**Universidad Nacional Autónoma de México**

**Universidad de Chile**

**Introducción a la Bioinformática**

**Unidad 6 Introducción a la genómica y secuenciación de siguiente generación**

**Tarea 2: Filtrado y alineamiento de secuencias**

**Daniel Enrique Acosta García**

Se realizó un alineamiento de la secuencia 20 disponible en la carpeta del curso en formato sam contra el genoma de referencia “hg19\_reference”. A partir de este se obtuvieron las lecturas alineadas por posición, se anotó el alineamiento y se generó un reporte de calidad con la herramienta “qualimap”.

**Resultados**

% de lecturas mapeadas: **99,96%**

% de lecturas únicamente mapeadas: **61,05%**

% de lecturas efectivamente mapeadas:

Profundidad promedio

Cobertura: **0,004x**

Calidad de mapeo (QMAP por lectura): **3.49**

Distribución de tamaños de inserto: **5,150.81**

% de lecturas en el blanco: **0,17%**

Enriquecimiento de región blanco:

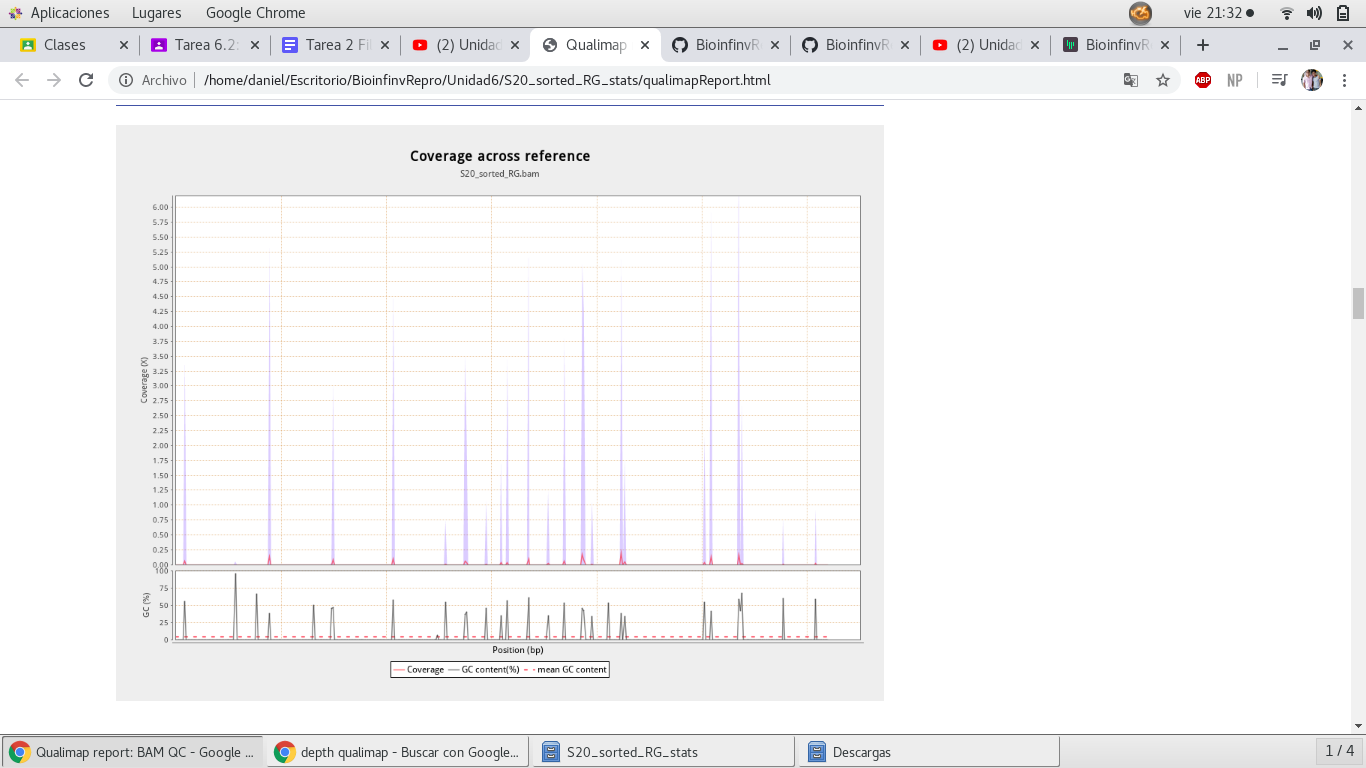


Fig. 1 Cobertura a lo largo de la referencia. Este gráfico muestra la cobertura de color rojo comparada con la nube de color azul que representa el porcentaje de GC a lo largo de la secuencia contra la que se mapeo (eje x)/.

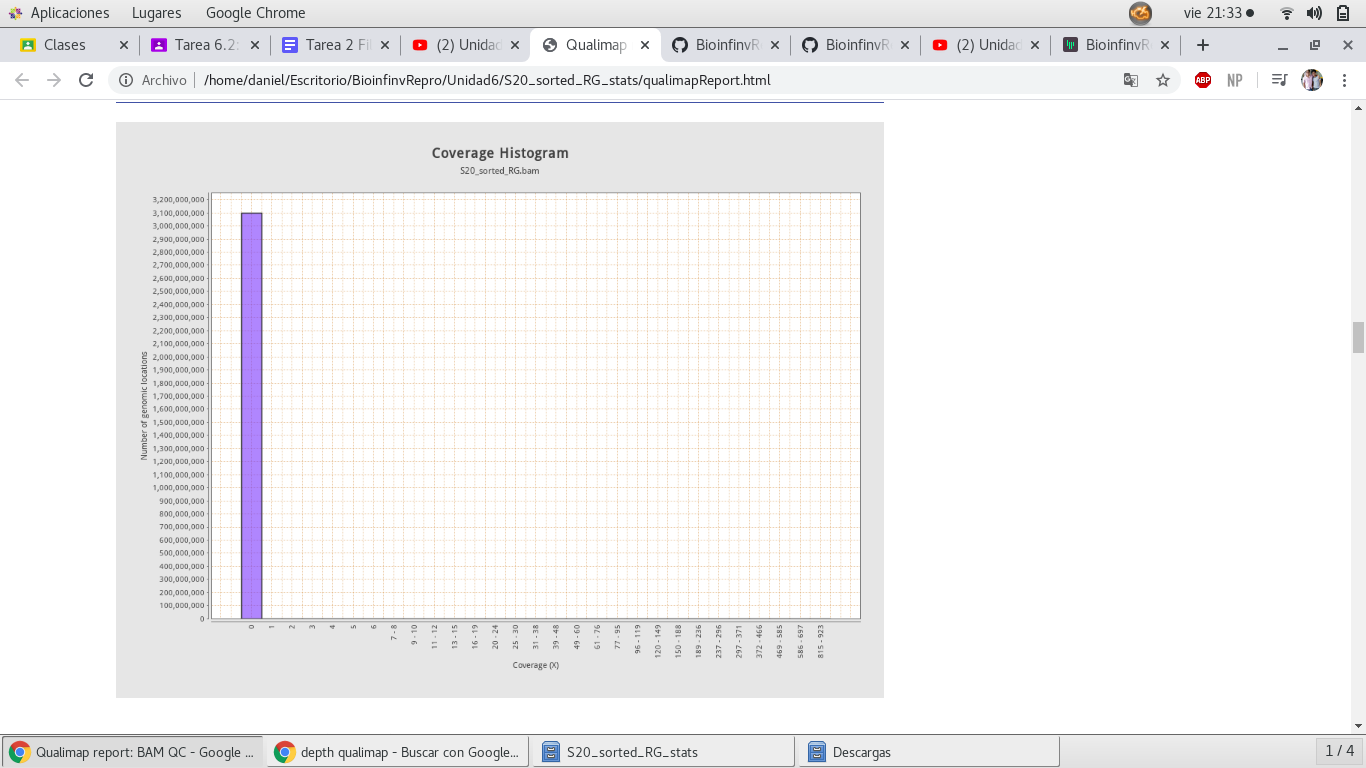


Fig. 2 Histograma de cobertura: Presenta los valores de cobertura en función de las locaciones del genoma contra el que se alineó. Al presentar un valor de 0,04% de cobertura significa que está representado solamente una porción muy pequeña de la referencia por lo que el histograma se presenta sólo en el principio del eje.

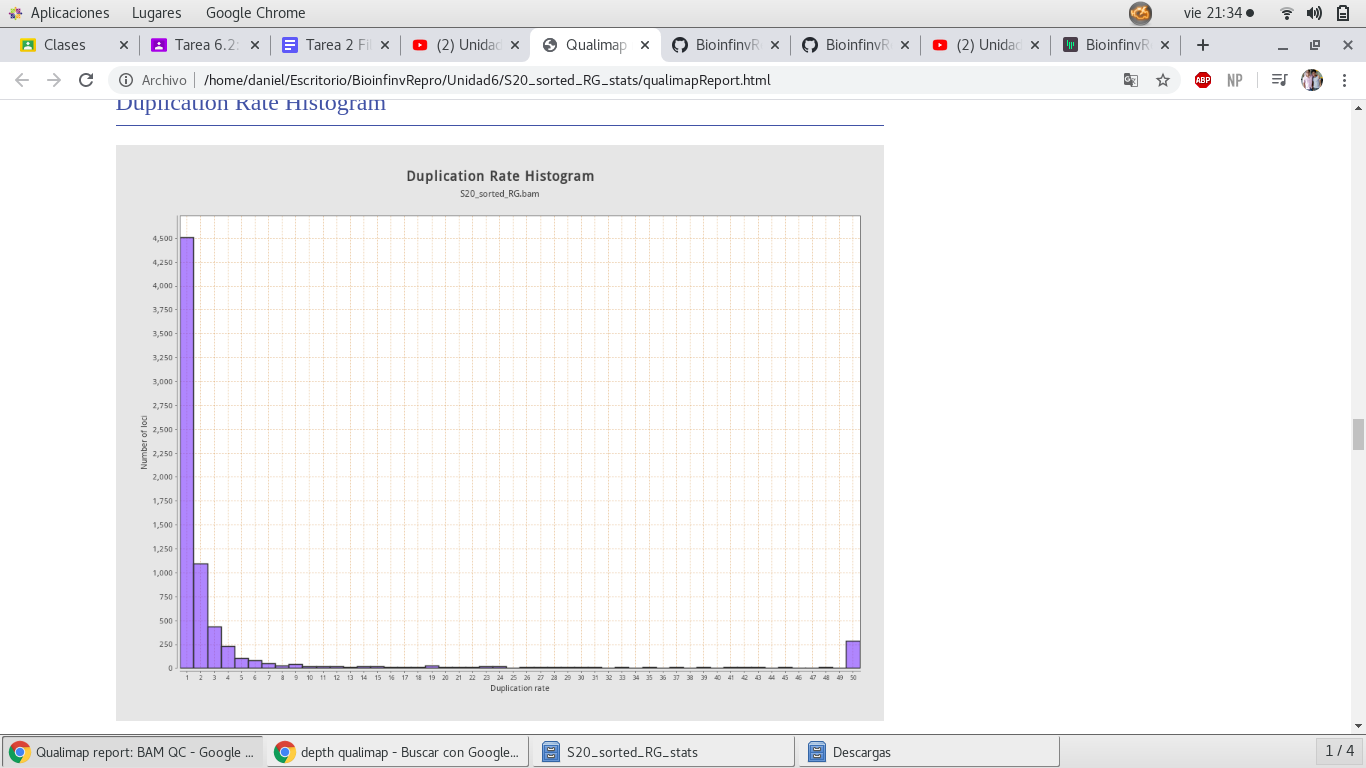
****

Fig. 3 Histograma de duplicación: Presenta el número de reads que presentan cierto tasa de divulgación. El alineamiento actual presenta el valor de loci con una tasa de duplicación de uno, con pocos loci con una tasa de dos a diez y otros pocos con una tasa alta de cincuenta, lo que indica que el control de calidad con el que se evaluaron nuestras secuencias se hizo correctamente antes.

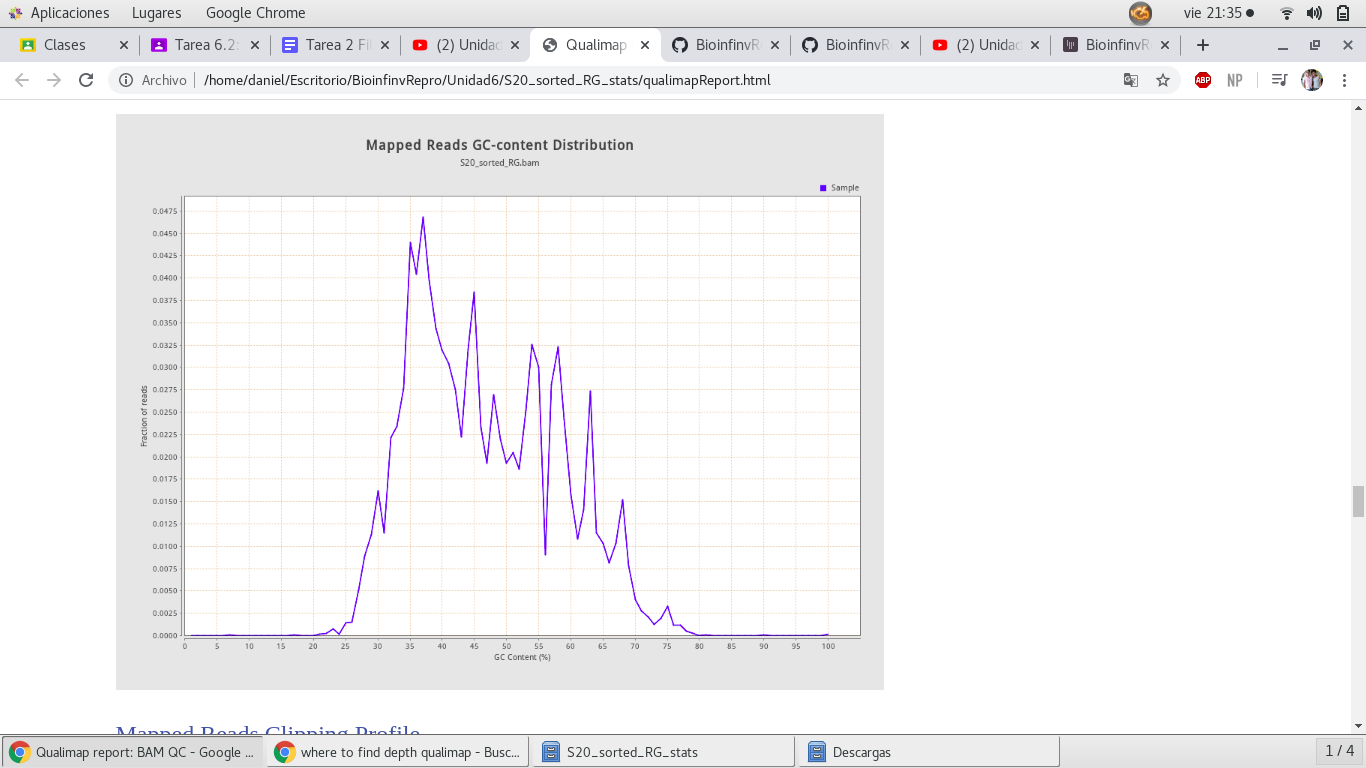
****

Fig. 4 Distribución de GC’s de los reads mapeados. Presenta una distribución general del porcentaje de GC’s de aquellos reads que se alinearon correctamente contra la secuencia de referencia. Habitualmente se espera una distribución normal para descartar errores de secuenciación.