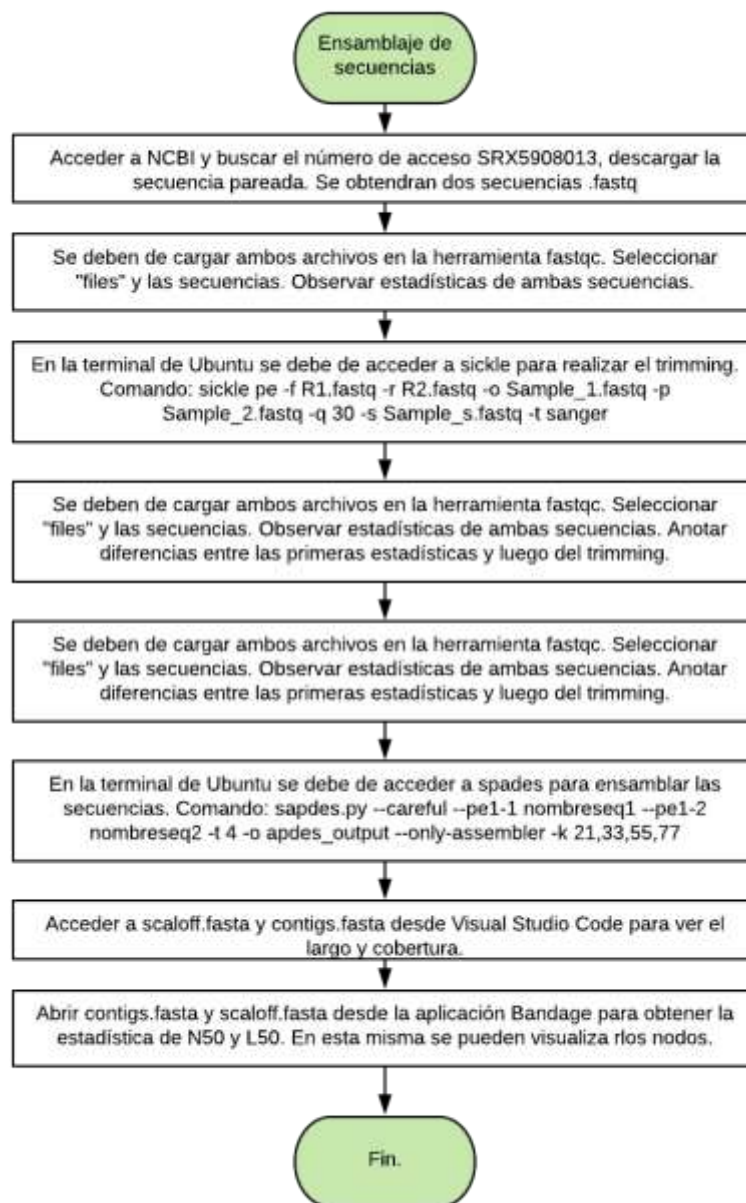


Universidad del Valle de Guatemala
Andrea Paniagua #18733
Bioinformática.

Informe de Ensamblaje

- Diagrama de flujo



- Datos trabajados

Run	Spots	Bases	Size	GC content	Published	Access Type
SRR9134252	972.8k	292.7Mbp	142.2M	37.3%	2019-05-28	public

Imagen 1. Metadata de las secuencias SRR9134252

- Nombre de secuencias
 - SRR9134252_2.fastqc
 - SRR9134252_1.fastqc
- Número de acceso SRA
 - SRX5908013
- Plataforma de secuenciación
 - En NCBI Sequence Read Archive se obtuvieron ambas muestras. Con el SRA toolkit versión 2.10.8 para Windows x64 se extrajeron las secuencias a la computadora utilizada (Asus Vivobook). Luego en FastQC versión 0.11.9 se cargaron los archivos para analizar estadísticamente el control de las muestras.
- Tamaño de lecturas
 - 357,593 KB, total de secuencias 972750 y 357,577 KB , total de secuencias 972750.
- Formato de archivos
 - Fastqc

- Resultados

- Control de calidad lecturas en crudo

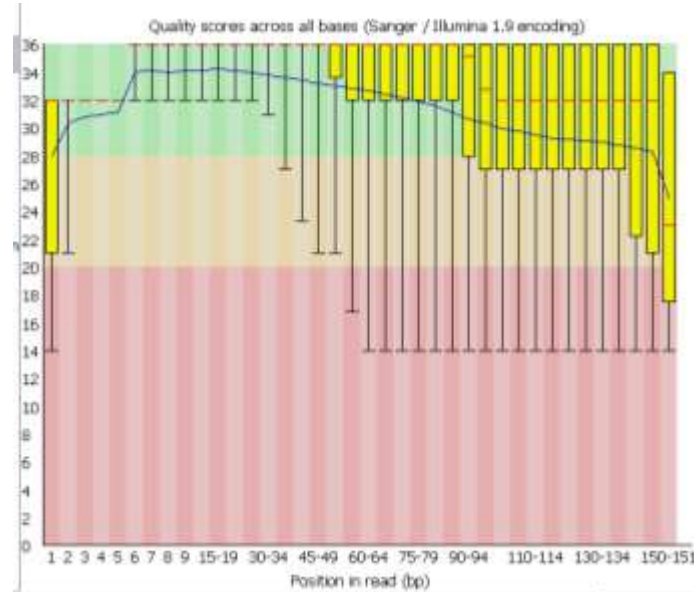


Imagen 2. Distribución de puntaje de calidad por base de secuencia SRR9134252_1

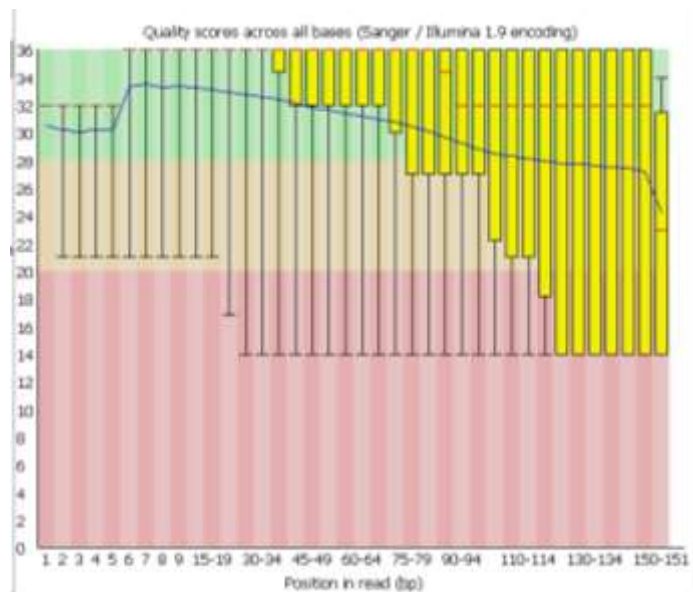


Imagen 3. Puntaje de calidad por base de secuencia SRR9134252_2

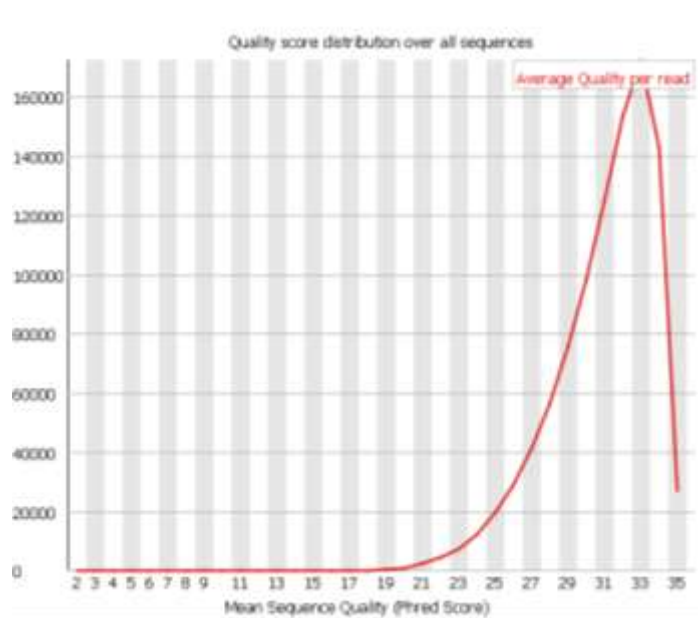


Imagen 4. Distribución del promedio de puntaje de calidad de secuencia SRR9134252_1

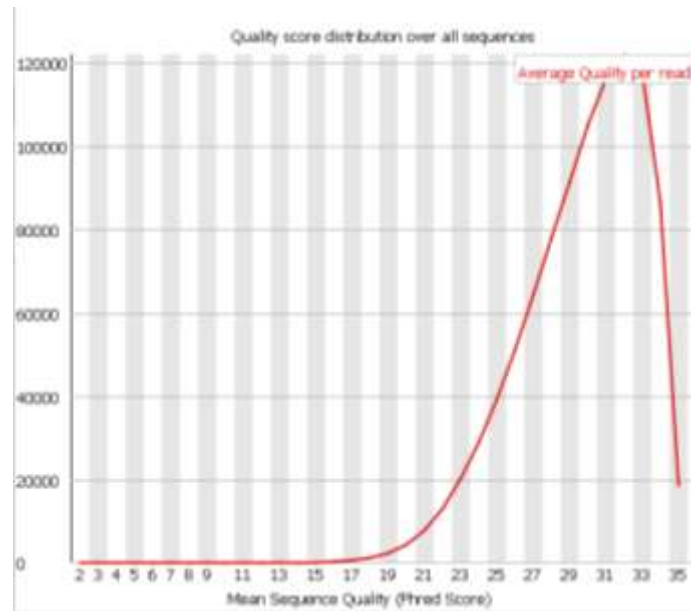


Imagen 5. Distribución del promedio de puntaje de calidad de secuencia SRR9134252_2

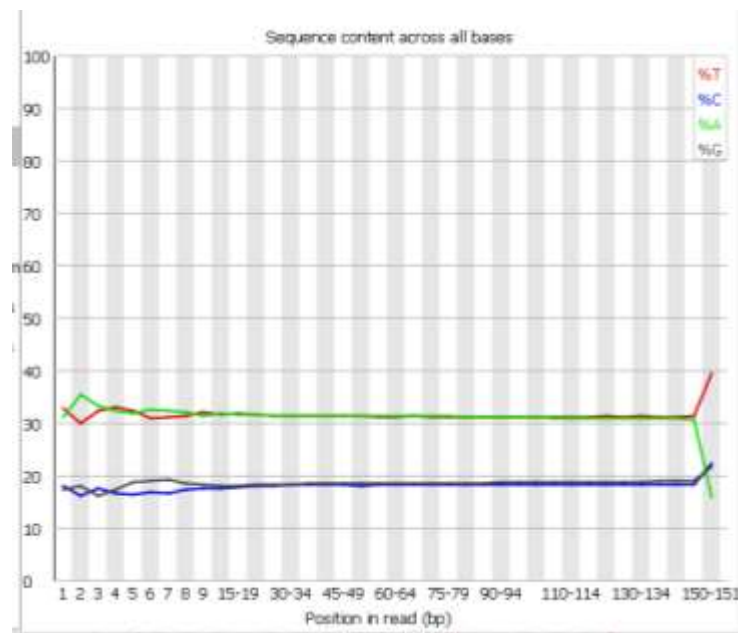


Imagen 6. Distribución de bases de secuencia SRR9134252_1

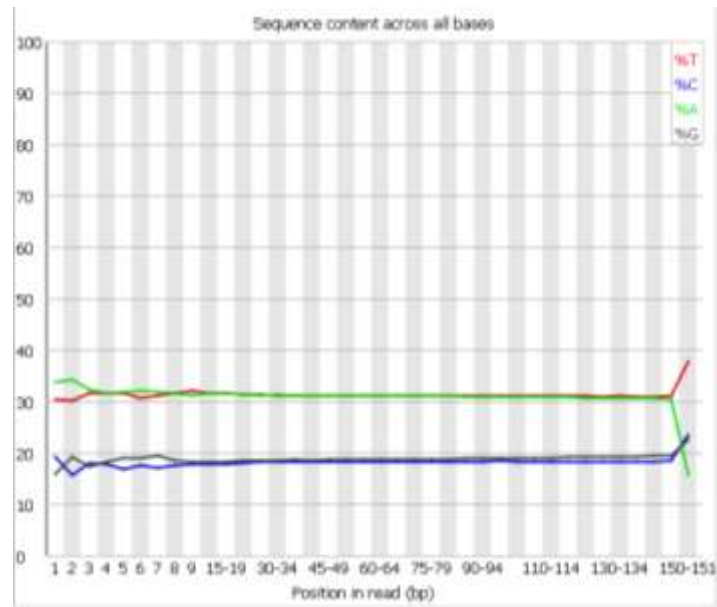


Imagen 7. Distribución de bases de secuencia SRR9134252_2

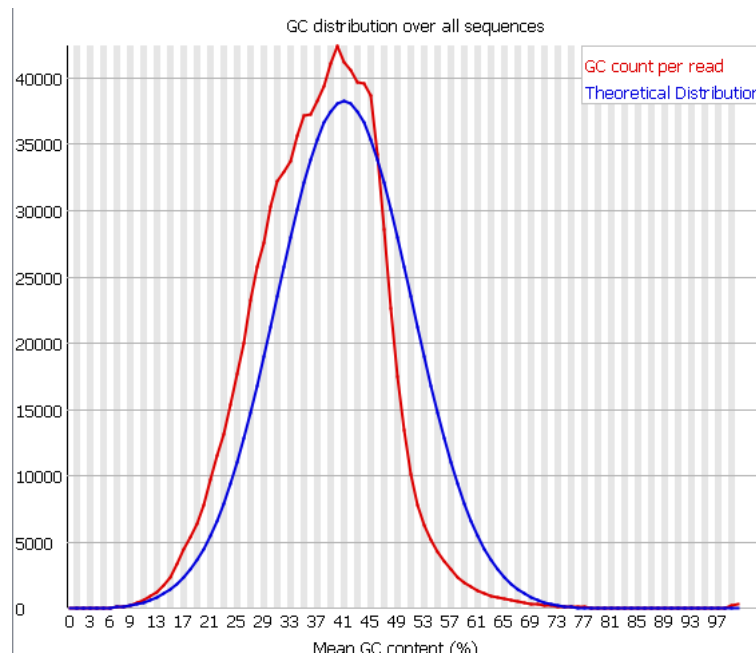


Imagen 8. Distribución de GC por secuencia SRR9134252_1

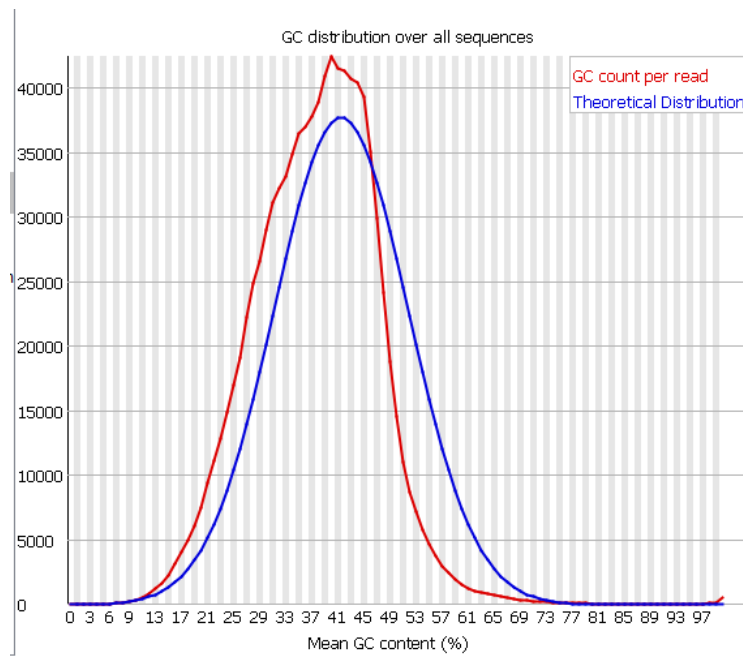


Imagen 9. Distribución de GC por secuencia SRR9134252_2

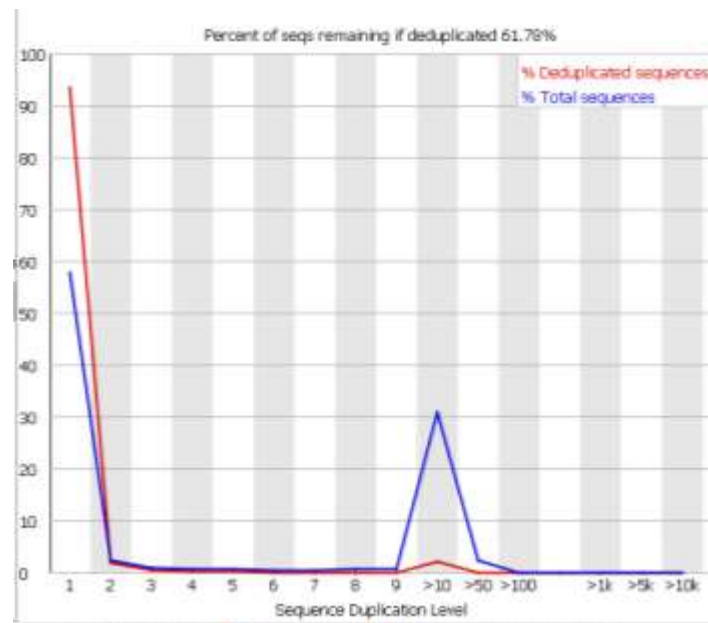


Imagen 10. Porcentaje de duplicación en secuencia SRR9134252_1

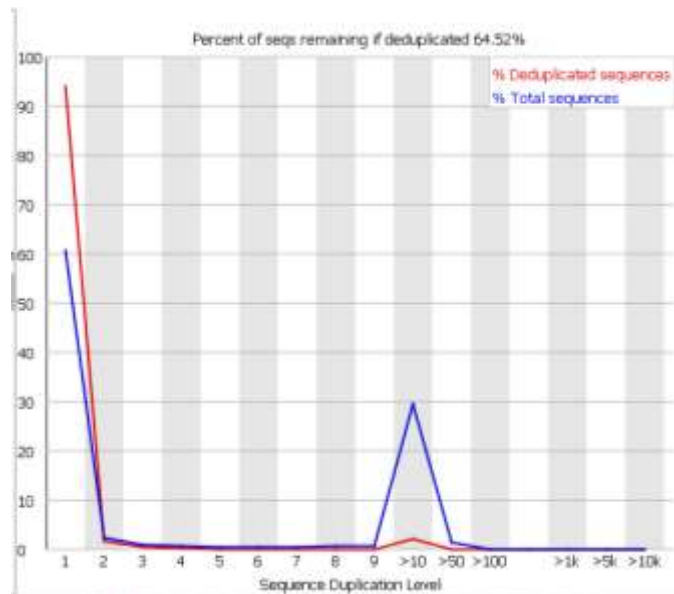


Imagen 11. Porcentaje de duplicación en secuencia SRR9134252_1

○ Resultados de trimming

```
andreapaniagua@LAPTOP-TU09DUJ1:/mnt/c/Users/Andrea Paniagua/Desktop$ sickle p
PE forward file: SRR9134252_1.fastq
PE reverse file: SRR9134252_2.fastq

Total input FastQ records: 1945500 (972750 pairs)

FastQ paired records kept: 1413502 (706751 pairs)
FastQ single records kept: 178068 (from PE1: 146499, from PE2: 31569)
FastQ paired records discarded: 175862 (87931 pairs)
FastQ single records discarded: 178068 (from PE1: 31569, from PE2: 146499)
```

Imagen 12. Resultado de trimming

- Control de calidad post trimming

Basic sequence stats	
Measure	Value
Filename	Sample_1.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	706751
Sequences flagged as poor quality	0
Sequence length	20-151
%GC	36

Imagen 13. Estadísticas básicas de muestra SRR9134252_1 luego de trimming.

Basic sequence stats	
Measure	Value
Filename	Sample_2.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	706751
Sequences flagged as poor quality	0
Sequence length	20-151
%GC	36

Imagen 14. Estadísticas básicas de muestra SRR9134252_2 luego de trimming

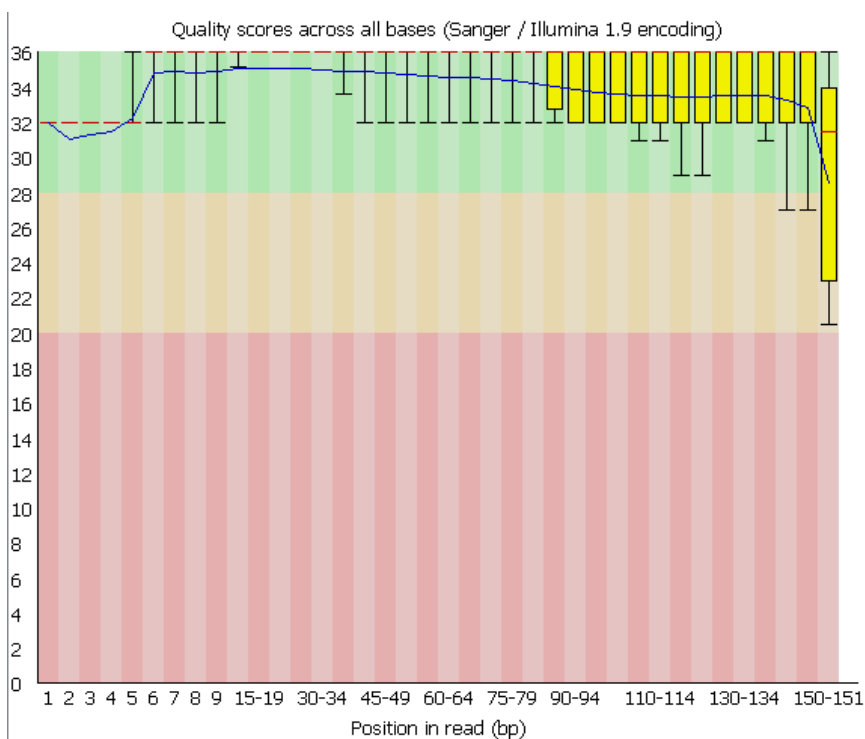


Imagen 15. Distribución de puntaje de calidad por base de secuencia SRR9134252_1

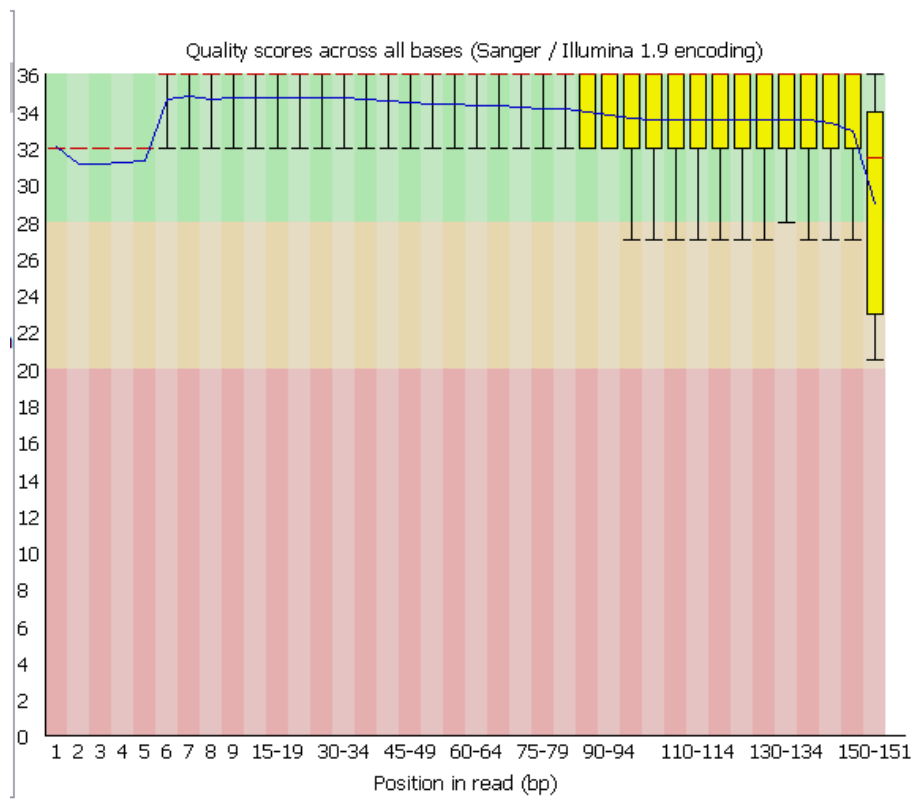


Imagen 16. Puntaje de calidad por base de secuencia SRR9134252_2

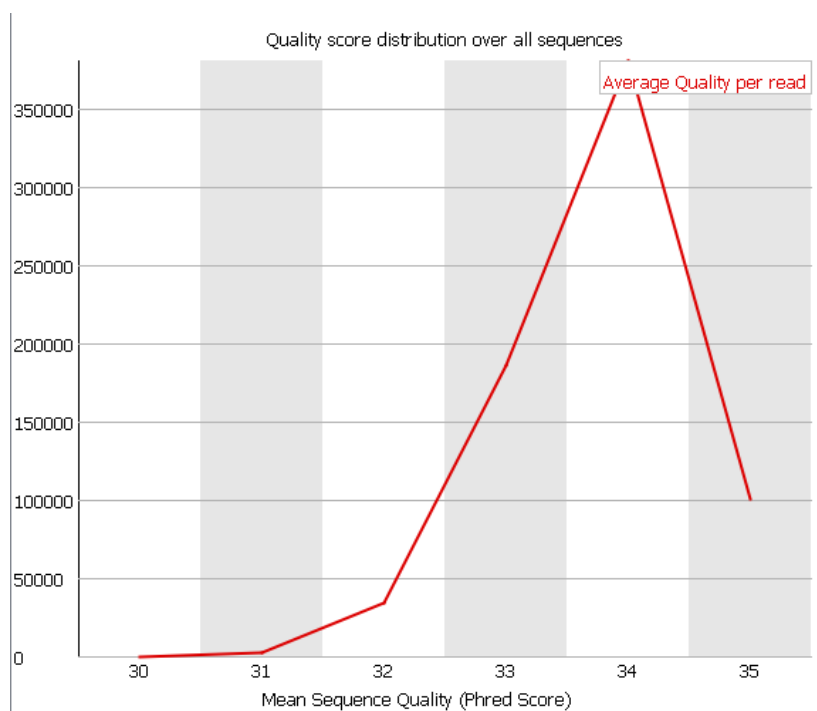


Imagen 17. Distribución del promedio de puntaje de calidad de secuencia SRR9134252_1

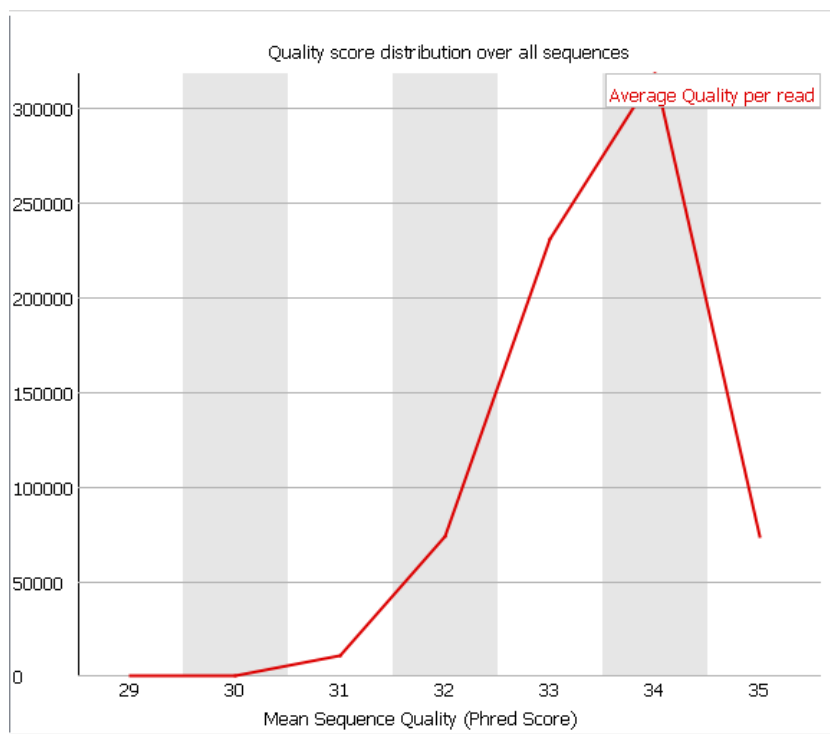


Imagen 18. Distribución del promedio de puntaje de calidad de secuencia SRR9134252_2

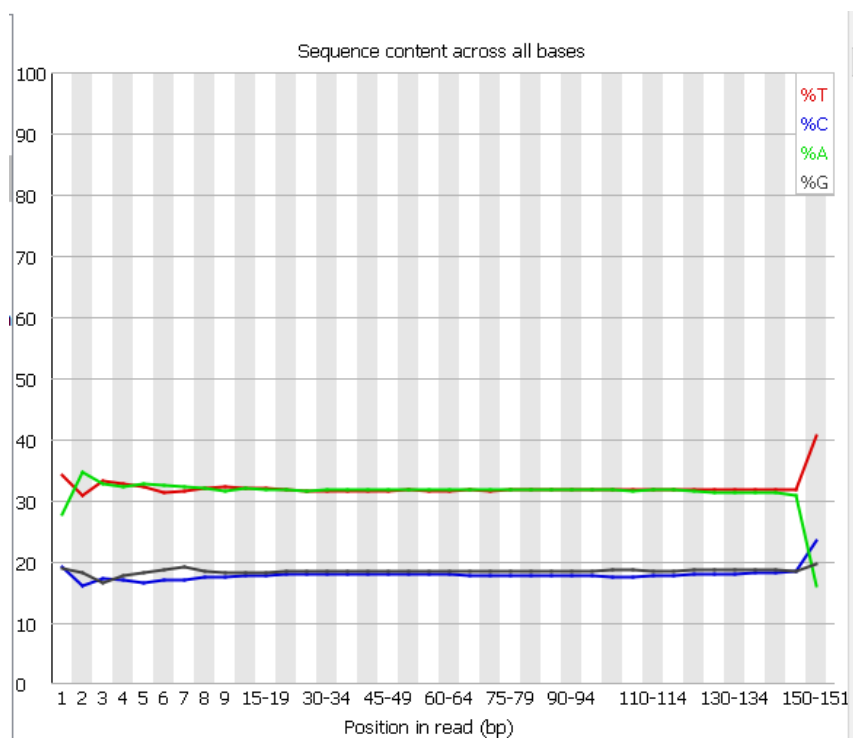


Imagen 19. Distribución de bases de secuencia SRR9134252_1

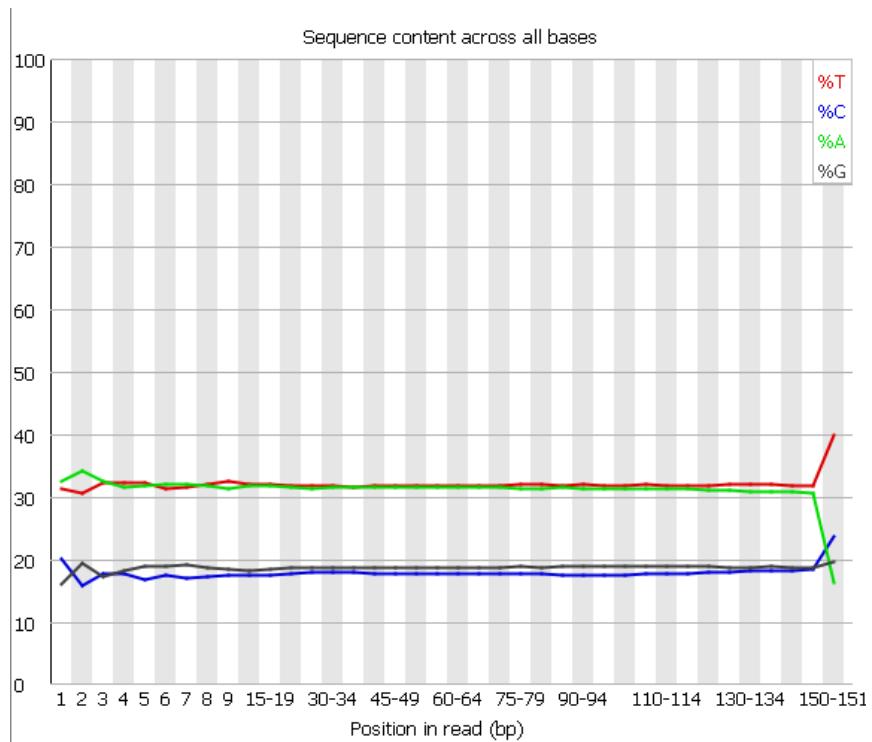


Imagen 20. Distribución de bases de secuencia SRR9134252_2

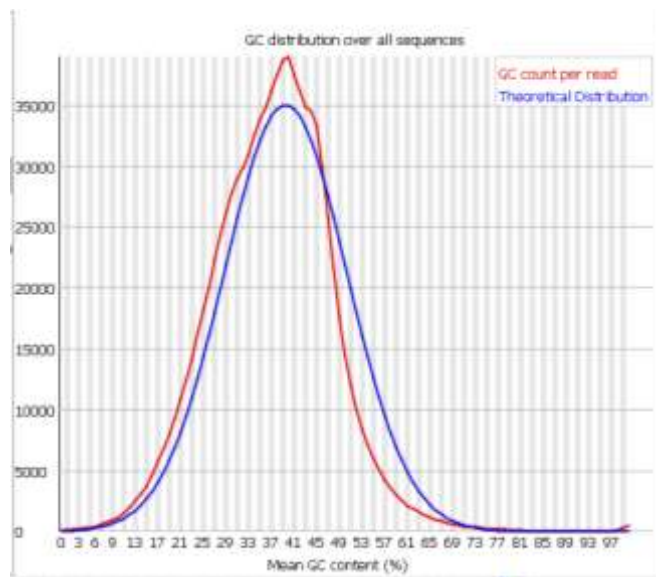


Imagen 21. Distribución de GC por secuencia SRR9134252_1

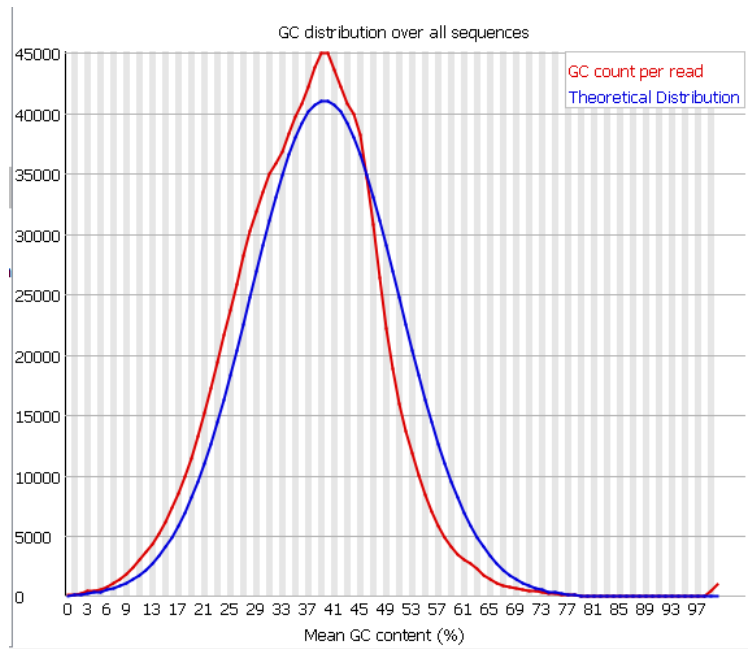


Imagen 22. Distribución de GC por secuencia SRR9134252_2

- Ensamblaje

- Número de contigs : 15,540,729

Node count:	46,107
Edge count:	0
Edge overlaps:	n/a
Total length:	15,540,729 bp
Total length (no overlaps):	15,540,729 bp

Imagen 23. Largo de contigs

- Número de scaffolds: 15,541,000

Graph size	
Node count:	46,033
Edge count:	0
Edge overlaps:	n/a
Total length:	15,541,600 bp
Total length (no overlaps):	15,541,600 bp

Imagen 24. Largo de Scaloff

- Estadística

- Coverage

```
NODE_1_length_86002_cov_29.150399
NODE_2_length_18364_cov_34.457188
NODE_3_length_16557_cov_45.762439
NODE_4_length_9135_cov_83.508059
NODE_5_length_9131_cov_29.051469
```

Imagen 25. Cobertura primeros 5.

- N50



N50:

309 bp

Imagen 26. Estadística de N50

- %GC
- | | |
|-----|----|
| %GC | 36 |
|-----|----|

Imagen 27. Estadística de %GC