

Datos del estudiante

Nombre y apellidos	Ana Sofia Santos Tedim Sousa Pedrosa
Fecha de entrega	13/01/2026

Análisis de datos metagenómicos con QIIME2

Pregunta 1: ¿Qué profundidad de muestreo debemos seleccionar? Observa cuántas secuencias tiene 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.

Para determinar una profundidad de muestreo adecuada para el proceso de rarefacción, se procedió a la visualización de los principales ficheros generados durante el procesamiento de los datos (paso 7), en particular `table.qzv` (Tablas S1, S2 y S3 en el anexo), `denoising-stats.qzv` (Tabla S4) y `rep-seqs.qzv` (Texto 1).

El conjunto de datos analizado está compuesto por 54 muestras y 1088 características (features, Tabla S1). La profundidad de secuenciación por muestra mostró cierta variabilidad, con un mínimo de 0 lecturas, un máximo de 2090, una media de 1126,4 y una mediana de 1077 secuencias por muestra (Tabla S2). Esta variabilidad se considera moderada y adecuada para la estimación de la diversidad microbiana, después de aplicar un proceso de normalización mediante rarefacción.

Las estadísticas del proceso de denoising, obtenidas mediante DADA2 (Tabla S4), refleja un proceso de denoising eficiente y robusto, adecuado para los análisis posteriores de diversidad.

Con base en la información de la tabla S2 y siguiendo los principios de rarefacción descritos en el Tema 7, se seleccionó una profundidad de muestreo de 1077 secuencias, correspondiente a la mediana del número de secuencias por muestra. Este valor permite normalizar la profundidad de secuenciación entre muestras, maximizando el número de muestras retenidas y asegurando una representación adecuada de la diversidad microbiana para los análisis de diversidad alfa y beta posteriores.

Pregunta 2: ¿Qué variable categórica de los metadatos está asociada con más fuerza con las diferencias en riqueza de la comunidad? ¿Y con igualdad?

A partir de las secuencias representativas obtenidas tras el proceso de denoising (`rep-seqs.qza`), se construyó un árbol filogenético (Figura S1), que describe las relaciones evolutivas entre las variantes exactas de secuencia presentes en el conjunto de datos. Este árbol se generó como indicado en el enunciado de la actividad y se visualizó mediante la herramienta Interactive Tree Of Life (iTOL). Este árbol filogenético se empleó en el cálculo de métricas de diversidad que incorporan información evolutiva (diversidad filogenética de Faith y UniFrac).

Para evaluar la riqueza de la comunidad microbiana se emplearon las métricas de número de características observadas y diversidad filogenética de Faith, las cuales cuantifican la riqueza taxonómica y filogenética presente en cada muestra, respectivamente (Tabla 1).

La variable categórica *Vegetation* (presencia o ausencia de vegetación) es la que se asocia con mayor fuerza a las diferencias en la riqueza de la comunidad microbiana, ya que mostró diferencias estadísticamente significativas tanto en la diversidad filogenética de Faith como en el número de características observadas. Este resultado sugiere que la presencia de vegetación influye de manera directa en el número y diversidad evolutiva de los taxones microbianos presentes en las muestras.

Por otro lado, la variable *Site-name* (local donde se recogió la muestra) presentó diferencias estadísticamente significativas únicamente en la métrica de diversidad filogenética de Faith. Sin embargo, al no mostrar diferencias significativas en el número de características observadas, su asociación con la riqueza total de la comunidad resulta menos consistente.

Para evaluar la igualdad de la comunidad microbiana entre las diferentes muestras se empleó del índice Evenness, que cuantifica la uniformidad con la que se distribuyen las abundancias relativas de taxones en cada muestra (Tabla 1).

La variable categórica *Transect-name* es la que presenta la asociación más fuerte con la igualdad de la comunidad microbiana, indicando que la distribución relativa de abundancias entre los taxones varía principalmente a lo largo del gradiente espacial representado por los transectos. Este resultado sugiere que, más allá del número de taxones presentes, la estructura interna de la comunidad microbiana cambia en

función de la localización espacial. La variable *Vegetation* también mostró una asociación estadísticamente significativa con la igualdad de las comunidades bacterianas, si bien el nivel de significación fue menor que el observado para *Transect-name*, lo que indica una influencia secundaria de esta variable sobre la distribución de abundancias.

Tabla 1. Resultados de los índices de diversidad alpha

Variable	Diversidad Filogenética de Faith		Número de Características Observadas		Evenness	
	H	p-value	H	p-value	H	p-value
Vegetation	10.656	0.001*	9.976	0.002*	6.762	0.009*
Transect	0.021	0.8836	2.601	0.107	7.467	0.006*
Extract group	11.605	0.0714	10.246	0.115	9.523	0.146
Site name	24.180	0.030*	21.100	0.071	18.893	0.126

* valores estadísticamente significativos

Pregunta 3: ¿Qué pasa si evaluamos algunas de las secuencias con BLAST? ¿Son las clasificaciones taxonómicas diferentes a las de QIIME2? ¿En qué nivel taxonómico surgen las diferencias?

Debido a problemas de incompatibilidad entre versiones de QIIME2 y los clasificadores taxonómicos disponibles, se entrenó un clasificador basado en la base de datos SILVA compatible con la versión qiime2-amplicon 2024.10 utilizada en este estudio, siguiendo una solución propuesta en el foro de la actividad por el compañero Samuel Salazar Diaz. El código empleado se incluye en el anexo (Texto S2).

Para evaluar la robustez de la asignación taxonómica realizada mediante QIIME2, se analizaron tres secuencias representativas de forma individual utilizando BLAST, comparando los resultados obtenidos mediante BLASTn y BLASTx (Texto S3).

Como primer ejemplo, la secuencia 1062a45d18c0f6573d213a79070031b5 fue clasificada por QIIME2 como perteneciente al filo Bacillota (anteriormente Firmicutes), clase Bacilli, orden Bacillales, familia Bacillaceae y género *Bacillus*, con un alto valor de confianza (0,93). El análisis mediante BLASTn mostró resultados concordantes en los niveles taxonómicos superiores, confirmando su pertenencia a Bacillaceae. Sin embargo, se observaron múltiples coincidencias con secuencias de uncultured *Bacillus* sp. y otras bacterias cercanas, lo que impidió una asignación

Secuenciación y Ómicas de Próxima Generación

inequívoca a nivel de género o especie, a diferencia de QIIME2, que permitió alcanzar el nivel de género para esta secuencia (Figura S2).

En el caso de la secuencia 03e53d40c9bffca932b2fe3f68c38f6b, QIIME2 asignó la taxonomía hasta una especie no cultivada dentro de *Actinobacteriota* (orden y familia 0319-7L14), con una confianza elevada (0,90). El análisis mediante BLASTn mostró principalmente coincidencias con bacterias no cultivadas o muestras ambientales de Actinobacteria, sin una resolución clara a nivel de género o especie, en concordancia con QIIME2 (Figura S3).

Por su parte, la secuencia 071b01e2aa5340be4e48c99169d93e05 fue clasificada por QIIME2 como perteneciente a *Actinobacteriota*, clase *Actinobacteria*, orden *Streptomycetales*, familia *Streptomycetaceae* y género *Streptomyces*, con un valor de confianza muy alto (0,99). En este caso, BLASTn confirmó de forma consistente la afiliación al género *Streptomyces* (Figura S4), aunque mostró múltiples coincidencias con múltiples especies del género.

El análisis mediante BLASTx (Figuras S5 - S7) produjo asignaciones taxonómicas menos coherentes y biológicamente menos informativas. Dado que BLASTx traduce secuencias nucleotídicas cortas ($\pm 240\text{nt}$) a secuencias proteicas aún más cortas, este enfoque no resulta adecuado para este tipo de análisis, limitando la identificación taxonómica a niveles muy generales.

En conjunto, la comparación entre BLAST y QIIME2 indica que ambas aproximaciones son altamente concordantes en los niveles taxonómicos superiores (filo, clase y orden), mientras que las discrepancias aparecen principalmente en los niveles de género y especie. BLASTn resulta una herramienta complementaria útil para la exploración puntual y la confirmación de resultados de QIIME2, siendo BLASTx inadecuado para este tipo de análisis.

Pregunta 4: ¿Qué géneros presentan diferencias significativas en abundancia entre grupos de muestras?

El análisis de composición diferencial mediante ANCOM permitió identificar los géneros que presentan diferencias estadísticamente significativas en abundancia relativa entre los grupos de muestras definidos por las variables categóricas analizadas.

Secuenciación y Ómicas de Próxima Generación

En el caso de la variable *Site-name* (Figura S8, Tabla S5), se detectaron diferencias significativas en la abundancia de varios taxones a nivel de género: *Candidatus_Udaeobacter*, *Ammoniphilus*, *Marmoricola*, *Longimicrobiaceae* y *Nitriliruptoraceae*, además de dos grupos clasificados como bacterias no cultivadas. Para la variable *Vegetation* (Figura S9, Tabla S6), ANCOM identificó diferencias significativas en la abundancia relativa de varios taxones a nivel de género: *Nitrosphaeraceae*, *Candidatus_Udaeobacter*, *Ralstonia*, *Nitriliruptoraceae*, bien como un género bacteriano no cultivable.

Por el contrario, para la variable *Extract-group-no* (Figura S10, Tabla S7), el análisis ANCOM solo detectó diferencias significativas en abundancia únicamente en dos géneros de bacterias no cultivables. Más aún no se identificaron géneros con diferencias significativas asociadas a la variable *Transect-name* (Figura S11).

Estos resultados indican que la abundancia de géneros bacterianos, es decir la composición microbiana, de las muestras claramente cambia entre los dos sitios de muestreo y con la presencia o ausencia de vegetación.

Conclusiones:

La selección de una profundidad de muestreo adecuada permitió normalizar la profundidad de secuenciación entre muestras garantizando la comparabilidad de los análisis posteriores. El análisis de diversidad alfa mostró que la presencia de vegetación es el factor que ejerce una mayor influencia sobre la riqueza microbiana, mientras que la igualdad de la comunidad se ve más fuertemente asociada a la variable *Transect-name*, lo que indica que diferentes procesos ecológicos pueden afectar de forma diferente a la estructura de las comunidades.

La comparación de la asignación taxonómica obtenida mediante QIIME2 con resultados de BLAST confirmó la robustez del enfoque utilizado, mostrando una alta concordancia en los niveles taxonómicos superiores y discrepancias principalmente en niveles más finos. En este contexto, QIIME2 demostró ser una herramienta conservadora y reproducible para el análisis global de comunidades, mientras que BLASTn resultó útil como método complementario para la validación puntual de secuencias.

Secuenciación y Ómicas de Próxima Generación

Finalmente, el análisis de abundancia diferencial mediante ANCOM permitió identificar géneros bacterianos cuya abundancia varía significativamente entre grupos de muestras, principalmente en función del local de muestreo y de la presencia o ausencia de vegetación. Estos resultados evidencian que tanto factores ambientales como espaciales desempeñan un papel clave en la estructuración de las comunidades microbianas analizadas.

En conjunto, esta actividad pone de manifiesto la utilidad de los enfoques basados en variantes exactas de secuencia, métricas filogenéticas y análisis de composición diferencial para el estudio de la diversidad microbiana, así como la importancia de una interpretación integrada de los resultados estadísticos y biológicos.

ANEXO

1. TABLAS SUPLEMENTARIAS

Tabla S1 – Tabla resumen de los datos usado en esta actividad.

Resumen estadístico	Valor
Número de muestras	54
Número de características	1088
Frecuencia total	60,826

Tabla obtenida de la visualización de table.qzv

Tabla S2 – Tabla resumen de los datos usados en esta actividad por muestra analizada.

Resumen Estadístico	Frecuencia
Frecuencia mínima	0
Primer cuartil (Q1)	831,5
Mediana	1077
Tercer cuartil (Q3)	1379,5
Frecuencia máxima	2090
Frecuencia media	1126,4

Tabla obtenida de la visualización de table.qzv

Tabla S3 - Tabla resumen de los datos usados en esta actividad por característica analizada.

Resumen Estadístico	Frecuencia
Frecuencia mínima	2
Primer cuartil (Q1)	11
Mediana	22
Tercer cuartil (Q3)	48,2
Frecuencia máxima	1436
Frecuencia media	55,9

Tabla obtenida de la visualización de table.qzv

Secuenciación y Ómicas de Próxima Generación

Tabla S4 – Table resumen de las estadísticas del proceso de denoising

Identificador de la muestra	No secuencias entrada	No secuencias filtradas	Porcentaje secuencia que pasan filtrado	Secuencias corregidas	Secuencias fusionadas	Porcentaje de secuencias fusionadas	No Secuencias no químéricas	Porcentaje de lecturas no químéricas
#q2:types	numeric	numeric	numeric	numeric	numeric	numeric	numeric	numeric
BAQ2420.1.1	2314	2041	88.2	1674	831	35.91	831	35.91
BAQ2420.1.2	2154	1924	89.32	1675	758	35.19	758	35.19
BAQ2420.1.3	2044	1865	91.24	1614	755	36.94	755	36.94
BAQ2420.2	1986	1749	88.07	1613	833	41.94	833	41.94
BAQ2420.3	1993	1685	84.55	1547	777	38.99	777	38.99
BAQ2462.1	2839	2398	84.47	2236	1382	48.68	1382	48.68
BAQ2462.2	1892	1592	84.14	1493	982	51.9	982	51.9
BAQ2462.3	1581	1402	88.68	1274	639	40.42	639	40.42
BAQ2687.1	2734	2491	91.11	2147	842	30.8	806	29.48
BAQ2687.2	2067	1689	81.71	1551	1194	57.76	1194	57.76
BAQ2687.3	2820	2347	83.23	2053	1001	35.5	1001	35.5
BAQ2838.1	1881	1621	86.18	1370	556	29.56	556	29.56
BAQ2838.2	1734	1474	85.01	1246	476	27.45	476	27.45
BAQ2838.3	1152	961	83.42	807	369	32.03	369	32.03
BAQ3473.1	3532	3072	86.98	2639	1202	34.03	1202	34.03
BAQ3473.2	3122	2802	89.75	2416	1092	34.98	1092	34.98
BAQ3473.3	3956	3610	91.25	3342	1813	45.83	1813	45.83
BAQ4166.1.1	3255	2826	86.82	2413	1096	33.67	1055	32.41
BAQ4166.1.2	4185	3580	85.54	3096	1454	34.74	1356	32.4
BAQ4166.1.3	3139	2770	88.24	2474	1351	43.04	1351	43.04
BAQ4166.2	3920	3445	87.88	2951	1346	34.34	1346	34.34
BAQ4166.3	3953	3406	86.16	2989	1342	33.95	1342	33.95
BAQ4697.1	2676	2432	90.88	2267	1721	64.31	1721	64.31
BAQ4697.2	2513	2275	90.53	2071	1540	61.28	1540	61.28
BAQ4697.3	3454	3060	88.59	2797	2090	60.51	2090	60.51
YUN1005.1.1	2711	2392	88.23	2313	1945	71.74	1945	71.74
YUN1005.3	1290	1175	91.09	1149	1080	83.72	1080	83.72
YUN1242.1	1761	1570	89.15	1554	1422	80.75	1372	77.91
YUN1242.3	2334	2037	87.28	1971	1616	69.24	1559	66.8
YUN1609.1	2332	1941	83.23	1890	1393	59.73	1393	59.73
YUN2029.2	2729	2456	90	2365	1901	69.66	1884	69.04
YUN3153.2	1428	1233	86.34	1136	772	54.06	772	54.06
YUN3153.3	1815	1577	86.89	1471	1074	59.17	1074	59.17
YUN3259.1.1	180	162	90	46	0	0	0	0
YUN3259.1.2	2091	1887	90.24	1671	964	46.1	964	46.1
YUN3259.1.3	890	775	87.08	688	334	37.53	334	37.53
YUN3259.2	3090	2601	84.17	2213	908	29.39	908	29.39
YUN3259.3	3415	2984	87.38	2630	1207	35.34	1187	34.76
YUN3346.1	2133	1900	89.08	1789	1090	51.1	1070	50.16
YUN3346.2	617	544	88.17	446	182	29.5	182	29.5
YUN3346.3	2434	2067	84.92	1852	1085	44.58	1085	44.58
YUN3428.1	3181	2772	87.14	2447	981	30.84	979	30.78
YUN3428.2	4606	4052	87.97	3648	1890	41.03	1880	40.82
YUN3428.3	3460	3044	87.98	2713	1039	30.03	1039	30.03
YUN3533.1.1	2877	2588	89.95	2222	1240	43.1	1240	43.1
YUN3533.1.2	2853	2581	90.47	2197	664	23.27	664	23.27
YUN3533.1.3	2493	2258	90.57	1925	900	36.1	900	36.1
YUN3533.2	3846	3494	90.85	3157	1890	49.14	1890	49.14
YUN3533.3	4017	3672	91.41	3173	1629	40.55	1629	40.55
YUN3856.1.1	3305	2858	86.48	2502	1020	30.86	1020	30.86
YUN3856.1.2	3023	2748	90.9	2460	1666	55.11	1666	55.11
YUN3856.1.3	2301	2066	89.79	1888	1250	54.32	1250	54.32
YUN3856.2	3107	2739	88.16	2409	1058	34.05	1058	34.05
YUN3856.3	3858	3441	89.19	3042	1535	39.79	1535	39.79

Tabla obtenida de la visualización de denoising-stats.qzv

Tabla S5 – Resultados del análisis de composición diferencial a nivel de género según la variable *Site-name* (ANCOM)

Taxonomía	W	Rechazar la hipótesis nula
d_Bacteria;p_Actinobacteriota;c_Acidimicrobia;o_uncultured;f_uncultured;g_uncultured	238	True
d_Bacteria;p_Actinobacteriota;c_Actinobacteria;o_Euzebyales;f_Euzebyaceae;g_uncultured	237	True
d_Bacteria;p_Verrucomicrobiota;c_Verrucomicrobiae;o_Chthoniobacterales;f_Chthoniobacteraceae;g_Candidatus_Udaeobacter	233	True
d_Bacteria;p_Firmicutes;c_Bacilli;o_Paenibacillales;f_Paenibacillaceae;g_Ammoniphilus	230	True
d_Bacteria;p_Actinobacteriota;c_Actinobacteria;o_Propionibacteriales;f_Nocardioidaceae;g_Marmoricola	225	True
d_Bacteria;p_Gemmatimonadota;c_Longimicrobia;o_Longimicrobiales;f_Longimicrobiaceae;g_Longimicrobiaceae	222	True
d_Bacteria;p_Actinobacteriota;c_Actinobacteria;o_Nitriliruptorales;f_Nitriliruptoraceae;g_Nitriliruptoraceae	220	True

*Solo se muestran los resultados con significación estadística.

Tabla S6 – Resultados del análisis de composición diferencial a nivel de género según la variable *Vegetation* (ANCOM)

Taxonomía	W	Rechazar la hipótesis nula
d_Archaea;p_Crenarchaeota;c_Nitrososphaeria;o_Nitrososphaerales;f_Nitrosphaeraceae;g_Nitrososphaeraceae	231	True
d_Bacteria;p_Verrucomicrobiota;c_Verrucomicrobiae;o_Chthoniobacterales;f_Chthoniobacteraceae;g_Candidatus_Udaeobacter	230	True
d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Burkholderiales;f_Burkholderiaceae;g_Ralstonia	220	True
d_Bacteria;p_Actinobacteriota;c_Actinobacteria;o_Nitriliruptorales;f_Nitriliruptoraceae;g_Nitriliruptoraceae	219	True
d_Bacteria;p_Gemmatimonadota;c_Gemmatimonadetes;o_Gemmatimonadales;f_Gemmatimonadaceae;g_uncultured	208	True

*Solo se muestran los resultados con significación estadística.

Tabla S7 – Resultados del análisis de composición diferencial a nivel de género según la variable *Extract-group-no* (ANCOM)

Taxonomía	W	Rechazar la hipótesis nula
d_Bacteria;p_Actinobacteriota;c_Actinobacteria;o_Euzebyales;f_Euzebyaceae;g_uncultured	232	True
d_Bacteria;p_Actinobacteriota;c_Acidimicrobia;o_uncultured;f_uncultured;g_uncultured	230	True

*Solo se muestran los resultados con significación estadística.

2. FIGURAS SUPLEMENTARIAS

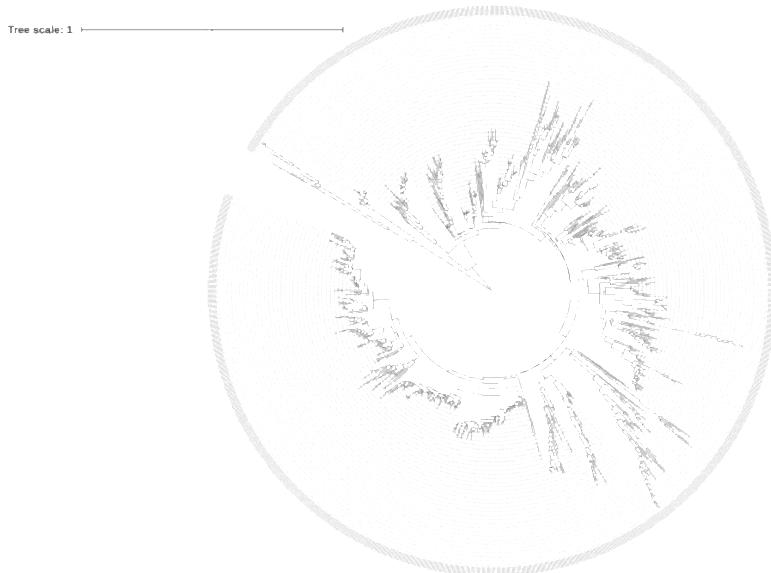


Figura S1 - Árbol filogenético construido a partir de las variantes exactas de secuencia. El árbol filogenético se visualizó con Interactive Tree Of Life (iTOL).

Taxonomy			
Taxonomy	Number of hits	Number of Organisms	Description
100 sequences selected			
root	100	8	
Bacteria	91	6	
environmental samples	66	2	
uncultured bacterium	63	1	uncultured bacterium hits
uncultured soil bacterium	3	1	uncultured soil bacterium hits
Bacillota	25	4	
Bacillaceae	21	3	
uncultured Bacillus sp.	18	1	uncultured Bacillus sp. hits
Nallia circulans	1	1	Nallia circulans hits
uncultured Bacillaceae bacterium	2	1	uncultured Bacillaceae bacterium hits
uncultured Bacillota bacterium	4	1	uncultured Bacillota bacterium hits
environmental samples	15	2	
uncultured organism	11	1	uncultured organism hits
uncultured prokaryote	4	1	uncultured prokaryote hits

Figura S2 - Resultados del análisis de la secuencia 1062a45d18c0f6573d213a79070031b5 mediante BLASTn (captura de pantalla de la tabla Taxonómica).

Taxonomy			
Taxonomy	Number of hits	Number of Organisms	Description
100 sequences selected			
root	100	7	
Bacteria	95	6	
environmental samples	73	2	
uncultured bacterium	72	1	uncultured bacterium hits
uncultured soil bacterium	1	1	uncultured soil bacterium hits
Actinomycetes	22	4	
environmental samples	19	2	
uncultured Actinomycetes bacterium	18	1	uncultured Actinomycetes bacterium hits
uncultured bacterium #0319-2F1	1	1	uncultured bacterium #0319-2F1 hits
uncultured Actinomycetales bacterium	2	1	uncultured Actinomycetales bacterium hits
Yinghuangia sp.	1	1	Yinghuangia sp. hits
uncultured prokaryote	5	1	uncultured prokaryote hits

Figura S3 - Resultados del análisis de la secuencia 03e53d40c9bffca932b2fe3f68c38f6b mediante BLASTn (captura de pantalla de la tabla Taxonómica).

Secuenciación y Ómicas de Próxima Generación

Taxonomy			
Reports	Lineage	Organism	Taxonomy
100 sequences selected			
Taxonomy	Number of hits	Number of Organisms	Description
Bacteria	109	48	
Actinomycetota	105	46	
Actinomycetes	104	45	
Streptomyces	101	42	
Streptomyces exfoliatus	1	1	Streptomyces exfoliatus hits
Unclassified Streptomyces	57	15	
Streptomyces sp.	43	1	Streptomyces sp. hits
Streptomyces sp. MBRC-17	1	1	Streptomyces sp. MBRC-17 hits
Streptomyces sp. A10(2020)	1	1	Streptomyces sp. A10(2020) hits
Streptomyces sp. BB47a	1	1	Streptomyces sp. BB47a hits
Streptomyces sp. KPS-E003	1	1	Streptomyces sp. KPS-E003 hits
Streptomyces sp. WW1-Bc	1	1	Streptomyces sp. WW1-Bc hits
Streptomyces sp. X1-2	1	1	Streptomyces sp. X1-2 hits
Streptomyces sp. SCAUEB23	1	1	Streptomyces sp. SCAUEB23 hits
Streptomyces sp. E3N368	1	1	Streptomyces sp. E3N368 hits
Streptomyces sp. ME-80	1	1	Streptomyces sp. ME-80 hits
Streptomyces sp. E2N459	1	1	Streptomyces sp. E2N459 hits
Streptomyces sp. DR-BS15	1	1	Streptomyces sp. DR-BS15 hits
Streptomyces sp. SC-JG031807	1	1	Streptomyces sp. SC-JG031807 hits
Streptomyces sp. SCA2-2	1	1	Streptomyces sp. SCA2-2 hits

Figura S4 - Resultados del análisis de la secuencia 071b01e2aa5340be4e48c99 169d93e05 mediante BLASTn (captura de pantalla de la tabla Taxonómica).

Taxonomy			
Clusters	Graphic Summary	Alignments	Taxonomy
Reports	Lineage	Organism	Taxonomy
100 clusters selected			
Taxonomy	Number of hits	Number of Organisms	Description
cellular organisms	100	38	
Bacteria	97	36	
Bacillales	61	23	
Actinomycetes	33	2	
Streptomyces griseocarneus	32	1	Streptomyces griseocarneus hits
Williamsia muralis	1	1	Williamsia muralis hits
Bacillales	27	20	
Bacilli	21	15	
Lactobacillales	11	7	
Lactobacillaceae	5	4	
Lacticaseibacillus rhamnosus GG	1	1	Lacticaseibacillus rhamnosus GG hits
Lactobacillus	3	2	
Lactobacillus helveticus	2	1	Lactobacillus helveticus hits
Lactobacillus delbrueckii	1	1	Lactobacillus delbrueckii hits
Furulilactobacillus curtus	1	1	Furulilactobacillus curtus hits
Streptococcus	6	3	
Streptococcus pneumoniae	4	1	Streptococcus pneumoniae hits
Streptococcus anginosus DORA_7	1	1	Streptococcus anginosus DORA hits
Streptococcus salivarius K12	1	1	Streptococcus salivarius K12 hits
Fuscales	10	8	

Figura S5 - Resultados del análisis de la secuencia 1062a45d18c0f6573d213a79 070031b5 mediante BLASTx (captura de pantalla de la tabla Taxonómica).

Taxonomy			
Clusters	Graphic Summary	Alignments	Taxonomy
Reports	Lineage	Organism	Taxonomy
100 clusters selected			
Taxonomy	Number of hits	Number of Organisms	Description
cellular organisms	100	48	
Bacteria	54	19	
Pseudomonadales	20	10	
Pseudomonadota	19	9	
Alphaproteobacteria	4	4	
Acetobacter pomorum DM001	1	1	Acetobacter pomorum DM001 hits
Rhodospirillales	2	2	
Magnetospirillum cryophilswaldense MSR-1	1	1	Magnetospirillum cryophilswaldense hits
Azospirillum sp. CAG-260	1	1	Azospirillum sp. CAG-260 hits
Brucella suis bv. 4 str. 40	1	1	Brucella suis bv. 4 str. 40 hits
Gammaproteobacteria	13	3	
Vibrio parahaemolyticus	11	1	Vibrio parahaemolyticus hits
Enterobacteriales	2	2	
Klebsiella pneumoniae subsp. pneumoniae	1	1	Klebsiella pneumoniae subsp. pneumoniae hits
Pantoea ananatis	1	1	Pantoea ananatis hits
Burkholderiales	2	2	
Thiomonas sp. CB2	1	1	Thiomonas sp. CB2 hits
Taylorella asinigenitalis 14/45	1	1	Taylorella asinigenitalis 14/45 hits
uncultured Desulfovobacterium sp.	1	1	uncultured Desulfovobacterium sp. hits
Branntiales	33	8	

Figura S6 - Resultados del análisis de la secuencia 03e53d40c9bffca932b2fe 3f68c38f6b mediante BLASTx (captura de pantalla de la tabla Taxonómica).

Secuenciación y Ómicas de Próxima Generación

Taxonomy			
Reports	Lineage	Organism	Taxonomy
100 clusters selected			
Taxonomy	Number of hits	Number of Organisms	Description
cellular organisms	100	38	
Bacteria	96	35	
Bacillales	55	18	
Acanthomycetes	40	4	
Streptomyces	37	2	
Streptomyces griseocarnneus	36	1	Streptomyces griseocarnneus hits
Streptomyces himastatinicus ATCC 53653	1	1	Streptomyces himastatinicus ATC
Williamsia muralis	1	1	Williamsia muralis hits
Cutibacterium acnes HL053PA2	2	1	Cutibacterium acnes HL053PA2
Bacillota	14	13	
Bacilli	8	8	
Lactobacillales	5	5	
Lactobacillaceae	2	2	
Lacticaseibacillus rhamnosus GG	1	1	Lacticaseibacillus rhamnosus GG
Lactobacillus helveticus	1	1	Lactobacillus helveticus hits
Streptococcus	3	3	
Streptococcus salivarius K12	1	1	Streptococcus salivarius K12 hit
Streptococcus anginosus DORA_7	1	1	Streptococcus anginosus DORA
Streptococcus ruminantium	1	1	Streptococcus ruminantium hits
Bacillales	3	3	

Figura S7 - Resultados del análisis de la secuencia 071b01e2aa5340be4e48c99 169d93e05 mediante BLASTx (captura de pantalla de la tabla Taxonómica).

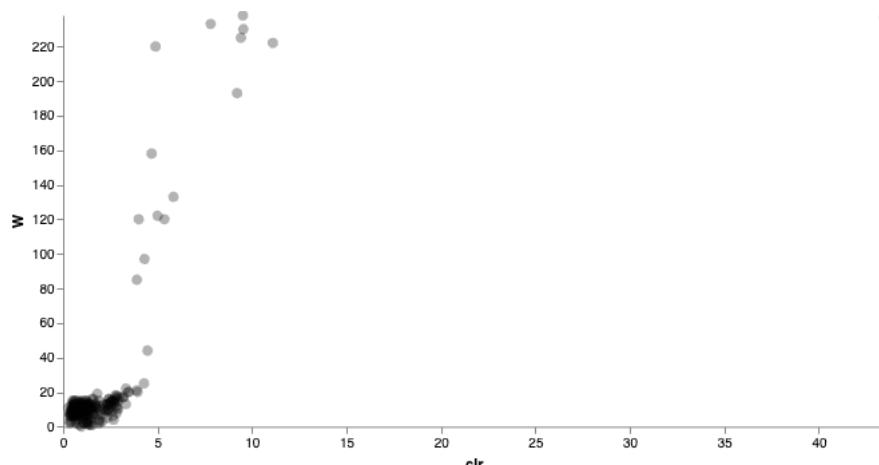


Figura S8 – Volcano plot representando el análisis de composición diferencial a nivel de género según la variable *Site-name* (ANCOM)

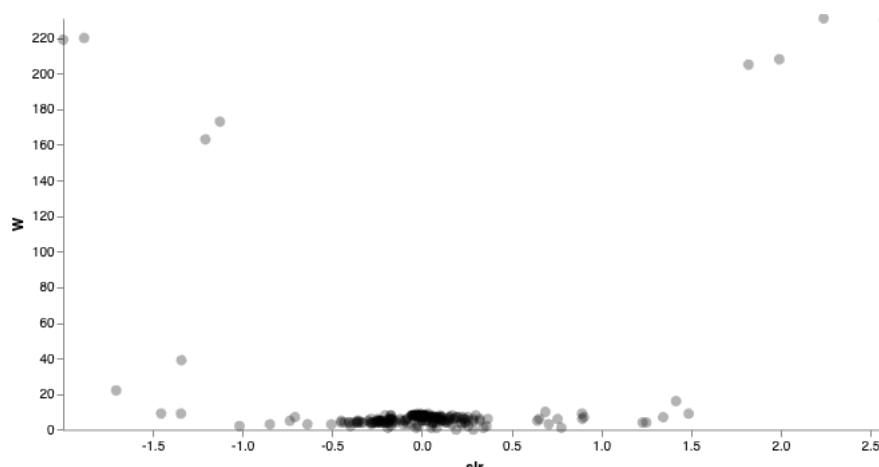


Figura S9 – Volcano plot representando el análisis de composición diferencial a nivel de género según la variable *Vegetation* (ANCOM)

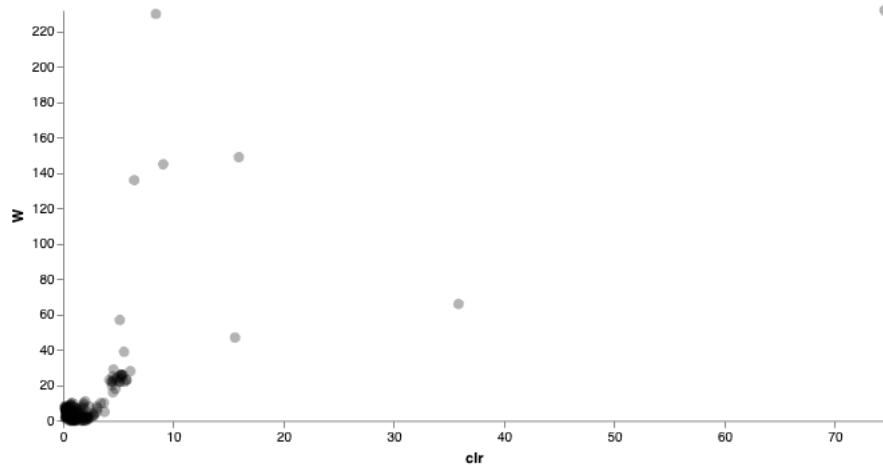


Figura S10 – Volcano plot representando el análisis de composición diferencial a nivel de género según la variable *Extract-group-no* (ANCOM)

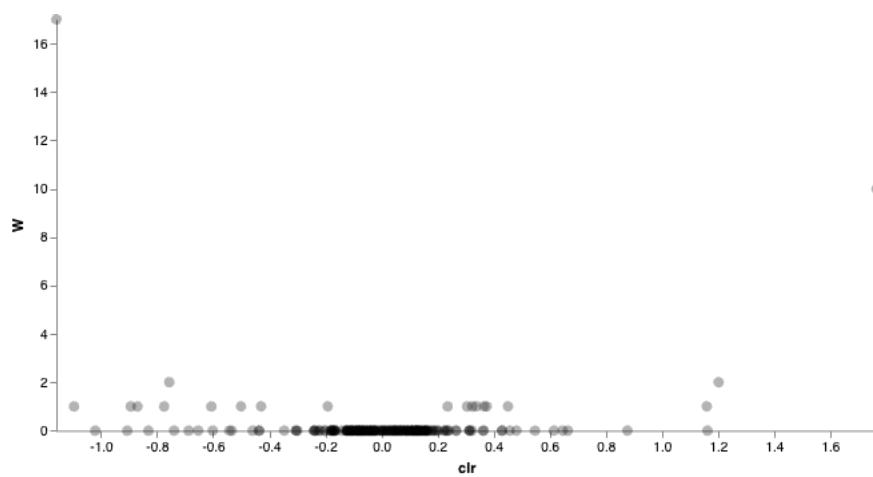


Figura S11 – Volcano plot representando el análisis de composición diferencial a nivel de género según la variable *Transect-name* (ANCOM)

3. TEXTO SUPLEMENTARIO

Texto S1 – Secuencias representativas almacenadas en el archivo rep-seqs.qza

```
>73da966289f204bee2ec120628a02256
AGCGTTAACGGAAATTACTGGCGTAAAGGGCGCTAGGCCTGGTAAGTCGGGTGTGAAAGC
CCTGGGCTTAACCTGGGAATGGCATTGAGACTACCTAGCTAGAGTCCTGGTAGAGGGAAAGTGGAA
TTCCGGTGTAGCGGTGAAATGTGTAGATATCGGAAGGAACACCAGTGGCGAAGGCGACTCCTG
GACCAAGACTGACGCTGAGGCGCAAAGCGTGGGATCAAACAGG
>f77e2c66e9994b2519c523d5215d0768
AGCGTTAACGGAAATCACTGGCGTAAAGGGCGCTAGGCCTGGTAGGTAAGTCGGATGTGAAAGC
CCTGGGCTTAACCTGGGAATGGCATTGAGACTGTCTATCTAGAGTCCTGGTAGAGGGAAAGTGGAAAT
TTCCGGTGTAGCGGTGAAATGTGTAGATATCGGAAGGAACACCAGTGGCGAAGGCGACTCCTG
ACCAAGACTGACGCTGAGGCGCAAAGCGTGGGATCAAACAGG
>c2d032970cd56687ab4647fb256255db
AGCGTTGTCCGGATTATTGGCGTAAAGAGCTCGTAGGCCTGGTAGTCGGGTGTGAAAGC
CCGAGGCTCAACCTCGGAATTGCATTGAGACTGTCTGGCTTGAGGCAGGTAGGGGAGGATGGAA
TTCCGGTGTAGCGGTGAAATGCGCAGATATCGGAGGAACACCTGCGCGAAGGCGGTCTG
GGCCTGTCCTGACGCTGAGGAGCGAAAGCGTGGGAGCGAACAGG
>b94c3645e470d8dfe0d2cb66e7320651
AGCGTTGTCCGGATTATTGGCGTAAAGAGCGTGTAGGCCTGGTAGTCGGGTGTGAAAGC
CTGGGCTCAACCCGGCGTGCCTGGATACTGCGCTAGAGGATGGTAGAGGCGAGTGGAA
ATTCCCGGTGTAGCGGTGAAATGCGCAGATATCGGAGGAACACCAGTAGCGAAGGCGGTCTG
GGGCCATTCTGACGCTGAGACGCAAAGCTAGGGGAGCGAACAGG
```

Texto S2 – Código usado para entrenar el clasificador SILVA.

1. Descargar SILVA

```
curl -sL "https://data.qiime2.org/2024.10/common/silva-138-99-seqs.qza" > silva-138-99-
seqs.qza
curl -sL "https://data.qiime2.org/2024.10/common/silva-138-99-tax.qza" > silva-138-99-
tax.qza
```

2. Extraer las regiones 515-806

```
qiime feature-classifier extract-reads --i-sequences silva-138-99-seqs.qza --p-f-primer
GTGYCAGCMGCCGCGTAA --p-r-primer GGACTACNVGGGTWTCTAAT --p-trunc-len 0 --o-
reads silva-138-99-515-806.qza
```

3. Entrenar el clasificador con Naive Bayes

```
qiime feature-classifier fit-classifier-naive-bayes --i-reference-reads silva-138-99-515-
806.qza --i-reference-taxonomy silva-138-99-tax.qza --o-classifier silva-138-99-515-806-
nb.qza
```

4. Realizar la clasificación taxonómica (ejemplo)

```
qiime feature-classifier classify-sklearn --i-classifier silva-138-99-515-806-nb.qza --i-reads
rep-seqs.qza --o-classification taxonomy.qza
```

Texto S3 – Secuencias usadas para el análisis de BLAST y resultados obtenidos con el análisis de QIIME2

Secuencia en formato fasta

```
>1062a45d18c0f6573d213a79070031b5
AGCGTTGTCCGGATTATTGGCGTAAAGCGCGCGCAGGCCTTAAGTCCTGATGTGAAAGCC
CACGGCTAACCGTGGAGGGTCACTGGAAACTGGGGACTTGAGTACAGAAGAGAAGAGTGGAA
```

Secuenciación y Ómicas de Próxima Generación

TTCCACGTGTAGCGGTGAAATGCGTAGAGATGGAGGAACACCAGTGGCGAAGGCGACTCTTG
GTCTGTAAGTGCAGCTGAGGCGCAAAGCGTGGGAGCAAACAGG

Resultados obtenidos tras el análisis taxonómico con QIIME2 visualizando el fichero taxonomy.qzv

Secuencia número: 1062a45d18c0f6573d213a79070031b5

Identificación taxonómica: d_Bacteria; p_Firmicutes (Bacillota); c_Bacilli; o_Bacillales; f_Bacillaceae; g_Bacillus

Confianza: 0.9311703874749722

Secuencia en formato fasta

>03e53d40c9bffca932b2fe3f68c38f6b

AGCGTTGCCGGATTATTGGCGTAAAGAGCGCGTAGGTGGTTGAAAGTCGGATGTGAAAGTC
CAGGGCTCAACCCCTGGGATGCCATCCGATACTGCAGACTAGAGCATGGTAGGGGGATCTGAATT
CCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCAGGGATCCTGG
GCCATTGCTGACACTGAGGCGCGAAAGCTGGGGAGCGAACAGG

Resultados obtenidos tras el análisis taxonómico con QIIME2 visualizando el fichero taxonomy.qzv

Secuencia número: 03e53d40c9bffca932b2fe3f68c38f6b

Identificación taxonómica: d_Bacteria; p_Actinobacteriota; c_Actinobacteria; o_0319-7L14; f_0319-7L14; g_0319-7L14; s_uncultured_bacterium

Confianza: 0.9019760804186704

Secuencia en formato fasta

>071b01e2aa5340be4e48c99169d93e05

AGCGTTGCCGGATTATTGGCGTAAAGAGCTCGTAGGGCGCTTGTACGTCGGTTGAAAGCC
CGGGGCTTAACCCCGGGTCTGCAGTCGATACGGGCAGGCTAGAGTCGGTAGGGGAGATCGGAAT
TCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCAGGGATCTCTGG
GCCGATACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAGG

Resultados obtenidos tras el análisis taxonómico con QIIME2 visualizando el fichero taxonomy.qzv

Secuencia número: 071b01e2aa5340be4e48c99169d93e05

Identificación taxonómica: d_Bacteria; p_Actinobacteriota; c_Actinobacteria; o_Streptomycetales; f_Streptomycetaceae; g_Streptomyce

Confianza: 0.999985305046216