

# Título

## Subtítulo

## Subtítulo

**Darihana Linares Laureano**     *Estudiante de Licenciatura en Geografía mención recursos naturales y ecoturismo, Universidad Autónoma de Santo Domingo (UASD)*

---

Resumen del manuscrito

*Keywords:* palabra clave 1, palabra clave 2

---

### 1 Introducción

Desde mediados del siglo XVII es posible visualizar el interés del hombre a estudiar la flora, la fauna y el medio en el que están en conjunto con las interacciones que se producen entre ellos, pero no es hasta mediados del siglo XIX cuando se introduce el término Ecología y su definición, que se empieza a englobar en este tipo de estudios en una categoría (De la Llata Loyola, 2003). A partir de este punto se reconocieron distintos campos de estudios y se implementaron nuevos métodos de análisis, entre los cuales destacan la ecología numérica y métodos como el análisis multivariático. Según P. Legendre & Legendre (2012), la ecología numérica no es más que una de las disciplinas de la ecología cuantitativa, la cual a la vez es una de las divisiones de la ecología matemática.

La ecología numérica se concentra en el estudio y análisis de conjuntos de datos ecológicos, a fin de poder detallar y comprender la configuración de los conjuntos de datos, combinando diversas perspectivas numéricas y disciplinas, procedentes de la Geografía, las matemáticas físicas, taxonomía numérica, parámetros estadísticos y otros más (P. Legendre & Legendre, 2012). El análisis de conjuntos de datos ecológicos, especialmente de la flora, resulta importante tanto para la ciencia, como para la economía o el sector de salud, por eso se han establecido distintas parcelas permanentes de medición y monitoreo forestal donde se colectan datos sobre la diversidad forestal, su estructura, el crecimiento y su productividad (Pineda, 2014). En el estudio de las plantas a través de la Ecología Numérica se usan diversas técnicas que permiten obtener información sobre el rango de asociación, agrupamiento, ordenamiento, diversidad, autocorrelación, etcétera, diferentes herramientas para usar las técnicas.

La medición de la asociación es un coeficiente que sirve para medir y asociar los datos de variables cualitativos y cuantitativos. La medición de estas variables se puede hacer por dos modos el  $Q$  y el  $R$ , el primero consiste en hacer una comparación de un dúo de objetos y el segundo consiste en realizar una descripción de un par de objetos y luego compararlos. Para el modo  $Q$  se miden la asociación según la similaridad o disimilaridad de un par de objetos. Mientras que para el modo  $R$  se mide el grado de dependencia existente entre las variables, entre los cuales se puede mencionar la covarianza o el coeficiente de correlación (Borcard, Gillet, Legendre, & others, 2011).

En cambio, el análisis de agrupamiento o clúster análisis es una técnica que consiste en separar un conjunto de datos y luego estructurarlos, sin dejar uno fuera de lugar, como subconjuntos con distintas categorías o jerarquías de acuerdo a sus características. La finalidad de hacer un agrupamiento es identificar los pequeños grupos dispersos en un espacio discreto pero constante,

este agrupamiento divide en conjunto de objetos a estudiar, por lo que es necesario e importante que cada objeto agrupado en otros subgrupos no se encuentre en otros (Borcard et al., 2011).

En el caso del análisis de ordenamiento, son técnicas que consisten en simplificar la magnitud de los datos. Todas estas técnicas muestran las predisposiciones esenciales de variabilidad de cada dato que se encuentran en un campo de dimensiones simplificadas, organizando los ejes con rangos decreciente de varianza explicada en cada uno de los ejes sucesivo, de forma convencional. Estos tipos de análisis pueden ser tanto no restringido o simple, como restringido o canónico. En donde para el primero, las tendencias o predisposiciones del grupo que interesa no está restringida por otro grupo. Entre sus técnicas principales de análisis están los de componentes principales (PCA) basado en un vector propio y se utiliza en datos cuantitativos sin tratamiento preservando la distancia euclídea, los de correspondencia (CA) que se usa en datos frecuentes con dimensiones uniformes y positivos, y los de coordenadas principales (PCoA) que se concentra en organizar las matrices de disimilaridad, usualmente, con el modo Q en vez de tablas de sitio por variables (Borcard et al., 2011).

Mientras que, para el segundo análisis de ordenamiento, restringido o canónico, a diferencia de la simple, es una técnica que relaciona dos o más conjunto de datos en el proceso de organización u ordenación. Algunas de sus técnicas principales son análisis de redundancia (RDA) que consiste en la combinación de la regresión y PCA que funciona como una extensión de diversos análisis que muestran la respuesta multivariante de datos, y el análisis de correspondencia canónica (CCA) que funciona como un aproximado de una regresión Gaussiana multivariante, además, este se caracteriza por organizar las especies en todos los ejes canónicos acorde a su configuración ecológica óptima (Borcard et al., 2011).

En cuanto a la diversidad, según Borcard et al. (2011) y Magurran (1988), esto alude a la variedad y cantidad de especies en un espacio determinado, esta variedad también se produce a nivel de comunidad. La diversidad puede ir desde la diversidad local, hasta heterogeneidad espacial de esta diversidad. La diversidad de especies considerada como un número único puede ser medida por la riqueza o la rarefacción de especies usando la notación  $q$  o por la presencia o ausencia de los datos de esta. Según Whittaker (1972), el entendimiento de la diversidad y los cambios que conllevan, asociados a la configuración del relieve los componentes *alpha*, *beta* y *gamma* serían de mucha utilidad. Donde la primera se refiere a la riqueza de las especies existentes en una comunidad determinada, considerada homogénea. La segunda, se trata del rango de cambios que se producen en la estructura de especies que están en diferentes comunidades en un espacio. Y la última, se refiere a la riqueza de especies de forma conjuntiva que hay en una comunidad y que integra un espacio determinado.

La autocorrelación, según Borcard et al. (2011), forma parte de los análisis espaciales aplicados a datos ecológicos, que se produce por distintos procesos y que mide puntos cercanos para afirmar si estos poseen valores similares o distintos, por lo que la correlación puede ser una correlación positiva o negativa.

La *Myrtaceae* son una familia de plantas de árboles y arbustos bastante numerosas compuestas caracterizadas por ser leñosas. Esta familia de plantas pertenece al orden de los *Myrtales*, teniendo a nivel mundial 129 géneros y aproximadamente 5330 especies (Perez & R, s.f.). De sus características físicas se destacan sus hojas simples y opuestas, tienden a ser perennes por lo que resulta poco común ver individuos caducifolios, sus flores son hermafroditas y comúnmente son de color blanco con simetría radial, según la especie de esta familia los frutos tienen forma de bayas o cápsulas secas. De esta familia el género más conocido es el *Eucalypto Eucalyptus* por sus propiedades medicinales y su madera dura. Esta familia está distribuida en todos los continentes, pero predominan en América, África y Oceanía, en climas templados, tropicales y subtropicales. Muchos consideran que la importancia de esta familia está en lo económico, como la producción de fru-

tas para venta de zumos y mermeladas, producción de madera, producción de papel, de carbón, además de ser usadas en la industria farmacéutica, se usa en la cosmetología y en la producción de especias (Almeida, 2019; Britannica, 2016; Lorenzo-Cáceres, s.f.).

El estudio de la biodiversidad de la flora en especial de una especie vegetal es importante, debido a que permite conocer sus características propias, ecosistémicas, su distribución, su capacidad productiva, su potencial de uso y aportes a los humanos como a su ecosistema. Por todo lo anterior, es importante determinar el objetivo de este estudio, el cual se fragmenta así: *a)* identificar si los grupos de mi familia, *Myrtaceae*, se organizan de forma discontinua y acorde a la composición de las especies; *b)* indagar si existe algún tipo de patrón que sea o no sea consistente con alguna variable ambiental o atributo; *c)* determinar cuántas especies indicadoras hay o si hay alguna con preferencia por ciertas condiciones ambientales o atributos; *d)* investigar, en un espacio bidimensional, si hay una tendencia de ordenación visible de las especies de *Myrtaceae*; *e)* conocer si las tendencias de ordenación se asocian con las variables ambientales o atributos; *f)* determinar, de acuerdo a la estimación de riqueza, si mi familia está bien representada, tomando como buena representación un 85%; *g)* investigar si existe una alguna asociación entre la diversidad Alpha y las variables ambientales o atributos y cuales son estas; *h)* conocer si hay algún tipo de contribución local o de alguna especie a la diversidad Beta; *i)* identificar si hay alguna especie que presente algún patrón aglomerado, cuál es y si presenta alguna asociación con las variables ambientales; *j)* y determinar si los modelos de distribución de especies (*SDM*) predicen adecuadamente las ocurrencias de las especies.

## 2 Área de estudio

El área usada para este estudio se encuentra en *BCI* (Isla de Barrio Colorado), que es una isla perteneciente a Panamá, la cual se formó cuando las aguas del río Chagres fueron represadas para construir el Canal de Panamá. Es considerada como una colina con una superficie de 1,500 hectáreas sobresaliente a 137 m en el lago Gatún, y que está localizada entre los 9° 09' N y 79° 51' W. Esta isla se caracteriza por sus suelos arcillosos con profundidades entre los 50 cm y 1 m, y su clima es común de las tierras bajas tropicales con temperatura promedio anual de 27° C, y variación diurna de 9° C. También tiene unas precipitaciones anuales de 2,600 mm, con su estación lluviosa (mayo a diciembre), y su estación seca para el resto de meses (Pérez et al., 2005; Sugasti, Eng, & Pinzón, 2018). Según la clasificación de zonas de vida de Holdridge, *BCI* se encuentra ubicada en la zona de vida de bosque húmedo tropical; Claver Farías & others (1984), dice que conforme a la clasificación de climática de Köppen esta tiene un clima tropical húmedo con precipitaciones abundantes; y de acuerdo con el mapa de la ANAM (2000) su vegetación se caracteriza por tener bosques semicaducifolio tropical de tierras bajas (Rodríguez-Castro & Flores, 2021).

El área de estudio, específicamente, es una parcela permanente forestal para su medición y monitoreo de 50 hectáreas, establecida en 1980 por Stephen Hubbell y Robin Foster, en el bosque húmedo tropical de *BCI*. Esta parcela se caracteriza por su forma rectangular con 1000 m de largo y 400 m de ancho, además, esta se ubica en la parte central de una meseta en la isla (Hubbell, Foster, & Condit, 2005; Pérez et al., 2005). Los datos extraídos de las especies forestales de la parcela se obtuvieron a través de múltiples censos desde 1982 hasta 2015 (un total de 8 censos), en los cuales se hicieron identificación, marcado y monitoreo de los tallos leñosos individuales de más de 10 mm de diámetro a altura del pecho. Durante los 35 años se censaron más de 350,000 árboles individuales (Hubbell et al., 2005). Ver figura 1

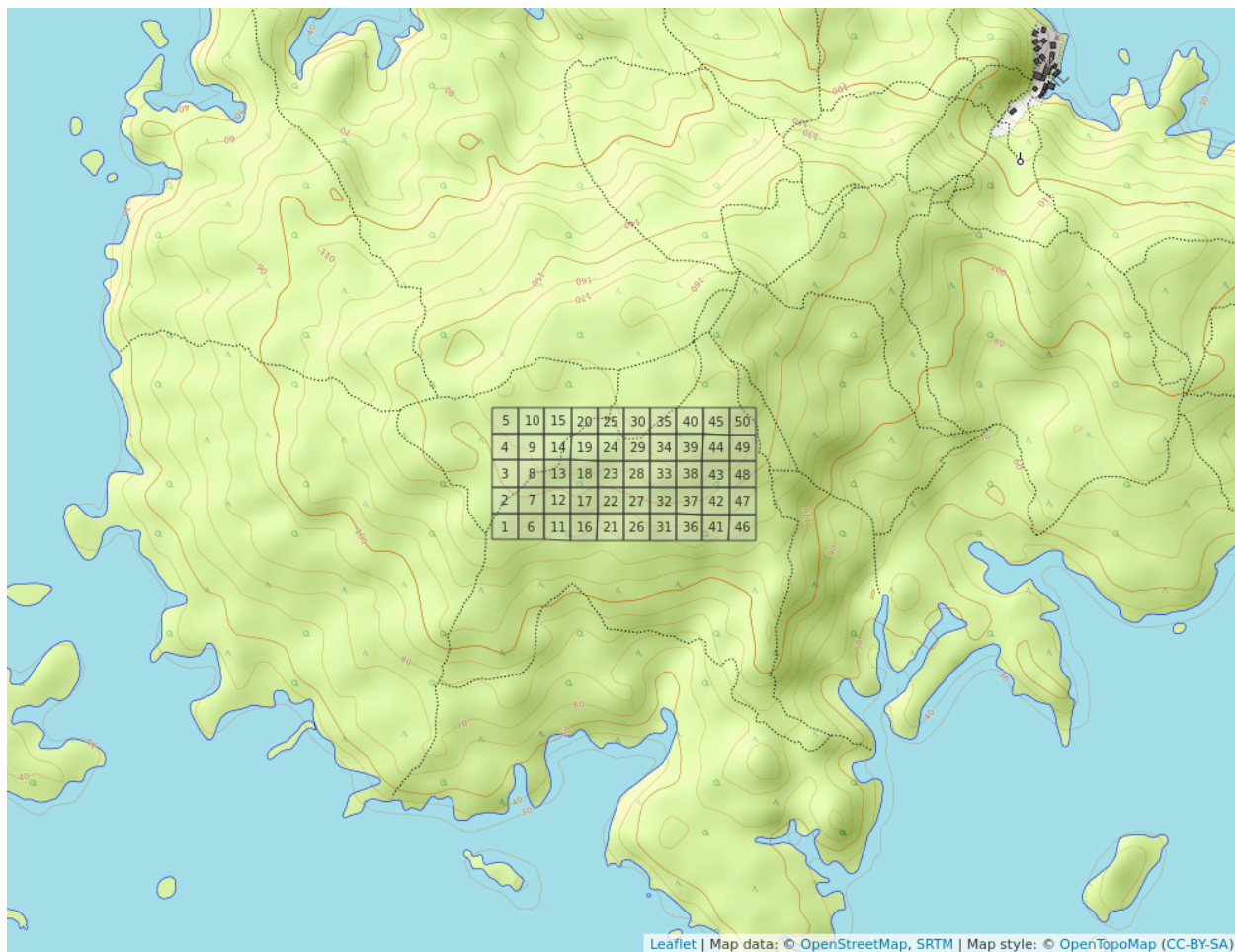


Figure 1: Mapa de la Isla Barro Colorado y la parcela permante

### 3 Metodología y materiales

Para el estudio de la biodiversidad de la familia *Myrtaceae* se usaron los datos censales de las 50 parcelas forestales de *BCI* que están presentes en el repositorio *scripts-de-analisis-BCI* (Batlle, 2020), de los cuales fueron tratados y procesados con distintos algoritmos y técnicas de medición, cálculo, análisis e interpretación por medio del entorno de desarrollo integrado y libre, *RStudio* (R Core Team, 2020). En cuanto a los gráficos presentes se obtuvieron por medio *R*, usando paquetes y funciones (ver tabla 1).

Para iniciar los análisis estadísticos de datos con dimensiones diversas, Buzai & Baxendale (2009) y Borcard et al. (2011), recomiendan realizar un análisis exploratorio de datos (EDA), debido a que es una herramienta imprescindible para conseguir una primera intermediación, información genérica de los datos, realizar transformaciones a variables de ser necesarias, y poder conducir más análisis. Se usaron los paquetes *vegan*, *tidyverse*, *sf*, *mapview* y *RColorBrewer* para cargar los datos censales, matriz de comunidad y matriz ambiental, y así poder extraer los datos de la familia *Myrtaceae*, de las variables ambientales y transformarlos en tablas, matrices, e incluso generar gráficos de mapas. También, se usaron las librerías *psych* y *ez*, en conjunto con las anteri-

ores para ver correlaciones de la *Myrtaceae* con distintas variables ambientales con la función `cor`, y generar mapas de variables ambientales por cada cuadro con `ggplot2`.

La medida de los coeficientes de asociación de la familia *Myrtaceae* con los modos *Q* y *R* [] se realizó mediante funciones pertenecientes a las librerías `vegan`, `adespatial`, `broom`, `tidyverse`, `sf`, `cluster` y `gclus`. En el modo *Q* se mide la abundancia de las especies de *Myrtaceae* con la función `dist.ldc` que calcula la distancia euclídea entre dos sitios y se usa también el método de transformación *Hellinger* [] y así obtener el grado de disimilaridad entre sitios. También, se calculó la distancia de *Jaccard* *Dj* [] utilizando la función `vegdist` que convierte la matriz de comunidad en una de presencia/ausencia para calcular la matriz de distancia. En cuanto a la asociación de las variables ambientales, se hizo un nuevo ambiente con la función `env` para las variables transformadas con la función `scale`. De igual manera, se usó la función `env_mix` para asociar las variables ambientales heterogeneidad ambiental, hábitat y quebrada, generando una matriz de disimilaridad.

Con el modo *R* de medición de asociación, se mensuró el grado de asociación entre especies haciendo una transformación de la matriz de comunidad usando la estandarización *Chi* con la función `decostand`, luego se calcula la distancia con la función `dist` y se representa la matriz en un mapa de calor con la función `coldiss`. Se calculó la distancia de *Jaccard* [] entre especies usando la matriz de comunidad transpuesta convertida a una de presencia/ausencia con la función `vegdist`. En cuanto a la asociación de variables ambientales numéricas, usando el índice de *rho* de *Spearman* [] con los verbos `select`, `mutate` y `matches` de la herramienta `dplyr` y también la función `cor` para correlación.

Los análisis de agrupamiento o cluster análisis aplicados a la familia *Myrtaceae* fueron realizados usando las librerías `magrittr`, `broom`, `tidyverse`, `mapview` y `indicspecies`. Para el estudio se usó el agrupamiento jerárquico (*AJ*) con un enfoque aglomerativo por enlace simple [], completo [], promedio [] y el método de *Ward* de varianza mínima []. Se agruparon pares de objetos según la mayor similitud (vecino más próximo o mínima distancia), partiendo del agrupamiento por enlace simple, utilizando la matriz de comunidad transformada por medio del método de normalización con la función `decostand` y la distancia euclídea con la función `vegdist`. El agrupamiento jerárquico se hizo con la función `hclust` para método simple y se generó un dendrograma (gráfico) con la función `plot` del resultado de este *AJ*. Para el agrupamiento completo usando la menor similitud (máxima distancia o vecino más lejano) se empleó la función `hclust` y la matriz de distancia de cuerdas o *chord* [], con tal de generar un dendrograma con la función `plot`. En cuanto al agrupamiento por enlace promedio (UPGMA, WPGMA, UPGMC, WPGMC), se usó el UPGMA [], para maximizar la correlación entre las distancias co-fenética (coeficiente de correlación de Pearson), para esto se usó la función `hclust` y para obtener el gráfico la función `plot`. Finalmente, el agrupamiento por *Ward* [], que define los grupos de una forma donde la suma de cuadrados se minimice en cada uno de ellos usando las funciones `hclust` y `plot`.

De todos los métodos anteriores de agrupamiento, se han seleccionado ideal para la familia *Myrtaceae* se usó la función `map` y así copiar sus valores. Además, para la elección del número de clusters se usó la función `calcular_anchuras_siluetas` en base a la matriz de comunidad original, la matriz de distancias y objeto de clúster usando UPGMA y *Ward*, luego, se hizo un nuevo dendrograma con la función `reorder.hclust` y un mapa de calor con la función `heatmap`. Finalmente, se hizo una evaluación para ver si el número de clusters obtenidos por los métodos anteriores son los ideales usando el método *bootstrap* [], usando el paquete `pvcust` para generar dendogramas con trazos en rectángulos y líneas que dividen el gráfico acorde al número de grupos.

En cuanto al agrupamiento de variables se usaron los grupos obtenidos por UPGMA para evaluar la homogeneidad por medio de pruebas *t* [], basadas en la distribución *t* de *Student* [] y la suma de rangos de *Wilcoxon* [], a partir de esto se generaron gráficos de caja y mapas presentando

la ubicación por cuadrado de cada grupo en la parcela permanente. Y para la homogeneidad de los grupos Ward se usó el método ANOVA [] y Kruskal/Wallis [].

Cerrando con el clusters análisis, se obtuvieron las especies indicadoras y las especies con preferencias por hábitats; la primera mediante IndVal con las funciones `multipatt` y `strassoc` y su significancia o valor p con la función `p.adjust`; las especies con preferencias por hábitats se obtuvieron mediante el coeficiente de correlación biserial puntual [] con las funciones `multipatt` y `strassoc`.

#### 4 Resultados

Ver tabla ?? y figura ??

#### 5 Discusión

#### 6 Agradecimientos

#### 7 Información de soporte

Table 1: Materiales usados en el estudio

Materiales	Uso
RStudio	Redacción del manuscrito, procesamientos de datos censales de la familia <i>Myrtaceae</i> por medio de Scripts.
library vegan	Conjunto de herramientas para hacer análisis de diversidad, ordenación de comunidad y análisis de disimilaridad.
library tidyverse	Colección de paquetes que permiten transformar, importar, visualizar, modelar y presentar distintos datos.
library sf	Creación de simple features, ampliando objetos tipo data.frame con una columna de lista de características simples.
library mapview	Para ver objetos espaciales de forma interactiva sobre un mapa base.
library RColorBrewer	Para crear paletas de colores para mapas temáticos.
library ez	Permiten una visualización y análisis de datos simples y con especificaciones consistentes.
library psych	Conjunto de herramientas para hacer análisis de datos multivariados.
library tmap	Para visualizar, con mapas temáticos, la distribución de datos espaciales.
library adespatial	Herramienta para hacer análisis espaciales, a distintas escalas, de datos multivariados.
library broom	Para resumir información de objetos estadísticos en tablas.
library cluster	Para el clúster análisis o de agrupamiento, que permiten encontrar grupos de datos.
library gclus	Ordena en matrices de diagramas, dispersión y coordenadas paralelas con un índice, los paneles.
library magittr	Paquete que permite, mediante mecanismos, cadenas de comandos con el operador pipa.

Materiales	Uso
library pvclust	Paquete que permite implementar un remuestreo multiescala para evaluar inconsistencia en análisis de agrupamiento jerárquico.
library indicspecies	Estima el valor estadístico de la relación presencia-abundancia de especies y sus sitios.
library plyr	Conjunto de herramientas que permiten separar, aplicar y combinar datos para generar resúmenes estadísticos de ellos.
library SpadeR	Estima diversos índices de biodiversidad y medidas de similitud de datos individuales tomados de diversas comunidades.
library iNEXT	Paquete que permite calcular y trazar la rarefacción y extrapolación de diversidad de especies.
library vegetarian	Para calcular la diversidad por comunidad en un conjunto de datos.
library ape	Paquete que permite hacer análisis filogenéticos y evolutivos de árboles.
library spdep	Conjunto de funciones para crear matrices de ponderaciones espaciales de puntos de patrones polígonos, entre otros.
library ade4	Herramientas para análisis de datos multivariados.
library adegraphics	Sirve para hacer representaciones gráficas de datos multivariados.
library gridExtra	Ofrece funciones para poder trabajar con gráficos en <i>grid</i> y crear diversos trazados en una página y dibujar tablas.
library grid	Reescribe los gráficos, sus capacidades y da soporte a la interacción.
library gtable	Herramientas que permiten trabajar más fácil con tablas.

## 8 Script reproducible

...

## Referencias

- Almeida, S. (2019). *Myrtaceae, familia*. Retrieved from <https://knoow.net/es/ciencias-tierra-vida/biologia-es/myrtaceae-familia/>
- Batlle, J. R. M. (2020). biogeografia-master/scripts-de-analisis-BCI: Long coding sessions (Version v0.0.0.9000). <https://doi.org/10.5281/zenodo.4402362>
- Borcard, D., Gillet, F., Legendre, P., & others. (2011). *Numerical ecology with r* (Vol. 2). Springer.
- Britannica, E. (2016). *Myrtaceae*. Retrieved from <https://www.britannica.com/plant/Myrtaceae>
- Buzai, G. D., & Baxendale, C. A. (2009). Análisis exploratorio de datos espaciales. *Geografía Y Sistemas de Información Geográfica, No 1*,(2009).
- Claver Farías, I., & others. (1984). Guía para la elaboración de estudios del medio físico: Contenido y metodología. *Ministerio de Obras Públicas Y Urbanismo. Madrid*.
- De la Lata Loyola, M. D. (2003). *Ecología y medio ambiente*. Editorial Progreso.
- Hubbell, S., Foster, R., & Condit, R. (2005). *Barro colorado forest census plot data*. Retrieved from <http://ctfs.si.edu/webatlas/datasets/bci/>
- Legendre, P., & Legendre, L. (2012). *Numerical ecology*. Elsevier.
- Lorenzo-Cáceres, J. M. S. de. (s.f.). *Familia myrtaceae*. Retrieved from <https://www.arbolesornamentales.>

es/Myrtaceae.htm

Magurran, A. E. (1988). *Ecological diversity and its measurement*. Princeton university press.

Perez, R., & R, C. (s.f.). *Tree atlas of panama*. Retrieved from <http://ctfs.si.edu/PanamaAtlas/famdescr.php?Family=Myrtaceae>

Pérez, R., Aguilar, S., Condit, R., Foster, R., Hubbell, S., & Lao, S. (2005). Metodología empleada en los censos de la parcela de 50 hectareas de la isla de barro colorado, panamá. *Centro de Ciencias Forestales Del Tropico (CTFS) Y Instituto Smithsonian de Investigaciones Tropicales (STRI)*, 1–24.

Pineda, P. (2014). *Análisis del sistema de parcelas permanentes de medición en los bosques de guatemala. informe final*. Guatemala: Proyecto" Sistemas de información sobre la productividad de los ....

R Core Team. (2020). *R: A language and environment for statistical computing*. Retrieved from <https://www.R-project.org/>

Rodríguez-Castro, L., & Flores, N. (2021). Colonización de la hepática epífila leptolejeunea elliptica (lehm & lindenb) schiffn.(Lejeuneaceae), sobre dos especies de arbustos de la isla barro colorado (bci). *Tecnociencia*, 23(1), 82–103.

Sugasti, L., Eng, B., & Pinzón, R. (2018). Medición continua de flujo de co2 ensuelo en una parcela de bosque tropical en isla barro colorado, canal de panamá. *Universidad Tecnológica de Panamá*, 1–7.

Whittaker, R. H. (1972). Evolution and measurement of species diversity. *Taxon*, 21(2-3), 213–251.