



## Recursos de Investigación en ciencias de la salud: Métodos de trabajo en bioinformática y genética aplicados al estudio del vector de la Enfermedad de Chagas

Helen Jineth Rincón Galvis, Bibiana Matilde Bernal Gómez <https://orcid.org/0000-0002-4897-3368>, Fred Gustavo Manrique Abril <https://orcid.org/0000-0002-0723-3240>

### Artículo original

**Recibido:** 10 de agosto, 2021

**Aceptado:** noviembre 10 de 2021 **Publicado** 2022

### Como citar este artículo:

Rincón Galvis, H. J., Bernal Gómez, B. M., & Manrique Abril, F. G. (2021). Recursos de Investigación en ciencias de la salud: Métodos de trabajo en bioinformática y genética aplicados al estudio del vector de la Enfermedad de Chagas. *Salud y Sociedad Uptc*, 6(2). ISSN(online) E-2744-953X doi 10.19053/uptc.19092407.15323

\* Grupo de Investigación biomédica y y Tecnológica de Colombia [gibp@uptc.edu.co](mailto:gibp@uptc.edu.co)



### RESUMEN

Por medio de un ejemplo concreto de uso de métodos de trabajo en bioinformática se han diseñado las etapas para obtener dos posibles resultados con base en un taller de estudio dirigido de investigación en ciencias biológicas aplicadas a resolver los problemas de la salud pública. Se menciona el uso de estos y la importancia de su aplicación. La investigación en bioinformática y genética ha contribuido a la comprensión de los vectores de la enfermedad de Chagas, especialmente el *Trypanosoma cruzi*. Los reads, secuencias cortas de nucleótidos, pueden

generarse mediante plataformas como Illumina y Oxford Nanopore para obtener o completar información genómica o transcriptómica. La diversidad genética del *T. cruzi* es crucial para entender su patogenicidad y respuestas al tratamiento. Los estudios transcriptómicos revelan aspectos de la fisiología y respuestas inmunitarias de los vectores triatominos, esenciales para estrategias de control vectorial y los reads repetidos pueden usarse para identificarlos por medio de genética de poblaciones y para entender, la interacción huésped vector parásito, así como probables usos posteriores en inmunogenicidad o inmunodiagnóstico.

### **MATERIALES Y MÉTODOS:**

Bajo la dirección del docente Andrés Mauricio Gómez, se llevó a cabo una actividad práctica de aula en la asignatura de bioinformática para explorar secuencias derivadas de secuenciación masiva de genoma y transcriptoma, y luego su análisis, con el objetivo de delimitar y cuantificar los parámetros de calidad de los reads obtenidos. El análisis inicial se centró en evaluar la calidad de los reads en las plataformas Illumina y 454-Roche, donde los puntajes de calidad (Q-value) de los nucleótidos varían entre -40 y 40 en Illumina y entre 0 y 93 en 454-Roche, considerándose óptimos aquellos superiores a 40. Para este análisis, se utilizó la base de datos Sequence Read Archive (SRA) del NCBI y el programa FastQC, evaluando los parámetros de calidad de los reads del organismo *Rhodnius prolixus*, un triatomino vector de la enfermedad de Chagas.

La exploración de datos consistió en evaluar los datos leídos del organismo *Rhodnius prolixus*, un vector triatomino de la enfermedad de Chagas, utilizando FastQC para evaluar los parámetros de calidad de las lecturas obtenidas en las plataformas mencionadas.

### **RESULTADOS**

Un valor Q de 20 indica una probabilidad de error del 1%, lo que sugiere una alta confianza en la precisión de las lecturas, ya que la longitud promedio de las secuencias varía significativamente entre los ejemplos. El análisis de la distribución de nucleótidos mostró una representación uniforme del 25% para la adenina, la timina, la citosina y la guanina, aunque se observaron desviaciones debido a la contaminación y a defectos preanalíticos. Se destacó la presencia de lecturas repetidas,

con niveles aceptables en algunos ejemplos, pero con tasas de repetición en otros, lo que indica posibles errores de secuenciación. Se utilizaron representaciones gráficas de los datos para ilustrar la distribución de nucleótidos y la frecuencia de lectura, lo que ayudó a evaluar rápidamente la calidad y la fiabilidad de la secuenciación.

**PALABRAS CLAVE** reads bioinformática genética recursos de investigación en ciencias de la salud.

**ABSTRACT**

**SUMMARY**

Using a concrete example of bioinformatics work methods, stages were designed to obtain two possible results based on a directed study workshop in biological sciences research applied to solving public health problems. The use and importance of these methods are highlighted. Bioinformatics and genetic research have contributed to understanding the vectors of Chagas disease, especially *Trypanosoma cruzi*. Reads, short nucleotide sequences, can be generated using platforms like Illumina and Oxford Nanopore to obtain or complete genomic or transcriptomic information. The genetic diversity of *T. cruzi* is crucial for understanding its pathogenicity and treatment responses. Transcriptomic studies reveal aspects of the physiology and immune responses of triatomine vectors, essential for vector control strategies. Repeated reads can be used for population genetics to understand host-vector-parasite interactions and potential uses in immunogenicity or immunodiagnosis.

**MATERIALS AND METHODS**

Under the guidance of Professor Andrés Mauricio Gómez, a practical classroom activity in the bioinformatics course was conducted to explore sequences derived from massive genome and transcriptome sequencing and their subsequent analysis, aiming to delimit and quantify the quality parameters of the obtained reads. The initial analysis focused on evaluating the quality of reads on Illumina and 454-Roche platforms, where nucleotide quality scores (Q-

values) range from -40 to 40 on Illumina and from 0 to 93 on 454-Roche, with scores above 40 considered optimal. For this analysis, the Sequence Read Archive (SRA) database from NCBI and the FastQC program were used to evaluate the quality parameters of reads from the organism *Rhodnius prolixus*, a triatomine vector of Chagas disease. Data exploration involved evaluating reads from *Rhodnius prolixus* using FastQC to assess the quality parameters of reads obtained from the mentioned platforms.

## RESULTS

A Q-value of 20 indicates a 1% error probability, suggesting high confidence in read accuracy, as the average sequence length varies significantly between samples. Nucleotide distribution analysis showed a uniform representation of 25% for adenine, thymine, cytosine, and guanine, although deviations were observed due to contamination and pre-analytical defects. The presence of repeated reads was noted, with acceptable levels in some samples but higher repetition rates in others, indicating possible sequencing errors. Graphical representations of the data were used to illustrate nucleotide distribution and read frequency, aiding in the rapid assessment of sequencing quality and reliability.

## Keywords

reads, bioinformatics, genetics, health sciences research resources.

## INTRODUCCIÓN

La investigación en bioinformática y genética ha avanzado significativamente en la comprensión de los vectores de la enfermedad de Chagas, en particular mediante el estudio del *Trypanosoma cruzi* y su dinámica de transmisión. En secuenciación, los **reads** son secuencias cortas de nucleótidos obtenidas a partir de fragmentos de ADN o ARN. Para obtener reads, se prepara una biblioteca añadiendo adaptadores a los fragmentos de ADN o ARN, que luego se cargan en una máquina de secuenciación. Dependiendo de la tecnología (como Illumina o Oxford Nanopore), los reads se generan mediante métodos como la incorporación de nucleótidos fluorescentes o la

detección de cambios en la corriente eléctrica. Estos reads se ensamblan y analizan para reconstruir la secuencia original y obtener información genómica o transcriptómica relevante.

### **Diversidad genética y tipificación**

La diversidad genética del *T. cruzi* es crucial para comprender su patogenicidad y las respuestas al tratamiento. Estudios recientes han desarrollado bases de datos de referencia para tipificar cepas de *T. cruzi* mediante la secuenciación de amplicones de la región hipervariable del minicírculo, lo que mejora la sensibilidad y la especificidad de la identificación de las cepas (Rusman et al., 2023). Este conocimiento genético ayuda a identificar las unidades discretas de tipificación (DTU) más prevalentes en diversas poblaciones, como el predominio del TCI en los vectores selváticos (Valença-Barbosa et al., 2022).

### **Enfoques transcriptómicos**

Los estudios transcriptómicos han proporcionado información valiosa sobre la biología de los vectores triatómicos, revelando aspectos de su fisiología y respuestas inmunitarias que son esenciales para las estrategias de control vectorial (Borsatto et al., 2022). Estos estudios mejoran la comprensión de la forma en que estos vectores interactúan con el *T. cruzi*, lo que podría servir de base para intervenciones específicas.

### **Cepas bioluminiscentes para la investigación**

La creación de cepas bioluminiscentes y fluorescentes de *T. cruzi* permite monitorizar in vivo de forma no invasiva la dinámica de la infección, lo que facilita el estudio de las interacciones huésped-parásito y para medir la eficacia de los medicamentos (Olmo et al., 2024). Este enfoque innovador puede acelerar los procesos de detección de fármacos y garantizar que los nuevos tratamientos sean eficaces en diversas cepas de *T. cruzi*.

La integración de metodologías genéticas, transcriptómicas e innovadoras de bioluminiscencia proporciona un marco integral para avanzar en la investigación de la enfermedad de Chagas, aunque sigue habiendo desafíos para abordar la complejidad de la enfermedad y las funciones ecológicas de los vectores

## **METODOLOGÍA**

Por medio de la guía de trabajo de la asignatura bioinformática propuesta por el docente Andrés Mauricio Gómez (Gómez, 2021 ) se realizó una actividad de aula para explorar secuencias derivadas de secuenciamiento masivo de genoma y transcriptoma, con base en la delimitación y cuantificación de los parámetros de calidad de los reads obtenidos. El análisis inicial en el proceso de secuenciación de genomas y transcriptomas se centró en la delimitación y cuantificación de los parámetros de calidad de los reads obtenidos. En la plataforma Illumina, los puntajes de calidad (Q-value) de los nucleótidos tienen una variación entre -40 y 40. Por otro lado, en la plataforma 454-Roche, los puntajes de calidad oscilan entre 0 y 93, considerándose óptimos aquellos a partir de 40.

## **Herramientas y Bases de Datos**

Para el análisis de calidad, se usó la base de datos Sequence Read Archive (SRA) del NCBI y el programa FastQC.

## **Procedimiento**

**Exploración de Datos:** Se realizó una exploración de los datos de read sobre el organismo de interés, *Rhodnius prolixus*, un triatomino vector del Chagas.

**Análisis de Calidad:** Utilizando FastQC, se evaluó los parámetros de calidad de los reads obtenidos de las plataformas mencionadas.

## **RESULTADOS**

Los métodos empleados para analizar los datos de secuenciación, particularmente mediante el examen de la calidad de la lectura de reads, la distribución de nucleótidos y la presencia de lecturas repetidas, se basaron en las siguientes medidas:

**Evaluación de la calidad de las lecturas:** la evaluación de la calidad de lectura utilizó valores Q, que indican la probabilidad de error asociada a cada nucleótido. Por ejemplo, un valor Q de 20 corresponde a una probabilidad de error del 1%, lo que sugiere una alta confianza en la precisión de las lecturas [1: *Se tomaron tres ejemplos de read y se evidenció que el tamaño promedio de las secuencias en el primer ejemplo estuvo dentro de los 699 hasta 800, y tamaños de 1400 hasta 1349; para el segundo ejemplo nos muestra varios picos, el cual registra variedad de tamaños, uno de los más altos varia de 520 hasta 539, el segundo que le sigue varia de 379 a 440 y el siguiente varia de 120 a 139; el color del resultado nos indica que no está mal pero debemos tener cuidado porque puede estar pasando algo, según la plataforma 454 GS FLX de Titanium, permite longitudes de lectura de hasta 1000 bases (con una medida de 700 pb) y 1.000000 de lecturas por carrera lo que significa un throughput de 700 MB por carrera en 23 horas aproximadamente.*].

**Análisis de la longitud de las secuencias:** se identificó las longitudes medias de las secuencias obtenidas en la plataforma de secuenciación. Algunas longitudes de lectura esperadas para la plataforma 454 GS FLX Titanium pueden alcanzar hasta 1000 bases, con un promedio de alrededor de 700 bases. Esto es crucial para comprender el rendimiento y la eficiencia del proceso de secuenciación [2: *En las graficas 1 y 2 sobre el contenido de Guanina y Citocina podemos observar que su distribución es muy similar (esperada o teórica y observada) con los mismos picos de concentración para el primer ejemplo 52% y para el segundo 45%, mostrándonos una escala más homogénea en el segundo ejemplo durante toda la secuencia.*].

**Análisis de la composición de nucleótidos:** La distribución de los nucleótidos (adenina, timina, citosina y guanina) en las lecturas se correlacionó con una distribución uniforme del 25% para cada nucleótido, con una secuenciación aleatoria e imparcial. Las desviaciones con respecto a esta distribución esperada se basan en datos obtenidos con contaminación y otros defectos preanalíticos

**Examen de las lecturas repetidas:** se destaca la importancia de identificar las lecturas repetidas en los datos de secuenciación pues un alto porcentaje de lecturas idénticas puede sugerir problemas, con interpretaciones erróneas de los datos. Por ejemplo, uno de los ejemplos mostró que casi el 90% de las lecturas se hacían con una o dos repeticiones, lo que se considera aceptable, mientras que otro ejemplo indicaba más de 10 repeticiones, lo que indica un posible error [5: *A partir de los datos de baja calidad, podemos observar 5 picos, el valor Q de 14 (Con un total de casi 3 reads), 15 (con 7 reads), 17 (con 18 reads) y 19 (con 23 reads) que nos indica una probabilidad del error asociado, indicándome si ese nucleótido es verdaderamente el nucleótido y su 50% dentro de este grupo que sería el valor de Q asociado 20, registraría un error de probabilidad de 0.01 con una correcta probabilidad de 0.99% de ser el nucleótido correcto el cual reporta más o menos 17 el número de reads*].

**Representación gráfica de los datos:** El artículo utiliza representaciones gráficas para ilustrar la distribución de los nucleótidos y la frecuencia de las lecturas repetidas. Estas herramientas visuales ayudan a evaluar rápidamente la calidad y confiabilidad de los datos de secuenciación [6: *Las anteriores graficas nos muestra el porcentaje de cada nucleótido con la posición del reads, donde podemos evidenciar, en el primer ejemplo, las primeras 9 posiciones de los read registran un 100% de la base de cada uno de los nucleótidos el primero con Guanina, el segundo con timina, el tercero con Adenina y el cuarto con citocina, después va adenina, citocina, timina y guanina, luego se empiezan a regular los porcentajes, aunque con mayor proporción en la mayoría de los read con el nucleótido de guanina, uno esperaría que su composición fuera equivalente, por lo que fue establecido aleatoriamente a lo largo de todos los reads, el cual nos indica que algo está pasando, puede que las secuencia tengan los adaptadores o se registre una contaminación; por otro lado, el segundo ejemplo nos muestra una composición equidistante a lo largo de cada posición en todos los reads que varía entre 35 a 15%, aunque al final de la secuencia, podemos evidenciar unos porcentajes altos de guanina, seguido de adenina y citocina, aunque la mayor proporción en la*



*mayoría de las posiciones de los reads, fue el contenido de adenina-timina que vario de 35% a 25%, mientras que guanina y citocina vario entre 25% a 15%)*

Los resultados presentados en el documento se centran en el análisis de los datos de secuenciación, particularmente en lo que respecta a la calidad de las lecturas y la distribución de los nucleótidos.

**Calidad de las lecturas de secuenciación:** El análisis de la calidad de las lecturas muestra valores de Q variables, lo que indica diferentes probabilidades de error. Por ejemplo, en el conjunto de datos SRR1056546, los valores de Q oscilaron entre 14 y 19, con un pico notable en el Q20, lo que sugiere una probabilidad de error del 1%. Esto indica que es probable que aproximadamente el 50% de las lecturas sean precisas, con un nivel de confianza del 99% para esas lecturas.

**Evaluación de la calidad promedio:** En otro conjunto de datos, el SRR1057873, la calidad media de las lecturas fue del 29%. Esto sugiere que la mayoría de las lecturas son de buena calidad, lo que hace que los datos sean creíbles para un análisis posterior.

**Distribución de nucleótidos:** El artículo analizó la distribución de los nucleótidos, centrándose particularmente en el contenido de GC. En un ejemplo, se observó un contenido de GC del 52%, lo que se alinea estrechamente con la distribución teórica esperada en una secuencia genómica. Esto indica una representación equilibrada de los nucleótidos, con un 48% de adenina y timina [3: *En las gráficas vemos el número de repeticiones de read idénticos y su porcentaje, no olvidemos que estos fragmentos son aleatorios por lo cual se esperaría que no hubieran muchas repeticiones a lo largo de la secuencia, como podemos observar en el segundo ejemplo que su porcentaje es casi del 90% para una o dos repeticiones, siendo esto lo que se espera comparado con el primer ejemplo, la cual nos muestra que hay un número más de 10 repeticiones en reads idénticos con un 10 % registrado, también se puede evidenciar que más o menos 4, 5 y 6 repeticiones también presenta un porcentaje de read repetidos idénticos en las secuencias, mostrándonos un resultado rojo, un error.*].

**Presencia de lecturas repetidas:** Los resultados también destacan la presencia de lecturas repetidas en los datos de secuenciación. En un caso, casi el 90% de las lecturas tuvieron una o dos

repeticiones, lo que se considera aceptable. Sin embargo, en otro ejemplo, se observaron más de 10 repeticiones, lo que suscitó una posible fuente de errores de secuenciación. (4: Si el secuenciamiento fuera perfecto, esperaríamos que cada nucleótido ocupara  $\frac{1}{4}$ , es decir el 25% de Guanina, el 25% de citosina, un 25% de Adenina y un 25% de Timina, esta fuera la proporción a lo largo de cada posición en todos los reads ya que sería secuenciados a partir de fragmentos aleatorios)

**Representación gráfica:** Los resultados están respaldados por representaciones gráficas que ilustran la calidad y la distribución de los nucleótidos en los conjuntos de datos. Estas ayudas visuales ayudan a evaluar rápidamente la fiabilidad de los datos de secuenciación e identificar cualquier anomalía [3].

## CONCLUSIONES

Las lecturas de los reads de secuenciación son de alta calidad, y existe una distribución equilibrada de los nucleótidos y la importancia de supervisar las lecturas repetidas para garantizar su integridad. Los hallazgos son cruciales para comprender la confiabilidad del proceso de secuenciación y sus implicaciones para futuras investigaciones. La guía seguida con los métodos utilizados en este artículo hacen concluir que un análisis exhaustivo de los datos de secuenciación mediante la evaluación de la calidad, la distribución de los nucleótidos y la identificación de lecturas repetidas, con el apoyo de representaciones gráficas para facilitar su comprensión, son un método de investigación de uso en ciencias de la salud.

Los *reads* en biología molecular pueden modificar la manera de realizar diagnósticos en medicina incluyendo la parasitología, ya que al ser fragmentos cortos, que tienen gran diversidad pueden servir como biomarcadores para identificar microorganismos, particularmente en muestras de pacientes; pueden identificar cambios en el ADN, como mutaciones, que pueden estar asociadas con enfermedades o resistencia a medicamentos y se han usado para medir acciones/expresiones de un gen específico. Son usados en estudios poblacionales particularmente por la variabilidad genética dentro de una población, que sirve para entender tanto la simbiosis como la evolución de

los organismos. Pueden ser biomarcadores más precisos y contribuyen al desarrollo de inmunógenos, genes de resistencia, y la denominada epidemiología molecular. Son una herramienta fundamental en la investigación biomédica y parasitológica, ya que permiten analizar el material genético de una manera rápida y eficiente, lo que a su vez facilita el diagnóstico, el desarrollo de nuevas terapias y la comprensión de las enfermedades infecciosas.

[MT\\_24 h PBM3 \(quinto estadio\)](#)

1. 1 ejecución de ILLUMINA (Illumina NovaSeq 6000): 20,7 M de puntos, 6,2 G de bases, 777 Mb de descargas  
Adhesión: SRX10876917

[MT\\_24 h PBM2 \(quinto estadio\)](#)

2. 1 ejecución de ILLUMINA (Illumina NovaSeq 6000): 20 millones de puntos, bases 6G, 779,2 MB de descargas  
Adhesión: SRX10876916

[MT\\_24 h PBM1 \(quinto estadio\)](#)

3. 1 ejecución de ILLUMINA (Illumina NovaSeq 6000): 21,5 M de puntos, 6,4 G de bases, 807,6 Mb de descargas  
Adhesión: SRX10876915

[MT\\_3 h PBM3 \(quinto estadio\)](#)

4. 1 ejecución de ILLUMINA (Illumina NovaSeq 6000): 20,3 M de puntos, 6,1 G de bases, 987 Mb de descargas  
Adhesión: SRX10876914

[MT\\_3 h PBM2 \(quinto estadio\)](#)

5. 1 ejecución de ILLUMINA (Illumina NovaSeq 6000): 21,2 M de puntos, 6,4 G de bases, 1 Gb de descargas  
Adhesión: SRX10876913

[MT\\_3 h PBM1 \(quinto estadio\)](#)

6. 1 ejecución de ILLUMINA (Illumina NovaSeq 6000): 20,5 millones de puntos, bases de 6,1 G, descargas de 1 Gb  
Adhesión: SRX10876912

[MT\\_UFC3 \(quinto estadio\)](#)

7. 1 ejecución de ILLUMINA (Illumina NovaSeq 6000): 21,1 M de puntos, 6,3 G de bases, 956,7 Mb de descargas  
Adhesión: SRX10876911

[MT\\_UFC2 \(quinto estadio\)](#)

[SRX39722](#): 454 whole genome shotgun sequencing of genomic DNA library '454-5198' containing sample '4747.6.a.1'  
1 LS454 (454 GS FLX) run: 663 spots, 297,232 bases, 781,713b downloads

UUID: bcac4ecc-6b9e-414c-a4e5-ec11535b21ff

Design: 454 sequencing of Dengue virus via random polymerase chain reaction

Submitted by: Broad Institute (BI)

Study: Comparative Genomics of Dengue virus for Broad Institute Viral Genomics Initiative  
[PRJNA31235](#) • [SRP014724](#) • [All experiments](#) • [All runs](#)  
[show Abstract](#)

Sample: Generic sample from Dengue virus  
[SAMN02395399](#) • [SRS518161](#) • [All experiments](#) • [All runs](#)  
Organism: [Dengue virus](#)

Library:  
Name: 454-5198  
Instrument: 454 GS FLX  
Strategy: WGS  
Source: GENOMIC  
Selection: RANDOM PCR  
Layout: SINGLE

Spot descriptor:  


Experiment attributes: [show all 6 attributes...](#)

Pipeline: [show...](#)

Runs: 1 run, 663 spots, 297,232 bases, [781,713b](#)

Run	# of Spots	# of Bases	Size	Published
<a href="#">SRR1056546</a>	663	297,232	781,713b	2013-12-22

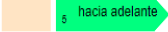
[SRX397927](#) : el pH del lago es 4.0  
1 LS454 (454 GS FLX Titanium) ejecución: 5.125 puntos, 4.2M bases, 10.4Mb descargas

Presentado por: INSTITUTO DE GEOGRAFÍA Y LIMNOLOGÍA DE NANJING, ACADEMIA CHINA DE CIENCIAS (INSTITUTO DE GEOGRAFÍA Y LIMNOLOGÍA DE NANJING, CHIN)

Estudio: Bacterioplankton a lo largo de gradiente de pH Locus dirigido (Loci)  
[PRJNA231216](#) • [SRP034651](#) • [Todos los experimentos](#) • [Todas las ejecuciones](#)  
[mostrar Resumen](#)

Muestra: Muestra genérica de la comunidad ambiental de bacterioplankton 20  
[SAMN02441977](#) • [SRS518506](#) • [Todos los experimentos](#) • [Todas las ejecuciones](#)  
Organismo: [metagenoma de agua dulce](#)

Biblioteca:  
Instrumento: 454 GS FLX Titanium  
Estrategia: AMPLICON  
Fuente: METAGENOMIC  
Selección: PCR  
Diseño: ÚNICO

Descriptor de spot:  



Carreras: 1 carrera, 5.125 lugares, 4.2M de bases, [10.4Mb](#)

Correr	# de puntos	# de Bases	Tamaño	Publicado
<a href="#">SRR1057873</a>	5.125	4.2M	10.4 Mb	2013-12-24

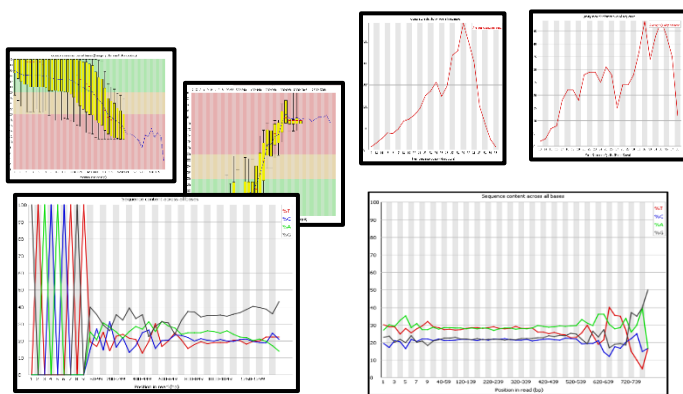
Gráfica 1. Búsquedas de secuencias (reads)

	Organismo	N° de read	N° de bases	Tamaño	Library	Plataforma
<b>SRR1057873</b>	<i>Bacterioplancton, metagenoma de agua dulce</i>	5,125	4.2M	10,4 Mb	<b>Instrumento:</b> 454 GS FLX Titanium <b>Estrategia:</b> AMPLICON <b>Fuente:</b> METAGENOMIC <b>Selección:</b> PCR <b>Diseño:</b> ÚNICO	454 GS FLX Titanium  <b>Codificación</b> Sanger / Illumina1.9
<b>SRR1056546</b>	virus del dengue	663	297,232	781,713b	<b>Nombre:</b> 454-5198 <b>Instrumento:</b> 454 GS FLX <b>Estrategia:</b> WGS <b>Fuente:</b> GENOMIC <b>Selección:</b> RANDOM PCR <b>Diseño:</b> ÚNICO	454 GS FLX  <b>Codificación</b> Sanger / Illumina1.9

Gráfica 2. Secuencias SRR1057873 y SRR1056546

 Basic Statistics	Total, de reads	Rango de tamaño de estos reads	Composición nucleotídica tienen (G/C)
<b>SRR1057873</b>	5125	61-1495	52
<b>SRR1056546</b>	663	66-798	43

Gráfica 3. Total, rangos y composición nucleotídica (G/C) de reads.



Gráfica 4. Diferentes análisis estadísticos y curvas del software de análisis

## REFERENCIAS BIBLIOGRÁFICAS

1. Gómez Palacio, Andrés Mauricio. (2021) Talleres y apuntes de clase Bioinformática y genética. Maestría en Ciencias biológicas Universidad Pedagógica y Tecnológica de Colombia.
2. Rincón Galvis, Helen Jineth (2021) Taller práctico Ensamblaje de secuencias biológicas: análisis de calidad de reads.
3. Forero Pineda, N., & Gómez Palacio, A. M. (2018, November). ANÁLISIS DEL VIROMA DE *Aedes aegypti* (Diptera: Culicidae) DE COLOMBIA. In XIII Encuentro Facultad de Ciencias-UPTC I Encuentro Internacional. disponible en: [https://repositorio.uptc.edu.co/bitstream/001/5698/1/Analisis\\_del\\_viroma\\_de\\_aedes\\_aegypti\\_diptera\\_culicidae\\_de\\_colombia.pdf](https://repositorio.uptc.edu.co/bitstream/001/5698/1/Analisis_del_viroma_de_aedes_aegypti_diptera_culicidae_de_colombia.pdf)
4. Rusman, F., Díaz, A. G., Ponce, T., Floridia-Yapur, N., Barnabé, C., Diosque, P., & Tomasini, N. Wide reference databases for typing *Trypanosoma cruzi* based on amplicon sequencing of the minicircle hypervariable region. *PLOS Neglected Tropical Diseases*. <https://doi.org/10.1371/journal.pntd.0011764>
5. Valença-Barbosa, C., Finamore-Araujo, P., Moreira, O. C., Alvarez, M. V. N., Borges-Veloso, A., Barbosa, S. E., Diotaiuti, L., & Souza, R. de C. M. de. High Parasitic Loads Quantified in Sylvatic *Triatoma melanica*, a Chagas Disease Vector. *Pathogens*. <https://doi.org/10.3390/pathogens11121498>
6. Borsatto, K. C., Coronado, M. A., Galvão, C., Arni, R. K., & Alevi, K. C. C. Transcriptomics Applied to the Study of Chagas Disease Vectors. *American Journal of Tropical Medicine and Hygiene*. <https://doi.org/10.4269/ajtmh.21-0636>
7. Olmo, F. J., Kelly, J. M., Kelly, J. M., Ward, A. I., Ward, A. I., Kelly, J. M., Kelly, J. M., Kelly, J. M., Ward, A. I., & Kelly, J. M. A panel of phenotypically and genotypically diverse bioluminescent:fluorescent *Trypanosoma cruzi* strains as a resource for Chagas disease research. *PLOS Neglected Tropical Diseases*. <https://doi.org/10.1371/journal.pntd.0012106>

