**ACTIVIDAD 2**

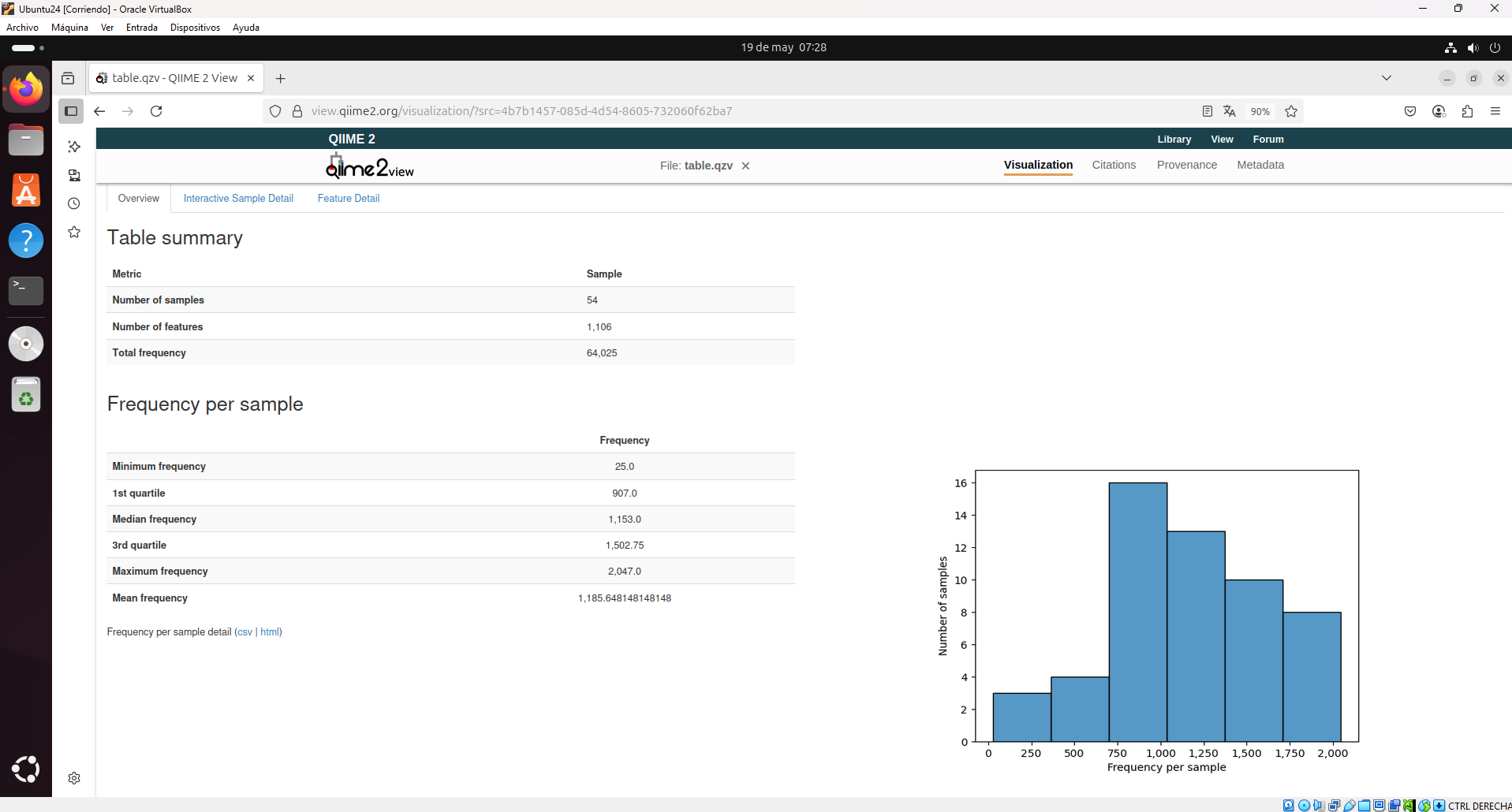
**ALUMNO/A: Andres Felipe Meneses Castro**

**Respuestas**

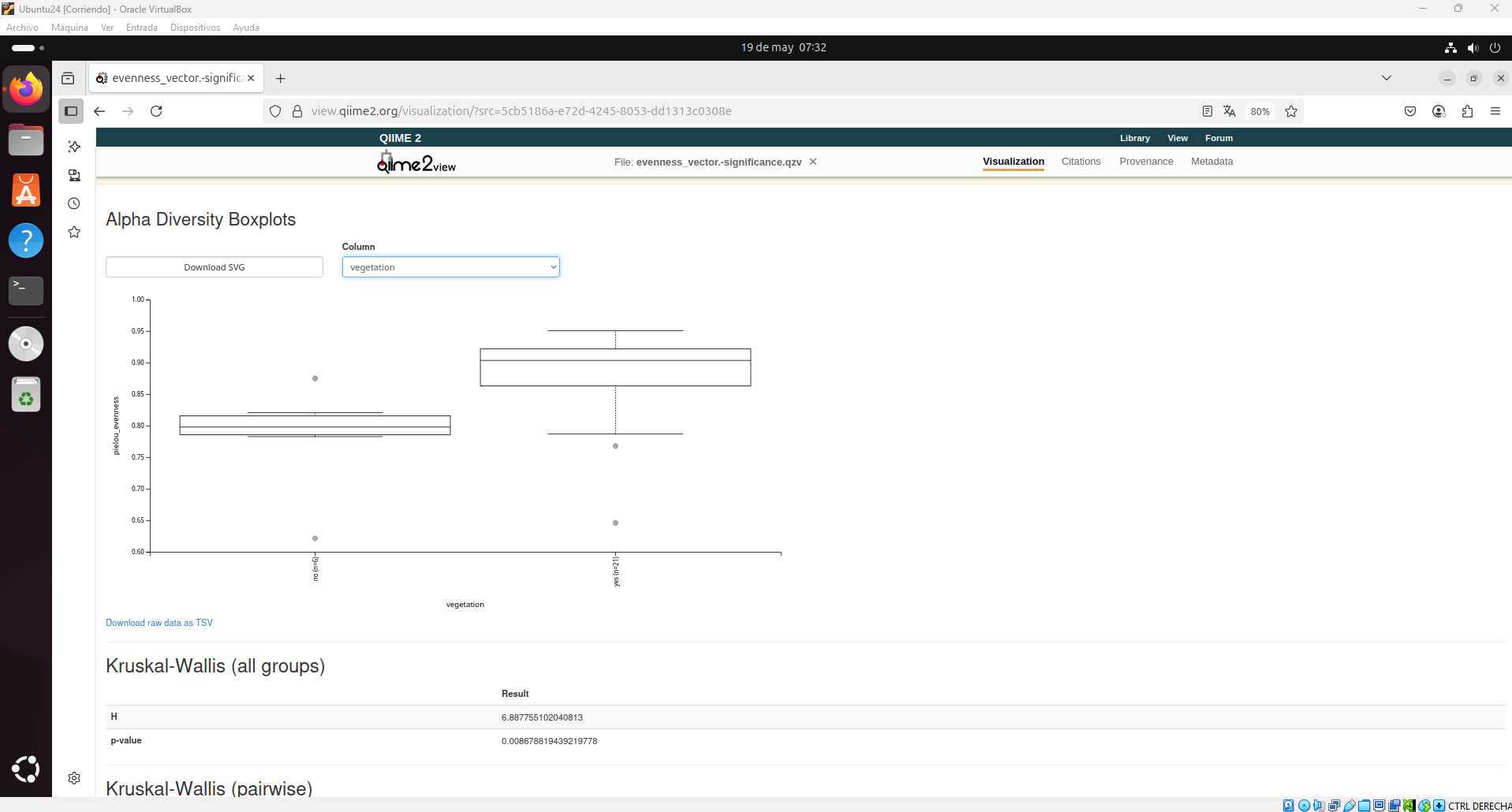
Para iniciar con la Actividad planteada se cambia la versión de Qiimen2 que se iba a usar debido a que por fallas en el sistema para procedimiento de la 2023.9 a la versión 2021.4 que si soporta el sistema usado y no presento fallas para concluir la actividad

1. ¿Qué profundidad de muestreo debemos seleccionar? Observa cuántas secuencias tienen cada muestra visualizando la tabla **table.qzv** y selecciona la profundidad de muestreo (pista: ¿Cuáles son los valores menores y mayores de secuencias en una muestra? ¿Cuál es la mediana? ¿Y el promedio? ¿Qué valor permitirá que mantengamos la mayor cantidad de muestras posibles?). En base a este número realizaremos el paso de rarefacción de muestras.

Se uso el valor de 1153 que la **mediana, la cual** representa el **valor central** de la distribución de frecuencias por muestra. Es el punto en el que el 50% de las muestras tienen más secuencias y el otro 50% tienen menos.



1. ¿Qué variable categórica de los metadatos está asociada con más fuerza con las **diferencias en riqueza** de la comunidad? ¿Y con **igualdad**?

Se realizaron pruebas de Kruskal-Wallis para evaluar la relación entre las variables categóricas de los metadatos y la diversidad alfa. En las métricas de **riqueza** no se encontraron diferencias significativas (p > 0.05). Sin embargo, para la métrica de **igualdad** (*evenness*), una variable mostró una diferencia significativa entre grupos (p = 0.0087). Esto indica que **esta variable está asociada con diferencias en la igualdad de la comunidad microbiana**, siendo la más relevante en este aspecto. 

1. ¿Qué pasa si evaluamos algunas de las secuencias con BLAST? ¿Son las clasificaciones taxonómicas diferentes a las de QIIME2? ¿A qué nivel taxonómico surgen las diferencias?

Cuando evaluamos algunas de las secuencias obtenidas con QIIME2 usando **BLAST**, podemos observar lo siguiente:

**Las clasificaciones taxonómicas pueden diferir ligeramente**, especialmente en niveles más específicos (género o especie). Esto ocurre porque:

* + QIIME2 usa bases de datos y clasificadores entrenados (como SILVA o Greengenes) con algoritmos como el clasificador Naive Bayes.
  + BLAST realiza una búsqueda por similitud de secuencia directamente contra una base de datos y da como resultado la mejor coincidencia, que no siempre se ajusta al modelo de QIIME2.

**Las diferencias suelen surgir a nivel de género o especie**, ya que:

* + QIIME2 puede ser más conservador y quedarse en niveles taxonómicos superiores si no hay suficiente confianza.
  + BLAST puede sugerir una identificación más específica, pero no siempre con alta certeza.

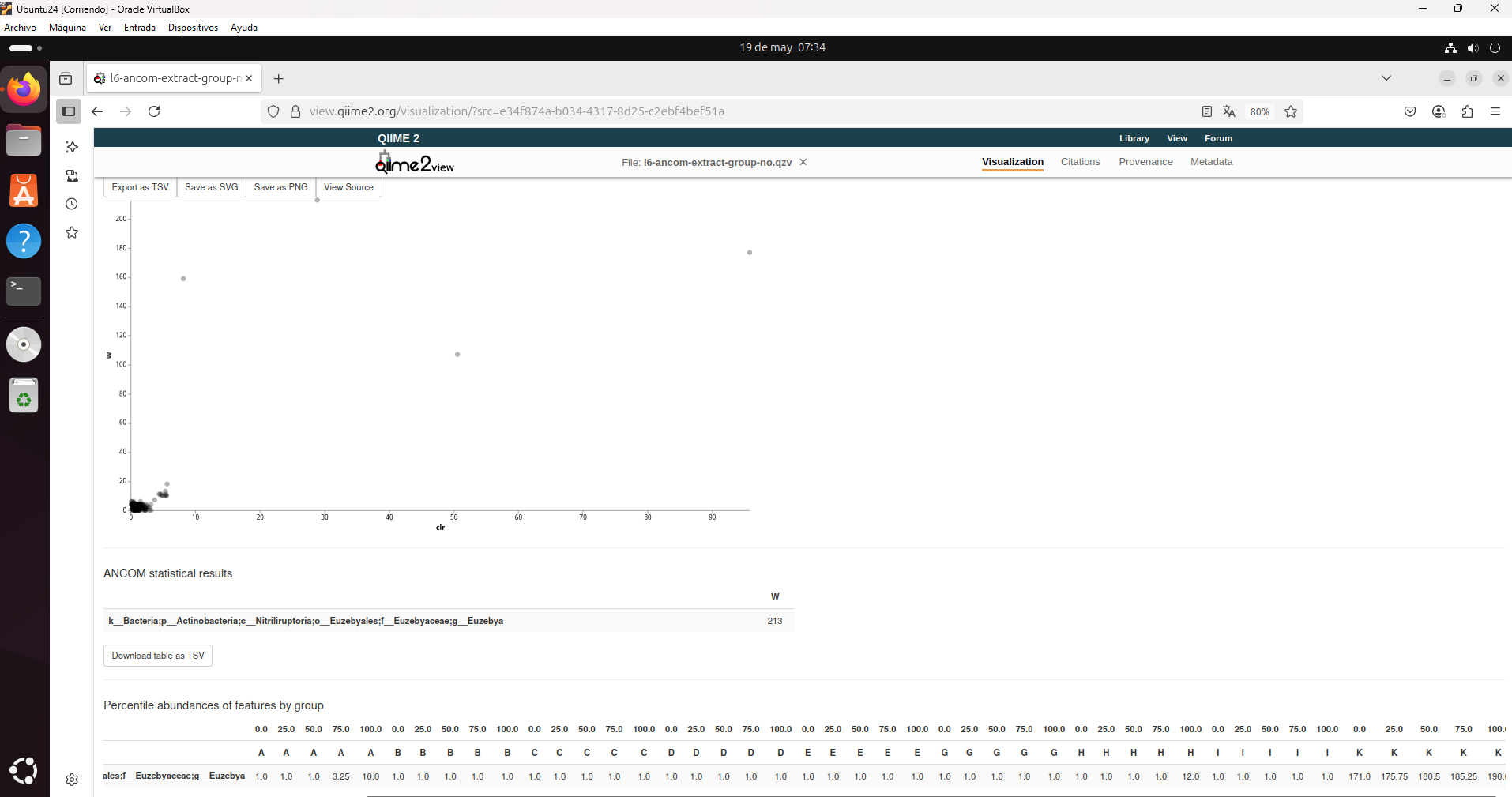
1. ¿Qué géneros presentan diferencias signifiativas en abundancia entre grupos de muestras?

### Interpretación:

* En todos los grupos excepto el grupo **K**, la abundancia relativa del género Euzebya es **baja y constante**, con mediana de **1.0**.
* En el grupo **K**, la mediana es **180.5**, lo cual indica una **diferencia clara y marcada** con respecto al resto de los grupos.

### Conclusión:

El género **Euzebya** presenta una **abundancia significativamente mayor** en el grupo **K** en comparación con los demás grupos. Esta diferencia sugiere que Euzebya podría estar asociada a una condición o variable específica del grupo **K**, y probablemente fue identificado como **significativo** en el análisis ANCOM.



1. **Interpretación y Conclusiones Generales**

La actividad permitió comprender el análisis de comunidades microbianas con QIIME 2, cubriendo desde la calidad de los datos hasta la interpretación biológica.

* **Importación y demultiplexado:** Organiza secuencias por muestra, base para todo el análisis.
* **Rarefacción:** Estándariza el número de secuencias para comparaciones justas; usar la mediana evita sesgos.
* **Análisis de diversidad:** Métricas como Shannon y Unifrac exploran riqueza y composición; pruebas estadísticas identifican diferencias significativas.
* **Clasificación taxonómica y BLAST:** Valida resultados y detecta discrepancias.
* **Análisis diferencial (ANCOM):** Identifica taxones con cambios relevantes entre grupos.
* **Visualización:** Archivos .qzv facilitan interpretación y reproducibilidad.

**Conclusión:**  
La actividad enseña tanto el uso técnico de QIIME 2 como el pensamiento crítico para interpretar datos microbiológicos, fomentando buenas prácticas en bioinformática y aportando a estudios en salud y ecología.