UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Identificación de elementos repetitivos en proyectos de secuenciación

Apellidos, nombre	Gramazio, Pietro **(piegra@upv.es)** Prohens, Jaime **(jprohens@btc.upv.es)** Herraiz, Javier **(fraherga@upvnet.upv.es)** Vilanova, Santiago **(sanvina@upvnet.upv.es)** Plazas, Mariola **(maplaav@btc.upv.es)**
Departamento	Departamento de Biotecnología
Centro	Universitat Politècnica de València

1 Resumen de las ideas clave

La última década ha supuesto la entrada oficial en la era de las ómicas. Este sufijo, que deriva del griego -oma (ωμα), se aplica a ciertas palabras para indicar el estudio de las mismas en un conjunto o en la totalidad. De esta manera el estudio a gran escala de proteínas se llama proteómica; el de los metabolitos, metabolómica; el de los genomas genómica, etc.

Esta última, la genómica, es la que probablemente ha avanzado más debido a la aparición de secuenciadores de nueva generación, que han permitido un abaratamiento considerable de los costes de secuenciación (Figura 1). Pero contrariamente a la secuenciación de Sanger, que genera secuencias alrededor de 800-1000 pares de bases (pb) aunque limitadas en número, las nuevas técnicas de secuenciación (NGS) generan millones de fragmentos pero más cortos, de 50-250 pb, lo que provoca un desafío técnico a la hora de lidiar con el ADN repetitivo.

Aunque haya diferentes tipos y complejidades, los elementos repetitivos que más problemas ocasionan, desde el punto de vista bioinformático, son los que miden más de 100-150 pb, se encuentran dos o más veces a lo largo del genoma y presentan una identidad de más de un 97%. Han aparecido nuevos programas informáticos para facilitar la identificación de los mismos en proyectos de secuenciación.

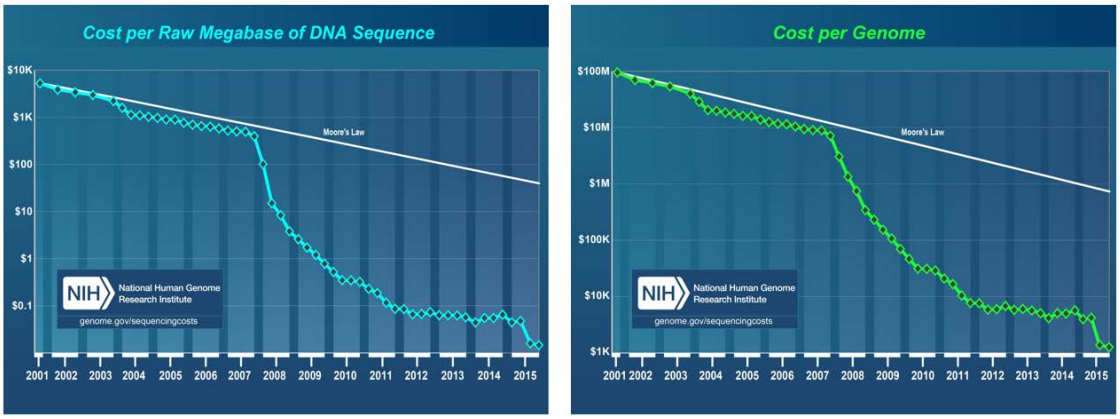


Fig. 1. Series históricas del coste por megabase (Mb) de ADN secuenciado (izquierda) y del coste por genoma humano secuenciado (derecha). Datos del NHGRI (National Human Genome Research Institute).

2 Objetivos

Una vez que el alumno haya estudiado con detenimiento este documento y los recursos de apoyo asociados, será capaz de:

- Identificar los diferentes tipos de ADN repetitivo.
- Describir las posibles funciones biológicas de los tipos de ADN repetitivo.
- Encontrar en bases de datos proyectos de secuenciación de interés.
- Analizar mediante programas adecuados, proyectos de secuenciación para la búsqueda de ADN
  Repetitivo.
- Analizar e interpretar los resultados obtenidos.

3 Introducción

El ADN repetitivo se ha encontrado en una amplia gama de especies, desde las bacterias hasta el ser humano. La abundancia de estos elementos es muy variable, desde pocas copias hasta a más de un 95% del genoma, siendo de casi un 60-70% en humanos.

Aunque se desconoce la función de la mayoría de ellos, tanto que durante mucho tiempo se les ha llamado ADN basura, se sabe que han tenido un papel muy importante a lo largo de la evolución creando nuevas funciones participando en la regulación y dando forma a los genomas modernos.

El ADN repetitivo ha surgido a partir de varios mecanismos, produciéndose copias adicionales de una secuencia e insertándose a lo largo de un genoma. La diversidad de tipos de secuencias y tamaño del repetitivo es asombrosa, la cual varía desde pocas hasta millones de pares de bases.

Hay muchas formas de poder clasificar los elementos repetitivos. En este artículo docente haremos una distinción en base a si el motivo repetido dentro de cada elemento es moderadamente o altamente repetido.

Los altamente repetidos, suelen presentar secuencias cortas (motivos) que se repiten de manera consecutiva unas detrás de otras (repetición en tándem) y en base a la longitud del elemento repetido se clasifican en microsatélites, minisatélites y satélites. El término satélite se debe a que cuando se centrifuga ADN genómico fragmentado aparecen bandas de diferentes densidad, si las secuencias satélites presentan una riqueza en nucleótidos A+T superior a la media del genoma.

Los **microsatélites** suelen presentar motivos repetidos de 2-8 nucleótidos y una longitud máxima de unas 400 pb. Las repeticiones dinucleótidicas y trinucleótidicas son las más frecuentes y dependiendo del organismo se encuentran diferentes motivos preferentes. Los microsatélites son bastante variables y están repartidos más o menos homogéneamente a lo largo de los genomas, al contrario de otros repetitivos, permitiendo la evaluación y comparación de perfiles genéticos de manera sencilla y rápida y la construcción de mapas genéticos. Su número varía mucho despendiendo de la especie; en humanos se estima haya unos 200.000 microsatélites.

Los **minisatélites** están compuestos por motivos repetidos de 10-60 nucleótidos y una longitud muy variable de pocos centenares hasta decenas de miles de pares de bases. Se encuentran preferentemente en los telómeros y son muy variables; como los microsatélites se suelen usar para perfiles genéticos. Se ha comprobado que parte de ellos son hipermutables, presentando una tasa de mutación muy elevada (entre 0,5 y 20%) y una inestabilidad de la más alta de todos los elementos genómicos. Parte de ellos está involucrado en la expresión génica como reguladores (transcripción, splicing alternativo, etc.). En humano se estima que existen unos 30.000 minisatélites.

Los **satélites** son los elementos más largos compuestos por repeticiones en tándem. Sus motivos pueden variar de 5 a cientos de nucleótidos, repitiéndose miles de veces y alcanzando tamaños que oscilan entre 100 kb hasta varias megabases. Se suelen situar en zonas de baja recombinación como los centrómeros y en algunos organismos son muy abundantes; en humanos se estima que ocupan un total de 250 Mb.

Dentro del grupo de los moderadamente repetidos hay muchos tipos de elementos diferentes, pero sin duda los más abundantes son los transposones.

Los transposones o elementos genéticos transponibles son secuencias de ADN capaces de moverse de forma autosuficiente de una zona a otra del genoma. La idea de la existencia de elementos móviles fue postulada por primera vez en los años cuarenta por Barbara McClintock, pero no fue validada hasta los años setenta, lo cual le valió el premio Nobel en 1983.

Los transposones han sido encontrados tanto en procariotas como en eucariotas, aunque sus tamaños, estructuras y funciones varían dependiendo del organismo. Los transposones eucariotas, a diferencia de los procariotas, suelen representar gran parte del ADN repetitivo de los genomas y aunque durante mucho tiempo han sido definidos como ADN basura o egoístas, cada vez más se les atribuyen funciones reguladoras y un papel muy importante en la diversidad evolutiva.

En el caso de procariotas los genomas son muy compactos, lo que significa que hay muy poco espacio para ADN que no codifique para una proteína esencial a la sobrevivencia del organismo mismo. De esta forma el número de elementos repetitivos es muy bajo, sobre todo de elementos transponibles, ya que la inserción de unos de estos elementos en un gen esencial podría comprometer la continuidad del organismo.

Los elementos más comunes que se han encontrado en procariotas son los IS y Tn. Los **IS** (insertion sequences) presentan una longitud de 750 a 1.425 pb, son generalmente silentes y están compuestos por una transposasa y una resolvasa, que son necesarias para la movilidad, y repeticiones invertidas flanqueantes a ambas extremidades. Además, a menudo los elementos se suelen introgresar en zonas de elementos repetidos del huésped (Figura 2). Un elemento **Tn** o transposón compuesto no es más que dos elementos IS cercanos y el ADN entre los dos, que generalmente es un gen de resistencia a antibióticos (Figura 2).

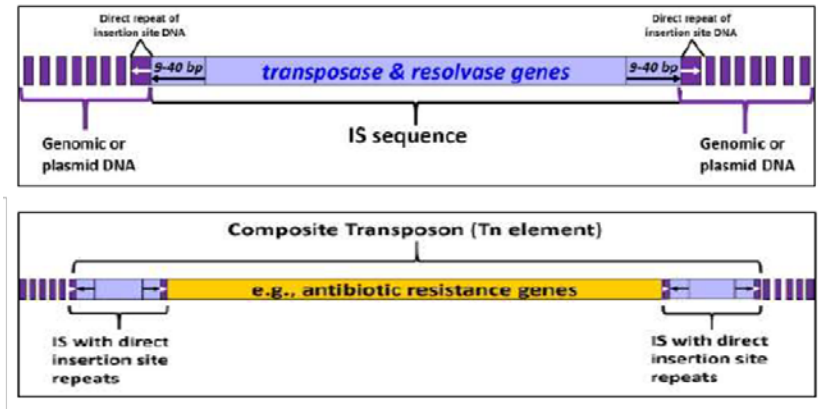


Fig. 2. En la imagen de arriba un transposón tipo IS y abajo un transposón compuesto. "Cell and Molecular Biology: What We Know & How We Found It" por Gerard Bergtrom licenciado bajo CC-BY 4.0

En eucariotas hay dos tipos de elementos transponibles.

Los **transposones de Clase I** o **retrotransposones** pasan a través de una fase intermedia de ARN para poder moverse en el genoma. Como su nombre indica, comparten el mismo ancestro que los retrovirus. Esta clase de transposones incluye los trasposones LTR y no-LTR.

Los **retrotrasposones LTR** (long terminal repeats) difieren de otros tipos de retrotrasposones por presentar, en las zonas flanqueantes a sus genes, secuencias repetidas de más de 300 pb (Figura 3). Un tipo de retrotrasposón LTR es el Ty de la levadura. Entre las secuencias flanqueantes hay varios genes necesarios para la maquinaria que permite el movimiento de una zona a otra del genoma.

El gen Gag codifica una proteína, una partícula vírica que protege el ADN una vez retrotranscrito. Además la región Pol contiene el gen RT, que codifica para la retrotranscriptasa que produce el ARN intermediario, el gen Ptr, que codifica para la proteasa que degrada la partícula vírica al entrar en el núcleo, y el gen Int, que codifica la integrasa que permite la introgresión del retrotrasposón en el núcleo.

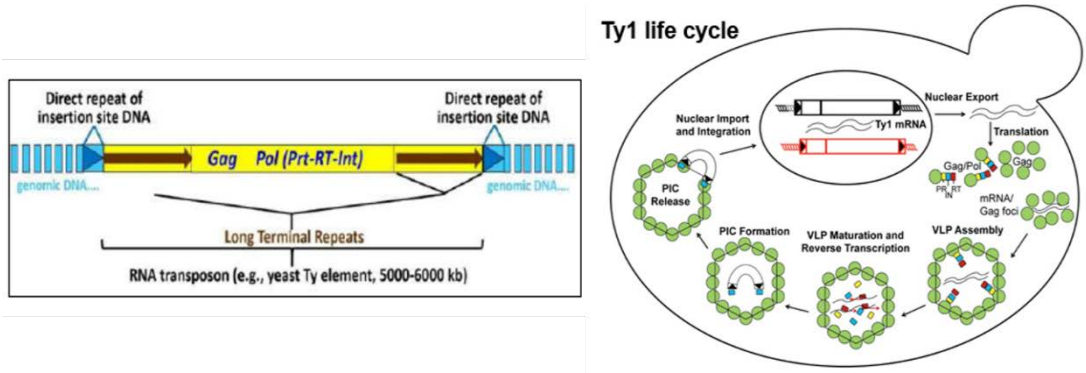


Fig. 3. En la imagen de la izquierda un retrotrasposon Ty de tipo LTR de levadura. Las flechas marrones indican las repeticiones flanqueantes los genes. A la derecha el ciclo de un transposón Ty. Fig. 2. En la imagen de arriba un transposón tipo IS y abajo un transposón compuesto. "Cell and Molecular Biology: What We Know & How We Found It" por Gerard Bergtrom licenciado bajo CC-BY 4.0

Los retrotrasposones no-LTR incluyen los elementos LINE y SINE.

Los **LINE** (Long Interspersed Nuclear Elements) presentan UTRs flanqueantes a los genes responsables de la transposición. La 5'UTR presenta un promotor reconocido por la ARN polimerasa, la cual sintetiza los genes (ORF1 y ORF2) necesarios para la inserción. La ORF1 codifica para la endonucleasa mientras que la ORF2 codifica para la transcriptasa inversa y la integrasa (Figura 4)

Los **SINE** (Short Interspersed Nuclear Elements) no tienen de los genes necesarios para la transposición, aunque presentan UTRs, y necesitan de otro retrotransposón para llevar a cabo la inserción, por eso también se les llama retroposón (Figura 4).

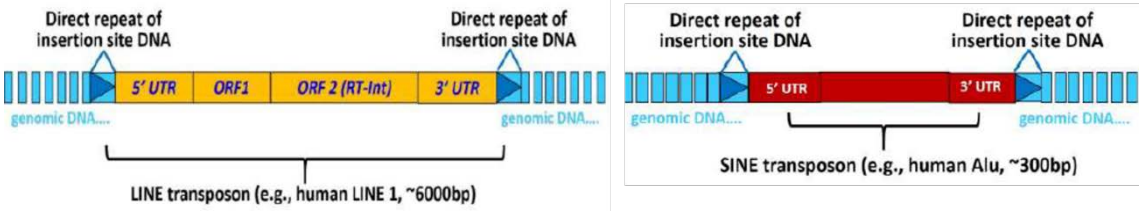


Fig. 4. A la izquierda un transposón LINE y al derecha un transposón SINE. "Cell and Molecular Biology: What We Know & How We Found It" por Gerard Bergtrom licenciado bajo CC-BY 4.0

Los **transposones de Clase II** o **ADN transposones** se mueven directamente de una posición a otra en el genoma usando una transposasa. Como los transposones procariotas, estos poseen repeticiones invertidas flanqueantes a ambas extremidades y se insertan en elementos repetidos del huésped (Figura 5).

Los ADN transposones usan dos mecanismos para insertarse en el genoma:

- El mecanismo corta y pega (cut-&-paste pathway), en el cual el transposón deja un sitio para insertarse en otro.
- El mecanismo replicativo (replicative mechanism), en el cual el transposón se queda en su posición inicial y produce una copia que se inserta en otro sitio.

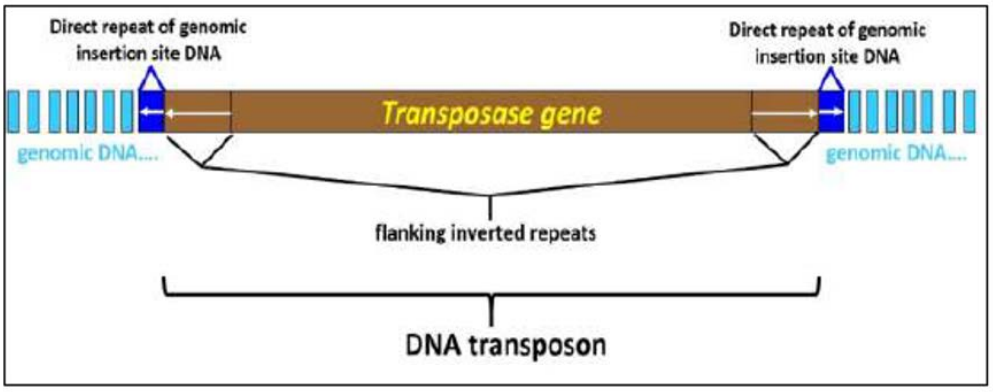


Fig. 5. Imagen de un ADN transposón. "Cell and Molecular Biology: What We Know & How We Found It" por Gerard Bergtrom licenciado bajo CC-BY 4.0

A continuación en la tabla 1 podemos hacernos una idea de la distribución y abundancia de los transposones en algunas especies

Especie	% transposón Clase I	% transposón Clase I respecto al genoma	% transposón Clase II	% transposón Clase II respecto al genoma	% todos transposones respecto al genoma
Bacteria (*E.coli*)	0	0	100	3	3
Levadura	100	3.5	0	0	3.5
Maíz (*Z.mays*)	50	30-45	50	40	70-95
Protozoo (*T. vaginalis*)	0	0	100	66	66
Rana (*R. esculenta*)	25	19	75	58	77
Ratón (*M. musculus*)	95	38	5	2	40
Mosquito (*A. aegypti*)	30	14	70	33	47
Homo sapiens	90	40	10	5	45
Nematodo (*C. elegans*)	5	0.5	95	12	12
Mosca (*D. melanogaster*)	80	3	20	1	4
Arroz (*O. sativa*)	15	1	85	5	6

Tabla 1. Comparación de la distribución y abundancia de elementos transponibles en diferentes organismos. "Cell and Molecular Biology: What We Know & How We Found It" por Gerard Bergtrom licenciado bajo CC-BY 4.0

Los datos de la tabla 1 nos llevan a las siguientes conclusiones:

- La abundancia de transposones no está correlacionada con una mayor complejidad evolutiva de los organismos.
- Las especies pueden presentar los mismos transposones aunque tengan una historia evolutiva diferente.
- Cuanto más transposones activos, más rápidamente cambian los genomas y más inestables son.

Para ampliar conocimientos en los mecanismos de transposición se aconseja el libro de Gerald Bergtrom cuya referencia se encuentra en la bibliografía.

4 Desarrollo

En la última década, los proyectos de secuenciación se han multiplicado de forma exponencial. Como hemos dicho anteriormente, esto ha sido posible gracias a la llegada de secuenciadores de nueva generación que han permitido un abaratamiento notable en los costes de secuenciación. Además recientemente han aparecido en el mercado secuenciadores de tercera generación, como Pacific Biosciences y Oxford Nanopore, que prometen secuencias más largas, más baratas y en menor tiempo.

Desafortunadamente en la actualidad no hay mucho personal investigador que sepa aprovechar al máximo todo el potencial de estos proyectos de secuenciación, los cuales en muchos casos se quedan sin analizar o se hace de forma básica e incompleta. Por este motivo hacen falta más documentos como este que ayuden en la complicada tarea del análisis.

Asimismo, para poder entender y analizar los resultados de estos proyectos de secuenciación es necesario tener conocimientos previos de Biología, Genética, Genómica, Filogenia y Bioinformática.

En este artículo docente se proporcionarán los conocimientos necesarios para poder entender y reproducir el contenido de la misma, y trasladar a un proyecto verdadero.

El contenido se ha estructurado en varios pasos y refleja uno de los análisis imprescindibles de un proyecto de secuenciación, que corresponde a la búsqueda y determinación de elementos repetitivos.

Paso 1. Elección de las secuencias de un proyecto se secuenciación

Para este primer paso vamos a necesitar las secuencias en formato FASTA de un proyecto de secuenciación. En bioinformática el formato FASTA es un fichero de texto utilizado para representar ácidos nucleicos, péptidos, aminoácidos, etc. Se compone de una línea de cabecera que empieza con el símbolo '>' que contiene la descripción de la secuencia misma seguida por líneas de datos de secuencia, donde cada ácido nucleico es representado por una letra (A, adenina, T, timina, C, citosina, G, guanina).

En el caso de que no tengamos nuestro propio proyecto de secuenciación, hay diferentes bases de datos donde poder conseguir secuencias de forma totalmente gratuita. Una de las bases de datos más grande y completa es la del GenBank (https://www.ncbi.nlm.nih.gov/genbank/). En esta base de datos se almacenan secuencias de todo tipo, desde un gen hasta cromosomas enteros, además de encontrar mucha información adicional respecto al organismo y a la secuencia misma. En uno de los repositorios del GenBank, el SRA (Sequence Read Archive) (https://www.ncbi.nlm.nih.gov/sra), se almacenan exclusivamente las secuencias bruta (sin procesar) y los alineamientos de proyectos de secuenciación. En el caso de que quisiéramos secuencias procesadas y analizadas podemos recurrir al repositorio ftp del mismo GenBank (ftp://ftp.ncbi.nih.gov/genomes/).

Paso 2. Descarga de las secuencias de un proyecto se secuenciación

Para este articulo usaremos la secuencia del cromosoma 1 del maíz (*Zea mays*). En este enlace se puede descargar la secuencia directamente. (ftp://ftp.ncbi.nih.gov/genomes/Zea_mays/CHR_01/zm_ref_B73_RefGen_v3_chr1.fa.gz)

Otra manera más eficiente de descargar la secuencia es hacerlo a través de una terminal (también llamada consola, shell o bash). Hoy en día para poder trabajar con secuencias, que en muchos casos ocupan decenas de gigabytes (GB), hay que dominar los programas bioinformáticos, que en la mayoría de los casos solo se pueden usar a través de una terminal (para más información pulsa el enlace,

https://es.wikipedia.org/wiki/Shell_de_Unix). La terminal hace que los programas sean mucho más rápidos y menos exigentes en memoria RAM, pero no presenta una interfaz gráfica que a veces dificulta su uso. Para aprender a moverse dentro de la terminal y a usar programas bioinformáticos se aconsejan los cursos del grupo bioinformático del COMAV-UPV (https://bioinf.comav.upv.es/courses.html).

Aunque para esta práctica no es estrictamente necesario el uso de la consola, se recomienda usarla para familiarizarse con la misma. Seguidamente se muestra como crear una carpeta para el proyecto y descargar la secuencia (Figura 6).

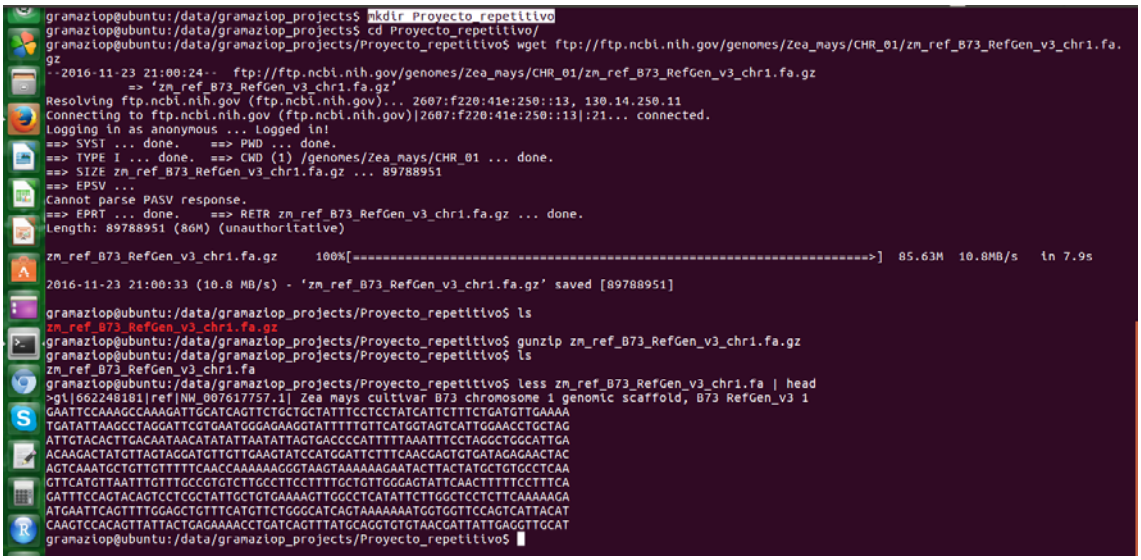


Fig 6. Pasos para descargar las secuencias usando una consola.

A continuación vamos a analizar los comandos, los cuales empiezan después del símbolo ($):

- *mkdir Proyecto_repetitivo*
Con el comando mkdir (MaKeDIRectory) creamos la carpeta que contendrá el fichero que vamos a descargar.

- *cd Proyecto_repetitivo*
Con el comando cd (Change Directory) accedemos a la carpeta creada.

- *wget ftp://ftp.ncbi.nih.gov/genomes/Zea_mays/CHR_01/zm_ref_B73_RefGen_v3_chr1.fa.gz*
Con el comando wget (web get) podemos descargar directamente la secuencia en nuestro directorio solo poniendo la url de la página web.

- *ls*
Con el comando ls (list) podemos listar los archivos de una carpeta.

- *gunzip zm_ref_B73_RefGen_v3_chr1.fa.gz*
Con el comando gunzip podemos descomprimir el archivo comprimido que hemos descargado.

- less *zm_ref_B73_RefGen_v3_chr1.fa | head*
Con el comando less podemos ver el contenido del archivo en formato FASTA, mientras que con el comando head solo visualizamos las primeras 10 líneas en vez de ver todo el archivo.

Paso 3. Seleccionar un subconjunto de las secuencias de un proyecto de secuenciación

En el caso de que quisiéramos analizar secuencias brutas, porque el organismo de nuestro interés todavía no ha sido analizado o porque estamos trabajando con nuestras propias secuencias, antes de poder someterlas a una búsqueda de elementos repetitivos, podemos seleccionar un subconjunto de ellas. Como hemos dicho anteriormente el resultado de la secuenciación son miles o millones de secuencias cortas (reads) de entre 50 y 250 pb, cuya suma nucleotídica puede ser de pocas hasta centenares de veces más grande del mismo tamaño de genoma.

Por ejemplo si el tamaño del genoma de un organismo es 1 GB, quiere decir que posee 1.000.000.000 de nucleótidos y se establecería como 1X o sea una vez el tamaño del genoma; la suma nucleotídica de todos las reads de un proceso de secuenciación podría ser tanto 100.000.000 (0.1X) como 100.000.000.000 (100X) o más.

Por tanto, para evitar prolongar innecesariamente el análisis y saturar el ordenador, servidor o programa on-line que sea podemos seleccionar un subconjunto al azar de las secuencias, el cual nos dará una idea del tipo de repetitivo que podemos encontrar en nuestras secuencias. En la siguiente captura de pantalla se muestra como seleccionar un subconjunto de un 10% de las secuencias.

- *seqtk sample archivo.fq 0.1 > archivo_reads_0.1.fq*

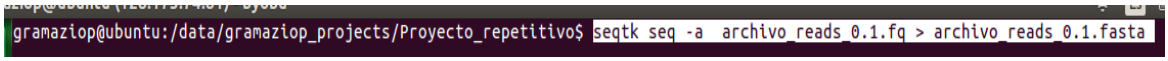
Con el programa *seqtk* (https://github.com/lh3/seqtk) y la opción *sample* podemos seleccionar el subconjunto que queramos usando el siguiente orden:

seqtk	nombre del programa.
sample	opción del programa *seqtk* que nos permite seleccionar el subconjunto de secuencias.
archivo_reads.fq	archivos con las secuencias en formato FASTQ.
0.1	porcentaje del total de las secuencia que queremos seleccionar, en este caso un 10%.
>	este símbolo es para redirigir el resultado de esta operación a un nuevo fichero.
archivo_reads_0.1.fq	nuevo fichero en formato FASTQ con el resultado de la operación.

El formato FASTQ es un formato de texto que además de la secuencia compuesta por nucleótidos almacena el valor de calidad de secuenciación de cada uno de ellos en formato ASCII y es proporcionado por las plataformas de secuenciación. A continuación se presenta un ejemplo de una secuencia en formato FATSQ.

@Secuencia_1
GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTTT
+
!''*((((***+))%%%++)(%%%%).1***-+*''))**55CCF>>>>>>CCCCCCC65

Como el software que vamos a usar solo permite analizar secuencias en formato FASTA tenemos que convertir nuestras secuencias a este formato.

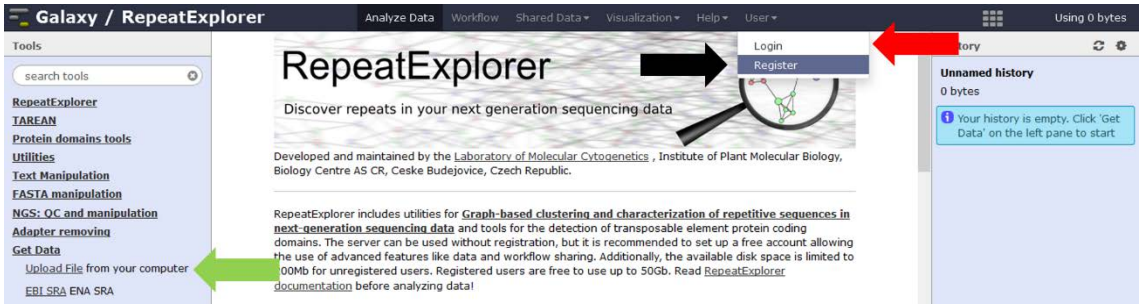


Con el mismo programa seqtk podemos convertir las secuencias de formato FASTQ en formato FASTA usando esta vez la opción *seq –a*.

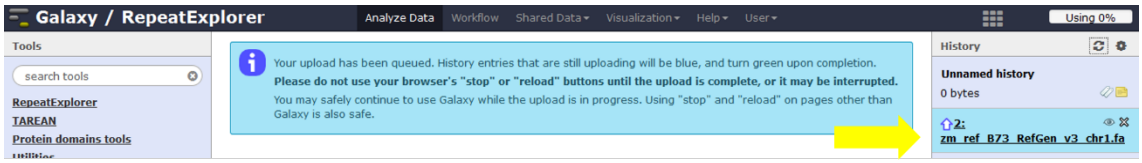
Paso 4. Análisis de elementos repetitivos de las secuencias de un proyecto se secuenciación

Finalmente podemos someter nuestras secuencias a la búsqueda de elementos repetitivos. Aunque hay muchos programas que lo hacen, para esta práctica hemos elegido la herramienta informática Repeat Explorer (http://www.repeatexplorer.org/). Usaremos la interfaz gráfica de la web, pero se aconseja instalar el programa localmente ya que dependiendo del tamaño de la secuencia el análisis puede tardar desde horas hasta semanas. Antes de empezar se recomienda leer con detenimiento el manual (http://repeatexplorer.umbr.cas.cz/static/html/help/manual.html), ya que se puede personalizar el tipo de análisis en función de nuestros objetivos.

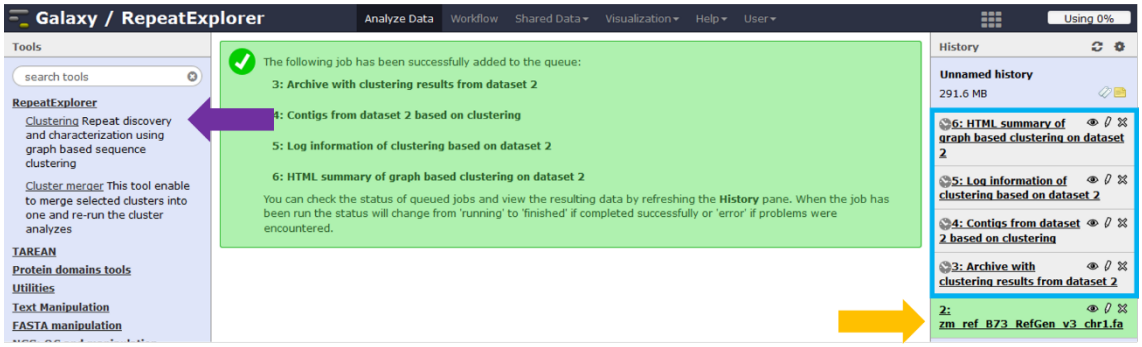
Como primer paso hay que crear un perfil de usuario para evitar la limitación de memoria de los usuarios no registrados (flecha negra). Seguidamente entramos como usuarios (flecha roja) y subimos nuestras secuencias (flecha verde).



El tiempo de subida del archivo al servidor de Repeat explorer es muy variable dependiendo de nuestra conexión a internet y del tamaño del archivo. Hasta que el recuadro a la derecha (flecha amarilla) no se ponga en verde quiere decir que la subida no ha terminado.



Una vez que el programa haya acabado de subir las secuencias a su servidor (flecha naranja), lanzamos el análisis del clustering (flecha morada). Automáticamente se generarán cuatro archivos (recuadro azul), que hasta que el color no cambie del gris al verde no estarán terminados. También la duración de este paso es muy variable, pues para secuencias de gran tamaño puede tardar incluso semanas.



Cuando haya terminado el análisis guardamos los cuatro archivos obtenidos en la carpeta que hemos creado y abrimos el archivo con los resultados (HTML summary of graph based clustering on dataset 2) (Figura 7). Este archivo nos proporcionará

información respecto al ADN repetitivo encontrado y cómo se agrupa en clusters. Además, al final del archivo hay otro enlace web donde hay un resumen en forma de tabla que se puede exportar a otra aplicación (editores de ficheros de texto, Microsoft Excel, etc.) para la elaboración de datos y estadísticas. Para más información respecto a cómo se forman los clusters debe leerse detenidamente el manual y la publicación asociada.

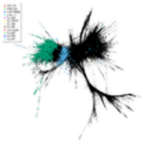
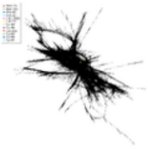
	cluster	total length [bp]	number of reads	Genome proportion[%]	cumulative GP [%]	Repeat Masker	Domain hits	Layout	Portion of similarity hits to other clusters[%]	Outside reads with similarity [%]
1	CL1	1.7e+07	119649	2.740	2.7	LTR.Gypsy (**70578hits, 50.5%**) Simple_repeat (233hits, 0.0478%) LTR.Copia (90hits, 0.0407%) LTR (39hits, 0.0204%) Low_complexity (67hits, 0.0111%) Organelle.Mitochondrion (28hits, 0.0089%)	Ty1-RT Ty1/copia Maximus/SIRE (35 hits 0.0293%) Ty3-INT Ty3/gypsy chromovirus (34 hits 0.0284%) DTM-CD1 NA NA (24 hits 0.0201%) Ty3-RT Ty3/gypsy chromovirus (19 hits 0.0159%) Ty3-RH Ty3/gypsy chromovi.....		0.5204	3.523
2	CL2	1.5e+07	108296	2.480	5.2	LTR.Gypsy (**21808hits, 13.9%**) LINE.CRE (456hits, 0.105%) Simple_repeat (248hits, 0.0781%) Low_complexity (206hits, 0.0346%) LTR.Copia (61hits, 0.0201%) DNA.hAT.Tip100 (71hits, 0.0157%)	Ty3-INT Ty3/gypsy chromovirus (7 hits 0.00646%) DTM-CD1 NA NA (6 hits 0.00554%) Ty3-RH Ty3/gypsy chromovirus (6 hits 0.00554%) DHH-CD1 NA NA (4 hits 0.00369%) LINE-RT NA NA (4 hits 0.00369%) Ty3-		0.3781	3.499

Fig. 7. Ejemplo de fichero de resultados obtenido a partir del análisis de clustering del repetitivo encontrado en la secuencia a examinar.

Los resultados de esta práctica nos indican que un elevadísimo porcentaje del genoma de maíz está constituido por transposones de diferentes tipos, aunque también los hay del tipo de repetición en tándem. Como se ha dicho anteriormente el genoma de maíz puede llegar a presentar hasta un 95% de elementos repetitivos, lo que le proporciona una enorme inestabilidad y una continua metamorfosis genómica, dificultando su estudio, comparación y aprovechamiento.

5 Cierre

A lo largo de este objeto de aprendizaje hemos aprendido lo importante que es la determinación de los elementos repetitivos de un genoma, desde un punto de vista biológico, genético, genómico y evolutivo, y la manera de determinarlo de forma sencilla. Ahora que has asimilado estos conceptos te animamos a repetir el análisis con otros organismos y comparar los resultados obtenidos.

6 Bibliografía[1]

Bergtrom, G: “Cell and Molecular Biology: What We Know & How We Found Out (Annotated iText)”, 2015. Cell and Molecular Biology i-Text. Book 3.

de Koning, A. J., Gu, W., Castoe, T. A., Batzer, M. A., & Pollock, D. D. (2011). Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet, 7(12), e1002384.

Hua-Van, A., Le Rouzic, A., Boutin, T. S., Filée, J., & Capy, P. (2011). The struggle for life of the genome's selfish architects. Biology direct, 6(1), 1.

1