

Pre Laboratorio 6: Análisis de secuencias génicas

Este es un tutorial que te permitirá por un lado, entender cómo se registran secuencias génicas o de ADN /ARN, en una de las bases de datos públicas de secuencias nucleotídicas. Al analizar en detalle ejemplos de registros de secuencias en esta base de datos y al buscar la información detallada contenida en dicho registro, te ayudará además a repasar y reforzar todos los conceptos que hemos visto en clase y en talleres en torno a la estructura de un gen, y las diferentes partes que garantizan su expresión. Al recorrer esta secuencia de extremo a extremo, podrás identificar qué fragmentos de secuencia se requieren para que pueda haber transcripción, para la regulación de la misma, identificar el marco de lectura abierto y la secuencia de proteínas producto de su traducción, y así lograr tener una visión de conjunto de sus ubicaciones respectivas en el fragmento de ADN. Esta vez, lo haremos trabajando sobre una secuencia real de nucleótidos, es decir sobre el mismo código de la vida... te invitamos a hacer toda la actividad de prelaboratorio mirando todos los tutoriales para aprovechar mejor la práctica de laboratorio en computador (no olvides traer tu portatil o tablet el día de la práctica, pues no se podrá trabajar con el celular). Nota: es posible que las versiones vigentes de los portales web de las bases de datos a las que accederán cambien un poco en apariencia con respecto a las versiones mostradas en los videos...pero el contenido y organización será similar.

WILSON TERAN PEREZ 20 DE MARZO DE 2020 23:15

Una breve introducción

Como sabemos los organismos celulares emplean a la molécula de ADN para almacenar y transmitir la información genética a su progenie. Pueden emplear uno o varios cromosomas para organizar esta información según el tipo de célula o especie, sin embargo algo común a todas, es que las unidades de información que le permiten sintetizar un producto con funcionalidad biológica (una proteína o un ARN) se denominan genes. Para que se produzca entonces la correcta expresión de los genes en forma específica y en el momento propio que la célula lo requiere, existen diversos elementos que estructuran las secuencias génicas en unidades transcripcionales, siendo la transcripción el primer paso de la expresión de todo gen. Estas unidades, deben contener secuencias diana que aseguran la correcta transcripción, y regulación específica de su expresión, así como, para el caso de los genes que codifican proteínas, la correcta traducción de la(s) respectiva(s) secuencia(s) codificantes.

En este taller-laboratorio tendremos la oportunidad de repasar todos los conceptos claves asociados con la estructura de un gen y que garantizan su expresión, pero partiendo desde el detalle de la secuencia misma de ADN, con el fin de identificar todos los elementos necesarios a dicha expresión y su correcta ubicación dentro de las secuencias génicas.

Emplearemos un ejemplo de secuencia de ADN depositada en uno de los repositorios públicos internacionales de secuencias de ADN. Esto nos dará la oportunidad de conocer cómo quedan registradas las secuencias de ADN depositadas por los investigadores de todo

el mundo en este tipo de bases de datos, y qué detalles sobre estas secuencias se pueden llegar a extraer.

Empecemos este prelaboratorio repasando conceptos...

A continuación, en cada uno de los siguientes párrafos podrás repasar y tener bien claros los diferentes elementos o partes génicas que hemos visto hasta ahora en clase..de eso se trata de repasar...leyendo y después haciendo 😊

1. Regiones promotoras y terminadoras

Estos elementos delimitan toda **unidad transcripcional**, tanto en procariotas como en eucariotas.

- **Los promotores** son las secuencias diana para la RNA polimerasa (bacterias) o para la unión previa de factores generales (GTFs) de la transcripción que van a reclutar a la RNA polimerasa (eucariotas). Estas secuencias organizadas en forma de **cajas** (por ej. La caja Pribnow o -10 bacteriana, o la caja TATA eucariota) permiten no sólo definir con precisión el inicio de la transcripción desde un punto de inicio (**+1 de transcripción**), sino que determinan la **dirección** de la transcripción.

Ya nos queda muy claro, por lo que hemos trabajado en clase y en talleres, que en cualquier fragmento de ADN, los genes pueden estar orientados en cualquier dirección y que dependiendo de esta dirección, la cadena de ADN que actuará como codificante no será la misma (lo mismo para la cadena molde: cualquiera de las dos

cadenas de ADN puede asumir este rol, dependiendo de la orientación misma del gen). Esto será muy importante tenerlo en cuenta, y lo volveremos a repasar en esta práctica.

- Los **terminadores** son secuencias diana que **al ser transcritas**, pueden formar sobre el ARN una estructura terminadora en forma de hebilla (*hairpin* en inglés) o servir de diana para proteínas terminadoras (por ej. Rho, en bacterias) o endoribonucleasas (en eucariotas), consiguiendo que se termine la transcripción por separación entre el mRNA y el DNA.

Así con los dos elementos anteriormente descritos, ya se define dentro de cualquier secuencia de ADN lo que se conoce como **unidad transcripcional**, que es el primer nivel de organización de la información genética: unidades de lectura por parte del complejo transcripcional.

La extensión de una unidad transcripcional va desde el promotor hasta el terminador de transcripción, produciendo **un único ARN**, cuyo tamaño va desde el **+1 de transcripción hasta la última base de la secuencia terminadora**.



Ojo no confundir en este punto, el +1 (transcripción) con el codón de inicio! Tampoco el terminador de transcripción con el codón STOP o de parada! Estos sirven al proceso de traducción que veremos más adelante...

2. Regiones reguladoras:

Si bien esto lo veremos con más detalle próximamente, ya hemos introducido la importancia de estas regiones para controlar la actividad de los genes: en procariotas, son regiones colindantes (cerca) a la región promotora, en ocasiones hasta se traslapan con el promotor, y constituyen, generalmente, **secuencias diana para proteínas reguladoras de la transcripción**, ya sean activadores o represores: llevan el nombre de **operadores**. Su presencia permite que los reguladores transcripcionales se unan y logren aumentar o disminuir la transcripción por parte de la RNAPol desde un determinado promotor.

En Eucariotas existe un mayor número y diversidad de estas secuencias o elementos reguladores y pueden estar tanto a proximidad del promotor como a gran distancia de los mismos, como veremos en próximas clases.



Hasta ahora, los anteriores elementos génicos garantizan la correcta transcripción y en gran medida, la regulación de la expresión transcripcional de los genes, es decir la obtención de ARN. Este ARN puede ser en si el producto funcional final (por ej. ARN reguladores, ARNt, ARNr) o puede ser un producto intermediario de la síntesis de proteínas (ARNm o preARNm). En este último caso, el preARNm (=pre-mRNA) de os eucariotas

deberá ser procesado o madurado en sus extremos, así como internamente con el fin de cortar los intrones y y empalmar los exones (splicing) que son fragmentos de secuencias codificantes. Finalmente, el ARNm será traducido por los ribosomas.

3. Secuencias diana para la traducción:

Una vez transcrito (y madurado en el caso de los eucariotas) el **ARNm**, debe portar las secuencias diana necesarias para la correcta **traducción** del marco de lectura abierto (**ORF: Open Reading Frame**) o **secuencia codificante (CDS)** de cada polipéptido. En el caso de genes *monocistronicos*, un mRNA portará un único ORF para un único polipéptido, pero en procariotas, los genes pueden agruparse en estructuras *policistronicas* llamadas **operones**, en donde un mismo ARNm puede portar varios ORFs, uno detrás de otro, cada uno codificando un polipéptido diferente.

3.1 Inicio de la traducción:

En cualquier tipo de células, la subunidad pequeña del ribosoma es siempre la encargada de recorrer el mRNA y reconocer en la región 5' no traducida (5'UTR), la secuencia diana para el inicio de la traducción en el mRNA. En clase vimos que hay unas pequeñas diferencias en dicho reconocimiento entre procariotas y eucariotas, sin embargo el objetivo de este acoplamiento entre subunidad menor ribosomal y ARNm, es identificar y ubicar correctamente el codón de inicio (AUG), es decir el principio del marco de lectura abierto que va a ser traducido.

Las secuencias de reconocimiento que guían a la subunidad menor del ribosoma junto con sus respectivos factores de iniciación de la traducción (IFs) hacia el codón de inicio son:

- Secuencia de **Shine Dalgarno** o de unión al ribosoma (**Ribosome Binding Site : RBS**) en procariotas: es una secuencia complementaria al ARNr 16S presente en la subunidad pequeña del ribosoma.

- **Caperuza o cap en 5'** del mRNA eucariótico a partir del cual la subunidad menor realiza un "escaneo" hasta hallar el codón de inicio, o alternativamente, entrada por secuencias denominadas **IRES** (Internal Ribosomal Entry Site) que forman estructuras secundarias y permiten acoplamiento sin escaneo desde el cap en 5'.

3.2 Elongación de la traducción.

Una vez localizado el codón de inicio, y ensamblado el ribosoma, este primer codón queda ubicado en el sitio P del ribosoma, mientras que en el sitio A queda el siguiente triplete o codón. Cada codón va a reclutar el aminoacil-tRNA correspondiente gracias a la complementariedad entre codón (en el mRNA) y anticodón (en el tRNA). Así el marco de lectura abierto o secuencia codificante (CDS) se define como la sucesión de tripletes o codones que existen entre el codón de inicio y el codón de parada o codón STOP; dicho marco siempre va a ser leído de 5' a 3', triplete tras triplete y la traducción se hará obedeciendo al *código genético*.

3.3. Terminación de la traducción.

Cuando aparece un codón STOP en el sitio A, ningún aminoacil-tRNA es capaz de reconocer ese codón, sólo es reconocido por un

factor de liberación (*Release Factor*), el cual es una proteína que al entrar al sitio A permite el desensamblaje del ribosoma, el polipéptido terminado y el mRNA.



No confundas organización policistrónica (Procariotas) con estructura exónica (presencia de varios exones e intrones) de los genes Eucariotas...recuerda que los genes eucarióticos son monicistronicos, independientemente del número de exones que contengan, porque despues del corte y empalme (splicing) el mRNA será monicistronico (un sólo ORF contenido)

¿Estás listo? ya lograste tener claridad sobre todos estos elementos o partes génicas? La práctica buscará que tengas más claridad sobre cómo se posicionan éstas y facilitan la correcta lectura de la información por parte de la maquinaria transcripcional y traduccional celular. A continuación te presentaremos la herramienta de trabajo...es decir la base de datos a la cual entraremos para trabajar con un ejemplo de secuencia. Son varios videos para darte a conocer uno de los varios portales que contiene muchísima información cruzada sobre secuencias biológicas y herramientas para analizar o trabajar con las mismas en investigación. Aprovecha para estirarte y buscar algo para hidratarte: bebida caliente o fría, lo que prefieras...

👁️ Visitémos el portal del National Center for Biotechnology Information (NCBI) y los otros dos portales equivalentes (7 min).

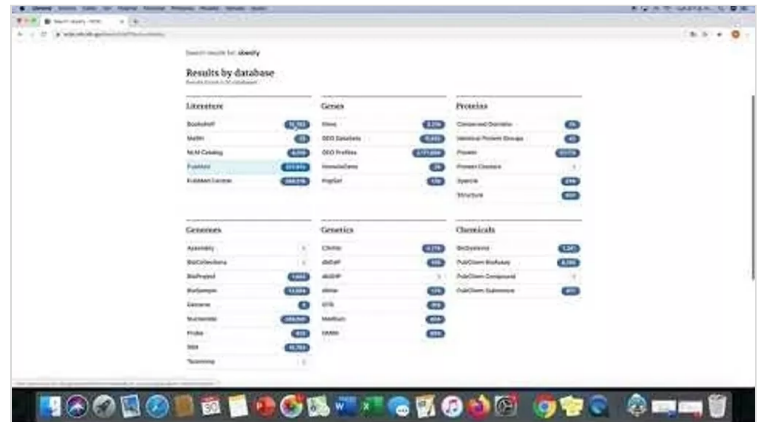
Entremos juntos al repositorio público que aloja la base de datos "genbank" y otras muchas bases de datos de interés...



1 Acceso NCBI, EBI, DDBK comp de BiologíaMolecularWT

YOUTUBE

Explorando el portal: Hagámos juntos una metabúsqueda en todas las bases de datos del NCBI (6.5 min)



2. Ejemplo de metabúsqueda en NCBI de BiologíaMolecularWT

YOUTUBE

Se necesita toda tu concentración en los próximos cuatro tutoriales!



Visitemos ahora la base de datos Nucleotide haciendo una búsqueda con términos (13 min)

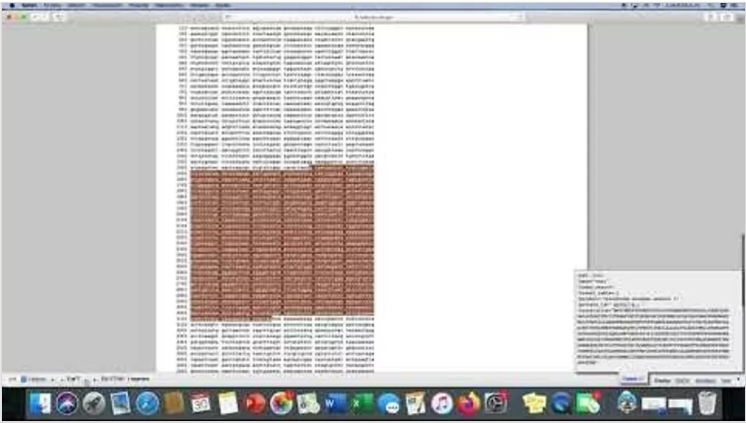


4. Accediendo a la base de que emplearemos en el taller: la base Nucleotide...

de BiologíaMolecularWT

YOUTUBE

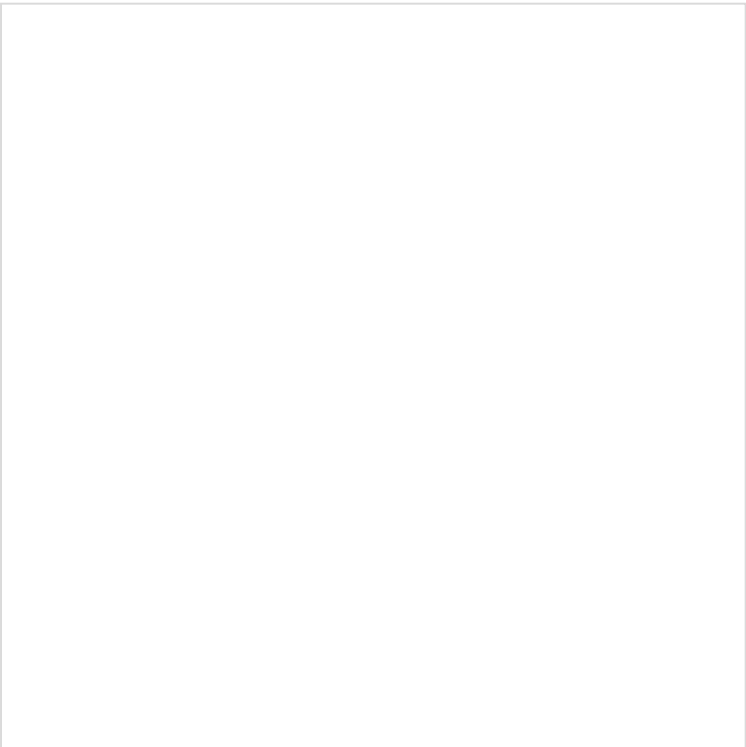
Exploreemos con más detalle cómo se organiza y qué contiene un registro o accesión en la base de datos Genbank o Nucleotide (6.5 min)



5. Detalle de la información contenida en una secuencia registrada en Nucleotide: features
de BiologíaMolecularWT
YOUTUBE

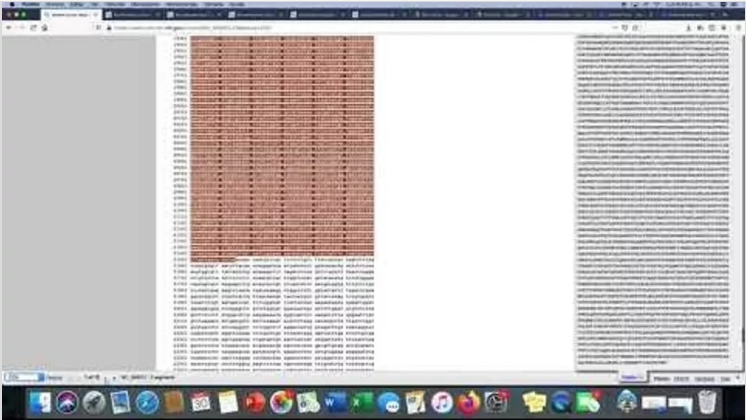
y si la cadena codificante es la complementaria a la que aparece...🤔 (3.5 min)

Después de ver cómo quedan registrados los detalles del contenido en elementos génicos en la cadena 5-3', visible al final del registro, ¿qué pasa con las coordenadas de elementos portados (o codificados) por la cadena complementaria (la que no estamos viendo directamente...)?? **haz click en el cuadro de arriba para ver el video...**



6. aclaración importante sobre la información contenida en la cadena complementaria
de BiologíaMolecularWT
YOUTUBE

Cómo encontrar entre las características (features) de la secuencia elementos génicos como promotor, terminador de transcripción, o secuencias de unión a proteínas: exploremos otros ejemplos (10 min).



7 Ejemplos otras features vf
de BiologíaMolecularWT
YOUTUBE

Ahora te toca a tí!

Tras haber sido guiado por todos estos tutoriales ya debes ser capaz de buscar una secuencia o accesión en "nucleotide", entrar al registro y reconocer donde quedan los diferentes tipos de

información como el organismo fuente, o información taxonómica , los autores que depositaron esta secuencia, artículos asociados, extraer la misma secuencia, así como explorar y buscar los detalles de su contenido incluyendo sus coordenadas. ¡No está nada mal! Entonces pongámoslo en práctica primero con una secuencia sencilla, antes de pasar a hacerlo con la secuencia de trabajo.

Autoevaluación prelaboratorio


1. Abre una pestaña nueva y accede al portal del NCBI. Una vez en la página principal selecciona la base de datos "nucleotide" y realiza una búsqueda con el siguiente ID: "MN908947" Haz clic en el siguiente enlace para contestar a unas preguntas muy rápidas que pondrán a prueba tu aprendizaje con los tutoriales, tranquil@, es muy sencillo y rápido, y es para practicar (autoevaluado).
Has click aquí: 📄

Laboratorio 6: análisis de la secuencia de operones bacterianos

Para el día de la práctica, recuerda descargar tanto la [guía/cuestionario](#) , como la [plantilla](#) que encontrarás en los siguientes enlaces.

Te invitaremos a trabajar en pareja con tu habitual [compañer@](#) de laboratorio para realizar todas las actividades e ir contestando el cuestionario juntos el día de la práctica: no olides llevar un portátil o tableta para poder trabajar!

Verás que la práctica te servirá mucho para repasar los conceptos de transcripción y traducción 😊, algo muy oportuno antes del segundo parcial.



BIOLOGÍA MOLECULAR - ID1261

Guía para laboratorio 6: Análisis de secuencias génicas en genbank

Para esta práctica debes haber realizado el prelaboratorio y visualizado los tutoriales. Entonces te resultará más fácil trabajar con tu compañero de laboratorio.

Debes descargar también la plantilla word que contiene la secuencia de trabajo, ya que sobre ella deberán resaltar con sus propios colores, letra cursiva, letra en negrilla, o letra subrayada los elementos solicitados en el presente cuestionario, conforme a sus coordenadas identificadas en la secuencia de trabajo en genbank. Esto debe hacerse en forma muy precisa...No olvides entonces al realizar el ejercicio, incluir una leyenda indicando a qué corresponde cada color o tipo de resaltado empleado.

PROCEDIMIENTO:

- Ingresen a la página principal del NCBI y desde ahí a la base de datos **Nucleotide**.
- Asegúrense que a la izquierda del cuadro de búsqueda figure la base de datos "nucleotide" para que en el cuadro de búsqueda por términos, escriban el siguiente número de acceso: **AF222792**
- Serán direccionados al registro de la secuencia que analizaremos en esta práctica. Tomense el tiempo para recorrerla e identificar las diferentes secciones que vimos en los tutoriales del prelaboratorio (encabezado, información general asociada, detalles de elementos identificados con sus coordenadas, y al final secuencia completa organizada por bloques de 10 nucleótidos).
- Abren el archivo con la plantilla word, la cual será también un documento de trabajo.

Ya tienen todo lo necesario para hacer la práctica, recuerda que pueden volver a mirar los videos tutoriales en el prelaboratorio, si necesitan recordar algo puntualmente.

Cuestionario:

- ¿Cuál es el nombre del organismo al que pertenece esta secuencia registrada?
- ¿Qué tipo de organismo es (procarionta/eucarionta)?
- ¿Cuántos más a cerca de su Taxonomía:
- ¿A qué tipo de molécula corresponde esta accesión? Y cuál es su tamaño?
- ¿Cuándo fue sometida y cuál fue la fecha de su última actualización?

Guia_lab_virtual_analisis_de_secuencias

Documento de Word

PADLET DRIVE

Plantilla

Laboratorio autónomo de análisis de secuencias génicas en genbank

Plantila para resaltar elementos génicos sobre la secuencia.

Secuencia con coordenadas para accesión AF222792:

1 gagcagcagg agcgcgaagg cgcggtcgcc gccgatcgtg gcgtgcgggc ttctccgctc
61 ggccgggaac gccgtgtcgc gggctgctgt cgtgggtcatg ggggggtccg ttccggcagt
121 tgccccgcgc cgtgggtgtg gtctgttccc ggccgcggcg ggtctgggcg ggggtgcggg
181 tcagcagcag tggccgtcgc tctgcggctc ggtggaggcc ggtgtgggga cgggtcgggg
241 tgcgcagcag tgcctccgtg cgcggcgga gggggctgtc ccggtggggg cgaagtggcg
301 ggccggtcgc tgggtgcggt ggccgacgc ggccgcgcc accgcgagga cgaaccaccg
361 ggccggtgac gccatggcgg ggtgtgacc ctggatccag gcggtggcgg tggcgagaaa
421 ggaggcgagg ccgcgcgcgg ggggtgcgct ggtggcgccg ccggtggcag cgggttaccg
481 gtaccaggcc aggtaggcgc cggtaggagc gaggacggcg gcggtgatgc ggggtgccgtg
541 ccgggccagg gcggtgatct tgcgggtgag ggccgcgcc ccggcgcccg tggtagccgc
601 gacgagcagc aggaaggcgg ccgatccggc ggctaggcgg gcgaagaccg cgaagaatcc
661 ggcaaatgtc gcgtagcct gggcctgggc gatgacggcg agcagcactc cgaagtgtca
721 cgacaggagc gcggcgccgt agccgacgcc gaagaccacc atgcgcggcg ccgtcgccgg
781 gcccgccgac cggcggtgca cgttgccag gtgcagccgc aggcagatg tgcggccggt
841 gaccatgagc gcgcgcagga gcagcaggat gatgccggtg gccagaccga gccaggggcg
901 ggcttggtatc agggcgcggg ctccggcgct gacgatgagc ccggcagcgg cgaagggtgc
961 ggcaagcccg agggtaggg cggcgccgga gccaggggcc ccggcgagcg gtaccggcag
1021 cggtagggtg tgcgtgtcgc cgagggcgga ggtgatccag gccggcagca gggcgaaacc
1081 gcagggggtc accggggcga gcatccggc ggcaaggcg agggcgagca ggccgttcac
1141 caagcgcccg ccttcttcag ccggtccggg atctggccgg ccgagggtcg ggtggccggg
1201 taggtgacct tgcggcgcg atcgacgagc atcagcgtgg acagcgccgc cacttggtac
1261 cgtctggaca gtgtgcgctt ctgttcgag gtggcgggca gactcgggcg cttgatgtac
1321 tgcaggaact gcatgatcgt cgtcttcgac tgcctcggat ccatgtcgac ccgaggaag
1381 ttggtttctt tcccgccctt gtccgaagcc ttccggccct tgtccaggtt cttggcgccg
1441 ccggcgacct caccgcagcc caccgagaag aagaacagcg ccgggggctc cttcgccggc
1501 accgcagcgt tgggtgtcgc gaggagggtg aggtgttcgg ctttcgagc ggccggtgtg
1561 ccgcgctggg gtctcggtgg ggggtgtgt gtctgtgtgt cgtgtccgga ggccgacagg
1621 gtgagcgccg cgttgccggc cgcggccaga gccaggcgcg tacggcgccg cagacgggtg
1681 cggcgggcgg cgggggaagg agaggtcatg ggtggtcgg gccttcccg gagccgggga
1741 ctcgatggag ggtcccggtc gccagggaac gaagggtgga gcgcgccgat gcccgacaga
1801 cgtcgggcgg ggccgtgtgg cggcgccggc ccgggtcagt gcttgcgag caggtgcgcc
1861 tctgtctggt cgttgccggt gccagggggg cagcagcgct ccgctgcgc ggtggggcgg
1921 cggcgagcgc accaggcgag accgcggcgc agcaggacgg ccgccggggc cagcagccag
1981 ggactgacca gcaccccgcc aaggcccgcc agggcgccgc tggccagcag caccggcccg
2041

Plantilla_ana_lisis_secuencias

Documento de Word

PADLET DRIVE

