Ensamble de genoma en mRNA de una célula infectada con el virus del VIH

José Saúl Villa Pérez
Ciencias Agrogenomicas
UNAM: Escuela Nacional de Estudios Superiores Unidad León

En este presente, se busca como objeto principal llevar a cabo el ensamble del genoma de Virus de Inmunodeficiencia Humana, ya que este Retrovirus ha tenido gran impacto en la sociedad. Aún no se encuentra un tratamiento eficaz, aunque grandes esfuerzos han sido validos en desarrollar algunos medicamentos como retrovirales, Aún faltan grandes avances en la investigación, en los cuales la disciplina de la Bioinformática puede contribuir activamente con el uso de algunas herramientas, para eliminar definitivamente esta afección.

Introducción:

A lo largo de la historia, hemos sido testigos de innumerables procesos ecológicos e interacciones bióticas, muchos de los seres vivos desde bacterias hasta los mamíferos, hemos interaccionado, con una gran cantidad de patógenos. En esta clasificación se encuentran los virus, aunque no se ha resuelto un debate en el campo de la Biología, si estos pueden ser considerados como seres vivos o no vivos, "*La naturaleza de los virus sigue siendo algo enigmático, dado que desafían la actual teoría celular*" (**Forterre P., 2006**) Una descripción cuantitativa sobre estos puede, que son un conjunto de ácidos nucleicos, proteínas, lípidos y azucares complejas.

En 1981 fueron los primeros casos encontrados en África de esta enfermedad, en ese entonces desconocida, y hasta en 1983 estudiada con gran preocupación en Estados Unidos por un gran aumento de número de casos de neumonía "*Peneumocytis carinii*" enfermedad que en la mayoría de los casos es subsecuente del contagio de este virus.

Actualmente en el mundo viven 37.9 millones personas con VIH según los informes estadísticos de *ONUSIDA* en el año 2021
Aunque que ahora se tiene mucha información disponible sobre este agente infeccioso, se sabe con exactitud que este es trasmitido por vía sexual, sanguíneas, por vía vertical o perinatal. También la literatura médica científica también se ha enriquecido con bastante conocimiento, que ahora se sabe que se trata de un lenovirus y que este pertenece a la familia de los retrovirus Al momento en que el agente infeccioso entra en contacto con el hospedero, este infecta directamente a los linfocitos, por medio de una paridad entre la capa de membranas de los receptores de glicoproteínas viron gp120 por parte del virus, y el receptor CXCR4 CD4 por parte de la célula T, en este momento dentro, se digiera a la envoltura nuclear, y liberara su material genético de Ácido Ribonucleico, y unas cuantas enzimas que serán esenciales en el proceso de integración y asimilación de este.

Aunque nuestra información genética está constituida por Ácido desoxirribonucleico y el del virus en ribonucleico,
La enzima transcriptasa inversa se encargara de sintetizar cadenas dobles de ADN a partir de una cadena monocatenaria de ARN, la enzima integrasa que es la encargada de integrar la cadena nueva creada de la transcriptasa inversa del virus en alguna parte de nuestro ADN para así, este pase desapercibido a la maquinaria de síntesis de nuestro cuerpo, al lograr este proceso de replicación se liberan cada vez más partículas virales infectando más células T repetitivamente, hasta que es infecta una cantidad considerable de

linfocitos en el cuerpo, el hospedador desarrolla una etapa de SIDA que se conoce como (síndrome de inmunodeficiencia adquirida) donde el paciente infectado será más susceptibles a enfermades oportunistas por su bajo nivel de defensas y este pueda causar un daño o incluso la muerte. A lo largo de este proceso ha existido mucho interés en poder revelar los enigmas que esta enfermedad guarda, aunque con éxito se ha desarrollado medicamentos retrovirales que inhiben la replicación del virus, y que han mejorado considerablemente la vida a las personas que sufren esta afección, si estos se dejan de administrarse, el virus de nuevo empieza su actividad replicante. No hay manera de eliminarlo definitivamente. Aunque actualmente con el avance de la computación, y el avance de nuevas tecnologías de punta, realizando tareas cada vez más exactas y complejas se ha dado el campo a la bioinformática, reuniendo los campos biológicos con la informática, esto ha abierto una brecha a muchas de las áreas estudiadas por la biología por ejemplo a la Biología molecular, donde se estudia el código genético y busca una mejor comprensión de los efectos que suceden como esta afección antes mencionada. La secuenciación de diferentes organismos ayuda mucho a la comprensión de su estructura genética tanto a las proteínas o regiones de interés. A lo largo de este documento se buscará la explicación de un ensamble de genoma, una práctica usual al secuenciar una estructura de genética de un organismo.

Metodología

Como antecedentes, tomamos un trabajo de secuenciación realizado en el año 2021 en hospital universitario de Kunming China, Este realizo la secuenciación individual de una célula humana de un paciente infectado con VIH y se habla que el objetivo de este trabajo hacer estudios transcriptomicos en reconstrucción de repuesta inmunitaria. Algunos datos técnicos relevantes sobre la secuenciación son los siguientes presentados en forma de tabla.

Nombre	**IR-14_S2_L001_R1**
Instrumentos	Ilumina NovaSeq
Estrategia	RNA-Seq
Fuente	Célula Única Transcriptómica
Selección	RT-PCR
Diseño	Emparejado

Para describir un poco estas especificaciones y quede lo mas claro posible.

- ***Ilumina NovaSeq***
  Esta es una tecnología de secuenciación por síntesis (SBS), se considera como una tecnología de "*next generation*" por ser generalmente muy popular por permitir secuenciación masiva en paralelo mediante métodos que detecta bases individuales en forma de que se incorporan las cadenas de ADN en crecimiento. De manera resumida en datos técnicos estas trabajan de una base terminadora reversible marcado por campo de florescencia a disposición que se le incorporan dNTP para así consecuentemente permitir la adicción de siguientes bases.

- ***Estrategia***
  La estrategia utilizada de RNA-Seq, se refiere a la técnica de secuenciación que permite analizar de manera específica de la proporción asociada, ligada los transcriptos de distintos genes que existen en la muestra biológica utilizada.

- ***Fuente***
  La fuente se refiere a la muestra tomada, en este caso es una Transcriptómica única celular, este se tomó de una célula y se busca establecer la interacción con los genes y sus procesos de transcripción, así como expresión génica.

- ***Selección***
  La parte de la selección especifica que fue en RT-PCR, la reacción de cadena

polimerasa con transcriptasa inversa en tiempo real.

- ***Diseño***
  La secuencia en extremos emparejados se refiere a la detención de dos fragmentos en ambas direcciones tanto como positivas y negativas, esto es ayudado por el software del mismo secuenciador, que hace una comparación para una mejor solución a problemas, como la duplicación de bases, la intercesión de fragmentos en la secuencia analizada.

Control de calidad

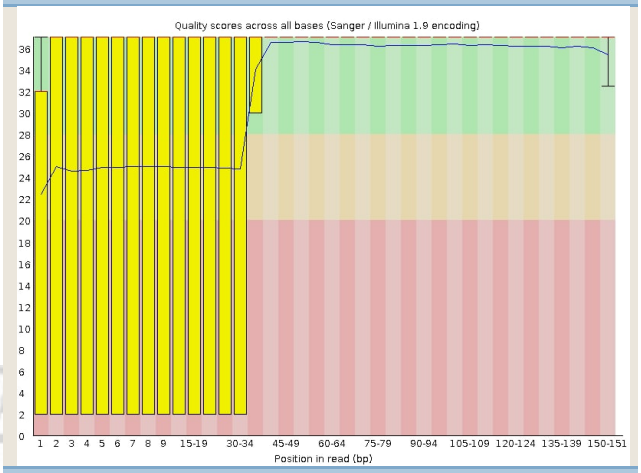
Continuamos con la validación de la calidad de los reads obtenidos de la ya antes mencionada, para esto realizamos unos programa ***FastaQC***, este tiene como objetivo principal revisar las características y realizar comprobaciones de control de calidad de los datos, de la secuencia sin procesar, este software se puede utilizar por medios disponibles en línea y también se puede instalar en un sistema GNU/Linux donde puede ser ejecutado con comandos de bash desde el Shell del equipo, algunas ventajas es que nos proporcionara

- Importación de datos en archivo ***Sam*** contiene la información de las lecturas ya alineadas y archivos ***Bam*** contiene la misma información, pero en formato binario, además que aporta un documento en HTML de rápida visualización.

- Es de rápida ejecución e indica una descripción general en áreas donde pudieron ocurrir algunos problemas en la secuenciación, además de también incluir graficas y tablas para una rápida compresión.

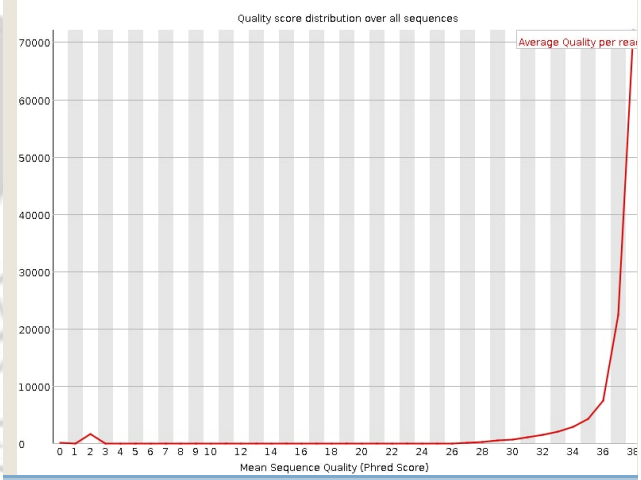
Al realizar este proceso nos descargamos los archivos que están en httml donde podemos hacer una rápida visualización, y aparecerán parámetros

Calidad por secuencia de bases



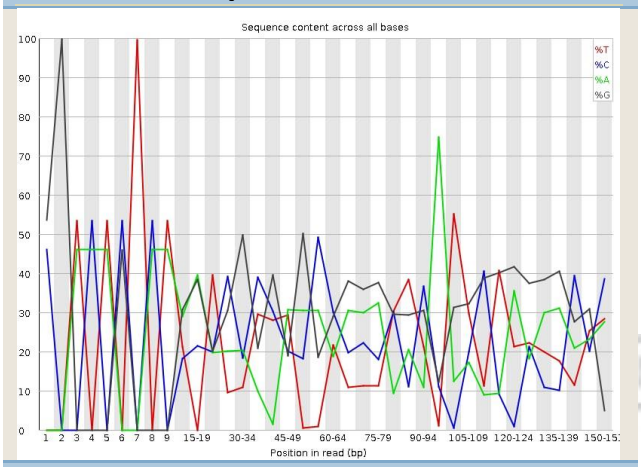
La calidad de secuencias de bases esta muy fuera del rango y no se considera aceptable Phreds score esta muy baja en las primeras secuencias, ya después se ve una estabilidad mas adecuada.

Puntuación de calidad por secuencia



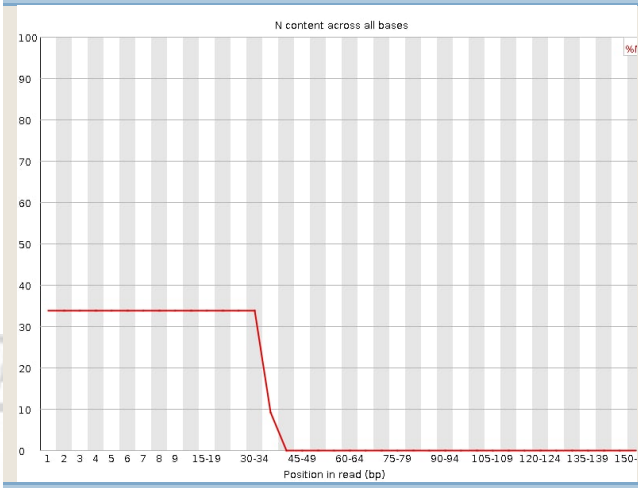
La calidad que se puede observar en esta gráfica se refiere a la calidad de la celda de flujo, aquí se muestra que no hay mucha desviación, entonces no hubo mucha perdida por bases

Contenido por secuencia de bases



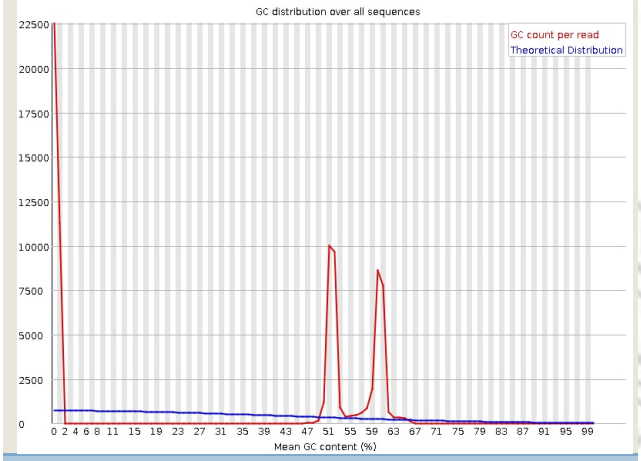
La calidad de los reads en los apartados de las bases, tienen muchas irregularidades, tienen muchos picos, en un parámetro optimo, todas deberían estar una frecuencia que no oscilara, teóricamente tendría que haber una misma cantidad de T

Por contenido N base



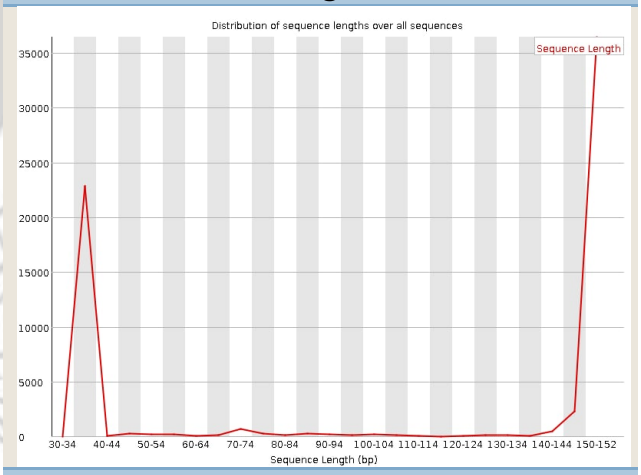
La tecnología de secuenciación, cuando no puede detectar una base, la cambia por la letra “N” en este caso, solo en las primeras posiciones estuvo muy elevado, a partir de punto alarmante se considera estar >20%

Contenido de GC por secuencia



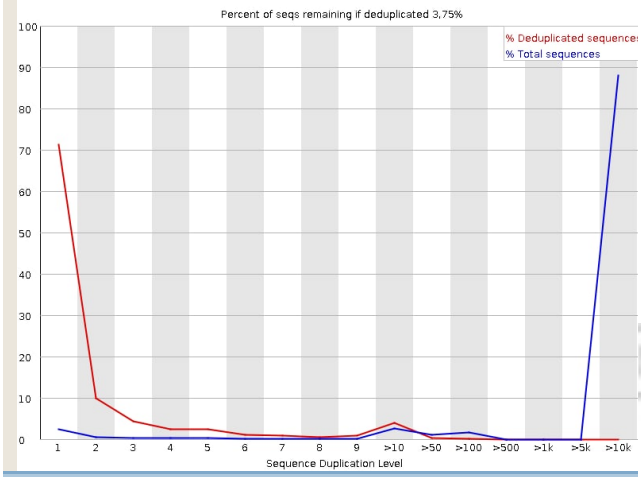
Generalmente esta característica, se toma que si hay uno alto o muy bajo nivel elevado de GC causa un problema para la secuenciación de ilumina, en el grafico se muestra número de reads y se comparan con una distribución teórica, esto debería ser uniforme en el caso de esta grafica se ve un poco elevado el patrón en dos puntos.

Distribución de la longitud de la secuencia



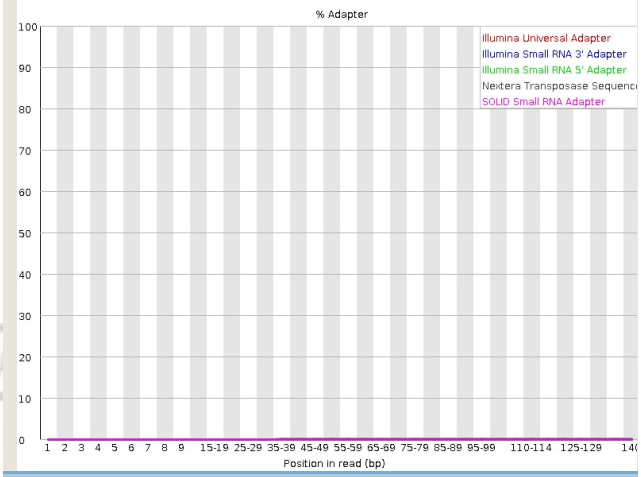
Se genero un alto rendimiento para generar fragmentos de secuencia de longitud uniforme en los reads que ha generado la secuenciación, al principio tuvo un pico, pero al final se elevó de manera muy estrepitosa

Secuencia Niveles duplicados



En esta reads obtenidos, se muestra que el porcentaje de secuencias se replicaron entre replicaron al principio después mantuvo el nivel estable.

Contenido de adaptadores



No se encontraron adaptadores, estos generalmente se utilizan en otra tecnología de secuenciación.

En el caso posterior que hubiera algunos adaptadores estos los tendríamos que eliminara ya que estos no son parte de la secuencia y solo se utilizaron con el fin de adaptar la secuenciación.

Secuencias sobrerrepresentarte

Sequence	Count	Percentage	Possible Sourc
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	22528	33.977346424747	No Hit
GGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTA	19725	29.7497850775983	No Hit
CGAAAGTAAAGCCAGAGGAGATCTCTCGACGCAGGACTCGGCTTGCTGAA	16179	24.401610786842223	No Hit
CGAAAGTAAAGCCAGAGGAGATCTCTCGACGCAGGACTCGGCTTGCTGAAGCGCGTATGGGCAAGCAGGGAG	265	0.3996802557953637	No Hit
GGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCCCTGGCTAACTA	247	0.3725321629488862	No Hit
GGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGACTAACTA	193	0.29108788440945355	No Hit
GGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCCCTCTGGCTAACTA	151	0.22774233443433933	No Hit
GGTCTCTCTGGTTAGACCAGACCTGAGCCTGGGAGCTCTCTGGCTAACTA	134	0.20210246896822165	No Hit
CGAAAGTAAAGCCAGAGGAGATCTCTCGACACAGGACTCGGCTTGCTGAA	114	0.1719379213610244	No Hit
CGAAAGTAAAGCCAGAGGAGATCTCTCGACGCAGGACTCGGCTTGCTGAAGCGCGCACGGCAAGCAGGGAG	101	0.15233096541634616	No Hit
CGAAAGTAAAGCCAGAGGAGATCTCTCGACGCAGGACTCGGCTTGCTGAAGCGTATGGGCAAGCAGGGAG	92	0.1387569189931074	No Hit
CGAAAGTAAAGCCAGAGGAGATCTCTCGACGCAGGACTCGGCTTACTGAA	89	0.13423223685202781	No Hit
GGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAACTCTCTGGCTAACTA	76	0.11462528090734958	No Hit
CGAAAGTAAAGCCAGAGGAGATCCCTCGACGCAGGACTCGGCTTGCTGAA	75	0.11311705352698974	No Hit
GGTCTCTCTGGTTAGACCCGATCTGAGCCTGGGAGCTCTCTGGCTAACTA	67	0.10105123448411082	No Hit

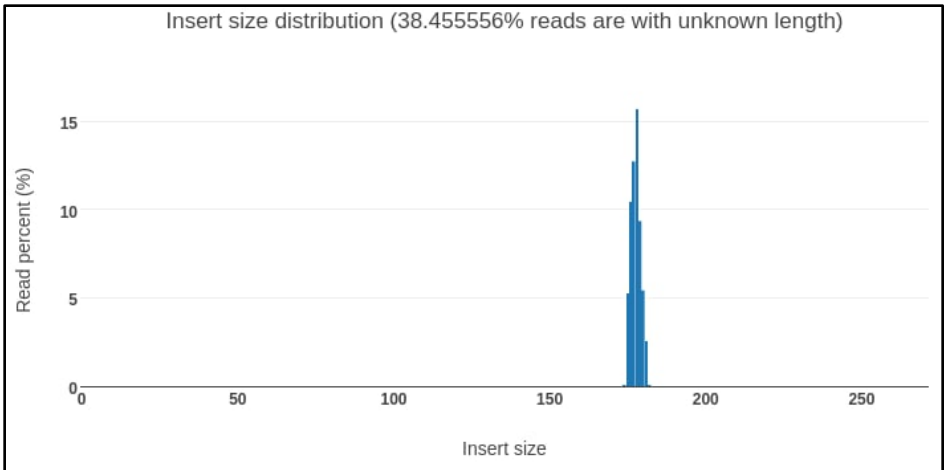
En esta secuencia se busca contaminantes que pudieran apareces durante la secuenciación. Las cuales si tuvo muchas secuencias de contaminación, pero no se encontró de la fuente que provenían, generalmente se contraminan de los cebadores, que en esta ocasión menciona que fueron específicos para esta secuencia

Filtrado

Es este paso se busca de una manera corregir los datos de los reads para que estos no tengan los problemas que anteriormente.

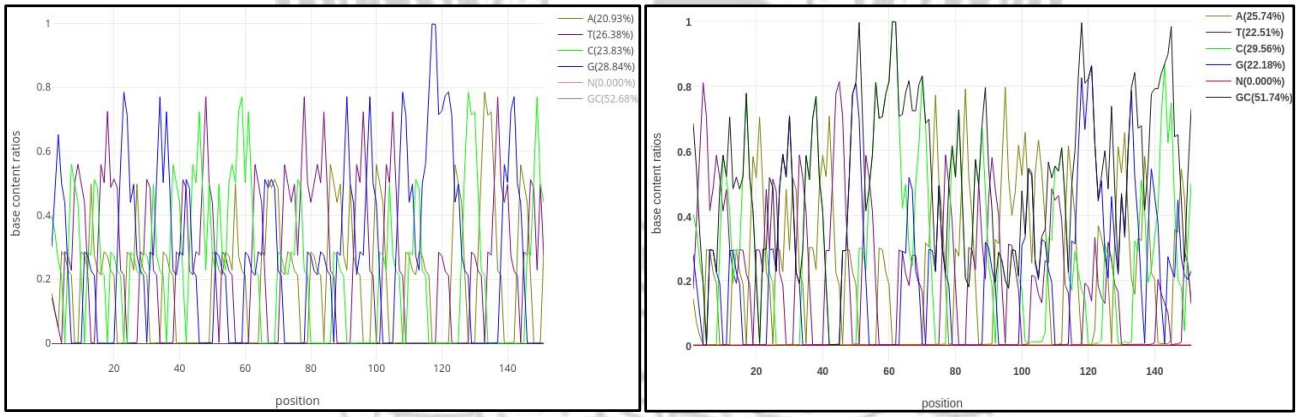
Para filtrar los datos del NGS, utilizamos el programa "***Trimmomatic***", unas características de este, es que es muy potente, tiene mucha popularidad tanto que se ha citado en 2810 veces.

Además, una gran aceptación en la comunidad que utiliza la tecnología Ilumina, algunas ventajas es que estos también se puede ejecutar de una manera muy sencilla con simples comandos en la terminal que incluyen muchos equipos con Linux servidores o algunas Shell que trabajen con Bash, este también realiza su trabajo de una manera muy rápida, puede soportar formatos Phres-33 y phreed-64.



En esta grafica hace una referencia a las superposiciones de los extremos emparejados, hay un 38.4 % de bases en esta situación, esto puede ser por errores en la secuenciación o que no hayan podido ser detectados.

A continuación, un par de graficas donde se puede observas el porcentaje del contenidos de bases antes y después de realizar el filtrado.



Se puede comparar la mejoría después de hacer el filtrado:

	Porcentaje de bases	
	Antes de filtro	Después del filtro
A	25.74%	20.93%
T	22.51%	26.38%
C	29.56%	23.83%
G	22.18%	28.8
CG	51.74%	52.68%
N	0%	0%

Before filtering: read2: KMER counting

Darker background means larger counts. The count will be shown on mouse over.

	AA	AT	AC	AG	TA	TT	TC	TG	CA	CT	CC	CG	GA	GT	GC	GG
AAA	AAAAA	AAAAT	AAAAC	AAAAG	AAATA	AAATT	AAATC	AAATG	AAACA	AAACT	AAACC	AAACG	AAAGA	AAAGT	AAAGC	AAAGG
AAT	AATAA	AATAT	AATAC		AATTA	AATTT	AATTC		AATCA	AATCT	AATCC					
AAC	AACAA	AACAT	AACAC	AACAG		AACTT	AACTC	AACTG	AACCA	AACCT	AACCC	AACCG			AACGC	AACGG
AAG	AAGAA	AAGAT	AAGAC	AAGAG	AAGTA	AAGTT	AAGTC	AAGTG	AAGCA	AAGCT	AAGCC	AAGCG	AAGGA	AAGGT	AAGGC	AAGGG
ATA	ATAAA					ATATT				ATACT						
ATT	ATTAA				ATTTA	ATTTT	ATTTC	ATTTG	ATTCA	ATTCT	ATTCC					
ATC	ATCAA	ATCAT	ATCAC	ATCAG		ATCTT	ATCTC		ATCCA	ATCCT	ATCCC			ATCGT		
ATG	ATGAA															
ACA	ACAAA	ACAAT	ACAAC	ACAAG		ACATT	ACATC		ACACA	ACACT	ACACC	ACACG	ACAGA		ACAGC	ACAGG
ACT						ACTTT			ACTCA	ACTCT	ACTCC		ACTGA	ACTGT		
ACC	ACCAA	ACCAT	ACCAC	ACCAG	ACCTA	ACCTT	ACCTC	ACCTG	ACCCA	ACCCT	ACCCC	ACCCG	ACCGA		ACCGC	ACCGG
ACG									ACGCA	ACGCT		ACGCG	ACGGA		ACGGC	ACGGG
AGA	AGAAA	AGAAT	AGAAC	AGAAG	AGATA	AGATT	AGATC	AGATG	AGACA	AGACT	AGACC	AGACG	AGAGA	AGAGT	AGAGC	AGAGG
AGT	AGTAA						AGTTC	AGTTG		AGTCT	AGTCC					
AGC	AGCAA		AGCAC	AGCAG	AGCTA	AGCTT	AGCTC	AGCTG	AGCCA	AGCCT	AGCCC	AGCCG	AGCGA	AGCGT	AGCGC	AGCGG
AGG	AGGAA	AGGAT	AGGAC	AGGAG	AGGTA	AGGTT	AGGTC	AGGTG	AGGCA	AGGCT	AGGCC	AGGCG	AGGGA	AGGGT	AGGGC	AGGGG
TAA	TAAAA	TAAAT	TAAAC	TAAAG					TAACA				TAAGA	TAAGT	TAAGC	TAAGG
TAT						TATTT										

After filtering: read1: KMER counting

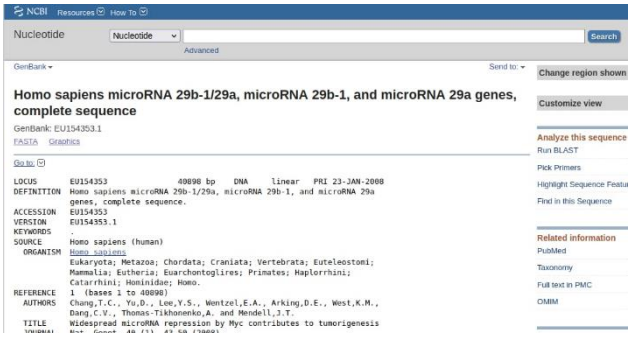
Darker background means larger counts. The count will be shown on mouse over.

	AA	AT	AC	AG	TA	TT	TC	TG	CA	CT	CC	CG	GA	GT	GC	GG
AAA					AAATA											
AAT		AATAT								AATCT						
AAC						AACTT							AACGA			AACGG
AAG											AAGCC					
ATA						ATATT										
ATT														ATTGT		ATTGG
ATC							ATCTC	ATCTG		ATCCT						ATCGG
ATG																
ACA																
ACT					ACTTA		ACTTC			ACTCT						
ACC			ACCAC												ACCGC	ACCGG
ACG		ACGAT		ACGAG					ACGCA				ACGGA			
AGA										AGACT			AGAGA			
AGT																
AGC			AGCAC		AGCTA					AGCCT	AGCCC					AGCGG
AGG		AGGAT														AGGGG
TAA												TAACG				
TAT								TATTG		TATCT						
TAC							TACTC		TACCA							TACGG
TAG									TAGCA							
TTA							TTATC			TTACT					TTAGC	
TTT																

En esta parte de KMER se utiliza de una forma de denotar la consecutivita de los nucleótidos de una manera en que se predice y se va dejando rastro del orden de los nucleótidos según su orden. Se puede observar que después del filtrado se ve un poco más limpio.

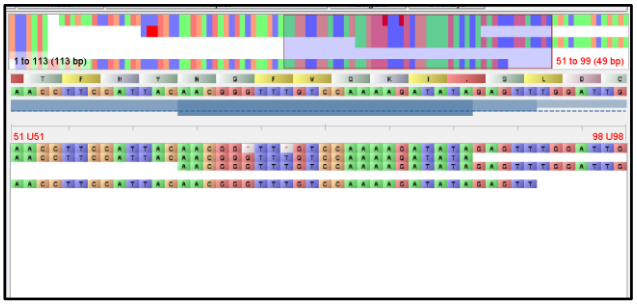
Mapeo de secuencia

Ahora para complementar el manera de complementar el mapeo del genoma, vamos en cuestión se tendría que descargar un genoma que ya se tenga de una manera ordenada, para esto entramos en la pagina del ncbi y nos dirigimos a la parte de nucleótidos, ponemos el nombre de nuestra especie con la que estamos trabajando y la descargamos en formato fasta. Como se mostrará a continuación.



Podemos mapear la calidad de nuestros Reads también de una manera muy fácil, con un simple comando y nos regresara un archivo. Sam.

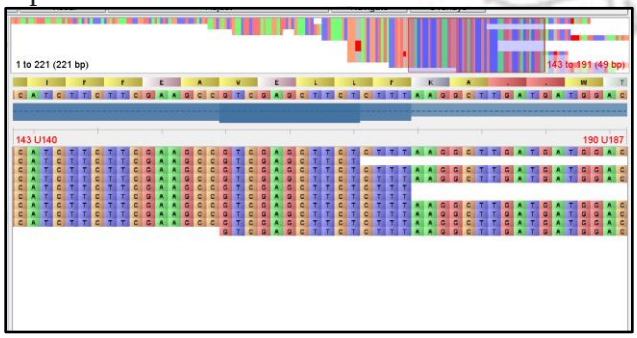
En este caso, cuando ya hayamos descargado la secuencia, la llevaremos a la terminal del Shell y ejecutaremos unos cuantos comando para hacer el índice adecuadamente, y ejecutaremos un comando y nos regresara 6 archivos en .bt2”



Cuando ya hayamos concluido este ejercicio proseguiremos con mapear las lecturas del genoma, allí nos indicar el porcentaje de la reads que se alinearon con el genoma, esto no lo mostrara en un archivo sam, para verificar que este adecuadamente podemos utilizar algún comando para ver las primeras líneas del contenido, no se recomienda ejecutarlo con les o cat ya que desplegaría toda la información y esto tardaría mucho tiempo, entonces podemos utilizar el comando head -n10 mapeo.sam y este mostrara las primeras 10 líneas de esta manera;

```
(base) [svilla@gala ~]$ head -n10 Mapeo_readsZv9.sam
@HD     VN:1.0  SO:unsorted
@SQ     SN:EU154353.1   LN:40898
@PG     ID:bowtie2      PN:bowtie2      VN:2.4.4        CL:"/app/anaconda3/envs/
bowtie2/bin/bowtie2-align-s --wrapper basic-0 --maxins 1000 -x Zv9 -S Mapeo_read
sZv9.sam -1 SRR14930883_1.fastq -2 SRR14930883_2.fastq"
SRR14930883.1   77      *       0       0       *       *       0       0       C
GTCGGATCCTTAGCACTTATCTGGGACGATCTGCGGAGCCTGTGCCTCTTCAGCTACCACCGCTTGAGAGACTTACTCTT
GATTGTAACGAGGATTGTGGAACTTCTGGGACGCAGGGGGTGGGAAGCCCTCAAATATTGGTGGAATCTC  CCCCCCBC
CFFFGGGGGGGGGGGHHHFHGGGHGHHGGGGGGHHHHGHHHHHHHHGHHHHHGGGGGGHHGHGHHHHHHHHHBHHHGH
GHGGGGEHHHGGGHHHHGHHHHHHHGGGGGGGFFGGHHGHHGHHHGFHHHHHHHGGHGHHEF YT:Z:UP
SRR14930883.1   141     *       0       0       *       *       0       0       T
GCCTTTAGTTCCTGACTCCAATACTGTAGGAGATTCCACCAATATTTGAGGGCTTCCCACCCCCTGCGTCCCAGAAGTTC
CACAATCCTCGTTACAATCAAGAGTAAGTCTCTCAAGCGGGGGTAGCTGAAGAGGCACAGGCCCCGCAGA  AABBBFFB
FFFFGGGGFGGGGGHHHHHHFHBFHHHFHHHHHGDHFGGHFGEBEEGFHHHHHGGGGFEEGGGAEHHHHFGHEHHHHFH
f43EG1FFDD4GGF33?4@FEBFFB4GGF3EEC/BC@E<?F1DD11?<A.0<F.<.<A<-@<- YT:Z:UP
SRR14930883.2   77      *       0       0       *       *       0       0       T
GGCGGATCCTTAGCACTTATCTGGGACGATCTGCGGAGCCTGTGCCTCTTCAGCTACCACCGCTTGAGAGACTTACTCTT
```

Esta información sea un poco difícil de comprender ya que esta en un coman do ascii, este lo podremos visualizar en un software especializado llamado Tablet.



ENSAMBLE DE GENOMA

En esta parte es necesario conocer la tecnología con la que se secuencio en mi caso como era Ilumina además de conocer que teóricamente este por decirlo de una manera coloquial se romperá en pedazos mas pequeños de como nos lo entregaron en la secuenciación, entonces nuestro trabajo o desempeño es unirlo de una manera en que menos posible que tengan errores, en este caso utilizaremos en ensamble de comparación dado que ya tenemos una secuencia similar con la que nos podrá servir como guía, pero hay que tener en cuenta que también existe en ensamble de Novo ósea una técnica que se basa en ensamblar el genoma sin alguna secuencia de referencia aquí utilizaremos unos scripts de Python adaptados en bash, para que este corte las secuencias en fragmentos más cortos después este los unirá para formar nuestro ensamble.

QUAST

Quality Assessment Tool for Genome Assemblies by CAB

07 December 2021, Tuesday, 01:30:31

View in Icarus contig browser

All statistics are based on contigs of size >= 500 bp, unless otherwise noted (e.g., "# contigs (>= 0 bp)" and "Total length (>= 0 bp)" include all contigs).

Statistics without reference	mRNA
# contigs	1
# contigs (>= 0 bp)	1
# contigs (>= 1000 bp)	1
# contigs (>= 5000 bp)	1
# contigs (>= 10000 bp)	1
# contigs (>= 25000 bp)	1
# contigs (>= 50000 bp)	0
Largest contig	40 898
Total length	40 898
Total length (>= 0 bp)	40 898
Total length (>= 1000 bp)	40 898
Total length (>= 5000 bp)	40 898
Total length (>= 10000 bp)	40 898
Total length (>= 25000 bp)	40 898
Total length (>= 50000 bp)	0
N50	40 898
N75	40 898
L50	1
L75	1
GC (%)	45.09
Mismatches	
# N's	0
# N's per 100 kbp	0

Conclusión:

Tuve muchos problemas e inconvenientes para realizar el ensamble satisfactoriamente, no tengo la certeza si era la calidad de las secuencias o la comparación del scaffolds era inadecuada, no pude concluir en una idea fija ya que no pude encontrar una manera firme de que llegar a los resultados en este últimos paso, tuve un problema de dificultad en primeras ocasiones en buscar índice de referencia adecuado, ya que no se contaba con un sustento de un genoma que involucrara la parte en la que estaba trabajando, Estas practica me gusto mucho y me gustaría mejorar mis habilidades con el manejo de este tipo de herramientas, y conocer los conceptos de una manera adecuada, en algunas partes del ejercicio, mi poca practica y experiencia, me hicieron tardarme un poco mas de tiempo.

Referencias:

VILLASIS KEEVER, Angelina.A 20 años del descubrimiento del VIH. *Rev. invest. clín.* [online]. 2004, vol.56, n.2, pp.122-123. ISSN 2564-8896.

Zhang, C., Zhang, B., Vincent, M., y Zhao, S.(2016). Bioinformatics Tools for RNA-seq Gene and Isoform Quantification. Journal Of Next Generation Sequencing Applications, 03 (03). doi: 10.4172/2469-9853.1000140

RNAseq-an introduction. (2020). Consultado el 6 diciembre 2020, de https://galaxyproject.org/tutorials/rb rnaseq/

Secuenciaci´on: tecnolog´ıa de Illumina. (2020). Consultado 7 diciembre 2020, de https://support.illumina.com/content/dam/illumina-support/courses/sequencing-illuminatechnology-wbt-esp/story html5.html?iframe