



**FACULTAD DE CIENCIAS AGRARIAS
UNIVERSIDAD NACIONAL DE ROSARIO**

**“Metabarcoding de comunidades bacterianas asociadas a poblaciones
Amazónicas de *Drosophila* afectadas por la perturbación de la selva primaria
por prácticas agrícolas”**

Lucía Lorenzi

**TRABAJO FINAL PARA OPTAR AL TITULO DE ESPECIALISTA EN
BIOINFORMÁTICA**

DIRECTORA: Flavia Krsticevic

**Metabarcoding de comunidades bacterianas asociadas a poblaciones
Amazónicas de Drosophila afectadas por la perturbación de la selva primaria
por prácticas agrícolas**

Lucía Lorenzi

Licenciada en Biotecnología – Universidad Nacional de Rosario

Este Trabajo Final es presentado como parte de los requisitos para optar al grado académico de Especialista en Bioinformática, de la Universidad Nacional de Rosario y no ha sido previamente presentada para la obtención de otro título en ésta u otra Universidad. El mismo contiene los resultados obtenidos en investigaciones llevadas a cabo en ____CIFASIS (Rosario)_____, durante el período comprendido entre __Octubre de 2015 y mayo de 2018_____, bajo la dirección de la Dra. Flavia Krsticevic.

Lucía Lorenzi

Nombre y firma del autor



Flavia Krsticevic

Nombre y firma del Director



Defendida: 3 de abrilde 2019.

Agradecimientos

A mi familia, que desde el momento en que decidí iniciar mi camino en esta carrera de Especialización me brindó su apoyo emocional constante y motivación para cumplir con mis objetivos.

Al grupo de Bioinformática y Agroinformática del CIFASIS, donde tuve la oportunidad de desarrollar la mayor parte de este trabajo y del cual me llevo además de una experiencia de trabajo de alta calidad científica, una experiencia de calidez humana inolvidable.

A Flavia, quien me brindó desde el primer momento todo su apoyo, en quien encontré además de una guía, una amiga excepcional, y quien depositando en mí su confianza desde el primer momento, transitó y compartió conmigo este lindo camino, si bien escabroso por momentos, hacia la conclusión de este trabajo.

Resumen

La selva primaria en la región amazónica cambia abruptamente debido al desarrollo de la agricultura. No está bien establecido cómo reaccionan a estos cambios las especies de drosófilas. El microbioma bacteriano puede estar relacionado a la adaptación de la mosca al nuevo ambiente contribuyendo en su nutrición y tolerancia inmune. Para evaluar el impacto de cambios permanentes al paisaje selvático en la diversidad microbiana de moscas neotropicales, realizamos un análisis metagenómico del gen 16S de especies bacterianas (*metabarcoding*) asociadas a dos especies de moscas, *Drosophila malerkotilana* (especie cosmopolita y exótica) y *Drosophila sturtevanti* (especie neotropical) encontradas en selva primaria y ambientes agrícolas del estado de Pará, Brasil. Como control adicional y para evaluar también la influencia del sustrato en la composición bacteriana de moscas, incluimos en el análisis muestras de especies de moscas característicamente asociadas a sustratos alternativos a los frutos, *Hirtodrosophila* sp. (asociada a hongos) y *Drosophila speciosa* (asociada a flores). Nuestros resultados sugieren que el ambiente es un determinante importante de la comunidad bacteriana, mientras que la especie de mosca no parece serlo. Adicionalmente, la diversidad bacteriana es mayor en ambientes modificados donde existiría una mayor oferta de sustratos. Esta mayor diversidad bacteriana puede estar favoreciendo la nutrición de moscas y en última instancia su adaptación al ambiente modificado.

Abstract

The primary forest in the Amazon region changes abruptly due to the development of agriculture. It is not well established how species of drosophils react to these changes. The bacterial microbiome can affect the adaptation of the fly to the environment by contributing to its nutrition and immune tolerance. To evaluate the impact of permanent changes to the jungle landscape on the microbial diversity of neotropical flies, we performed a metagenomic analysis of the 16S gene of bacterial species (metabarcoding) associated with two species of flies, *Drosophila malerkotilana* (cosmopolitan and exotic species) and *Drosophila sturtevanti* (neotropical species) found in primary forest and agricultural environments of the state of Pará, Brazil. As an additional control, and in order to evaluate the influence of the substrate on the bacterial composition of flies, we also included in the analysis samples of drosophilic species characteristically associated with substrates alternative to fruit, *Hirtodrosophila* sp. (associated with fungi) and *Drosophila speciosa* (associated with flowers). Our results suggest that the environment is an important determinant of the bacterial community, whereas the fly species does not seem to have an impact on it. Additionally, bacterial diversity is greater in modified environments where a greater supply of substrates would be found. This greater bacterial diversity may be favoring the nutrition of flies and ultimately their adaptation to the modified environment.

Índice

Agradecimientos	3
Resumen	4
Abstract	5
Introducción	7
Objetivos	10
Objetivos específicos	10
Materiales y Métodos	11
Muestras de especies de Drosophilidae utilizadas	11
Amplificación de las regiones variables V3-V4 y estrategia de secuenciación	12
Análisis bioinformático	12
Índices de diversidad y análisis estadístico	13
Resultados y discusión	16
Tratamiento de datos para el armado de las matrices y clasificación taxonómica	16
Error estimado	20
Análisis de alfa y beta diversidad	21
Análisis de diversidad alfa.....	21
Análisis de composición bacteriana	26
Análisis de diversidad beta	31
Conclusiones	43
Anexos	44
Bibliografía	46

Introducción

El desarrollo de actividades agrícolas en áreas selváticas primarias en la región Amazónica tiene un impacto directo en el ambiente natural y su biota asociada. Especies de moscas del género *Drosophila* (drosófilas) han sido anteriormente empleadas como indicadores biológicos de la degradación de hábitats naturales por las actividades humanas (Avondet et al., 2003). En la región amazónica, han sido observadas diferencias entre las poblaciones de drosófilas presentes en la selva primaria respecto a aquellas provenientes de ambientes modificados (Furtado et al., 2011). Por un lado, existen variaciones en la riqueza específica y abundancia de drosófilas entre los ambientes selváticos primarios y aquellos modificados por el hombre. La diversidad específica es estable y menor en ambientes selváticos primarios. El número de individuos característicos de ambientes selváticos primarios disminuye o es nulo en ambientes modificados y viceversa, y especies cosmopolitas son raras en los ambientes selváticos primarios. Además, se han observado diferencias significativas en el tamaño del cuerpo entre moscas adultas de los diferentes ambientes (Furtado et al., 2011). En general, las especies drosófilas de ambientes agrícolas son más grandes que aquellas de la selva primaria. Esto se observa incluso para especies de moscas que son típicas de la selva primaria, pero que pueden encontrarse eventualmente en ambientes modificados. Por ejemplo, *Drosophila sturtevanti* es una mosca neotropical y cuando eventualmente se encuentran individuos de esta especie en ambientes modificados, su tamaño es mayor al observado en individuos de la misma especie en su ambiente natural (Furtado et al., 2011). Más allá de estas observaciones, aún no se ha establecido cuáles son los factores determinantes de la adaptación de las moscas a estos cambios en su hábitat natural. Una hipótesis sobre el cambio en el tamaño de adultos en ambientes modificados es que las comunidades bacterianas asociadas a éstos ayudarían a contribuir a su nutrición y tolerancia inmune (Vodovar et al., 2005), teniendo un impacto directo en su resistencia al estrés ambiental (Wong et al., 2015) y permitiéndoles adaptarse a condiciones cambiantes e irreversibles.

En cultivos de *Drosophila melanogaster*, ciertas bacterias de las familias Acetobacteraceae, Lactobacillaceae, y Enterobacteriaceae promueven el crecimiento

y desarrollo de los adultos, protegen contra la acumulación excesiva de lípidos en dietas altas en azúcares y proporcionan vitaminas B importantes para la salud en dietas bajas en nutrientes (Broderick et al., 2012). Además, también en cultivos de *D. melanogaster* se encontró que los individuos criados a 13°C y resistentes al frío muestran una mayor abundancia de la bacteria endosimbionte *Wolbachia*. En tanto que, las moscas criadas a 31°C y resistentes al calor mostraron una abundancia mayor de *Acetobacter* (Moghadam, et al., 2017). Las observaciones más interesante de estos estudios son que *D. melanogaster* puede cambiar rápidamente su microbioma ante fluctuaciones ambientales (Jehrke et al., 2018) y también que ciertos genotipos pueden influir en la composición del microbioma (Early et al., 2017).

Las especies de drosófilas de la naturaleza están asociadas a un grupo taxonómico restringido de bacterias intestinales (Wong et al., 2013), siendo la dieta un determinante importante de esta composición (Chandler et al., 2011). La composición de bacterias asociadas con catorce especies de drosófilas de diferentes hábitats y ubicaciones geográficas distantes fue evaluada por Chandler et al. En este estudio se encontró que, en general, la diversidad de bacterias asociadas con poblaciones naturales de *Drosophila* es baja, existiendo cuatro familias de bacterias dominantes, Enterobacteriaceae, Acetobacteraceae, Lactobacillaceae y Enterococcaceae. Al mismo tiempo, encontraron que la riqueza a nivel de especie, la uniformidad y la diversidad varían ampliamente entre las diferentes especies de moscas analizadas. Por otro lado, en poblaciones de *D. nigrospiracula*, una mosca especialista asociada a dos especies de cactus y endémica del desierto de Sonora, se encontró que la microbiota es dependiente de la especie de cactus del cual se alimenta. Dentro de los grupos bacterianos asociados a *D. nigrospiracula*, se han encontrado órdenes raros como el de bacterias pertenecientes al orden Orbales y a los géneros *Serpens* y *Dysgonomonas* (Martinson et al., 2017). Es decir, la microbiota parece estar moldeada por el tipo de substrato del cual las moscas se alimentan.

A pesar de que el neotrópico es un bioma ecológicamente importante, los microbiomas de poblaciones de drosófilas neotropicales no han sido caracterizados hasta el momento. En este contexto, es interesante y novedoso el estudio y comparación de la diversidad bacteriana de diferentes especies de drosófilas presentes en ambientes primarios y modificados del Amazonas. El estudio de

metagenomas bacteriano de moscas permitiría evaluar si la composición específica está dominada por el componente ambiental o del hospedador y cómo esta composición influye en la adaptación de individuos de moscas a los diferentes ambientes.

La metagenómica es la aplicación de técnicas genómicas modernas para acceder al contenido genético de comunidades enteras de microorganismos, sin cultivo previo y directamente a partir de su entorno natural (Chen et al., 2005). Con el advenimiento de las tecnologías de secuenciación de nueva generación o NGS (Shendure et al., 2008), empezó a ser económicamente viable la secuenciación total del ADN proveniente de muestras ambientales. Desde entonces, la metagenómica comenzó a establecerse como una estrategia de preferencia para abordar estudios microbiológicos, incluyendo a una gran mayoría de especies de cultivo difícil o imposibles de cultivar. Las nuevas tecnologías de secuenciación pueden, de esta manera, ayudar a la identificación precisa de especies bacterianas asociadas con drosófilas. Una manera efectiva y simple de generar una gran cantidad de datos de biodiversidad es el denominado *metabarcoding*, que consiste en la amplificación selectiva por PCR y subsecuente secuenciación de una porción específica del ADN, que por su función esencial ha sido conservada evolutivamente a un ritmo comparable al de la especiación, funcionando de esta manera como un “reloj biológico”. Estos marcadores permiten identificar especies de bacterias, por lo que se los conoce como “códigos de barras” (*barcode*). En el caso de bacterias, el marcador más comúnmente utilizado es el gen de ARN ribosomal 16S. Este gen consiste de nueve regiones variables flanqueadas por regiones de secuencia más conservada. Las regiones variables V3 y V4 proveen amplia información para la clasificación taxonómica de comunidades microbianas.

Objetivos

El objetivo general de nuestro trabajo es determinar la diversidad de las comunidades bacterianas asociadas con especies de Drosophilidae en ambientes selváticos primarios y ambientes modificados por la agricultura en el estado de Pará en el Brasil amazónico.

Objetivos específicos:

- Caracterizar mediante *metabarcoding* el microbioma bacteriano de especies de moscas asociadas a diferentes substratos (fruto, flores y hongos).
- Caracterizar mediante *metabarcoding* el microbioma bacteriano de especies de moscas selváticas (*D. sturtervanti*) y cosmopolitas (*D. malerkotliana*), en ambientes selváticos y modificados.
- Evaluar el impacto de cambios permanentes al paisaje selvático en la diversidad microbiana de moscas neotropicales.

Materiales y Métodos

Muestras de especies de Drosophilidae utilizadas

Individuos de las especies *D. malerkotliana* y *D. sturtevanti* fueron recolectados de una manera estandarizada, utilizando trampas conteniendo 100 g de cebos de banana fermentada situados durante 36 h en diferentes localizaciones seleccionadas entre ambientes de selva primaria (selváticos) y agrícolas (modificados) de la región de Pará en el Amazonas. Adicionalmente, se incluyeron como control otras dos especies de moscas, *D. speciosa* e *Hirtodrosophila* (SB13011), caracterizadas por su asociación específica a flores y hongos, respectivamente. Debido a variaciones aleatorias propias del método de muestreo y a la diferente abundancia de moscas en los distintos ambientes, el número de individuos que pudo ser recolectado para cada muestra varió entre 2 y 12 individuos. Para todas excepto una de las muestras se pudo realizar al menos una réplica biológica y, como un control adicional de secuenciación y PCR, se seleccionaron 2 muestras al azar para las cuales se realizaron también réplicas técnicas. La información detallada sobre las muestras utilizadas se encuentra en el Anexo, Tabla A1.



Figura 1. Selva primaria. Punto de muestreo en el cual fueron colocadas las trampas con atractivo de banana en la Fazenda Marupiara del estado de Pará. Coordenadas: 2°47'46.3"S 48°32'35.3"W.



Figura 2. Paisaje modificado de forma irreversible, ambiente agrícola. Punto de muestreo en la cual fueron colocadas las trampas con atractivo de banana en la Fazenda Marupiara del estado de Pará. Coordenadas: 2°46'12.1"S 48°31'19.8"W.

Amplificación de las regiones variables V3-V4 y estrategia de secuenciación

La extracción de ADN se realizó en moscas adultas (machos y hembras), previamente lavadas en una solución de agua lavandina 10%, con el kit específico para microbiomas QIAamp® DNA Microbiome. Un fragmento del gen 16S ribosomal conteniendo las regiones variables V3 y V4, se amplificó por PCR usando los primers PE250, 515F: GTGCCAGCMGCCGCGGTAA y 806R: GGACTACHVGGGTWTCTAAT. La secuenciación se realizó en multiplexado en una plataforma Illumina HiSeq-2000 siguiendo las recomendaciones de Fadrosch et al. (2014). Junto con las muestras, se agregó una comunidad artificial de bacterias HN-782D (*Mock community*), como un control interno positivo de la clasificación específica y abundancia y para evaluar la calidad de la secuenciación. La *Mock community* contiene el ADN genómico de 20 cepas bacterianas representadas equimolarmente en 100,000 copias por organismo por μL (Genomic DNA from Microbial Mock Community B v5.1L, for 16S rRNA Gene Sequencing Catalog No. HM-782D).

Análisis bioinformático

Las secuencias generadas (*output*) se demultiplexaron y montaron en una única secuencia de 450 pares de bases con la herramienta de análisis metagenómico *mothur* (Schloss et al., 2009), siguiendo el protocolo de MiSeq-SOP (https://www.mothur.org/wiki/MiSeq_SOP). Las secuencias fueron filtradas para remover las secuencias duplicadas (*unique*) y las quimeras (*chimera.vsearch*). De esta forma, se redujo el ruido producido por los errores de secuenciación. Luego, se procedió al pre-agrupamiento (*pre.cluster*), obteniéndose los grupos de secuencias organizados por abundancia (Anexo Tabla A2), siendo identificadas las secuencias que están a al menos 2 nt una de la otra. Una vez montado el fragmento V3-V4 para cada una de las secuencias *paired-end* y habiendo filtrado los datos para la obtención de secuencias únicas, se procedió a la clasificación.

La clasificación taxonómica de las secuencias se efectuó con el programa *Centrifuge* (Kim et al., 2016) (utilizando los parámetros -k1 y paired-end), el cual funciona de forma rápida y bajo un uso efectivo de memoria RAM. *Centrifuge* facilita el procesamiento de grandes cantidades de datos con una sensibilidad y especificidad comparable con Blast. Cada una de las secuencias fue comparada con la base de datos de genomas bacterianos de RefSeq del NCBI, la cual contiene genomas de alta

calidad y no redundantes, creando así un *ranking* basado en la identidad de secuencias con las secuencias de referencia. Una vez clasificadas las secuencias en el nivel de jerarquía de especies bacterianas, mediante scripts *in-house* se ordenaron los datos específicos en matrices con el número de ocurrencias de cada especie en cada muestra (es decir, el número de lecturas de secuenciación, de ahora en adelante, *reads*). Los subsecuentes análisis de diversidad dentro y entre los diferentes tipos de muestras se realizaron en R, mediante scripts *in-house* en complementación con funciones de los paquetes *phyloseq* (McMurdie et al., 2013) y *vegan* (Oksanen et al., 2018), ambos especializados en análisis de datos censales de microbiomas.

Índices de diversidad y análisis estadístico

La diversidad específica de poblaciones de organismos y la diferenciación entre éstas se suele expresar a través de diferentes índices de diversidad. La diversidad alfa mide la riqueza a nivel local, en un hábitat o una muestra determinada. La diversidad beta mide la diferenciación entre estos hábitats o muestras y existen diversas maneras de definirla y medirla. La diversidad gamma es la medida de la diversidad total de especies en un área geográfica determinada. Esta diversidad está determinada a su vez por la diversidad alfa y beta ($\gamma = \alpha \times \beta$). Estos términos fueron acuñados por Whittaker (1960) en el contexto de un estudio de la vegetación en ambientes montañosos. En estudios metagenómicos y particularmente en estudios de microbiomas, donde es difícil definir ambientes o hábitats tales como los que dieron origen a estas medidas de diversidad, estos términos suelen adaptarse para responder a las preguntas de investigación y al diseño experimental particular, pero, en general, la diversidad alfa hace referencia a la diversidad que hay dentro de una muestra, mientras que la diversidad beta mide la variación entre muestras. En el estudio de microbiomas, una muestra suele definirse como un individuo en un lugar y tiempo determinados, sobre el que se mide la diversidad de especies bacterianas huéspedes. En nuestro caso, una muestra corresponde a una especie de mosca (*D. malerkotliana*, *D. sturtevantii*, *D. speciosa* o *Hirtodrosophila*) en un ambiente determinado (para el caso de *D. malerkotliana* y *D. sturtevantii* los ambientes muestreados corresponden a un ambiente selvático, “F11”, un ambiente modificado denominado “Casa” y un ambiente modificado denominado “Pasto”, mientras que *D.*

speciosa e *Hirtodrosophila* fueron colectadas en locaciones diferentes y en este caso el ambiente hace referencia al sustrato específico, “flor” y “hongo”, respectivamente). Cabe aclarar que, debido a una necesidad metodológica, nuestra definición de muestra contiene a más de un individuo de Drosophilidae. Esto se debe a que, para la generación de cada muestra (en un ambiente y para una especie de mosca determinados), el ADN de diferentes individuos fue procesado de manera conjunta debido a que la cantidad de ADN extraído de una mosca individual resulta muy escasa para obtener buenos resultados de secuenciación (Anexo Tabla A1).

La medida más simple de la riqueza dentro de una muestra (diversidad alfa) es el número total de especies observadas, pero también existen índices que tienen en cuenta la abundancia relativa de especies. En este trabajo utilizamos el índice de diversidad de Shannon (1948):

$$S = - \sum_{i=1}^n p_i \ln(p_i) ;$$

donde n es el número de especies bacterianas y p_i es la abundancia proporcional de la especie i.

Este índice mide la entropía de cada muestra teniendo en cuenta las abundancias proporcionales de las especies bacterianas presentes. La idea de este índice es que cuanto mayor sea el número de especies diferentes presentes en la muestra y cuánto más homogéneas sean sus abundancias proporcionales, más difícil será predecir correctamente qué especie bacteriana se encontrará con mayor probabilidad al tomar aleatoriamente un individuo de esta muestra. El índice de Shannon cuantifica esta incerteza (entropía) y sus valores varían entre 0 (en el caso hipotético de que sólo exista una especie bacteriana en la muestra) y $\ln(n)$ en el caso de que todas las especies bacterianas se encuentren en igual proporción en la muestra.

A su vez, también existen diversos métodos para calcular distancias entre muestras (diversidad beta) en base a datos de abundancia de especies bacterianas. Algunos métodos son cualitativos ya que tienen en cuenta sólo la presencia o ausencia de especies, mientras que otros son cuantitativos y tienen en cuenta también la abundancia de cada especie en cada muestra. Por otro lado, se distinguen los métodos que consideran las relaciones filogenéticas entre las especies de los que no hacen uso de esta información. UniFrac es un método especialmente desarrollado

para medir diferencias entre muestras ambientales (Lozupone et al., 2005) que incorpora información sobre la relación relativa de los miembros de la comunidad al incorporar las distancias filogenéticas entre los organismos observados en el cálculo. La implementación de este método puede ser tanto cuantitativa (UniFrac *weighted*) como cualitativa (UniFrac *unweighted*). Otro método comúnmente utilizado en estudios ecológicos es el método cuantitativo de Bray y Curtis (1957). En este trabajo utilizamos y comparamos estos tres métodos para la estimación de la diversidad beta. Para calcular las distancias UniFrac utilizamos el paquete *phyloseq*, el cual permite integrar árboles filogenéticos (los cuales pueden descargarse de bases de datos como NCBI) de las comunidades bacterianas incluidas en el análisis. Para calcular la distancia Bray-Curtis utilizamos la función *vegdist* del paquete *vegan*.

Resultados y discusión

Tratamiento de datos para el armado de las matrices y clasificación taxonómica

En total, entre las 27 muestras provenientes de ambientes diferentes (selva primaria y ambientes modificados “Casa” y “Pasto”), substratos diferentes (“frutos”, “flores” y “hongos”) y especies de drosófilas diferentes (*D. malerkotliana*, *D. sturtevantii*, *D. speciosa* e *Hirtodrosophila*) se obtuvieron 5.1 M de secuencias concatenadas y pre-agrupadas (Anexo Tabla A2). De estas secuencias, 1.5 M fueron clasificadas en un total de 89 especies bacterianas (incluyendo las especies presentes en la *Mock community*). En promedio, 54459 secuencias fueron clasificadas por muestra, con una mediana de 29944 secuencias clasificadas. No se observó correlación entre el número de individuos muestreados y el número de *reads* asignados (spearman=0.06, p-val=0.769). En la figura 3 se muestra el número de *reads* totales asignados a especies bacterianas en cada muestra. Este número es

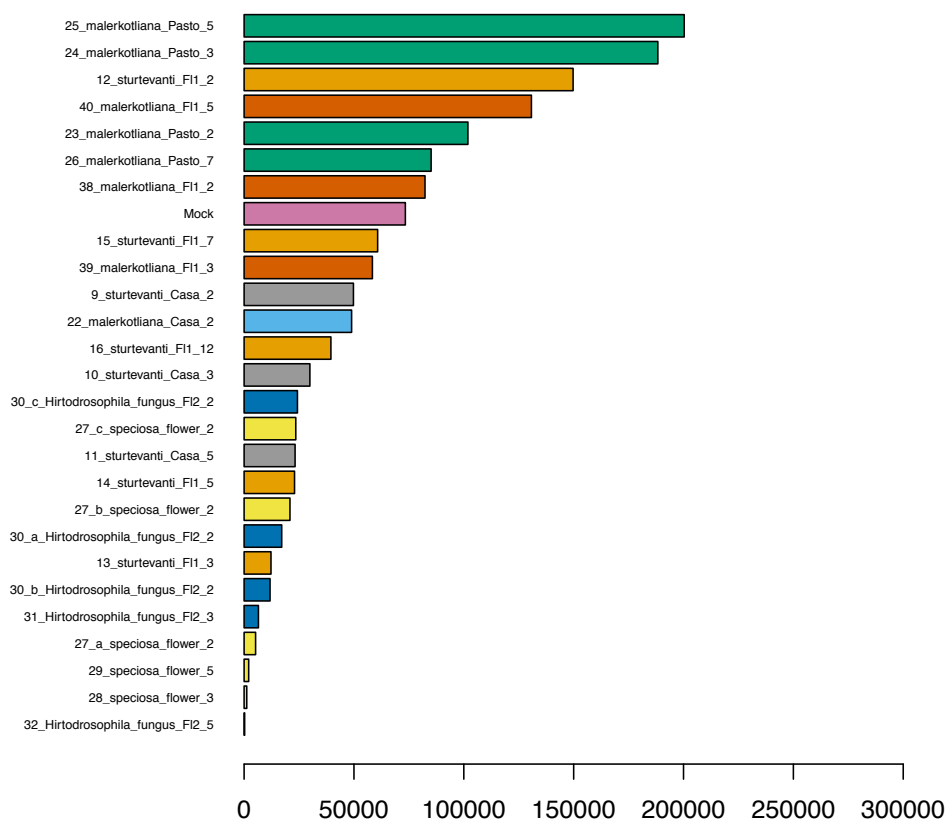


Figura 3. Número total de *reads* asignados a especies bacterianas para cada una de las 27 muestras procesadas. Los colores indican réplicas biológicas o técnicas de una misma muestra.

bastante variable, incluso entre réplicas técnicas, lo cual puede deberse a la variabilidad experimental introducida durante los diferentes pasos de amplificación por PCR y secuenciación. Es de notar que la variabilidad entre muestras en el número de secuencias concatenadas y pre-agrupadas es menor que la variabilidad en el número de éstas secuencias que finalmente fueron asignadas a especies bacterianas en cada muestra (coeficientes de variación: 0.43 y 1.03, respectivamente. Prueba asintótica para la igualdad de los coeficientes de variación, $p\text{-val}=0.003$).

El número de especies bacterianas observadas en cada muestra está correlacionado con el número total de secuencias asignadas en dicha muestra (Spearman: $\rho=0.69$, $p\text{-val}=5.024\text{e-}05$). Esta correlación es esperable, especialmente para muestras con bajo número de secuencias asignadas, donde es más probable que la cobertura de secuenciación sea insuficiente para capturar las especies menos abundantes en la muestra. Al analizar la correlación entre número de especies observadas y número de *reads* asignados para muestras con un número de *reads* asignados menor a 20000, observamos que la misma se mantiene alta (Spearman: $\rho=0.72$, $p\text{-val}=0.042$) (Figura 4 B), mientras que esta correlación se vuelve poco significativa para muestras con valores de *reads* asignados mayores a 20000 (Spearman: $\rho=0.29$, $p\text{-val}=0.232$) (Figura 4 C). El hecho de que no aumente significativamente el número de especies con el aumento en el número de *reads* implica que estarían bien representadas todas o la gran mayoría de las especies bacterianas (excepto aquellas que no se encuentren por una amplificación diferencial en la PCR).

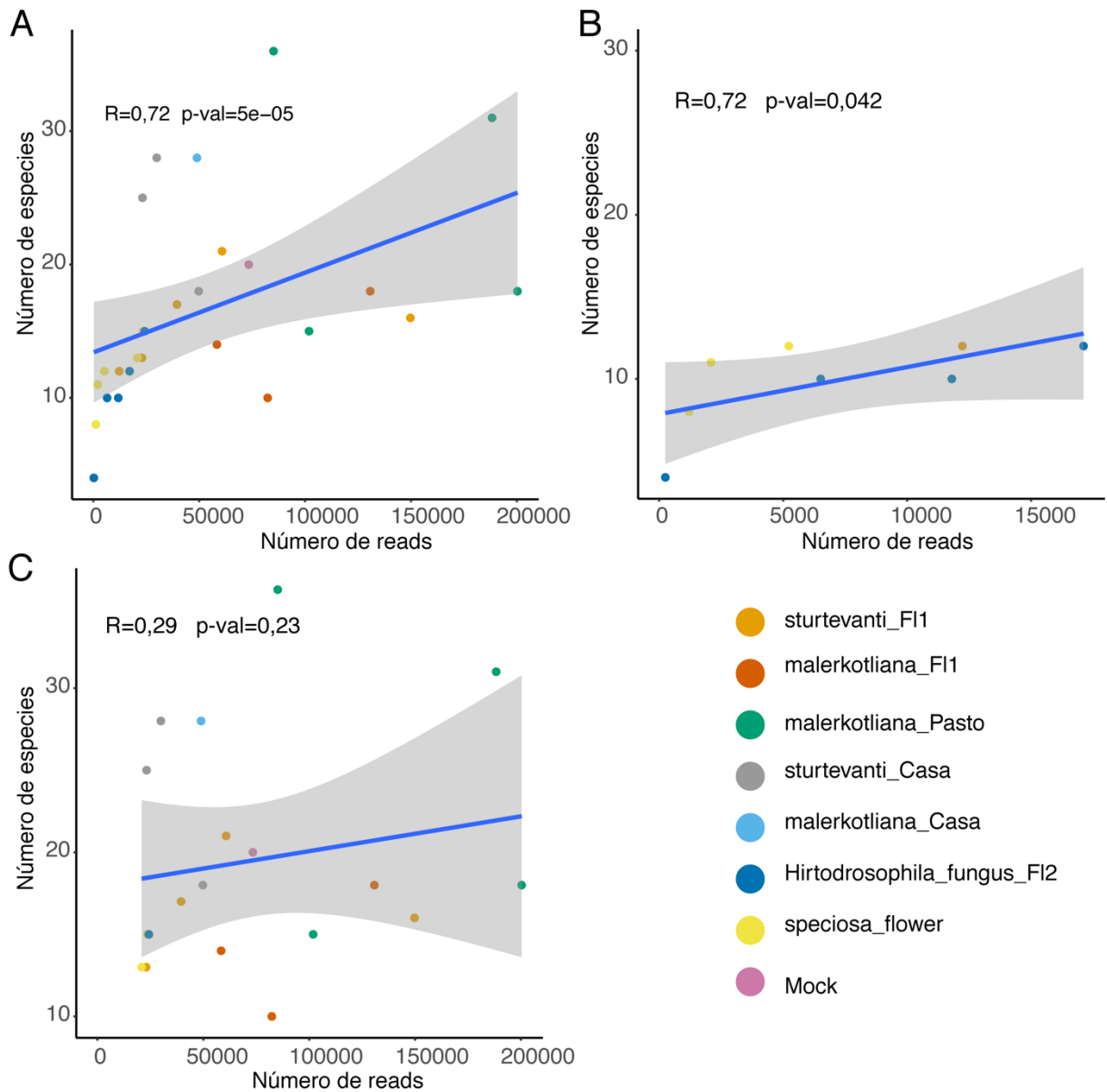


Figura 4. Correlación entre número de *reads* y número de especies observadas en cada muestra. **A.** Todas las muestras. **B.** Muestras con un número total de *reads* entre 0 y 20000. **C.** Muestras con un número total de *reads* entre 20000 y 200000

Por otro lado, si tomamos como ejemplo la muestra de *D. speciosa*, que cuenta con tres réplicas técnicas, vemos que el incremento en el número total de *reads* entre réplicas no parece tener un efecto proporcional en la cantidad de especies encontradas, incluso cuando el número de *reads* en la réplica menor es muy bajo. Así, vemos que a pesar de tener "27_b_speciosa_flower_2" y "27_c_speciosa_flower_2"

cerca de 4 veces más *reads* totales que la otra réplica técnica, "27_a_speciosa_flower_2", el número de especies observadas no aumenta proporcionalmente. Además, las tres réplicas técnicas tienen 10 especies en común y en los tres casos la especie dominante es *Tatumella citrea* (Figura 5). Estas observaciones junto con el hecho de que la biodiversidad bacteriana en moscas es relativamente baja, encontrándose generalmente dominada por un número de alrededor de 20 especies (Chandler et al., 2011), indican que, a pesar de haberse obtenido una gran variabilidad de secuencias asignables entre réplicas, la cantidad de *reads* totales obtenidos para (al menos una réplica de) cada muestra debería ser suficiente para capturar la biodiversidad dominante en cada caso. Cabe aclarar también que el objetivo de este trabajo no es describir en detalle la diversidad bacteriana de cada drósofila en cada ambiente, sino analizar y comparar el impacto de modificaciones antropogénicas en los diferentes ambientes sobre la microbiota de las moscas habitantes.

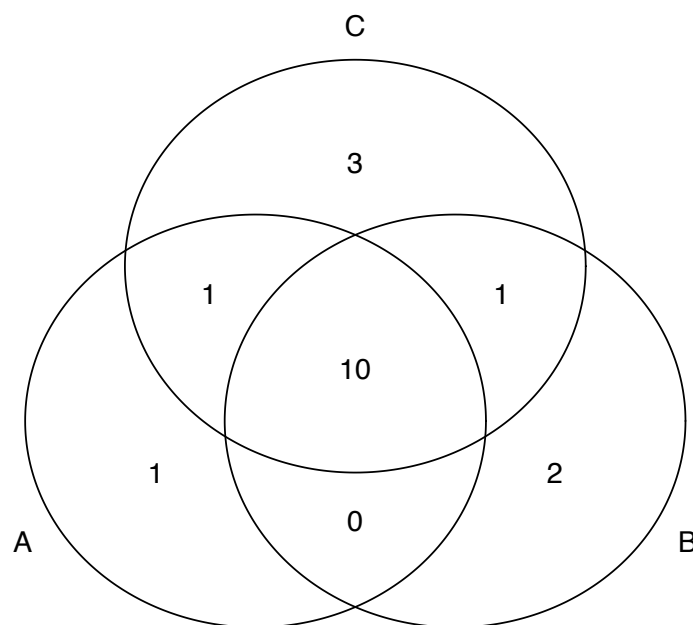


Figura 5. Diagrama de Venn para la muestra "27_speciosa_flower_2" (*D. speciosa*) mostrando el número de especies bacterianas halladas en común y únicamente en alguna(s) de las réplicas técnicas.

Asimismo, para todas las muestras analizadas hay al menos una réplica biológica con un número de *reads* mayor a 20000, por lo que, en general, obtuvimos una buena cobertura de *reads* asignados a especies. Estudios anteriores han utilizado números de *reads* totales significativamente menores que los obtenidos en este estudio. (Chandler et al., 2011).

Error estimado

El error de secuenciación se estimó con *Mothur*, utilizando la *Mock community* como referencia. El error medio obtenido para las secuencias fue del 0.2%, es decir un error a cada 500 pb. Este error fue corregido durante el pre-clustering de las secuencias. Adicionalmente, otro tipo de error fue detectado en la *Mock community*. De las 20 especies que deberían haber sido clasificadas, 16 fueron coincidentes, una fue mal clasificada (asignada a especies filogenéticamente cercanas) y tres fueron perdidas. Además, se obtuvo un número muy variable de secuencias asignadas a cada especie bacteriana cuando se esperaba obtener valores similares, al encontrarse la cantidad de ADN de cada especie en la comunidad Mock en proporciones iguales. Debido a que la composición de ADN en esta muestra control es conocida, estas diferencias pueden explicarse en su mayor medida por errores propios del proceso de PCR, resultantes en amplificación diferencial de secuencias (ver más abajo). En base a estos resultados se ajustó el valor de mínimo de abundancia entre todas las muestras requerido (al menos 0.0074% de la abundancia total) utilizado para filtrar especies bacterianas poco abundantes en los análisis subsiguientes.

Posibles explicaciones para los errores de secuenciación y en la *Mock community* y las diferencias observadas entre réplicas biológicas y técnicas sobre el número de *reads* asignados son: a) Variaciones en el proceso de extracción de ADN, las cuales son difíciles de monitorear y controlar. El proceso de extracción de ADN es un proceso mecánico y manual, en el cual es esperable observar variabilidad en la cantidad de ADN extraído en cada muestra. Por otro lado, debido al pequeño tamaño de las moscas, fue necesario el procesamiento del cuerpo de la mosca entero, por lo que el ADN de partida puede contener concentración variable de ADN proveniente de la mosca, factor también difícil de controlar. b) Diferencias introducidas en la

amplificación por PCR. Durante el proceso de PCR pueden ocurrir amplificaciones diferenciales de ciertas secuencias con mayor afinidad por los *primers* utilizados. También pueden ocurrir errores propios de la actividad enzimática de la polimerasa. Es de notar que estos errores pueden dar lugar a secuencias con nucleótidos modificados que ya no son capaces de ser alineadas al genoma de la especie bacteriana del cual provenía la secuencia original o incluso pueden llegar a ser asignados erróneamente a una especie cercana. El impacto de estos errores dependerá también del ciclo de PCR en que son introducidos, siendo los errores introducidos durante los primeros ciclos propagados exponencialmente, lo cual puede tener un impacto significativo en el resultado final. c) Variabilidad propia del proceso de secuenciación. Durante la secuenciación en las plataformas Illumina también ocurre un segundo paso de amplificación mediante PCR necesario para la amplificación de la señal, con las posibles causas de variabilidad asociadas discutidas en el punto anterior. Por otro lado, pueden ocurrir diferencias aleatorias en la carga del ADN en la celda de flujo, existen errores propios de la actividad enzimática de la polimerasa, y errores técnicos del secuenciador.

Análisis de alfa y beta diversidad

Una vez obtenidas las matrices de número de secuencias por especie y por muestra, procedimos al análisis de alfa y beta diversidad utilizando R y aprovechando en particular las funciones implementadas en las herramientas *phyloseq* y *vegan*.

Análisis de diversidad alfa

La riqueza total, es decir la cantidad de especies observadas entre todas las muestras, antes de filtrar es de 72 especies y 26 familias, sin incluir las especies y familias encontradas únicamente en la *Mock community*. De estas especies, 21 son especies raras o poco frecuentes ya que tienen en general una muy baja abundancia y se encontraron restringidas a una única muestra. Una única excepción dentro de este grupo respecto a la abundancia la constituye la especie *Bordetella trematum* que representa un 18% de los *reads* provenientes de una de las réplicas técnicas de la muestra 30_Hirtodrosophila_fungus_FI2_2, 30_c_Hirtodrosophila_fungus_FI2_2. Notablemente, esta especie bacteriana no fue hallada en las otras dos réplicas técnicas de esta muestra. Sin embargo, *B. trematum* es una especie filogenéticamente

cercana *Achromobacter xylosoxidans* que sí presentó una alta abundancia en estas muestras. Por lo tanto, podemos especular que la secuencia original que dio origen a estos *reads* en realidad corresponde a *Achromobacter xylosoxidans* y probablemente durante el proceso de amplificación por PCR ocurrió un error en esta secuencia que hizo que parte de los *reads* fueran asignados a *Bordetella trematum* en lugar de *Achromobacter xylosoxidans*.

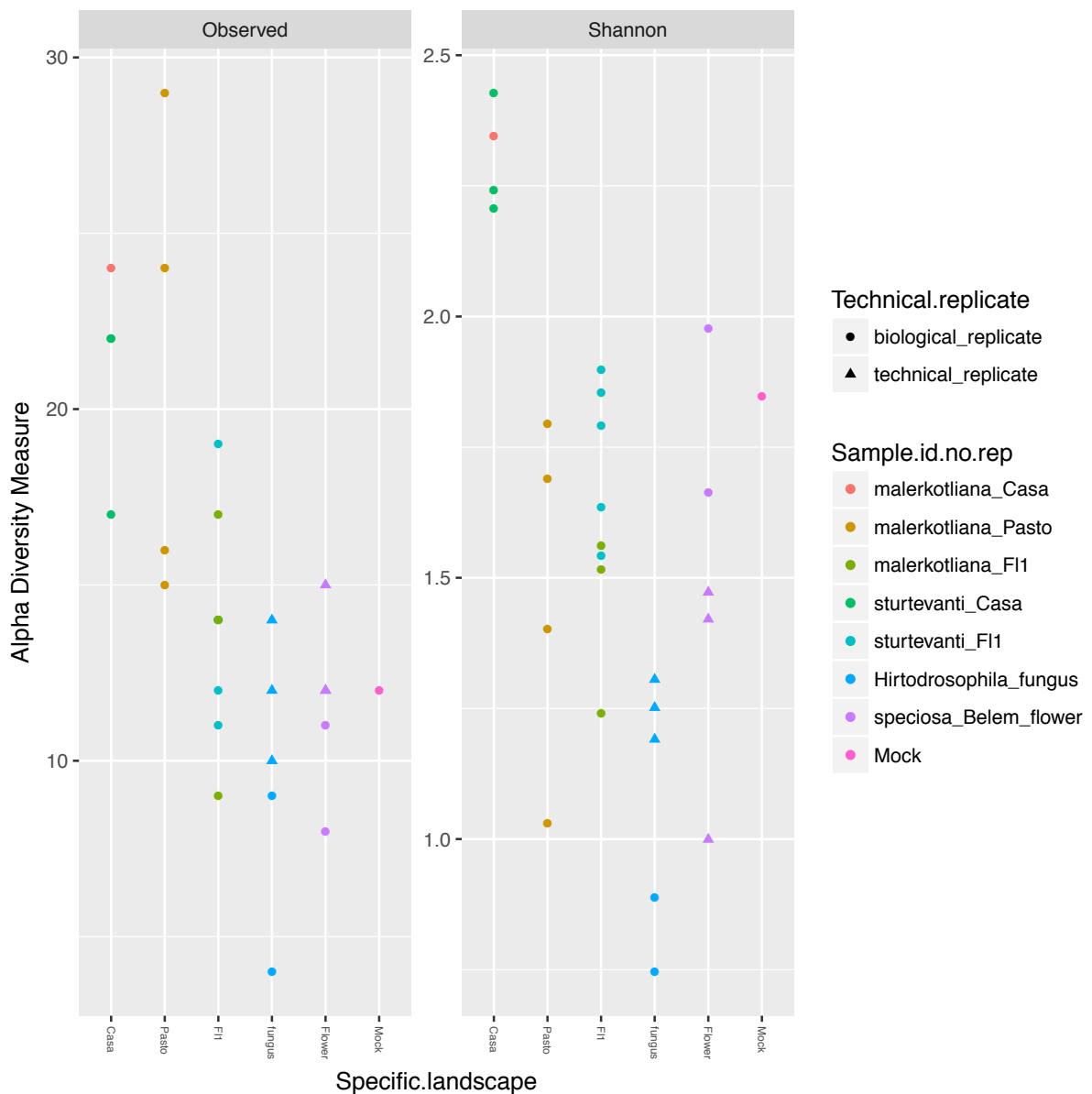


Figura 6. Diversidad alfa para cada muestra. Las muestras están agrupadas por el ambiente en que fueron colectadas. Las réplicas técnicas se indican con un triángulo. **Izquierda:** Riqueza observada (**Observed**). **Derecha:** Índice de Shannon (**Shannon**).

Antes de proceder al análisis de diversidad, realizamos un filtrado para eliminar las especies poco abundantes ya que éstas no son informativas y pueden añadir ruido al análisis. Como criterio de filtrado, teniendo en cuenta el error obtenido en la *Mock community* y utilizando los valores de abundancia proporcional al número de *reads* totales en cada muestra, retuvimos sólo las especies cuya abundancia sumada entre todas las muestras fuera de al menos 0.0074% de la abundancia total. Luego de este filtrado el número de especies totales se redujo a 49 y el número de familias a 20 (sin contar las especies y familias únicamente encontradas en la *Mock community*, que fueron 9 y 6, respectivamente). La figura 6 muestra la riqueza para cada muestra, luego del filtrado, medida como número de especies observadas y con el índice de diversidad de Shannon. En general, observamos que el índice Shannon varía menos entre réplicas biológicas y técnicas que el número de especies observado.

Para comparar la riqueza entre diferentes muestras, tomando la riqueza Shannon en réplicas biológicas, realizamos Test-T de a pares, ajustando los *p-values* por múltiples comparaciones con el método de Benjamini y Hochberg. Excluimos de este análisis a las muestras "malerkotliana_Casa" y "Mock" por no contar con réplicas biológicas para éstas. La especie *D. sturtevantii* de ambiente modificado "Casa" presenta una riqueza significativamente mayor en comparación con las demás muestras (*p-val* <0.05 para todas las comparaciones). Cabe aclarar que la muestra "malerkotliana_Casa" presenta una riqueza similar a la riqueza promedio de *D. sturtevantii* (2.36 y 2.3 respectivamente), si bien no fue incluida en las comparaciones por no poseer réplicas.

Del resto de las comparaciones, "speciosa_Belem_flower" y "Hirtodrosophila_fungus" mostraron diferencias significativas en el nivel de riqueza (1.67 en promedio contra 1.03 en promedio, respectivamente, *p-val*=0.021) y

Tabla 1. Comparaciones T-test de nivel de riqueza Shannon entre pares de muestras.

	Hirtodrosophila_fungus	malerkotliana_F11	malerkotliana_Pasto	speciosa_Belem_flower	sturtevantii_Casa
malerkotliana_F11	0.11119				
malerkotliana_Pasto	0.06381	0.82986			
speciosa_Belem_flower	0.02102*	0.34625	0.39131		
sturtevantii_Casa	0.00031*	0.00458*	0.00458*	0.02102*	
sturtevantii_F11	0.00582*	0.18074	0.19581	0.75604	0.02102*

"sturtevant_i_F11" presenta en promedio una riqueza mayor comparada con "Hirtodrosophila_fungus" (1.75 contra 1.03 en promedio, $p\text{-val}=0.005$). Las demás comparaciones no presentaron diferencias significativas (Tabla 1). Estas

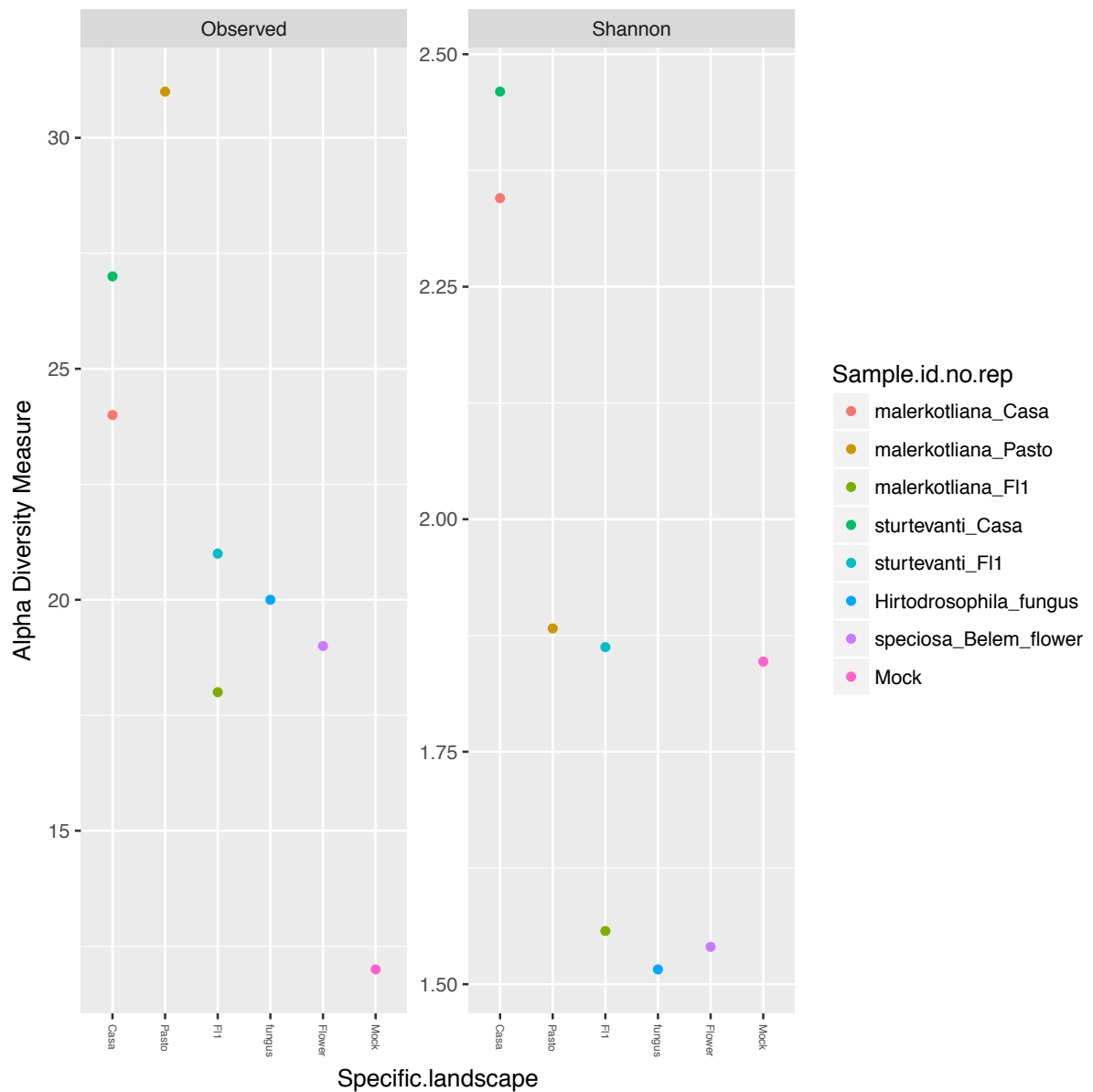


Figura 7. Diversidad alfa para cada muestra tomando un promedio entre réplicas técnicas y biológicas. Las muestras están agrupadas por el ambiente en que fueron colectadas **Izquierda:** Riqueza observada (**Observed**). **Derecha:** Índice de Shannon (**Shannon**).

observaciones preliminares indican que, como es de esperar, la riqueza de especies bacterianas es mayor en el ambiente modificado "Casa", donde las moscas están expuestas a mayor variedad de substratos, mientras que en ambientes sin modificar

la riqueza es un poco menor, existiendo diferencias en los niveles de riqueza asociados a diferentes substratos.

En la figura 7 se muestra la riqueza observada para cada especie de drosofilídeo en cada ambiente tomando el promedio de la abundancia de cada especie bacteriana entre todas las réplicas biológicas y técnicas en cada caso.

Análisis de composición bacteriana

Las especies dominantes son un buen parámetro para comparar las diferentes muestras basadas en su diversidad alfa, en general, las réplicas biológicas deberían compartir las mismas especies dominantes. Definimos el número de especies dominantes como el número de especies bacterianas que contribuyen al 80% de los *reads* observados en cada muestra. Vemos que el número de especies dominantes observadas en cada muestra presenta una correlación muy escasa con el número de *reads* totales (spearman= 0.077, p-val=0.7) y el número de especies dominantes en los ambientes selváticos (“FI1”, “Flower” y “fungus”) es bajo (valor medio 3.28

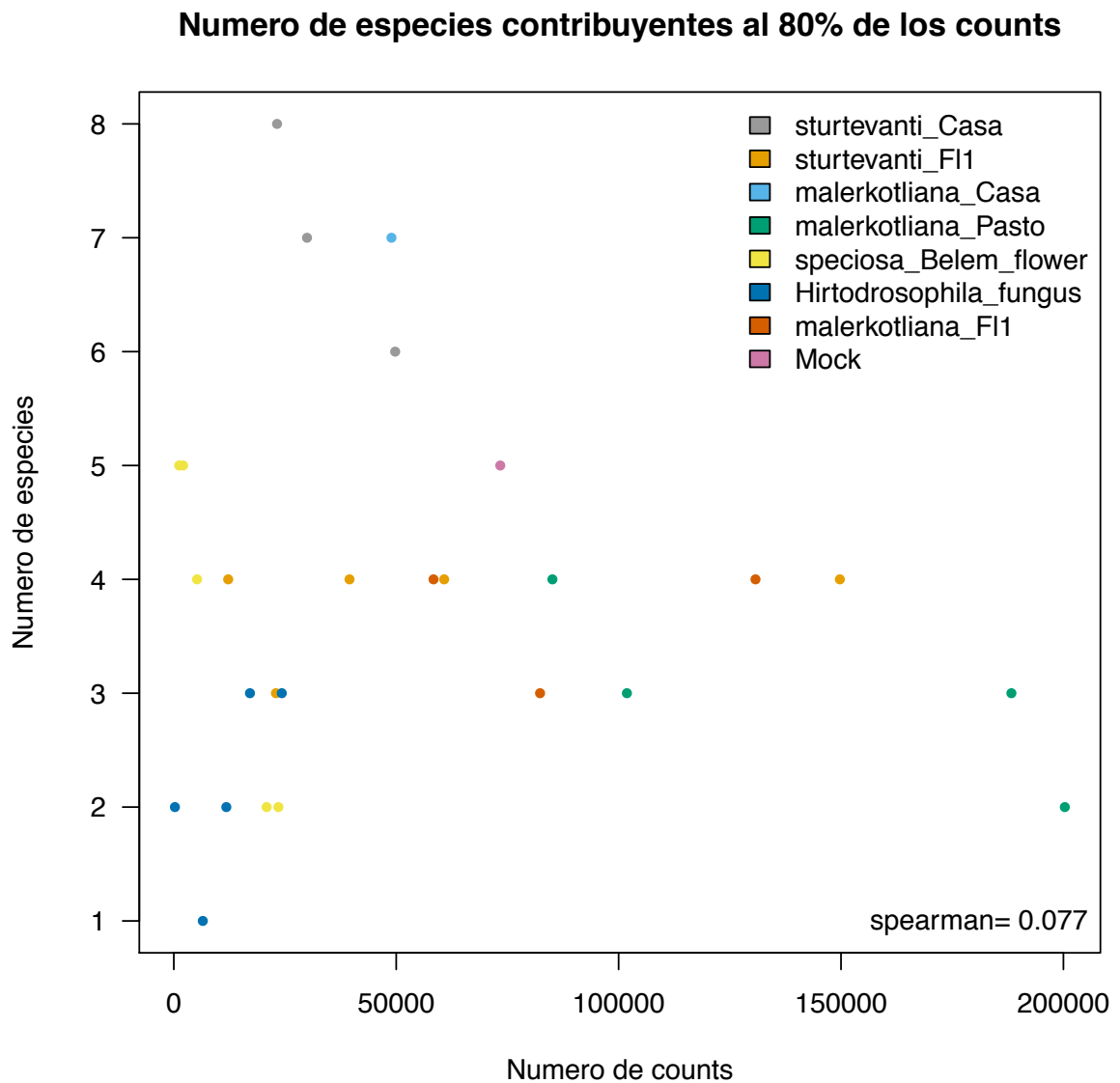


Figura 8. Correlación entre el número de reads totales y el número de especies dominantes en cada muestra.

especies) comparado con el ambiente modificado “Casa” (6.75 especies en promedio, Test-t, p-val=0.0014, Figura 8). Por otro lado, no se observó diferencia significativa entre el número de especies dominantes encontradas en el otro ambiente modificado, “Pasto” (valor medio 3) y las regiones selváticas (Test-t, p-val=0.59).

Si bien las especies bacterianas encontradas en *D. malerkotliana* y *D. sturtevantii*, ambas asociadas a “frutos”, tanto en ambientes modificados como selváticos, son similares a nivel de riqueza, las abundancias relativas de estas especies son distintivas de los diferentes ambientes. En la tabla 2 se muestran las cinco especies más abundantes para cada muestra, y en las figuras 9 y 10 se muestra la proporción de especies y familias bacterianas en cada muestra, respectivamente (por simplicidad, en estas figuras se muestran sólo las especies o familias que son dominantes en al menos una muestra). En *D. malerkotliana* encontrada en los ambientes modificados representados por las muestras de “Casa” y “Pasto”, la especie bacteriana dominante es *Acetobacter senegalensis*. Sin embargo, entre las réplicas biológicas de esta muestra, la muestra #25 de *D. malerkotliana* (Pasto), constituye una excepción, ya que presenta una clara dominancia (90% de las secuencias) de las especies *Providencia Rettgeri* y *Morganella morganii*, ambas de la familia Morganellaceae. Estas especies bacterianas aparecen en baja frecuencia y con una abundancia despreciable en otras muestras (Anexo Tabla A3). Otra especie dominante en estas muestras es *Acetobacter persici* y adicionalmente, la muestra de “Casa” presenta una abundancia importante de especies de las familias Lactobacillaceae (como *Lactobacillus brevis*) y

Leuconostocaceae (como *Leuconostoc mesenteroides* y *Weissella cibaria*). En este mismo ambiente de “Casa” estas últimas dos familias constituyen las familias dominantes en *D. sturtevantii* y también son abundantes en esta muestra, aunque en

Tabla 2. Cinco especies más abundantes en cada muestra

	Especie 1	Especie 2	Especie 3	Especie 4	Especie 5
Mock	<i>Helicobacter pylori</i>	<i>Streptococcus mutans</i>	<i>Rhodobacter sphaeroides</i>	<i>Lactobacillus gasserii</i>	<i>Staphylococcus epidermidis</i>
29 speciosa flower 5	<i>Lactococcus lactis</i>	<i>Pantoea vagans</i>	<i>Lactobacillus brevis</i>	<i>Cedecea neteri</i>	<i>Acinetobacter soli</i>
28 speciosa flower 3	<i>Klebsiella oxytoca</i>	<i>Lactobacillus brevis</i>	<i>Pantoea sp. PSNIH1</i>	<i>Cedecea neteri</i>	<i>Klebsiella pneumoniae</i>
27 c speciosa flower 2	<i>Tatumella citrea</i>	<i>Kosakonia cowanii</i>	<i>Klebsiella oxytoca</i>	<i>Pantoea vagans</i>	<i>Pantoea ananatis</i>
27 b speciosa flower 2	<i>Tatumella citrea</i>	<i>Klebsiella oxytoca</i>	<i>Pantoea vagans</i>	<i>Cedecea neteri</i>	<i>Pantoea sp. PSNIH1</i>
27 a speciosa flower 2	<i>Tatumella citrea</i>	<i>Klebsiella oxytoca</i>	<i>Pantoea vagans</i>	<i>Cedecea neteri</i>	<i>Pantoea ananatis</i>
32 Hirtodrosophila fungus FI2_5	<i>Cedecea neteri</i>	<i>Sphingomonas taxi</i>	<i>Pandoraea vervacti</i>	<i>Bradyrhizobium sp.</i>	<i>Acetobacter persici</i>
31 Hirtodrosophila fungus FI2_3	<i>Achromobacter xylosoxidans</i>	<i>Chryseobacterium indologenes</i>	<i>Pandoraea vervacti</i>	<i>Sphingomonas taxi</i>	<i>Stenotrophomonas maltophilia</i>
30 c Hirtodrosophila fungus FI2_2	<i>Cedecea neteri</i>	<i>Achromobacter xylosoxidans</i>	<i>Bordetella trematum</i>	<i>Morganella morganii</i>	<i>Chryseobacterium indologenes</i>
30 b Hirtodrosophila fungus FI2_2	<i>Cedecea neteri</i>	<i>Lactococcus lactis</i>	<i>Achromobacter xylosoxidans</i>	<i>Chryseobacterium indologenes</i>	<i>Morganella morganii</i>
30 a Hirtodrosophila fungus FI2_2	<i>Cedecea neteri</i>	<i>Lactococcus lactis</i>	<i>Achromobacter xylosoxidans</i>	<i>Morganella morganii</i>	<i>Chryseobacterium indologenes</i>
16 sturtevantii FI1_12	<i>Gluconobacter albidus</i>	<i>Acetobacter persici</i>	<i>Gluconobacter oxydans</i>	<i>Acetobacter senegalensis</i>	<i>Lactobacillus brevis</i>
15 sturtevantii FI1_7	<i>Gluconobacter oxydans</i>	<i>Gluconobacter albidus</i>	<i>Acetobacter persici</i>	<i>Acetobacter senegalensis</i>	<i>Lactobacillus heilongjiangensis</i>
14 sturtevantii FI1_5	<i>Gluconobacter albidus</i>	<i>Gluconobacter oxydans</i>	<i>Acetobacter persici</i>	<i>Acetobacter senegalensis</i>	<i>Weissella cibaria</i>
13 sturtevantii FI1_3	<i>Gluconobacter oxydans</i>	<i>Gluconobacter albidus</i>	<i>Acetobacter persici</i>	<i>Acetobacter senegalensis</i>	<i>Weissella cibaria</i>
12 sturtevantii FI1_2	<i>Gluconobacter albidus</i>	<i>Acetobacter persici</i>	<i>Gluconobacter oxydans</i>	<i>Acetobacter senegalensis</i>	<i>Lactobacillus plantarum</i>
11 sturtevantii Casa 5	<i>Leuconostoc mesenteroides</i>	<i>Weissella cibaria</i>	<i>Acetobacter senegalensis</i>	<i>Lactobacillus brevis</i>	<i>Enterococcus wangshanyuanii</i>
10 sturtevantii Casa 3	<i>Lactobacillus heilongjiangensis</i>	<i>Lactobacillus brevis</i>	<i>Lactobacillus fermentum</i>	<i>Enterococcus wangshanyuanii</i>	<i>Leuconostoc mesenteroides</i>
9 sturtevantii Casa 2	<i>Lactobacillus brevis</i>	<i>Acetobacter senegalensis</i>	<i>Leuconostoc mesenteroides</i>	<i>Lactobacillus fermentum</i>	<i>Weissella cibaria</i>
40 malerkotliana FI1_5	<i>Gluconobacter albidus</i>	<i>Acetobacter persici</i>	<i>Gluconobacter oxydans</i>	<i>Acetobacter senegalensis</i>	<i>Weissella cibaria</i>
39 malerkotliana FI1_3	<i>Gluconobacter albidus</i>	<i>Gluconobacter oxydans</i>	<i>Acetobacter persici</i>	<i>Acetobacter senegalensis</i>	<i>Lactobacillus brevis</i>
38 malerkotliana FI1_2	<i>Gluconobacter oxydans</i>	<i>Gluconobacter albidus</i>	<i>Acetobacter senegalensis</i>	<i>Weissella cibaria</i>	<i>Acetobacter persici</i>
26 malerkotliana Pasto 7	<i>Acetobacter senegalensis</i>	<i>Acetobacter persici</i>	<i>Gluconobacter oxydans</i>	<i>Providencia rettgeri</i>	<i>Weissella cibaria</i>
25 malerkotliana Pasto 5	<i>Providencia rettgeri</i>	<i>Morganella morganii</i>	<i>Acetobacter senegalensis</i>	<i>Acetobacter persici</i>	<i>Gluconobacter oxydans</i>
24 malerkotliana Pasto 3	<i>Acetobacter senegalensis</i>	<i>Acetobacter persici</i>	<i>Gluconobacter oxydans</i>	<i>Gluconobacter albidus</i>	<i>Providencia rettgeri</i>
23 malerkotliana Pasto 2	<i>Acetobacter senegalensis</i>	<i>Acetobacter persici</i>	<i>Gluconobacter oxydans</i>	<i>Gluconobacter albidus</i>	<i>Lactobacillus brevis</i>
22 malerkotliana Casa 2	<i>Acetobacter senegalensis</i>	<i>Lactobacillus brevis</i>	<i>Acetobacter persici</i>	<i>Gluconobacter oxydans</i>	<i>Leuconostoc mesenteroides</i>

menor medida, bacterias de la familia *Acetobacteraceae*, como *Acetobacter senegalensis* y *Gluconobacter oxydans*.

Por su parte, las muestras de *D. malerkotliana* y *D. sturtevantii* de ambientes selváticos presentan una composición bacteriana muy similar entre sí (distancia Bray-Curtis= 0.13), con una clara dominancia de la familia Acetobacteraceae. A nivel de especie, son dominantes en estas muestras las especies *Gluconobacter albidus* y *Gluconobacter oxydans*, y en menor proporción, aunque también abundantes, se encuentran las especies del género *Acetobacter*, *Acetobacter persici* y *Acetobacter senegalensis*.

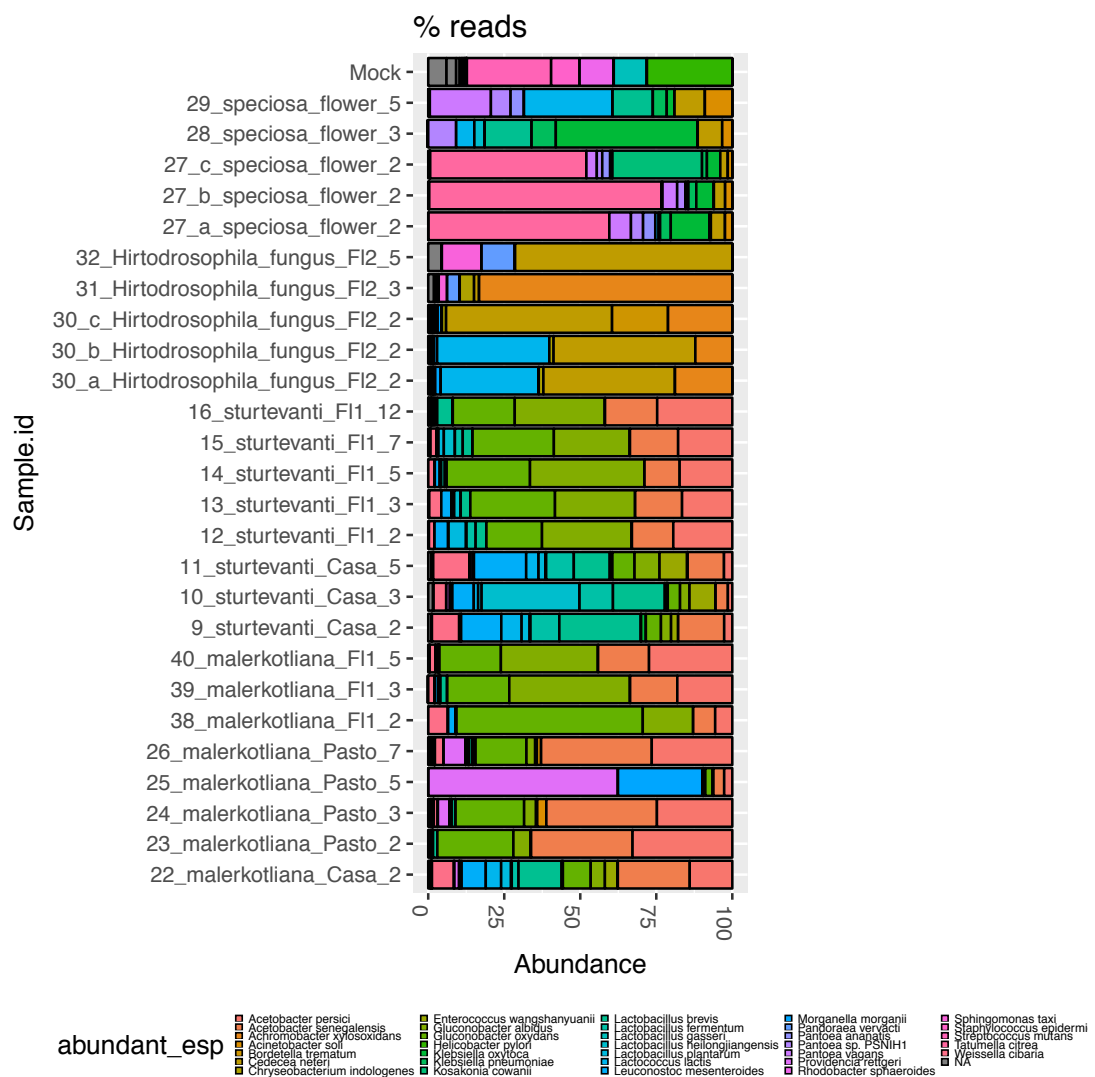


Figura 9. Composición de especies bacterianas en cada muestra. La abundancia se muestra en forma de porcentaje. Aquellas especies que no son dominantes para ninguna muestra se indican en gris (NA).

Para las muestras provenientes de sustratos de “Flores” y “Hongos”, la riqueza es de forma general menor en comparación con las especies provenientes del sustrato “Frutos”, pero se observan mayores variaciones entre muestras (distancia de Bray-Curtis media entre réplicas biológicas: 0.65 para *Hirtodrosophila* y 0.56 para *D. speciosa*). Para *D. speciosa*, las especies dominantes varían entre réplicas biológicas (#27, #28 y #29). Para las tres réplicas técnicas de esta muestra (#27A, #27B y #27C) la especie dominante es *Tatumella citrea*. Para las muestras #28 y #29, *Klebsiella oxytoca* y *Lactococcus lactis* respectivamente. La especie *Klebsiella oxytoca* es también relativamente abundante en la muestra #27, mientras que las especies *Lactobacillus brevis*, *Cedecea neteri* y especies del género *Pantoea* también se encuentran con una abundancia relativamente alta en al menos dos de las réplicas biológicas. La única diferencia cuantitativa entre las réplicas #27, #28 y #29 es el número de individuos usados para el pool de moscas, siendo de 2, 3 y 5 respectivamente. A nivel de familias, la familia dominante en las muestras #27 y #29 es Erwiniaceae. A esta familia pertenecen especies de los géneros *Pantoea* y *Tatumella*. La familia dominante en #28 (y segunda familia dominante en #27) es Enterobacteriaceae, familia que incluye los géneros *Klebsiella* y *Cedecea*.

En el caso de la composición bacteriana de muestras provenientes del sustrato “Hongos” representada por la mosca *Hirtodrosophila*, la especie más abundante en dos de las tres réplicas biológicas (muestras #30 y #32) es *Cedecea neteri* (familia Enterobacteriaceae). Mientras que *Achromobacter xylosoxidans* (familia Alcaligenaceae), es la especie dominante en la muestra #31 y la segunda especie dominante en la muestra #30. Notablemente, *Lactococcus lactis* (familia Streptococcaceae) es muy abundante en sólo dos de las réplicas técnicas de la muestra #30A y #30B. *Bordetella trematum* (familia Alcaligenaceae), como fue

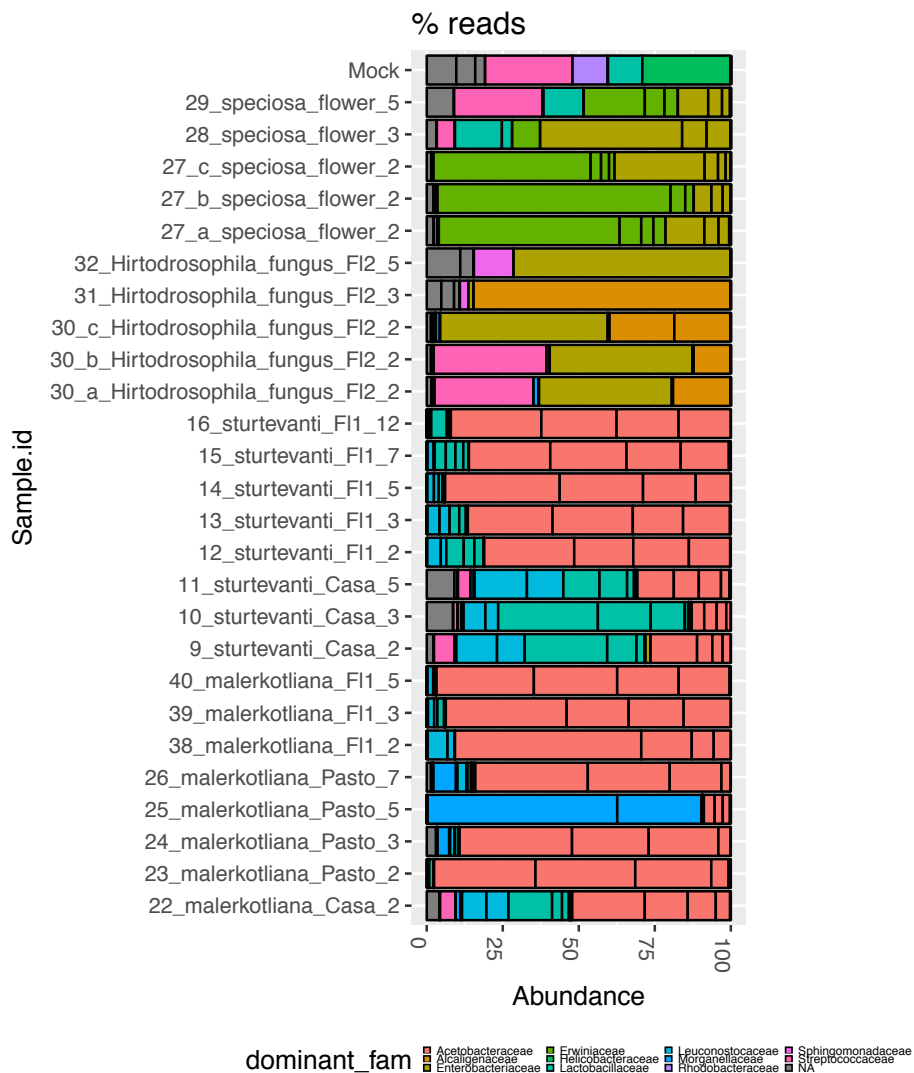


Figura 10. Composición de familias bacterianas en cada muestra. La abundancia se muestra en forma de porcentaje. Aquellas familias que no son dominantes para ninguna muestra se indican en gris (NA).

mencionado anteriormente, se encuentra únicamente en la otra réplica técnica #30A y #30C y en alta abundancia.

Análisis de diversidad beta

Para responder a nuestro tercer objetivo, el de evaluar cómo los cambios permanentes al paisaje selvático influyen en la diversidad microbiana de moscas neotropicales, realizamos un análisis de diversidad entre las muestras encontradas en los diferentes ambientes (diversidad beta).

Como se mencionó en la sección “Materiales y métodos”, calculamos las distancias entre pares de muestras utilizando tres métodos diferentes: las versiones

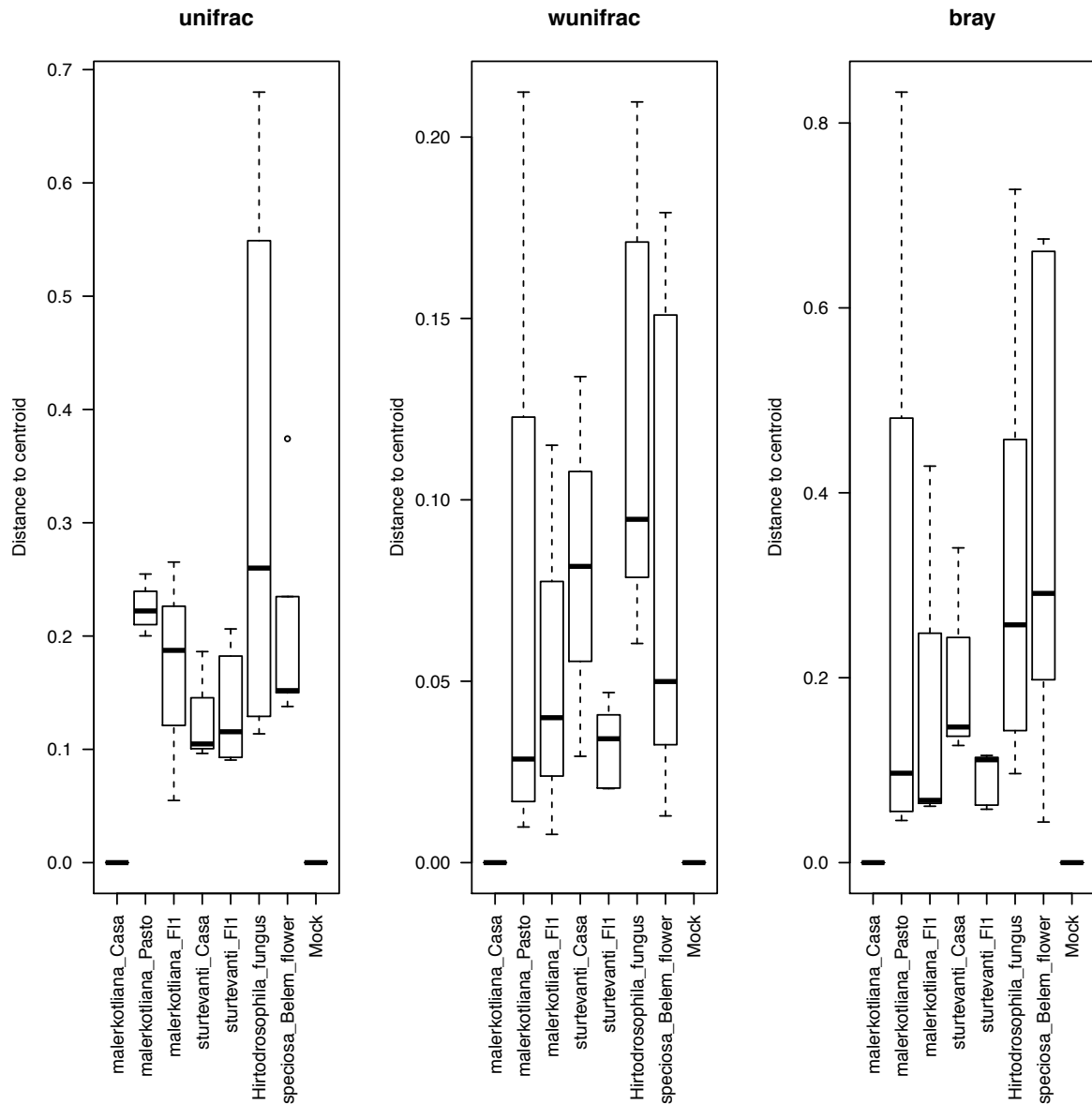


Figura 11. Análisis de la homogeneidad entre réplicas para cada especie *Drosophilidae* en cada ambiente. Se muestran gráficos de caja con la distribución de distancias UniFrac unweighted (**unifrac**), UniFrac weighted (**wunifrac**) y Bray-Curtis (**bray**) a la media espacial de cada grupo.

cuantitativas y cuantitativas de UniFrac, UniFrac *weighted* y *unweighted*, respectivamente y el método cuantitativo Bray-Curtis.

Una manera de analizar la homogeneidad dentro de cada grupo es el procedimiento implementado en la función *betadisper* del paquete *vegan*. Esta función

calcula la distancia promedio de los miembros del grupo al centroide del grupo o media espacial. Esto da una idea de la variabilidad existente dentro de cada grupo y entre réplicas. En la figura 11 se muestra para cada huésped y en cada ambiente la distribución de las distancias entre cada réplica biológica y técnica al centroide. En general, vemos que la variabilidad observada entre réplicas es mayor para las moscas asociadas a flores y hongos que para moscas provenientes de frutos. También vemos que, dependiendo del método de distancia, las relaciones entre las muestras varían. En todos los casos la distancia entre réplicas a la media parece ser mayor para el método cualitativo, UniFrac, que para los métodos cuantitativos. Esto se debe a que UniFrac es sensible a diferencias en especies poco frecuentes. Cuando se tiene en cuenta la abundancia, la distancia al promedio de todas las muestras se vuelve menor.

Para probar si uno o más grupos son más variables que otros, llevamos a cabo un análisis ANOVA de las distancias a los centroides. Esto es fácilmente implementado en R al aplicar la función *anova* a los resultados de *betadisper* para cada método de distancia. Según estos *tests*, no existen diferencias significativas en las variancias entre los distintos grupos para ninguno de los tres métodos de distancia utilizados (p-val: 0.1452 para UniFrac, 0.3774 para UniFrac *weighted* y 0.5321 para Bray-Curtis). Esto implica que la homogeneidad en cada ambiente (y o sustrato) y dentro de cada especie de mosca es similar entre ambientes y moscas. Para analizar visualmente cómo las muestras se agrupan en base a las distancias calculadas aplicamos el método de escalado multidimensional no métrico (NMDS por sus siglas en inglés *non-metric multidimensional scaling*) implementado en el paquete *vegan*. En la figura 12 se muestran los resultados de aplicar método de reducción dimensional sobre las matrices de distancia calculadas previamente.

En todos los casos se observan grupos distintivos por tipo de sustrato, “fruto”, “flor” y “hongo”. Notablemente, el método cualitativo parece agrupar mejor a las moscas por tipo de sustrato que los métodos cuantitativos. Esto se evidencia especialmente en el caso de las moscas asociadas a flores y hongos. En el caso de los métodos que tienen en cuenta la abundancia, se observa que mientras las réplicas técnicas siguen apareciendo cercanas, el agrupamiento de las réplicas biológicas se vuelve más disperso. Es importante tener en cuenta que para estas muestras el número de secuencias totales fue muy bajo para la mayoría de las réplicas, por lo que

no se podría confiar en que la abundancia obtenida para estas muestras sea una buena aproximación de la abundancia real. Por otro lado, si tenemos en cuenta sólo la ocurrencia de especies bacterianas, observamos claramente que la composición microbiana de moscas asociadas a flores y hongos es muy diferente a la de moscas asociadas a la fruta. Por lo tanto, el sustrato en el cual la mosca crece y se alimenta de levaduras tiene un papel importante en la determinación de esta microbiota específica, es decir las comunidades que provienen de moscas asociadas a flores

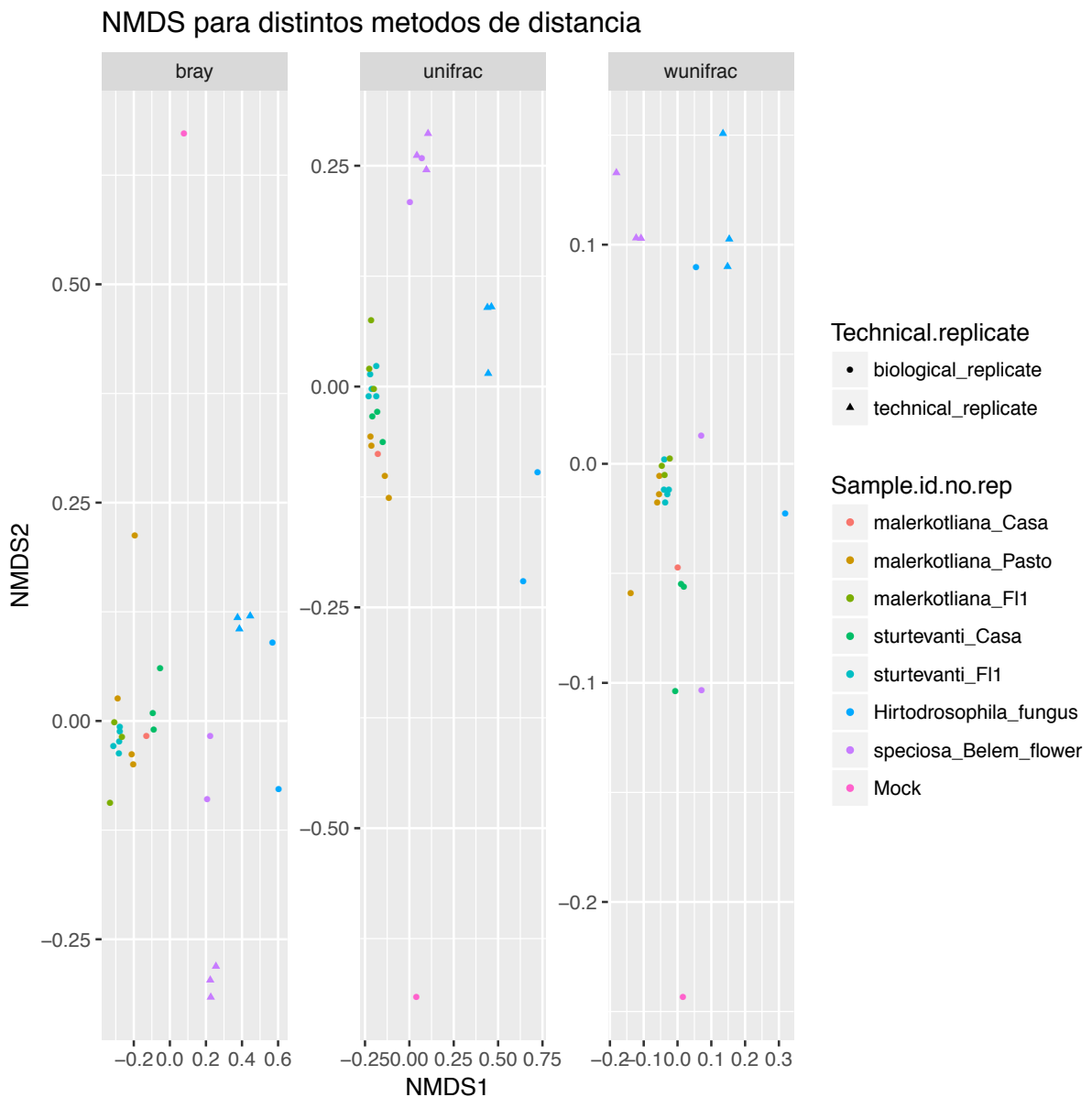


Figura 12. Agrupamiento de todas las muestras analizadas en base a las dos primeras componentes de NMDS aplicado sobre matrices de distancia calculadas con los diferentes métodos: Bray-Curtis (**bray**), UniFrac weighted (**unifrac**) y UniFrac unweighted (**wunifrac**).

tienen bacterias específicas de flores como *Tatumella citrea*, género que ya fue descrito en asociación a moscas de flores y en el caso de hongos, *Cedecea neteri*.

Debido a que nuestro principal objetivo es comparar las especies *D. malerkotliana* (cosmopolita) y *D. sturtevantii* (Neotropical) encontradas en ambientes modificados y de selva primaria, decidimos hacer un análisis más detallado teniendo en cuenta sólo estas especies. Para esto, eliminamos de la matriz de abundancia las especies de moscas asociadas a flores y hongos y la *Mock community* y volvimos a

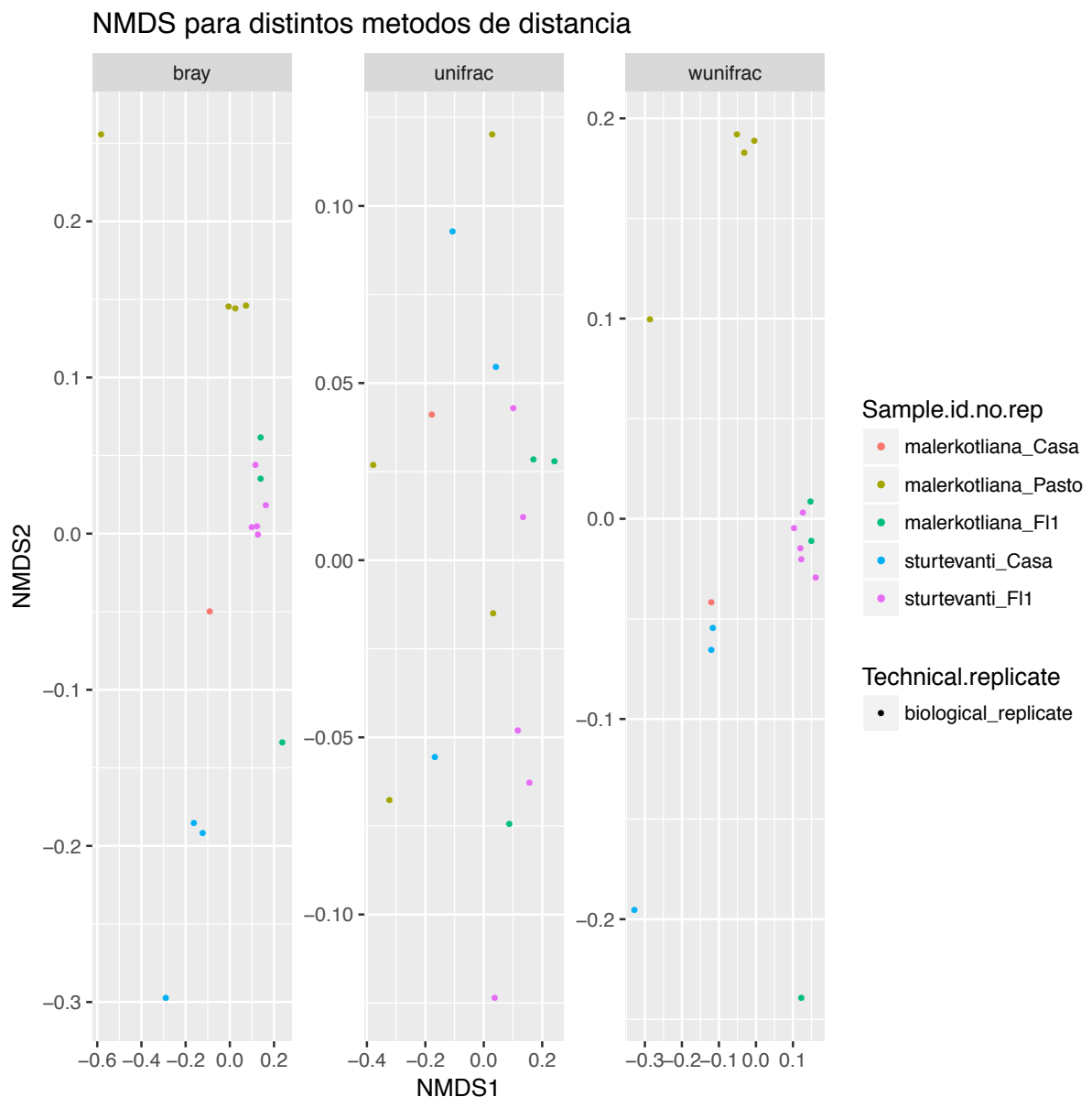


Figura 13. Agrupamiento de muestras de *D. malerkotliana* y *D. sturtevantii* en base a las dos primeras componentes de NMDS aplicado sobre matrices de distancia calculadas con los diferentes métodos: Bray-Curtis (**bray**), UniFrac weighted (**unifrac**) y UniFrac unweighted (**wunifrac**).

filtrar las especies bacterianas con el mismo criterio utilizado anteriormente, lo que resultó en un número final de 36 especies bacterianas provenientes de 15 familias.

En la figura 13 se muestra el resultado de aplicar el método NMDS sobre las distancias recalculadas para la nueva matriz de abundancia reducida a moscas asociadas a fruta. En este caso, a diferencia de lo que observamos para especies asociadas a la fruta, los métodos cuantitativos parecen agrupar mejor que el método cualitativo a las moscas de acuerdo al ambiente en el que se encontraron. Las especies encontradas en ambientes de selva primaria se agrupan muy cercanamente, y no parece haber una diferencia notable entre las dos especies de moscas, *D. malerkotliana* y *D. sturtevantii* en este ambiente específico. Para analizar en mayor detalle las diferencias y similitudes entre los distintos grupos de moscas, realizamos un análisis de la variancia multivariado permutacional (PERMANOVA) en base a las matrices de distancia, implementado en el paquete *vegan* con la función *adonis2*. Los resultados de este análisis fueron significativos para las tres medidas de distancia, UniFrac (p-val= 0.008), UniFrac weighted (p-val= 0.046) y Bray-Curtis (p-val= 0.002), para la combinación especie de drosófila (*D. malerkotliana* y *D. sturtevantii*) con ambiente específico (modificados “Casa” y “Pasto” y sin modificar “F11”). Esto implica que existen diferencias entre los grupos analizados. A continuación (Tabla 3) se muestra el resultado de esta función cuando agregamos en la fórmula la interacción entre ambiente específico y especie:

Tabla 3. Resultados de análisis PERMANOVA comparando todas las muestras de *D. malerkotliana* y *D. sturtevantii* en los diferentes ambientes

	Df	Sums of Sqs	MeanSqs	F.Mode	R2	Pr(>F)
Specific.landscape	2	1.35304	0.67652	8.1188	0.56947	0.001 ***
Species	1	0.03759	0.03759	0.4511	0.01582	0.797
Specific.landscape:Specie	1	0.06873	0.06873	0.8249	0.02893	0.469
Residuals	11	0.91660	0.08333	0.38578	2.37596	1

Observamos que, si bien hay diferencia significativa entre los distintos ambientes, no hay diferencias significativas a nivel de especie de mosca ni en la interacción entre especie y ambiente. Para analizar en mayor detalle entre qué grupos se encuentran diferencias, realizamos el mismo análisis tomando los distintos grupos de a pares. En la tabla se muestran los resultados de las comparaciones entre grupos de moscas por ambiente, obtenidos con la distancia Bray-Curtis. Los resultados para las otras dos medidas de distancia fueron similares por lo que no se muestran. Se indican con (*) los casos donde se observaron diferencias significativas con un nivel de significación del 0.05. Los resultados indican que *D. sturtevantii* es diferente a *D. malerkotliana* proveniente de “pasto” y también a *D. sturtevantii* proveniente de muestras de “casa” (Tabla 4).

Tabla 4. P-values obtenidos para las comparaciones de a pares entre todas las muestras de *D. malerkotliana* y *D. sturtevantii* en los diferentes ambientes.

	D.MALERKOTLI ANA/ PASTO	D.MALERKOTLI ANA /SELVA PRIMARIA	D.STURTEVAN TI/ CASA	D.STURTEVANTI / SELVA PRIMARIA
D. MALERKOTLIANA/ CASA	0,400	0,250	0,500	0,166
D.MALERKOTLIAN A /PASTO	NA	0.083	0,028*	0,005*
D.MALERKOTLIAN A/FL1	NA	NA	0.1	0.457
D.SURTEVANTI/CA SA	NA	NA	NA	0,015*

(*) diferencia significativa

En la tabla 5 se muestran los resultados obtenidos a comparar ambientes enteros (tomando todas las moscas independientemente de la especie). Los resultados

indican que las moscas provenientes de los ambientes modificados difieren en su diversidad bacteriana respecto a las que provienen de la selva primaria.

Tabla 5. P-values obtenidos para las comparaciones entre los ambientes “Pasto”, “Casa”, “FI1” y “Modificado” (“Pasto” y “Casa” tomados en conjunto).

	Pasto	Casa	Modificado
Selva primaria	0.005*	0.006*	0.001*
Pasto	NA	0.023*	0.354
Casa	NA	NA	0.273

(*) diferencia significativa

Finalmente, realizamos un agrupamiento en base a las distancias calculadas, utilizando el método jerárquico de Ward. En la figura 14 se muestra el *heatmap* correspondiente a la distancia Bray-Curtis. Los grupos obtenidos con las otras medidas de distancia fueron los mismos (no se muestra). Al observar este gráfico se evidencia que el agrupamiento es gobernado por el tipo ambiente y no por el tipo de especie de drosófila. También pueden observarse algunas diferencias más finas entre los distintos ambientes. Por ejemplo, si bien se observa que las especies de los géneros *Acetobacter* y *Gluconobacter* son especies dominantes para todas las moscas, en el *heatmap* puede verse cómo la abundancia relativa entre éstas cambia entre moscas de ambientes selváticos (dónde *Gluconobacter* es más abundante) y moscas de ambientes modificados (dónde *Acetobacter* es más abundante).

En la figura 15 se muestra el árbol filogenético de especies bacterianas encontradas en *D. malerkotliana* y *D. sturtevantii*. Por simplicidad se removieron las especies muy poco abundantes (cuya abundancia entre todas las muestras sumaba menos del 0.075%).

La mayoría de las especies abundantes se encuentran en más de una de las réplicas de todos los grupos. En *D. malerkotliana* de “Casa”, para la cual contamos con una sola réplica, se encontraron individuos de casi todas estas especies bacterianas, excepto *Lactobacillus crustorum*, que se encontró únicamente en *D. sturtevantii* de “Casa”. Las especies *Acinetobacter soli* y *Stenotrophomonas maltophilia*

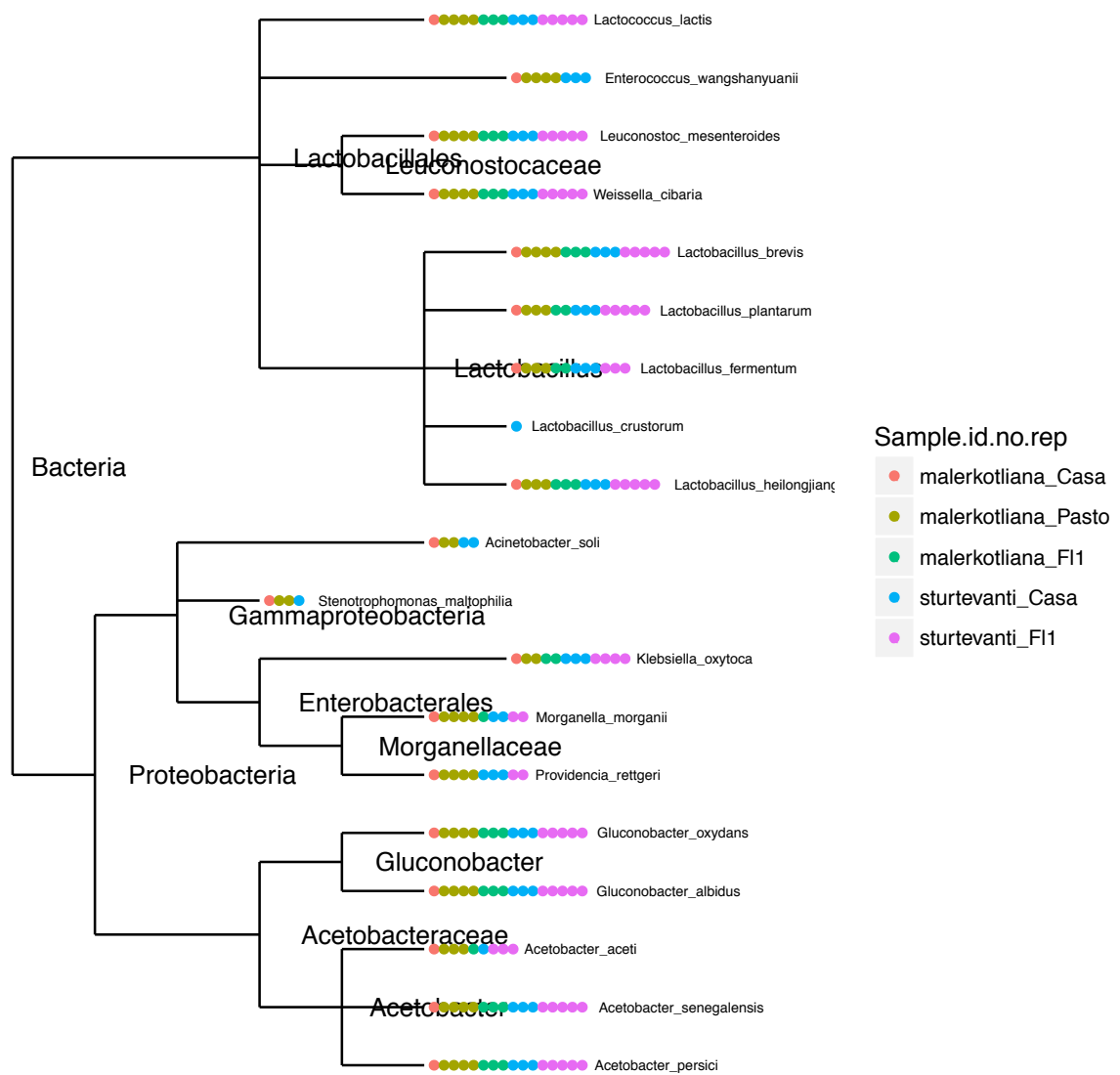


Figura 15. Árbol filogenético para las especies bacterianas más abundantes halladas en *D. malerkotliana* y *D. sturtevantii* en los diferentes ambientes analizados. Cada círculo corresponde a una réplica individual del grupo indicado por el código de color.

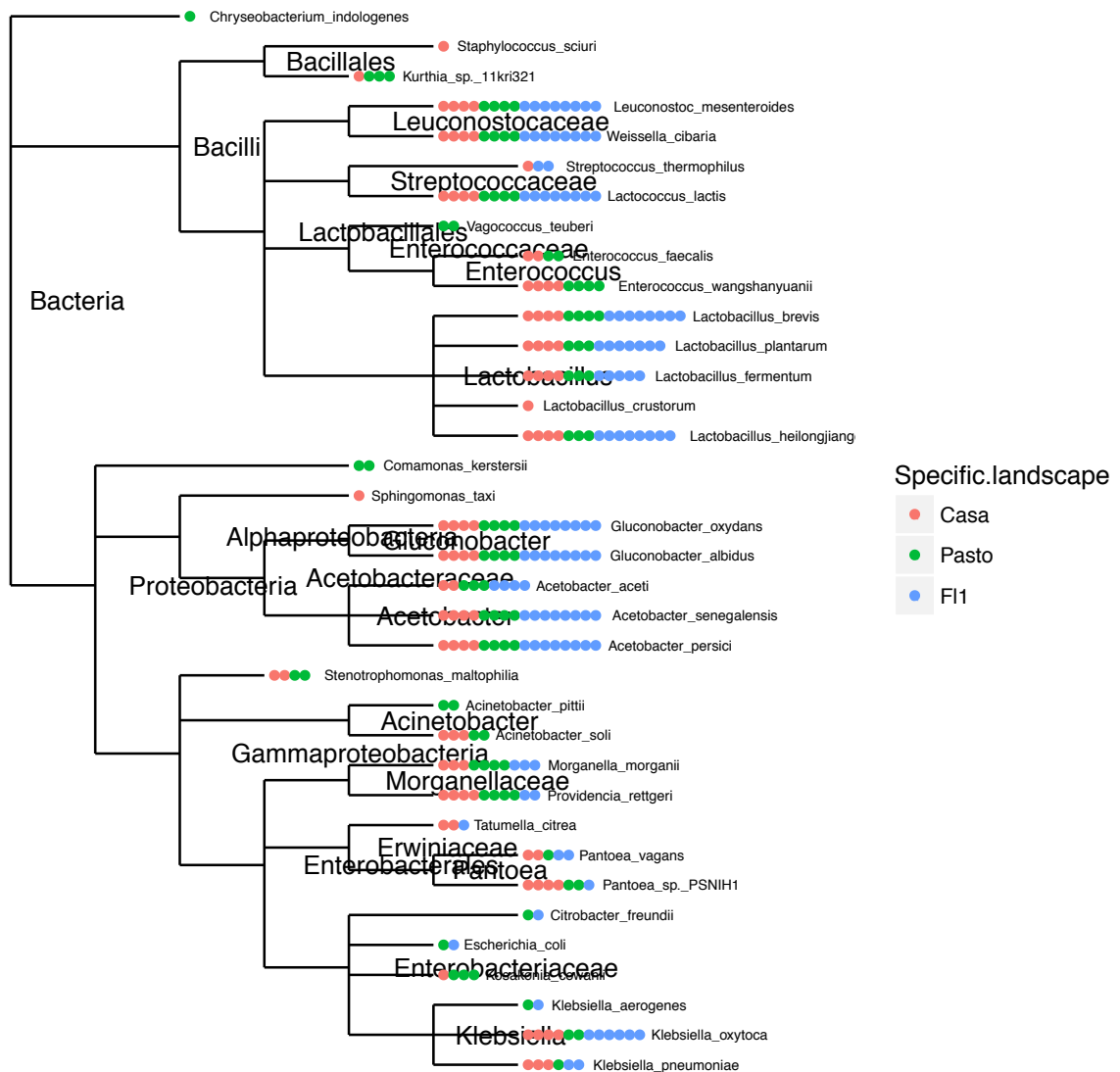


Figura 16. Árbol filogenético para todas las especies halladas en *D. malerkotliana* y *D. sturtevantii* en los diferentes ambientes analizados. Cada círculo corresponde a una réplica individual de uno de los ambientes (“Casa”, “Pasto” o “FI1”) indicado por el código de color.

Las comunidades bacterianas de moscas asociadas a frutos y de ambientes selváticos y modificados del estado de Pará son en general similares (Figura 14). Éstas están dominadas por los géneros *Acetobacter*, *Gluconobacter* y *Lactobacillus*, los cuales ya fueron reportados anteriormente como parte importante de la microbiota de drosófilas (Chandler et al., 2011). Lo que varía entre estos ambientes es la abundancia relativa de algunas especies en particular (principalmente de los géneros

Acetobacter y *Gluconobacter*) y, en general, vemos que la riqueza total observada es mayor en ambientes modificados, donde la dominancia específica no es tan fuerte, y se encuentra un mayor número de especies bacterianas representadas de forma relativamente balanceada. También son más abundantes en ambientes modificados especies de los géneros *Lactobacillus* y *Lactococcus*. Esta mayor diversidad bacteriana puede estar favoreciendo la nutrición de moscas y en última instancia la adaptación al ambiente modificado. Sin embargo, la adaptación a ambientes nuevos también puede estar influenciada por el ambiente genómico de la mosca. A pesar de que *D. sturtevanti* es una especie rara en ambientes modificados, los pocos individuos que llegan a estos ambientes poseen un tamaño mayor que aquéllos que se encuentran en su ambiente natural (Furtado et al., 2011) y una composición bacteriana idéntica a la de especies de moscas que normalmente habitan estos ambientes modificados. Esto sugiere que la microbiota particular del ambiente podría tener un impacto en la adaptación de la mosca a las nuevas condiciones y en su consecuente aumento de tamaño. Por otro lado, el componente genómico de la mosca podría ser el determinante primario en su capacidad de adaptación a los diferentes ambientes, explicando la preferencia de las distintas especies de moscas por los diferentes ambientes. En el caso opuesto, donde moscas típicas de ambientes cosmopolitas se encuentran en ambientes selváticos, observamos la misma situación. Los pocos individuos de *D. malerkotliana* que consiguen llegar a la selva primaria también necesitan adaptarse a un ambiente estable donde los recursos y especies se mantienen constantes y donde la diversidad bacteriana es menor. Estos individuos poseen un tamaño menor que aquellos que se encuentran adaptados al ambiente modificado y su microbiota parece ser idéntica a la de especies típicas del ambiente selvático.

Conclusiones

Los cambios permanentes al paisaje selvático debido a la deforestación con fines en el aprovechamiento de la tierra para la agricultura tienen un profundo impacto en las comunidades bacterianas asociadas a moscas. La modificación del ambiente de manera rápida y abrupta, influencia la composición del microbioma asociado a moscas drosófilas, que deben adaptarse rápidamente a estos cambios. Las comunidades bacterianas de moscas asociadas a frutos son cualitativamente similares entre ambientes selváticos y modificados del estado de Pará. La mayor diferencia entre estos ambientes reside en la abundancia relativa de algunas especies bacterianas y en la presencia de especies adicionales de menor abundancia en los ambientes modificados. Esta mayor diversidad bacteriana en ambientes modificados podría explicarse con el modelo central-marginal, donde la oferta de sustratos es fluctuante y diversa en la región marginal. Esta mayor diversidad bacteriana podría favorecer la nutrición de moscas y en última instancia su adaptación al ambiente modificado. En este trabajo caracterizamos por primera vez los microbiomas de las especies *D. malerkotliana* y *D. sturtevanti* en ambientes naturales y modificados. Nuestros resultados sugieren que el ambiente (selvático o modificado) es un determinante importante de la comunidad bacteriana, mientras que la especie de mosca (*D. malerkotliana* o *D. sturtevanti*) no parece serlo. Las comunidades bacterianas son específicas del tipo de sustrato al que están asociadas las moscas de la familia Drosophilidae (fruto, hongos y flores).

Anexos

Tabla A1. Muestras de moscas utilizadas en este trabajo

Sample.id	Species	Specific.landscape	#Individuals	Landscape.type	Technical.replicate	Sample.id.no.tech.rep	Sample.id.no.rep
9_sturtevant_i_Casa_2	sturtevant_i	Casa	2	modified	biological_replicate	9_sturtevant_i_Casa_2	sturtevant_i_Casa
10_sturtevant_i_Casa_3	sturtevant_i	Casa	3	modified	biological_replicate	10_sturtevant_i_Casa_3	sturtevant_i_Casa
11_sturtevant_i_Casa_5	sturtevant_i	Casa	5	modified	biological_replicate	11_sturtevant_i_Casa_5	sturtevant_i_Casa
12_sturtevant_i_F1_2	sturtevant_i	F11	2	forest	biological_replicate	12_sturtevant_i_F1_2	sturtevant_i_F11
13_sturtevant_i_F1_3	sturtevant_i	F11	3	forest	biological_replicate	13_sturtevant_i_F1_3	sturtevant_i_F11
14_sturtevant_i_F1_5	sturtevant_i	F11	5	forest	biological_replicate	14_sturtevant_i_F1_5	sturtevant_i_F11
15_sturtevant_i_F1_7	sturtevant_i	F11	7	forest	biological_replicate	15_sturtevant_i_F1_7	sturtevant_i_F11
16_sturtevant_i_F1_12	sturtevant_i	F11	12	forest	biological_replicate	16_sturtevant_i_F1_12	sturtevant_i_F11
22_malerkotliana_Casa_2	malerkotliana	Casa	2	modified	biological_replicate	22_malerkotliana_Casa_2	malerkotliana_Casa
23_malerkotliana_Pasto_2	malerkotliana	Pasto	2	modified	biological_replicate	23_malerkotliana_Pasto_2	malerkotliana_Pasto
24_malerkotliana_Pasto_3	malerkotliana	Pasto	3	modified	biological_replicate	24_malerkotliana_Pasto_3	malerkotliana_Pasto
25_malerkotliana_Pasto_5	malerkotliana	Pasto	5	modified	biological_replicate	25_malerkotliana_Pasto_5	malerkotliana_Pasto
26_malerkotliana_Pasto_7	malerkotliana	Pasto	7	modified	biological_replicate	26_malerkotliana_Pasto_7	malerkotliana_Pasto
27_a_speciosa_flower_2	speciosa_Belem_flower	Flower	2	modified	technical_replicate	27_speciosa_Belem_flower_Flower_2	speciosa_Belem_flower
27_b_speciosa_flower_2	speciosa_Belem_flower	Flower	2	modified	technical_replicate	27_speciosa_Belem_flower_Flower_2	speciosa_Belem_flower
27_c_speciosa_flower_2	speciosa_Belem_flower	Flower	2	modified	technical_replicate	27_speciosa_Belem_flower_Flower_2	speciosa_Belem_flower
28_speciosa_flower_3	speciosa_Belem_flower	Flower	3	modified	biological_replicate	28_speciosa_Belem_flower_Flower_3	speciosa_Belem_flower
29_speciosa_flower_5	speciosa_Belem_flower	Flower	5	modified	biological_replicate	29_speciosa_Belem_flower_Flower_5	speciosa_Belem_flower
30_a_Hirtodrosophila_fungus_F12_2	Hirtodrosophila_fungus	fungus	2	modified	technical_replicate	30_Hirtodrosophila_fungus_2	Hirtodrosophila_fungus
30_b_Hirtodrosophila_fungus_F12_2	Hirtodrosophila_fungus	fungus	2	modified	technical_replicate	30_Hirtodrosophila_fungus_2	Hirtodrosophila_fungus
30_c_Hirtodrosophila_fungus_F12_2	Hirtodrosophila_fungus	fungus	2	modified	technical_replicate	30_Hirtodrosophila_fungus_2	Hirtodrosophila_fungus
31_Hirtodrosophila_fungus_F12_3	Hirtodrosophila_fungus	fungus	3	modified	biological_replicate	31_Hirtodrosophila_fungus_3	Hirtodrosophila_fungus
32_Hirtodrosophila_fungus_F12_5	Hirtodrosophila_fungus	fungus	5	modified	biological_replicate	32_Hirtodrosophila_fungus_5	Hirtodrosophila_fungus
38_malerkotliana_F1_2	malerkotliana	F11	2	forest	biological_replicate	38_malerkotliana_F1_2	malerkotliana_F11
39_malerkotliana_F1_3	malerkotliana	F11	3	forest	biological_replicate	39_malerkotliana_F1_3	malerkotliana_F11
40_malerkotliana_F1_5	malerkotliana	F11	5	forest	biological_replicate	40_malerkotliana_F1_5	malerkotliana_F11
Mock	Mock	Mock	Mock	Mock	biological_replicate	Mock	Mock

Tabla A2. Número de secuencias concatenadas y pre-clusterizadas para cada muestra, antes de la clasificación

Sample	Counts
10_sturtevant_i_Casa_tai_3.primers_V34	186758
11_sturtevant_i_Casa_tai_5.primers_V34	158507
12_sturtevant_i_F11_tai_2.primers_V34	379968
13_sturtevant_i_F11_tai_3.primers_V34	140781
14_sturtevant_i_F11_tai_5.primers_V34	102643
15_sturtevant_i_F11_tai_7.primers_V34	320452
16_sturtevant_i_F11_tai_12.primers_V34	242772
22_malerkotliana_Casa_tai_2.primers_V34	312703
23_malerkotliana_Pasto_tai_2.primers_V34	328828
24_malerkotliana_Pasto_tai_3.primers_V34	398556
25_malerkotliana_Pasto_tai_5.primers_V34	351785
26_malerkotliana_Pasto_tai_7.primers_V34	195329
27_a_speciosa_Belem_flower_2.primers_V34	161909
27_b_speciosa_Belem_flower_2.primers_V34	538153
27_c_speciosa_Belem_flower_2.primers_V34	396534
28_speciosa_Belem_flower_3.primers_V34	281192
29_speciosa_Belem_flower_5.primers_V34	140655
30_a_Hirtodrosophila_SB13011_fungus_F12_tai_2.primers_V34	146444
30_b_Hirtodrosophila_SB13011_fungus_F12_tai_2.primers_V34	101959
30_c_Hirtodrosophila_SB13011_fungus_F12_tai_2.primers_V34	237822
31_Hirtodrosophila_SB13011_fungus_F12_tai_3.primers_V34	284018
32_Hirtodrosophila_SB13011_fungus_F12_tai_5.primers_V34	242004
38_malerkotliana_F11_tai_2.primers_V34	174805
39_malerkotliana_F11_tai_3.primers_V34	198565
40_malerkotliana_F11_tai_5.primers_V34	358146
9_sturtevant_i_Casa_tai_2.primers_V34	341826
Mock.primers_V34	288281

Bibliografía

1. Avondet, J., R.B. Blair, D.J. Berg & M.A. Ebbert. *Drosophila* (Diptera: Drosophilidae) Response to changes in ecological parameters across an urban gradient. *Environ. Entomol.* 2003; 32:347-358.
2. Bray JR and Curtis JT. An ordination of upland forest communities of southern Wisconsin. *Ecological Monographs.* 1957. 27:325-349.
3. Broderick N and Lemaitre B. Gut-associated microbes of *Drosophila melanogaster*. *Gut Microbes.* 2012 Jul 1; 3(4): 307–321.
4. Chandler JA, Morgan Lang J, Bhatnagar S, Eisen JA, Kopp A. Bacterial Communities of Diverse *Drosophila* Species: Ecological Context of a Host–Microbe Model System. *PLOS Genetics.* 2011, 7(9): e1002272. <https://doi.org/10.1371/journal.pgen.1002272>
5. Chen K, Pachter L. Bioinformatics for whole-genome shotgun sequencing of microbial communities. *PLoS Comput Biol.* 2005;1(2):0106-0112. doi:10.1371/journal.pcbi.0010024.
6. Early AM, Shanmugarajah N, Buchon N, Clark AG. *Drosophila* Genotype Influences Commensal Bacterial Levels. *PLOS ONE.* 2017.
7. Fadrosch WD, Ma B, Gajer P, et al. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome.* 2014, 2:6
8. Furtado et al., Tesis doctoral: http://repositorio.ufpa.br/jspui/bitstream/2011/8808/1/Tese_EfeitoPaisagens_Modificadas.pdf.
9. Jehrke L, Stewart F, Droste A & Beller M. The impact of genome variation and diet on the metabolic phenotype and microbiome composition of *Drosophila melanogaster*. *Sci Rep.* 2018 Apr 18;8(1):6215.
10. Kim D, Song L, Breitwieser FP, and Salzberg SL. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Research.* 2016
11. Lozupone C, and Knight R. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Appl. Environ. Microbiol.* 2005 December vol. 71 no. 12 8228-8235

12. Martinson VG, Carpinteyro-Ponce J, Moran NA & Markow TA. A Distinctive and Host-Restricted Gut Microbiota in Populations of a Cactophilic *Drosophila* Species. *Appl Environ Microbiol*. 2017 Nov 16;83(23). pii: e01551-17.
13. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE*. 2013. 8(4): e61217. <https://doi.org/10.1371/journal.pone.0061217>
14. Moghadam NN, Thorshauge PM, Kristensen T, et al. Strong responses of *Drosophila melanogaster* microbiota to developmental temperature. *Fly*. 2017. 12:1, 1-12.
15. Oksanen, JF, Blanchet G, Friendly M, et al. vegan: Community Ecology Package. R package version 2.5-1. 2018. <https://CRAN.R-project.org/package=vegan>
16. Schloss, PD, et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbio*. 2009. 75(23):7537-41.
17. Shannon C. A mathematical theory of communication. *Bell Syst. Tech. J*. 1948 27:379–423
18. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol*. 2008;26:1135-1145. doi:10.1038/nbt1486.
19. Vodovar N, Vinals M, Liehl P, Basset A, Degrouard J et. al. *Drosophila* host defense after oral infection by an entomopathogenic *Pseudomonas* species. *Proceedings of the National Academy of Sciences*. 2005 vol: 102 (32) pp: 11414-11419
20. Whittaker, R. H. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*. 1960. 30, 279–338.
21. Wong AC-N, Chaston JM, Douglas AE. The inconstant gut microbiota of *Drosophila* species revealed by 16S rRNA gene analysis. *The ISME Journal*. 2013;7(10):1922-1932. doi:10.1038/ismej.2013.86.
22. Wong AC, Luo Y, Jinga X, et al. The Host as the Driver of the Microbiota in the Gut and External Environment of *Drosophila melanogaster*. *Appl. Environ. Microbiol*. 2015