiente que se parece a las transiciones de fase de primero orden en la vecindad de la base 450 en la secuencia de bases acumulada de la gráfica 7 en la Fig. 2').

Además observamos cambios similares sobre la base 1400 del HUMCFVII y alrededor de la base 1000 para el ARNm del APC de ratón. Las gráficas respectivas del análisis genético manifiestan varios comportamientos peculiares para cada uno. Los arreglos de intrones y exones en las secuencias de bases

tienen una particular influencia en tales comportamientos. En particular hemos confirmado que se repiten mutuamente los arreglos de exones y el intrones preferentemente en el largo alcance donde el valor  $\alpha$  se mantuvo alto ( $\alpha \sim 1.3$ ) en la gráfica de HUMCFVII.

Como tercer método de análisis, fueron estudiadas las correlaciones de corto alcance, evaluando los medio espectros de potencia respectivos de cada sección corta (dividido en cada 100 bases) de las secuencias de bases. De tal análisis se puede examinar en que región de la secuencia de bases y en que tamaño existen repeticiones de las secuencias de bases. Encontramos una repetición distintiva en un conjunto de 17 bases dentro del rango de las 2000~2800 bases (parte del intrón) de la secuencia de HUMCFVII la cual se verifica por una repetición de CCCGGGGCGTGGGTGT en la secuencia de bases. También podríamos ver la aparición de una repetición de cerca de 30 bases en un análisis similar dentro del rango de las 4800~5400 bases (parte del intrón) para el espectro de potencia del Oncogén T24 de la Ampolla Humana. Para parte del exón que codifica para el neurofilamento del *Xenopus laevis* (NF-M2) podríamos confirmar la existencia de las periodicidades debido a los codones que contiene G y A.

También realizamos los análisis usando la función de información mutua. La comparación de esta cantidad para los datos de la muestra y los datos aleatorios mostró las diferencias significativas en las gráficas, particularmente en la región con las distancias más pequeñas entre dos bases. Enfatizamos que el método basado en esta cantidad puede proporcionar la información detallada adicional útil en las secuencias de bases. Debido a que el número de los genes analizado en este artículo no es lo suficientemente largo se puede extender el análisis mediante el incremento en el número de los genes y también explorando técnicas recientemente desarrolladas como la correlación entre dos bases (8) para conseguir una visión más profunda sobre la estructura de las secuencias de ADN. Se esperan que tales estudios contribuyan a la construcción de fundamentos en el desarrollo de la bioinformática.

## Apéndice. El método de la conversión de secuencias de ADN en secuencias numéricas y espectros de potencia.

Debido a que las secuencias de ADN están compuestas por cuatro bases con los caracteres A G C y T se tienen que convertir en numéricas para evaluar sus espectros de potencia. Hasta ahora se han propuesto tres métodos de conversión (4), (7), (8). Naturalmente no debe existir ninguna diferencia esencial en los espectros de potencia resultante adoptando estos tres métodos.

Ninguna prueba matemática directa se ha dado todavía para la identificación de los métodos. Sin embargo podemos verificar esto comparando los resultados numéricos directamente como sigue: El método de conversión debe cumplir con dos condiciones: 1) cuatro bases A, G, C y T están localizados en un vector unitario de manera simétrica, 2) Cada uno de los cuatro puntos vectoriales que forman el vértice del tetraedro están formados por cuatro bases y la suma de estos vectores son igual a cero. Hay tres métodos que satisfacen estas condiciones. Llámense A) B) y C), respectivamente.

Método A (4): Se asignan los vectores unitarios que apuntan al vértice del tetraedro en un espacio 3-dimensional. Entonces las bases A G C y T son expresados por los vectores unitarios 3-dimensionales respectivamente.

Símbolo	componente x,	componente y,	componente z
A:	{ 1 ,	0,	0 }
C:	{ -1/3	0,	$2\sqrt{2}/3$ }
G:	{ -1/3 ,	$-\sqrt{6}/3$ ,	$2\sqrt{2}/3$ }
T:	{ -1/3 ,	$\sqrt{6}/3$ ,	$2\sqrt{2}/3$ }

Tabla A.1. Vectores unitarios 3-dimensionales para las bases A, G, C y T.

Método B (7): Se asignan los vectores unitarios que apuntan al vértice del tetraedro en un espacio tetra-dimensional. Entonces los cuatro vectores unitarios (1,0,0,0), (0,1,0,0), (0,0,1,0), y (0,0,0,1) forman el tetraedro, corresponde a los cuatro bases respectivamente. Por consiguiente con este método de conversión se obtienen cuatro secuencias numéricas.

Método C (8): Como una versión modificada del método-B se resta 1/4 del componente respectivo de los vectores unitarios para que la suma de estos vectores se vuelva un cero nulo (0,0,0,0). Entonces los vectores unitarios:

(-1/4, 3/4, -1/4, -1/4), (-1/4, -1/4, 3/4, -1/4), y (-1/4, -1/4, -1/4, 3/4) corresponden a las cuatro bases respectivamente. Obteniendo las secuencias numéricas en este modo se pueden evaluar el espectro de potencia  $S_{\sigma}(f)$  con la ayuda de (3).

Desplegamos en la Fig. A.1 las gráficas de los espectros de potencia promedio  $S_{\sigma}(f)$  contra evaluando de esta manera los tres métodos de conversión. Está claro que la silueta de las gráficas mostradas en A), B), y C) de Fig. A.1 coincide entre sí. La frecuencia en la posición del pico  $f = 0.059$  en la región de alta frecuencia también coincide. Acerca de los valores absolutos de los espectros de potencia media se puede observar una ligera discordancia entre aquéllos obtenidos por tres métodos. Los resultados de los métodos B) y C) coincide pero el del método A) difiere de los anteriores como se muestra en la

Fig. A.2. Tal tendencia se ve a lo largo de toda la región de frecuencias. Sin embargo debido a que las siluetas de espectros de potencia promedio coinciden, la pendiente no difiere entre ellos. La misma situación se mantiene para la correlación de corto alcance (periodicidad) y las siluetas de los gráficos para la función de información mutual.

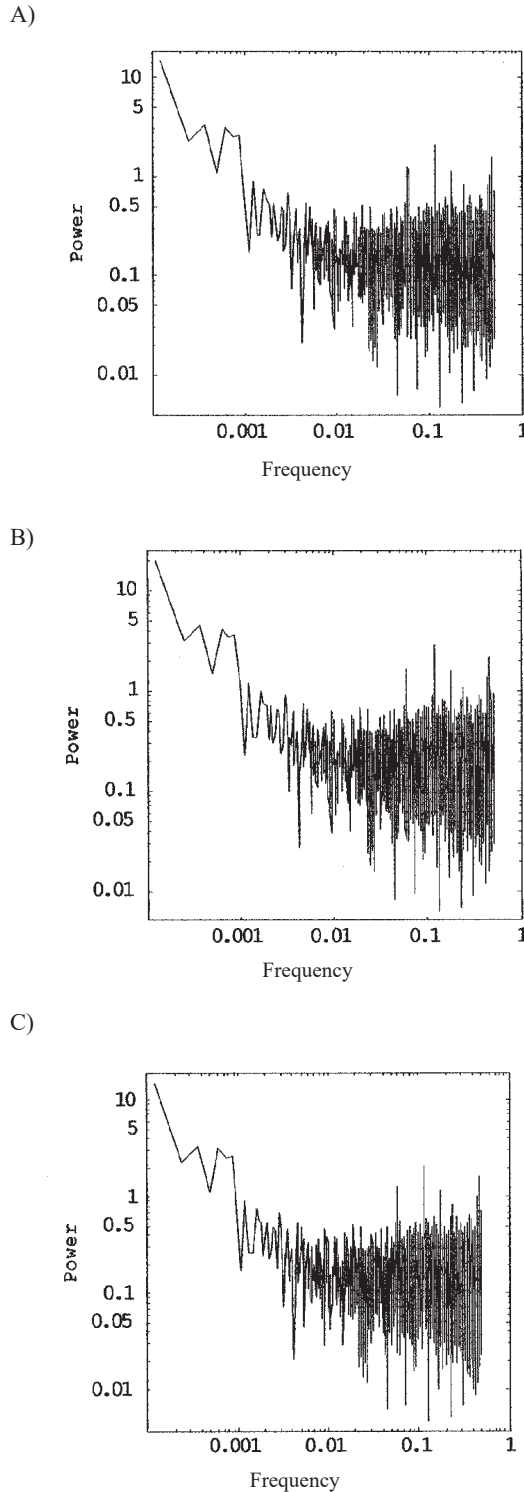


Fig. A.1. Espectro de potencia promedio respectivo para la secuencia del HUMCFVII transformado en datos numéricos según los métodos A), B) y C). El eje vertical corresponde a la densidad del espectro de potencia, mientras que el eje horizontal para  $f_k = k / N$  ( $N$  siendo el número total de la secuencia de bases,  $k = 1, 2, \dots, (N-1)/2$ ) que satisface la condición  $f_k \leq 0.5$ .

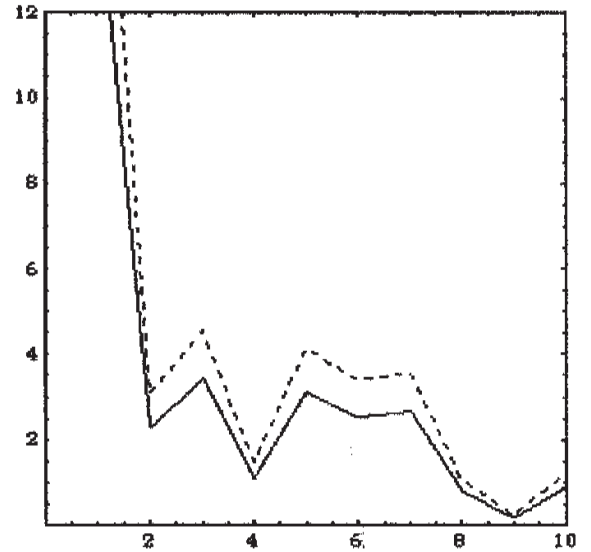


Fig A.2 Comparación entre los espectros de potencia promedio en la región de baja frecuencia. La línea sólida corresponde al método de la conversión B) y C), mientras la línea punteada corresponde al A). El eje vertical corresponde a la densidad del espectro de potencia, mientras que el eje horizontal a  $f_k = k / N$ .

Yasuhiko Isohata\* Masaki Hayashi\*

Honorio Vera Mendoza\*\*

\* University of Pharmacy P. Lite Science Hachioji,  
Tokyo, Japan.

\*\* Universidad Tecnológica de Puebla

## Referencias

- (1) OHNO, S.  
1970 Evolution by Gene Duplication, Springer-Verlag.
- 2) LI, W.  
1989 Europhys. Lett. n. 10, pp. 395-400.
- 3) LI, W.  
1991 Phys. Rev. n. A43, pp. 5240-5260.
- 4) LI W. Y KANEKO, K.  
1992 Europhys. Lett. n. 17 pp. 655-660.
- 5) PENG, K. S., BULDYREV, V., GOLDBERGER, HAVLIN A. L., SCIORTINO, F., SIMONS, M. Y STANLEY, H. E.  
1992 Nature, n. 356, pp. 168-170.
- 6) VOSS, R.  
1992 Phys. Rev. Lett., no. 68, pp. 3805-3808.
- 7) TAKUSHI, E., YOGI M. Y YAMADA C.

- 1998 Bulletin of the Faculty of Science, University of the Ryukyus, n. 65, pp. 31-38.
- 8) TAKUSHI E. Y MIZOUCHI, N.
- 1999 Bulletin of the Faculty of Science, University of the Ryukyus, n. , pp. 29-41.
- 9) TAKUSHI, E. Y MIYAGI, H.
- 2000 Bulletin of the Faculty of Science, University of the Ryukyus, n. 70, pp. 43-46.
- 10) GUHARAY, S., HUNT, B. R, YORKE J.A. Y WHITE, O.R.
- 2000 Physica no. D146, pp. 388-396.