

Asociación y abundancia relativa de especies de la familia Rubiaceae en la parcela permanente Isla Barro Colorado

J. Alberto Meléndez Juan Universidad Autónoma de Santo Domingo (UASD)

Resumen El estudio del grado de ordenación y riqueza de las comunidades de especies depende en gran medida de interacciones entre las distintas especies y su hábitat. En este trabajo se estudian los patrones de asociación entre las especies de la familia Rubiaceae y como varía la diversidad con respecto a las características del hábitat. Se encontró que la comunidad presenta disparidad en la abundancia, con especies muy dominantes y muchas otras pobremente representadas. Las ocurrencias de especies como *Psychotria horizontalis* y *Psychotria marginata* están asociadas con la cantidad de aluminio y fósforo en el suelo. La especie *Alseis blackiana* presenta asociación con el contenido de los metales manganeso, hierro y cobre. Los resultados sugieren que la especialización a ciertas condiciones puede verse reflejada en la asociación al hábitat incluso en una escala muy local. Más allá de su significado ecológico, la distribución de las poblaciones de estas plantas tolerantes podría servir para indicar ocurrencias anormales en la concentración de metales pesados en el ambiente.

Keywords: Especies dominantes, Diversidad, Rubiaceae

1 Introducción

Las comunidades vegetales de los bosques neotropicales ejemplifican la diversidad y complejidad ecológica de la región tropical. El estudio continuo de su riqueza y abundancia relativa permite identificar las especies raras, las cuales son más vulnerables a los cambios en su hábitat y por lo tanto propensas a extinguirse localmente (Volkov, Banavar, Hubbell, & Maritan, 2003). Adicionalmente, monitorear la diversidad como propiedad de las comunidades vivientes resulta de suma importancia para analizar el efecto que tiene la transformación de los ecosistemas en las comunidades naturales. Claramente existe entonces la necesidad de conocer como se encuentran asociadas las especies entre sí dentro de las comunidades ecológicas para ayudar a comprender los factores que inciden en su conservación (Moreno, 2001).

La familia Rubiaceae es un importante grupo de plantas vasculares de distribución cosmopolita con una marcada diversidad en regiones tropicales y subtropicales (Davis et al., 2009). Muchas de las especies que componen esta familia se encuentran adaptadas a la vida en la penumbra, y prosperan bajo la sombra del dosel selvático. En estas selvas tropicales, el grado de ordenación y riqueza de las comunidades que componen el sotobosque, depende en gran medida de interacciones entre las distintas especies (Torres-Leite et al., 2019). También, este grado de ordenación se asocia con factores ambientales de su hábitat, ya que muchas de estas especies están adaptadas a rangos elevados de acidez y otras condiciones específicas de los componentes del suelo, como la concentración de distintos metales (Jansen, Robbrecht, Beeckman, & Smets, 2000).

Es preciso señalar que trabajos anteriores (Condit et al., 2002) sobre el bosque tropical panameño y el grado de reemplazo entre especies de distintas comunidades o diversidad beta, sugieren que la disimilaridad tiende a aumentar en función de la distancia a la cual se encuentran separadas en el espacio. Sin embargo, estos trabajos no restan importancia a la variabilidad del hábitat y se toman en cuenta en este estudio, ya que un acercamiento inicial a los datos de abundancia de las

distintas especies de Rubiaceae en Barro Colorado arrojó indicios de posibles patrones acerca de su distribución, y se plantea la posibilidad de que existan especies con algún grado de asociación respecto a las variables ambientales que allí imperan.

Aunque la distribución de la abundancia de las especies actualmente es en buena parte atribuida a los mecanismos que definen a una comunidad en particular, como la prevalencia de especies dominantes, relativamente más abundantes en comparación con las especies raras. Las medidas para la distribución de la abundancia relativa se encuentran sujetas a interacciones que aún no se conocen del todo, ni en qué grado inciden en la estructura de la comunidad (Néda, Horvat, Toháti, Derzsi, & Balogh, 2008).

El presente estudio evalúa la relación entre abundancia relativa de especies de la familia Rubiaceae y su distribución en una porción de bosque tropical en la parcela permanente Barro Colorado Island (en lo adelante BCI), Colón, Panamá. Las parcelas permanentes, como BCI, son una excelente fuente de datos demográficos y posibilitan el estudio continuo de la diversidad a nivel local y contribuyen a medir el aporte de la familia Rubiaceae a la diversidad de su comunidad. En ese sentido, este trabajo aprovecha la información disponible (Hubbell, Condit, & Foster, 2021), y herramientas de libre acceso (Martínez Batlle, 2020), para conocer posibles patrones de asociación entre estas especies, como varía la diversidad con respecto a las características del hábitat en el cual crecen estas poblaciones de plantas, y otras condiciones ambientales mediante análisis estadístico de datos de los censos realizados en BCI.

2 Metodología

2.1 *Ámbito de estudio*

BCI es una estación de censo permanente administrada por el Center for Tropical Forest Science ubicada en el centro de la isla artificial Barro Colorado, con las coordenadas 09° 09'N, 079° 50'O. La parcela consiste en un polígono de 50 hectáreas cuadradas en el cual se han contabilizado todos los árboles con más de 10 mm de diámetro a la altura del pecho cada cinco años desde 1985 (Condit, Chisholm, & Hubbell, 2012; Condit, Pérez, Lao, Aguilar, & Hubbell, 2017; Hubbell & Foster, 1983; Hubbell et al., 1990). En este estudio se utilizaron los datos del censo realizado en el año 2015.

Los datos referentes a cada uno de los 50 cuadrantes de una hectárea que componen BCI, fueron manejados en R (R Core Team, 2020), partiendo de su disposición en dos matrices: de comunidad y ambiental (Martínez Batlle, 2020). Estas matrices contienen datos de las variables ambientales, tales como condiciones edáficas, tipo de hábitat, topografía del lugar, clasificación etaria del bosque, y datos demográficos y georeferenciación espacial de todos los individuos censados. Se adaptaron *scripts* reproducibles recuperados de Martínez Batlle (2020), utilizando la colección de paquetes multifuncionales Tidyverse (Wickham, 2017), paquetes gráficos y de procesamiento de datos espaciales para la representación de mapas y figuras como mapview (Appelhans, Detsch, Reudenbach, & Woellauer, 2019) y simplefeatures (Pebesma, 2018); y herramientas de análisis estadístico, como vegan (Oksanen et al., 2019), indicpecies (De Cáceres & Legendre, 2009), entre otros (ver información suplementaria: 4).

A fin de conocer las características distintivas de los datos conservados en las matrices de comunidad y ambiental, se realizó un análisis exploratorio de los mismos que incluyó la visualización de gráficos, tablas, mapas de los cuadrantes de una hectárea y paneles de correlación lineal entre variables de ambas matrices, esto permitió obtener una perspectiva general y ayudó a determinar los procedimientos posteriores que se detallan a continuación.

2.2 Medición de asociación

Para las pruebas de medición de asociación, se calculó la distancia euclídea entre los cuadrantes considerados como objetos. Para esto, fue requerida la transformada de la matriz de comunidad por el método de Hellinger, el cual consiste en la radicación al cuadrado de la abundancia relativa y_{ij} (cociente resultante de cada valor de abundancia entre la suma de los sitios) como muestra la fórmula 1. Donde j refiere a cada especie o columna en la matriz, i es la fila o cuadrante e $i+$ representa la suma de filas de la matriz de la i -ésima fila (Legendre & Gallagher, 2001). Además, la distancia euclídea entre cuadrantes en cuanto a la ocurrencia de especies fue evaluada aplicando el índice de disimilaridad de Jaccard de la matriz normalizada, con valores de abundancia convertidos en valores binarios (Borcard, Gillet, & Legendre, 2018). De la misma manera, se utiliza la métrica de Jaccard aplicada a la matriz de comunidad transpuesta y convertida a datos de presencia/ausencia para medir el grado de asociación entre especies.

$$y' = \sqrt{\frac{y_{ij}}{y_{i+}}} \quad (1)$$

Para poder comparar la relación entre las especies según su abundancia numérica, se utilizó estandarización ji -cuadrado de la matriz de comunidad transpuesta (Legendre & Gallagher, 2001). La ocurrencia entre las especies y su distribución en BCI fue examinada por medio de el coeficiente de correlación entre rangos de Spearman para medir el grado de asociación entre las variables riqueza numérica de especies y la abundancia con las variables ambientales geomorfológicas y la composición química del suelo (Borcard et al., 2018).

2.3 Análisis de agrupamiento

Tanto el método jerárquico aglomerativo de asociación entre pares de cuadrantes (según la composición de especies) bajo el criterio de enlace completo, y el método Ward basado en la varianza mínima, fueron utilizados en un acercamiento preliminar al análisis de agrupamiento para contrastar su eficacia en conseguir grupos consistentes y con significado ecológico (Borcard et al., 2018). Estos generaron dendrogramas que posteriormente son comparados con la matriz de distancia de cuerdas (Legendre & Gallagher, 2001), utilizando correlación cofenética entre ambos para determinar el número ideal de grupos. Además, se utilizó remuestreo bootstrap y bootstrap multiescalar para conocer la probabilidad de éxito en la inferencia del número de grupos y la identidad de sus componentes (Borcard et al., 2018). Las reparticiones se basaron en una probabilidad de 91% o más de acierto para el método bootstrap y de un 95% para bootstrap multiescalar.

Puesto que se encontraron patrones consistentes en la composición y número de grupos entre los métodos examinados (ver figura 1), los análisis de agrupamiento posteriores se basaron en los producidos mediante enlace completo, para el cual se consideraron dos grupos compuestos por 20 y 30 cuadrantes, respectivamente.

Para conocer cuales especies son características o se encuentran asociadas a cada grupo, se utiliza la métrica del “valor indicador” o IndVal (Dufrene & Legendre, 1997), la cual está basada en permutaciones aleatorias de los sitios según la ocurrencia de las especies y su abundancia. Así mismo, se estudia el grado de asociación de las especies con cierta preferencia por las combinaciones de cuadrantes consideradas como grupos, indicado por el coeficiente de correlación biserial puntual (Borcard et al., 2018). Se llevó a cabo un acercamiento parecido al anterior durante las pruebas estadísticas de la hipótesis nula, sobre la base de que las especies que se encontraban en cuadrantes pertenecientes a un determinado grupo lo hacían por obra del azar. Esta prueba

se logró mediante reordenación aleatoria de los valores de abundancia y comparando su distribución con los valores obtenidos anteriormente (Cáceres & Legendre, 2009). Para estas pruebas de asociación y las subsiguientes se utilizó un nivel de significancia $\alpha = 0.05$.

2.4 Ordenación

Las características de la varianza en los datos ambientales en BCI fueron estudiados mediante análisis de sus componentes principales (PCA) (Borcard et al., 2018). Este método permitió resumir la multidimensionalidad de las variables, explicar la varianza y los posibles patrones que estos podrían seguir. Esto se realizó también para la matriz de comunidad, con valores de abundancia normalizados por la transformada Hellinger, además de un análisis de correspondencia (CA) de la misma matriz. De manera alternativa, se realiza un análisis de las coordenadas principales (PCoA) para ayudar a conocer la relación entre las especies, utilizando el coeficiente de disimilitud de Jaccard como medida, y a su vez usando los promedios ponderados por los valores de abundancia para permitir su representación en los diagramas *biplot* (Borcard et al., 2018).

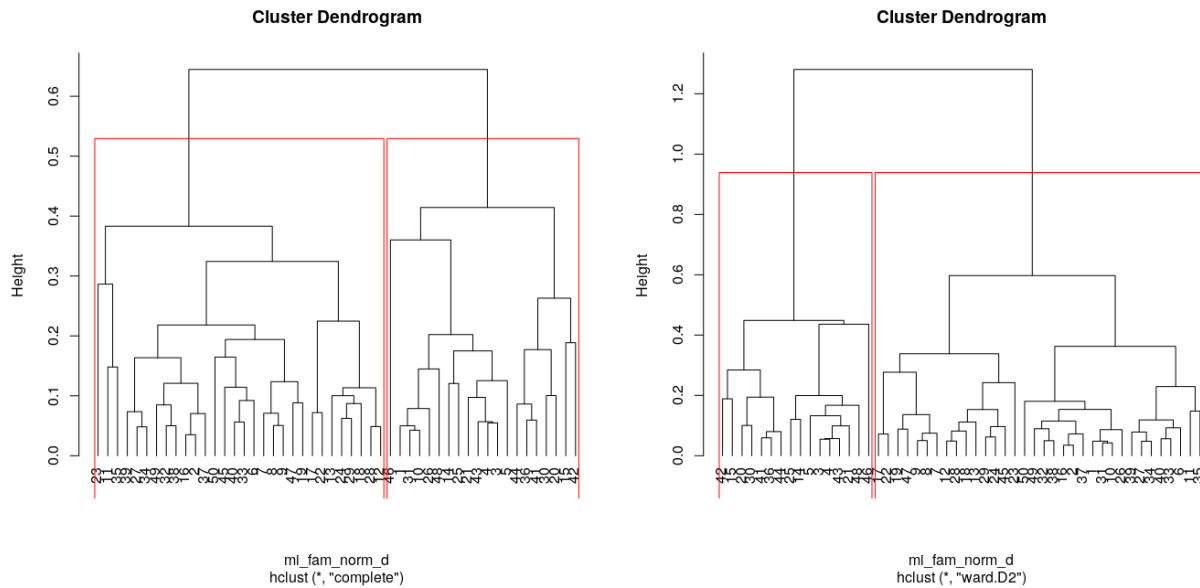


Figure 1: Dendrogramas de los grupos producidos por Ward y Complete.

Se realizó un acercamiento restringido de la ordenación para probar el grado de dependencia de los datos de la matriz de comunidad con la matriz ambiental, mediante ajuste lineal en un análisis de redundancia (RDA) (Borcard et al., 2018), utilizando la matriz de comunidad transformada por Hellinger. Se seleccionaron las variables que presentaron cierto grado de asociación, y (a discreción y de manera secuencial) se excluyeron algunas de estas, con el objetivo de reducir el grado de colinealidad entre las variables independientes restantes.

2.5 Análisis de la diversidad

Con la finalidad de asignar medidas apropiadas a la diversidad de especies, y aprovechando su relación con los índices de Shannon y Simpson, se empleó la serie de números de diversidad de Hill, la fórmula de la entropía de Rengi y el índice de equidad de Pielou. Se examinó la posible correlación entre estas medidas y las variables ambientales que aparentaron tener algún efecto en la riqueza y equidad de la comunidad (Borcard et al., 2018). Además, mediante interpolación por rarefacción al número de individuos del cuadrante con la menor abundancia, se compara el valor esperado de riqueza para todos los sitios. Adicionalmente, se estima la riqueza de la familia Rubiaceae que resultaría de aumentar al doble el muestreo realizado en BCI, mediante los métodos de extrapolación incluidos en las colecciones de funciones SpadeR y iNEXT (Chao, Ma, Hsieh, & Chiu, 2016; Hsieh, Ma, & Chao, 2020), modificadas por Martínez Batlle (2020).

La variación en la composición y abundancia de especies de la comunidad, es medida utilizando el índice multiplicativo de diversidad beta basado en los números de Hill (Borcard et al., 2018). Finalmente, se examina la contribución a la diversidad beta por parte de las especies y los sitios en BCI, al analizar la varianza de la abundancia y riqueza de los cuadrantes y las especies. Estos valores fueron comparados con la varianza promedio, y se considera entonces que las especies cuya varianza promedio supera la mitad de la varianza promedio total, presentan una contribución importante a la diversidad beta de la comunidad (Borcard et al., 2018).

3 Resultados

La familia Rubiaceae en Barro Colorado se encuentra representada por 31 especies y 20 géneros. El género *Psychotria* presenta la mayor riqueza (8 especies). La tabla S1 resume las abundancias de las especies de toda la comunidad, que en total suman 41,838 individuos, con una abundancia media de 65 individuos y mediana ubicada en los 1,350 individuos. El gráfico de mosaicos de la figura 2 presenta la riqueza numérica de las especies por cuadrante. En el mismo se observa la marcada diferencia entre las especies en cuanto a su incidencia. Además, el número de individuos de las especies más abundantes, como *Faramea occidentalis*, se mantiene prácticamente constante en todos los cuadrantes. Por otro lado, la abundancia de toda la comunidad muestra un aparente patrón en la parte centro-occidental de BCI, donde se encuentran los sitios con la mayor abundancia (ver figura 3).

análisis sugiere una posible relación entre la abundancia numérica de especies y la composición del suelo, mostrando relación positiva con valores altos de aluminio y fósforo. Así como negativa, para valores altos de pH y concentraciones de otros elementos.



Figure 4: Panel de correlación de Spearman entre los datos de la comunidad y las variables edafológicas.

Las pruebas de correlación entre los grupos 1 y 2 formulados por completo, resultaron significativas respecto a la variable fósforo. Por otro lado, el contenido de cobre y la abundancia global promedio, es decir, la media correspondiente a todas las plantas en BCI, son significativamente diferentes entre los sitios de ambos grupos, para un nivel de significancia de $\alpha = 0.1$ (ver 5).

El grupo 2 contiene los sitios con tendencia a presentar valores altos de acidez y contenido de aluminio. Es probable que las especies indicadoras del grupo con un mayor contenido de cobre estén mostrando preferencia por estas condiciones ambientales. Indicios de esto, además, pudieron observarse en los valores del índice de correlación de Spearman, el cual indicaba una relación negativamente significativa entre la abundancia y el pH (ver mapas de las figuras 6 y 7). No obstante, el pH y la mayoría de componentes del suelo en BCI tienen valores bastante homogéneos, y más bien se presentan pequeños gradientes entre los cuadrantes, lo cual evita que este tipo de acercamiento sea concluyente.

Las especies *Alseis blackiana* y *Psychotria limonensis* fueron las que obtuvieron un valor alto de confianza al examinar su potencial como especies indicadoras del grupo 1. Para el caso del grupo 2, las especies indicadoras fueron *Faramaea occidentalis*, *Psychotria horizontalis* y *Coussarea curvigemma*. La ocurrencia de *A. blackiana* y *Pentagonia macrophylla* indica su preferencia por el grupo 1. Por otra parte, la muy dominante *F. occidentalis*, *Psychotria deflexa*, *P. racemosa*, *P. horizontalis*, *Posoqueria latifolia*, *Alibertia edulis* y *Coussarea curvigemma* resultan de interés por ocurrir de manera sistemática en el grupo 2.

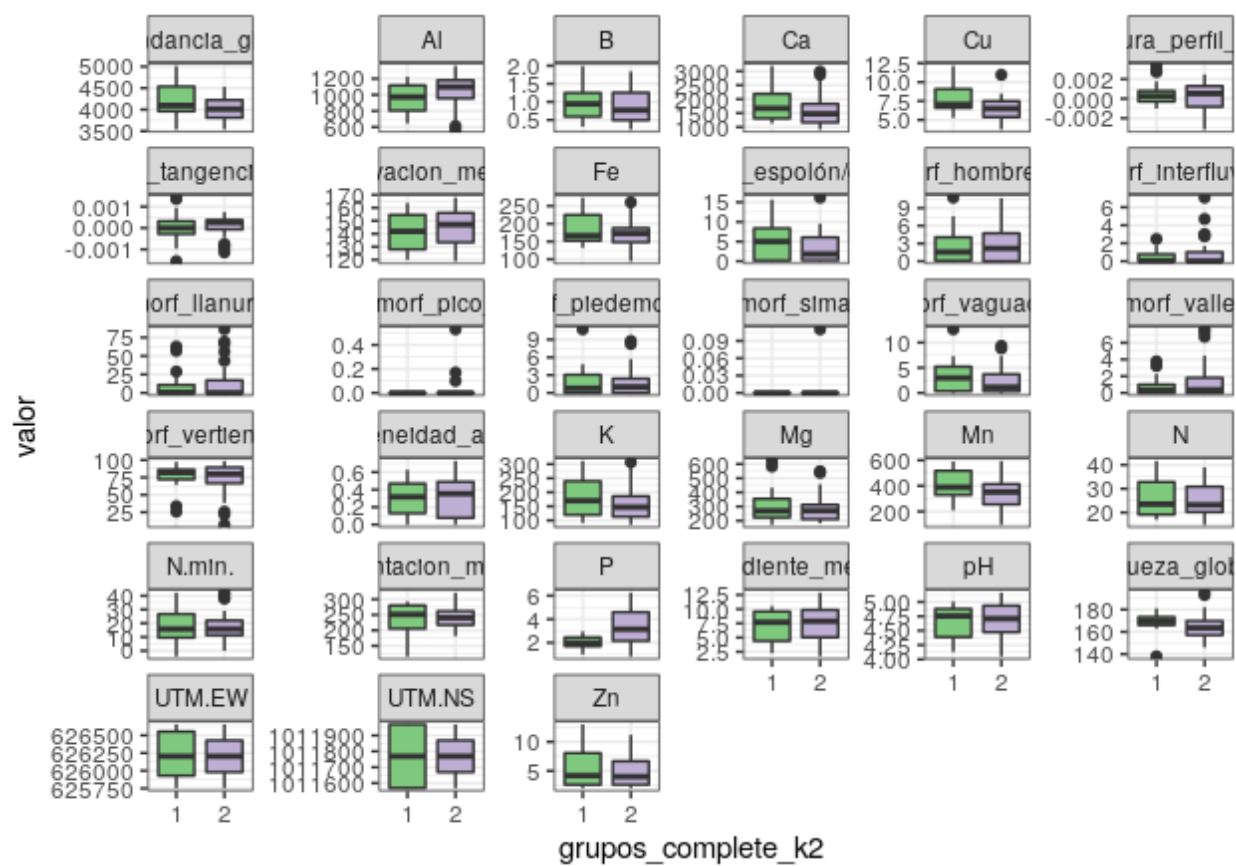


Figure 5: Diagramas de caja de las variables que tuvieron un efecto, según las pruebas de igualdad de medias. Los valores que presentan las medianas de la variable fósforo son muy distintos entre el grupo 1 (verde) y el grupo 2 (gris).

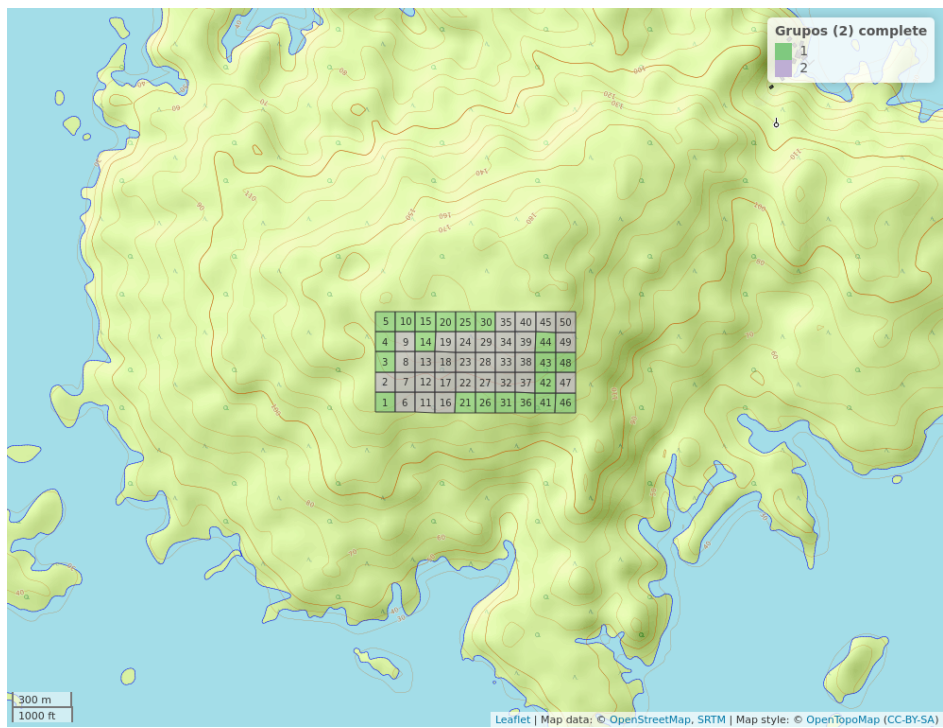


Figure 6: Mapa en el que se presenta la repartición de sitios en los grupos formulados por enlace completo.

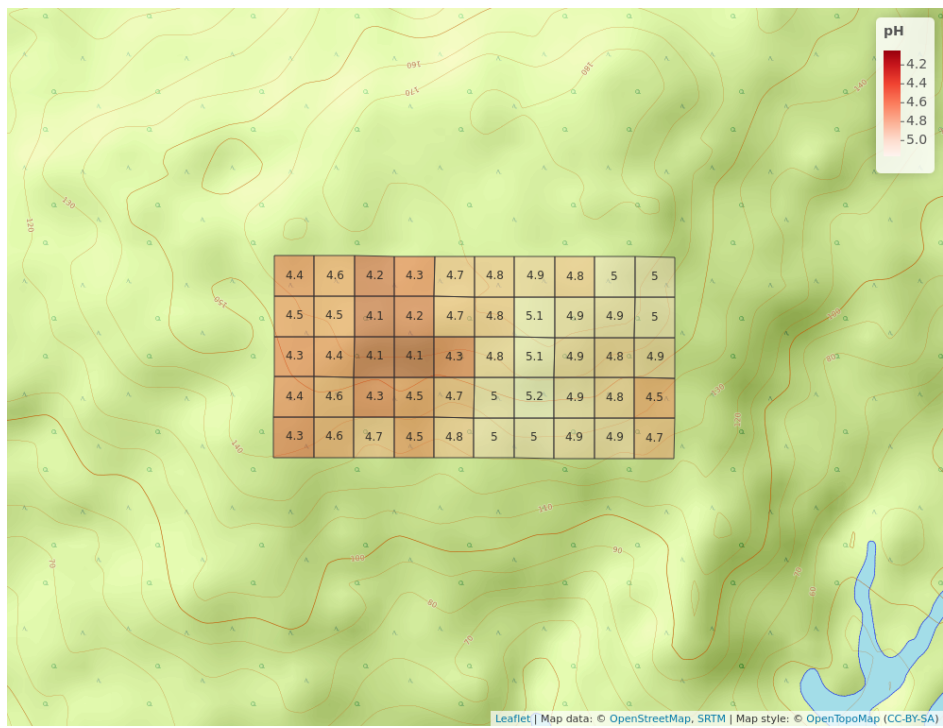


Figure 7: pH del suelo en los cuadros de 1ha

En el diagrama rotulado como escalamiento 1 de la figura 8, se observan tres grupos de cuadrantes diferenciados entre sí. Un grupo de sitios con un alto grado de acidez y contenido en aluminio, otro grupo caracterizado por la presencia de elementos metálicos, y un tercero, con una cantidad de fósforo, nitrógeno y valor de pH mayor. En el caso de las variables geomorfológicas, algunos sitios están asociados a un alto porcentaje de llanura y hombrera, aunque la mayoría se encuentra más cerca del origen formado por los ejes de los componentes principales 1 y 2 (ver figura S2).

Los resultados del PCA de los datos de la matriz de comunidad se encuentran resumidos en los diagramas de la figura 9. El escalamiento 1, muestra muchos de los cuadrantes dispuestos alrededor del origen formado por los ejes, lo que indica una contribución a la varianza relativamente equitativa por parte de las especies. Sin embargo, aparecen también unos cuantos cuadrantes con valores atípicos y más alejados. Se nota como las especies *A. blackiana*, *C. curvigemmia*, *P. marginalis* y *P. horizontalis*, presentan una contribución desproporcionada a la varianza total, en comparación con el resto de las especies.

El escalamiento 2 de la figura 10 en el análisis de correspondencia mostró que las especies *Psychotria graciliflora* y *Psychotria grandis* se encuentran asociadas. Ambas especies, además, tienen valores de abundancia parecidos dentro de la comunidad (65 y 57 individuos, respectivamente). Casi todas las demás especies se encuentran próximas al punto de intersección, salvo aquellas que presentaron una abundancia reducida, y en consecuencia, aparecen cercanas a los pocos cuadrantes en los que se encuentran representadas. La disparidad en la incidencia de las especies se refleja en su disposición en el diagrama. Sin embargo, estos resultados no coinciden del todo con los arrojados por el PCA de la matriz de distancias.

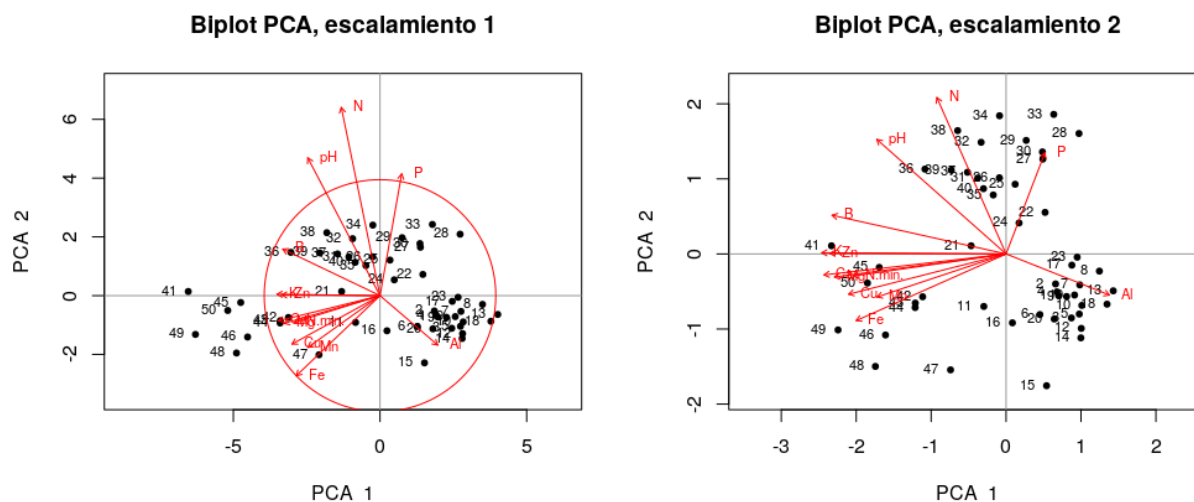


Figure 8: Biplots generados en el PCA de las variables de suelo. Se observa que las variables nitrógeno, fósforo y pH aportan la mayor parte de la varianza explicada. La relación entre las variables se encuentra debidamente representada en el recuadro del escalamiento 2, por medio de los ángulos que forman sus vectores.

Se observa asociación entre el contenido de cobre, manganeso y fósforo con algunos sitios y especies, en el *biplot* del análisis de coordenadas principales de la figura S3. También, algunas de las especies menos abundantes de la comunidad se presentan asociadas a algunos sitios comunes entre las mismas. Es probable que esta aparente asociación aparezca debido a la combinación de una incidencia restringida por parte de estas especies y al alto grado de autocorrelación entre los cuadrantes.

Se destacan los patrones encontrados en los resultados de las pruebas de asociación y los arrojados por el análisis de redundancia. En los *triplots* producidos por RDA (figura 11), se observó que *Psychotria horizontalis* y *Psychotria marginata* se encuentran asociadas a un grupo de cuadrantes, los que a su vez, constituyen un microhabitat encontrado sobre una pequeña meseta en la parte nor-oriental de BCI, y se diferencian además, por una elevada concentración de fósforo y aluminio en el suelo. Por otro lado, la especie *Alseis blackiana* presenta asociación con el contenido de metales como manganeso, hierro y cobre. Esto último coincide con los resultados vistos en el PCA, lo que demuestra cierto grado de preferencia hacia estas condiciones ambientales por parte de esta especie.

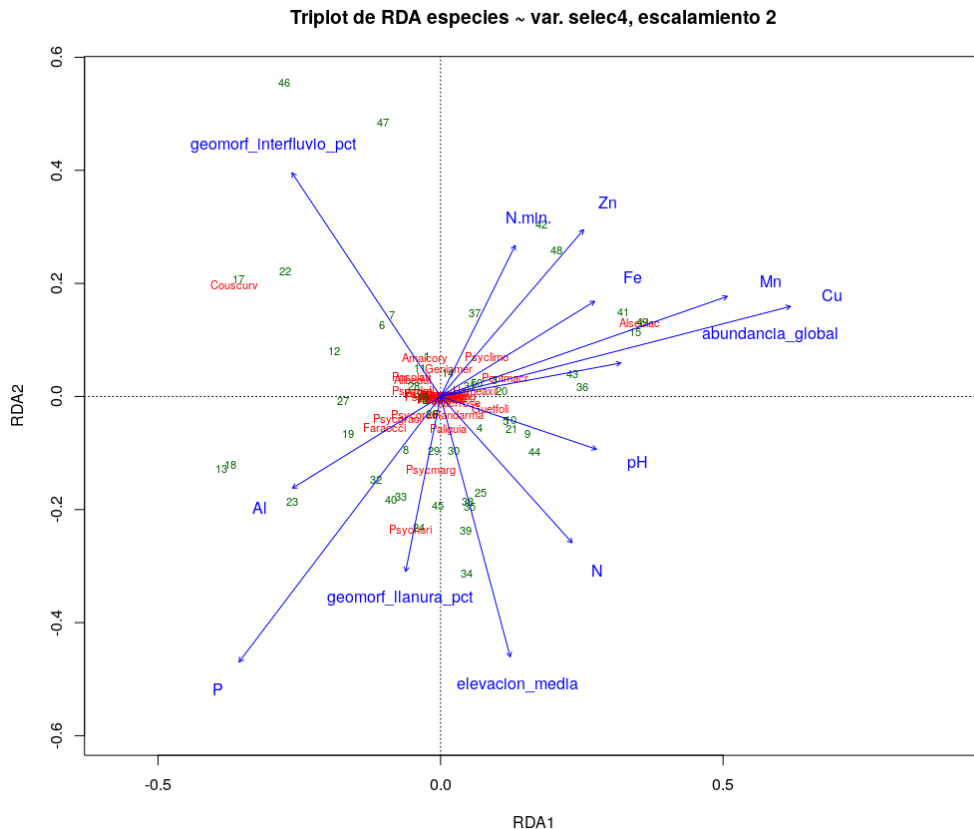


Figure 11: Triplot del RDA de los valores de abundancia transformados. Las variables ambientales que se observan son aquellas que fueron seleccionadas para restringir la varianza explicada de los datos de comunidad.

La riqueza de la familia Rubiaceae aumenta en función del contenido de hierro, nitrógeno y nitrógeno mineralizado. La equidad pareciera también estar asociada con estas variables, además de que la misma es mayor hacia el norte de BCI (13).

Aparte del cuadrante #50, los sitios con la mayor riqueza presentan también mucha abundancia. El cuadrante #1, con la menor riqueza, presenta una de las mayores dominancias. Sucede algo similar en el cuadrante #28, aunque en este caso el valor riqueza de especies presentado no es tan bajo, sino que en comparación, contiene una desorbitada abundancia (1414 individuos).

Las especies que contribuyen de manera significativa a la diversidad beta, fueron las que también resultaron ser indicadoras debido a su incidencia en los grupos 1 y 2. Con la excepción de la especie *Macrocnemum roseum*, la cual tiene una abundancia mucho menor que la de las demás especies que resultan de este análisis.

La riqueza de especies (número de Hill 0 en el gráfico de la figura 12) presenta la mayor contribución a la medida de diversidad beta para la comunidad. Así sucede, ya que la abundancia de las especies dominantes no varía drásticamente entre cuadrantes, y es precisamente su composición de especies lo que diferencia a los sitios entre sí.

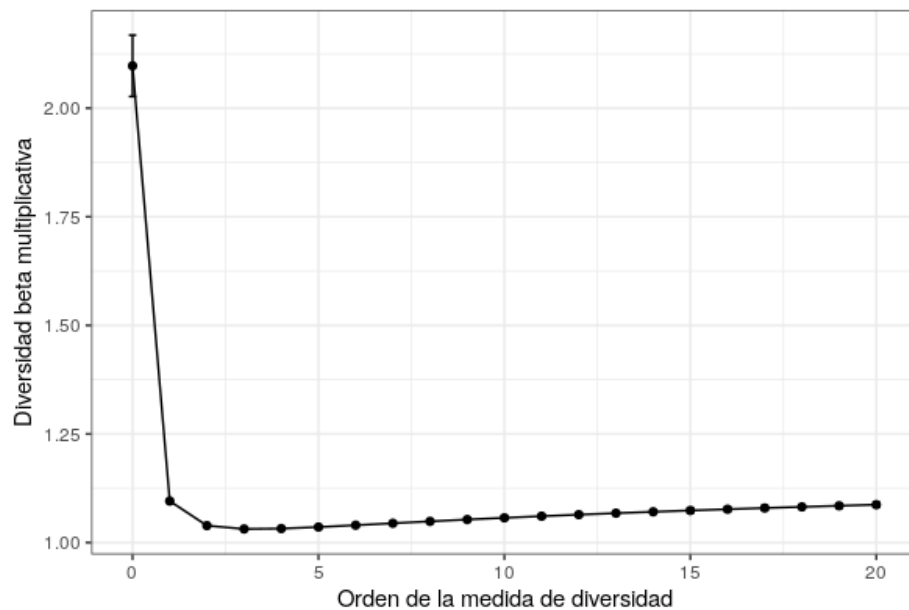


Figure 12: Gráfico de la diversidad beta multiplicativa en función de los números de diversidad en la serie de Hill.

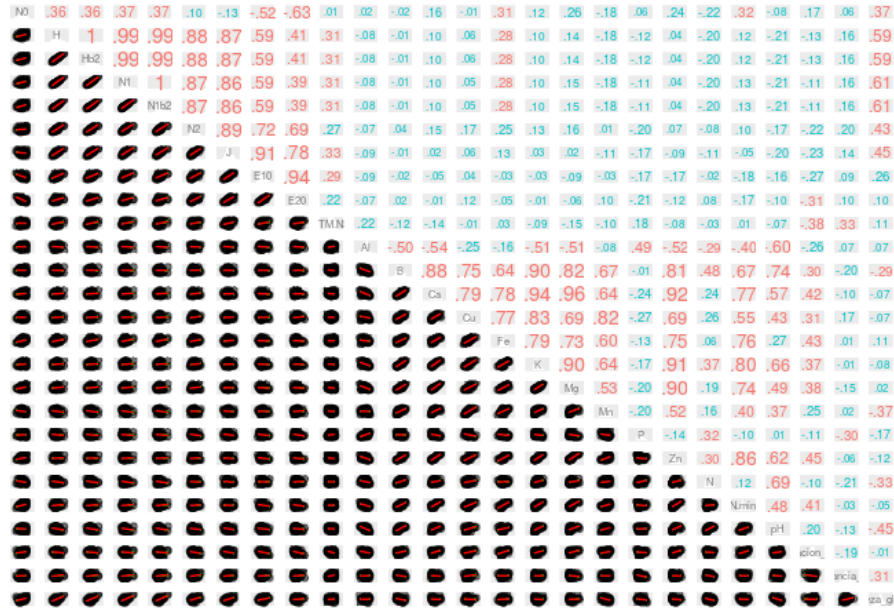


Figure 13: Correlación entre diversidad/equidad y algunas de las variables ambientales destacadas. N0: riqueza de especies; H: entropía de Shannon; Hb2: entropía de Shannon con 2 como base del logaritmo; N1 y N2: Números de Hill; N1b2: Número de Hill 1 en base log 2; J: Equidad de Pielou; E10 y E20: ratios de Hill 1 y 2.

4 Discusión

Se confirma que variables ambientales como el contenido de aluminio y fósforo en el suelo, tienen un efecto al caracterizar la abundancia relativa de las especies de la familia Rubiaceae. Adicionalmente, esta comunidad podría dividirse en dos grupos más o menos diferenciados entre sí por su preferencia, o más bien tolerancia, al contenido de cobre en el suelo. Otros autores se han referido con anterioridad a la tolerancia a los metales pesados que presentan algunas de las especies de esta familia (McAlister, Kolterman, & Pollard, 2015). Lo que hace remarcable el hecho de que pruebas de este fenómeno puedan apreciarse a un nivel muy local.

Algunas de las pruebas estadísticas empleadas en este estudio requieren del supuesto de independencia entre muestras, supuesto que es incumplido por los cuadrantes en BCI debido a su contigüidad. Sin embargo, esta condición permitió un acercamiento comparativo con las técnicas apropiadas para el análisis de muestras con cierto grado asociación entre sí. Además, en BCI, tanto las variables ambientales como las comunidades de plantas se encuentran correlacionadas espacialmente. El área correspondiente a la parcela es bastante homogénea en cuanto a sus condiciones físicas, ya que fué diseñada para que así fuera (Baillie, Elsenbeer, Barthold, Grimm, & Stallard, 2006). Como la misma se encuentra situada sobre el dorso de una cuesta de andesita, las variables edafológicas aquí presentes en su mayoría siguen un gradiente a lo largo de todo BCI, resultado de la meteorización de la roca original (Patino, Velbel, Price, & Wade, 2003). Por lo tanto, es razonable considerar que los patrones encontrados entre métodos contrastados son evidencia del grado de importancia que presentan las condiciones ambientales (especialmente las relacionadas con la composición química del suelo) en la estructuración de esta comunidad de especies en BCI.

Las especies con preferencia por el grupo 1 evaden sistemáticamente al grupo 2 y viceversa.

Esta diferenciación en la ocurrencia de las especies menos comunes no se explica del todo con las variables ambientales medidas en BCI y es posible que exista algún otro factor que determina la ordenación de esta comunidad de plantas. La cobertura del dosel arbóreo podría ser una de estas variables a considerarse, puesto que las especies dominantes en la comunidad son a la vez tolerantes a la sombra (Grandtner & Chevrette, 2013).

Los resultados expuestos en este trabajo muestran como el estudio de la estructura de las comunidades ecológicas puede presentar complicaciones cuando nos enfrentamos a taxones muy diversos, para los cuales las medidas de abundancia relativa convencionales podrían ver su eficacia limitada (Ricotta, 2004). Además, los análisis de agrupamiento aquí realizados pudieron haberse visto sesgados por la heterogeneidad morfométrica que presentan las especies de Rubiaceae. Estas plantas pueden tener diversos hábitos de crecimiento, desde porte herbáceo y arbustivo a árboles relativamente grandes. Esto hace que el hecho de que se incluya el criterio de dap de 10 mm en el momento de ser censadas podría estar excluyendo especies clave en el rompecabezas.

Información de soporte

Table S1: Abundancia total por especie.

Latin	n
<i>Faramea occidentalis</i>	24989
<i>Alseis blackiana</i>	7928
<i>Psychotria horizontalis</i>	2453
<i>Coussarea curvigemma</i>	2010
<i>Palicourea guianensis</i>	1118
<i>Randia armata</i>	937
<i>Psychotria marginata</i>	761
<i>Alibertia edulis</i>	417
<i>Pentagonia macrophylla</i>	306
<i>Guettarda foliacea</i>	252
<i>Hamelia axillaris</i>	128
<i>Macrocnemum roseum</i>	87
<i>Posoqueria latifolia</i>	73
<i>Psychotria limonensis</i>	70
<i>Genipa americana</i>	67
<i>Psychotria graciliflora</i>	65
<i>Psychotria grandis</i>	57
<i>Psychotria deflexa</i>	38
<i>Amaioua corymbosa</i>	19
<i>Psychotria chagrensis</i>	16
<i>Psychotria acuminata</i>	14
<i>Tocoyena pittieri</i>	8
<i>Psychotria racemosa</i>	7
<i>Psychotria cyanococca</i>	4
<i>Chimarrhis parviflora</i>	3
<i>Coutarea hexandra</i>	3
<i>Psychotria brachiata</i>	3

Latin	n
<i>Appunia seibertii</i>	2
<i>Borojoa panamensis</i>	1
<i>Psychotria hoffmannseggiana</i>	1
<i>Rosenbergiodendron formosum</i>	1

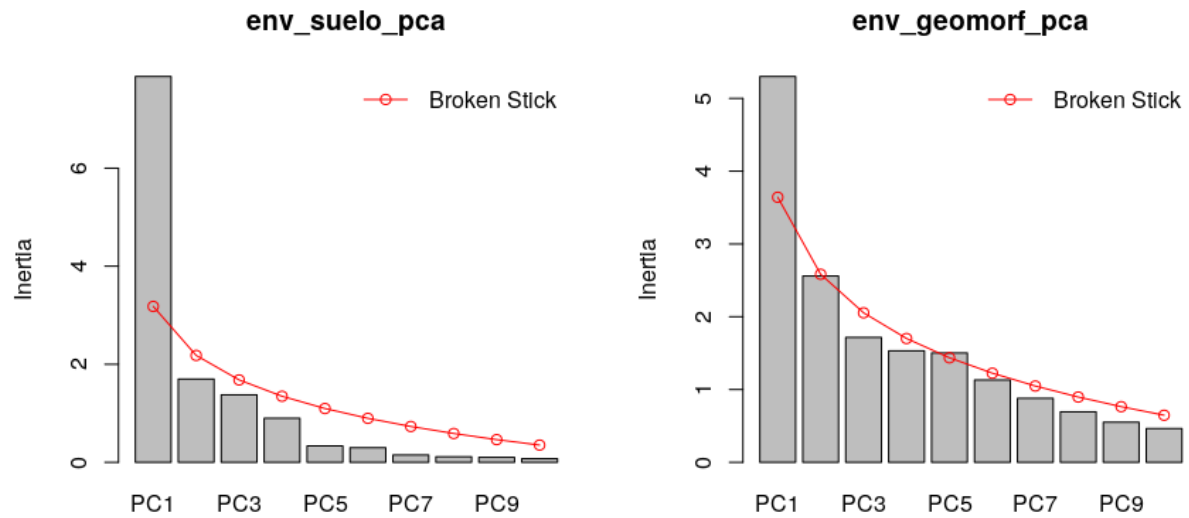


Figure S1: Componentes principales de la varianza en las variables de suelo y geomorfología en BCI. En estos gráficos se incluye el comportamiento de la varianza explicada, predecido por el modelo de vara quebrada, representado por la línea roja formando la curva. (La escala denominada “Inertia” representa la suma de los cuadrados de toda la varianza).

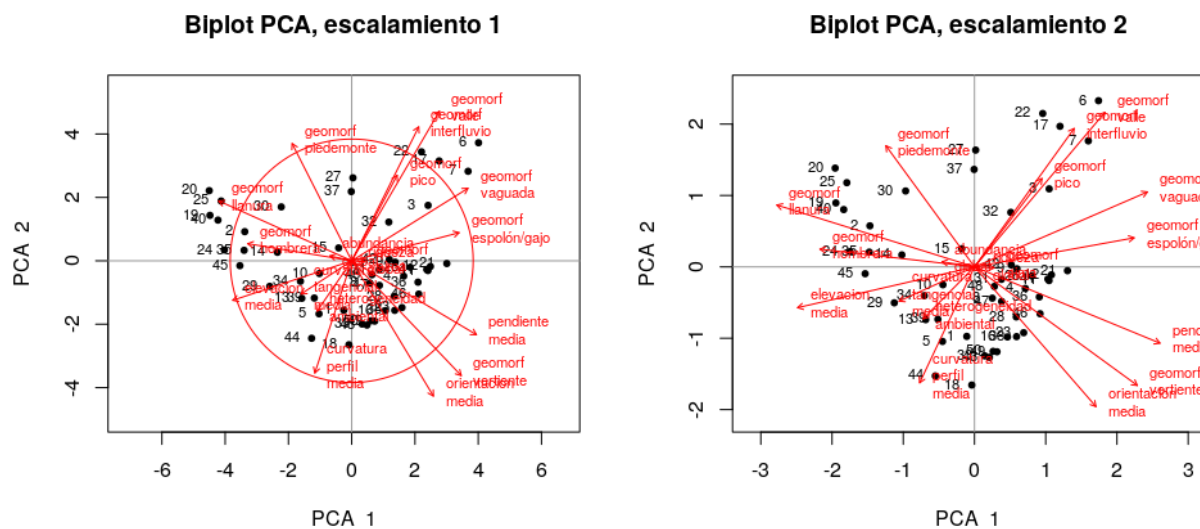


Figure S2: Biplots generados por PCA de las variables geomorfológicas.

PCoA con promedios ponderados de especies

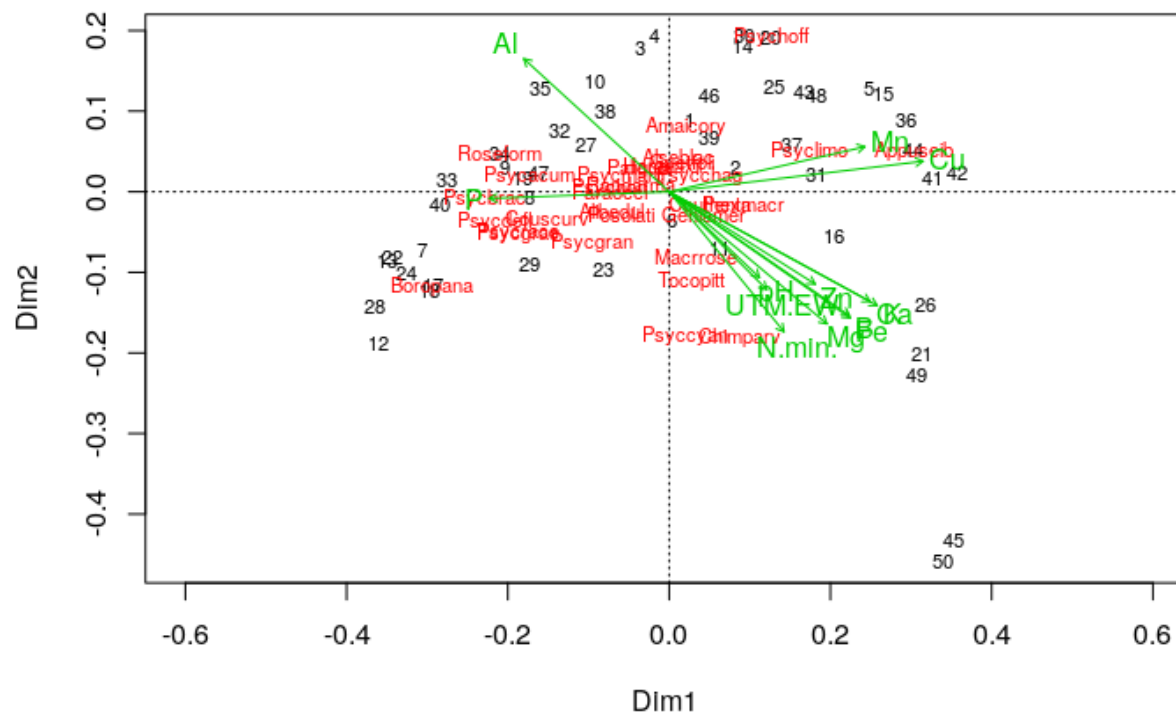


Figure S3: Biplot de PCoA de las distancias de Jaccard. Las distancias entre especies están ponderadas en base a sus valores de abundancia.

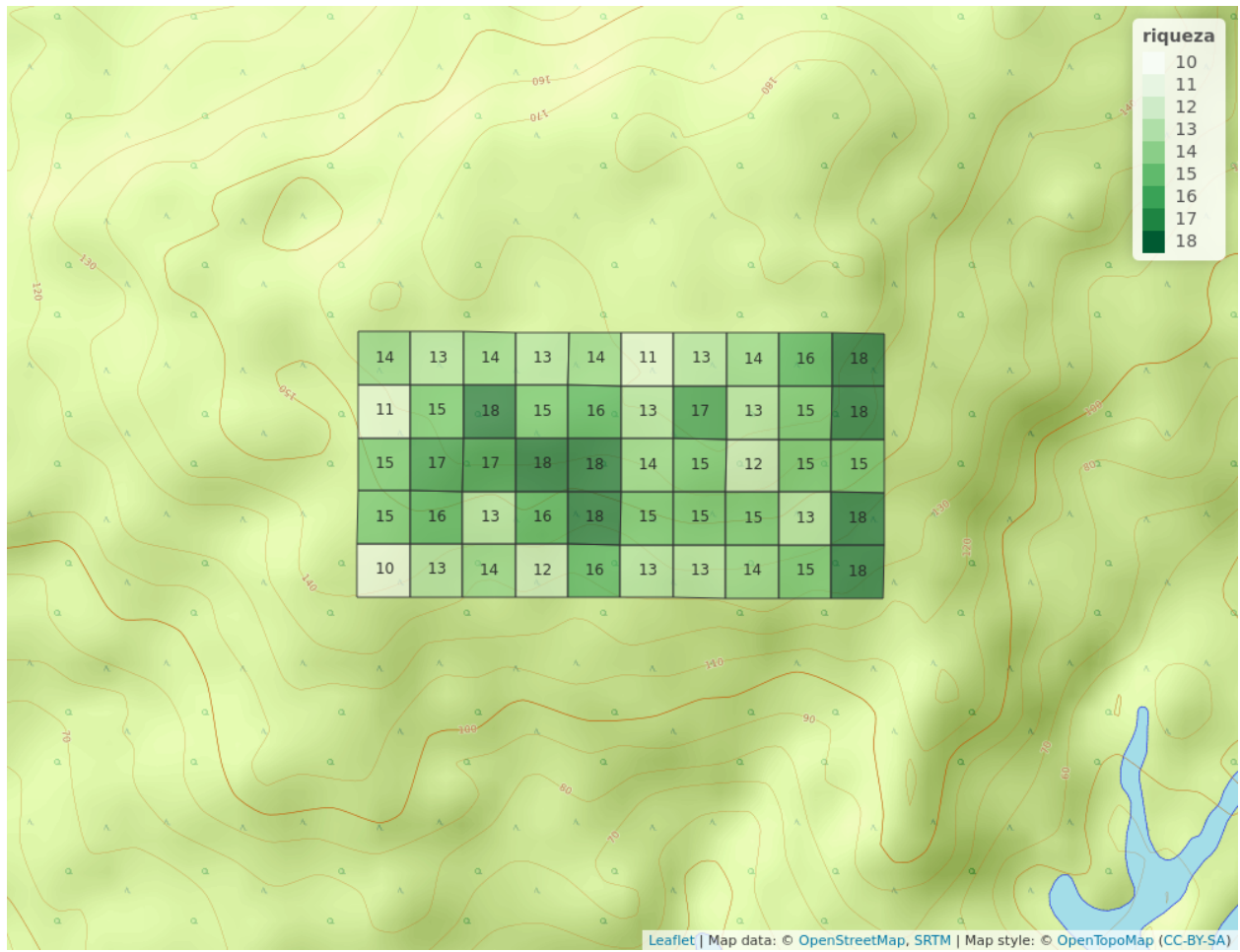


Figure S4: Distribución de la riqueza de rubiaceas en BCI

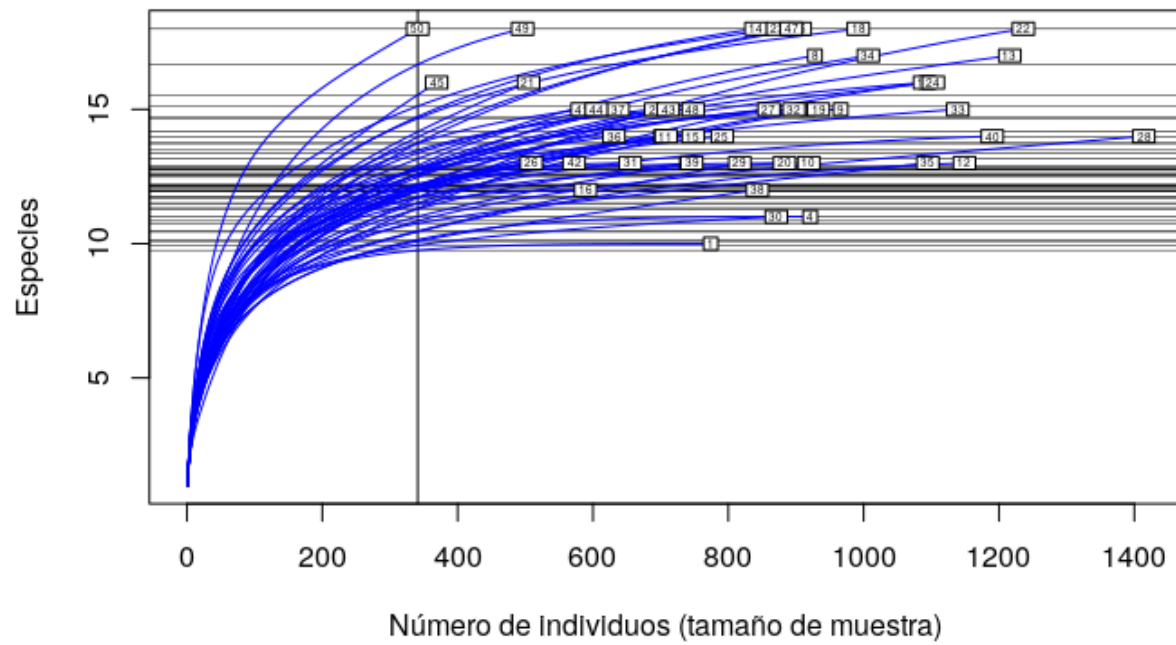


Figure S5: Curva de rarefacción. Las barras horizontales muestran el valor de la riqueza esperada en cada sitio al interpolar las muestras hacia la abundancia mínima observada (341 individuos), presentada por el cuadrante #50.

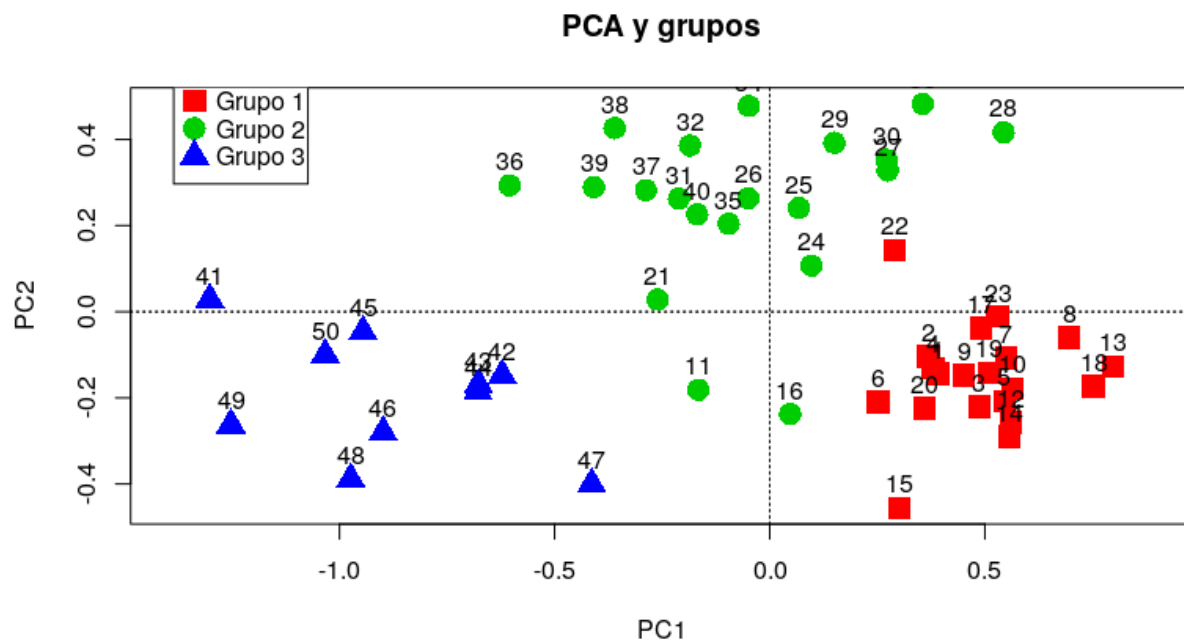


Figure S6: Diagrama PCA de los grupos de cuadrantes según variables de suelo

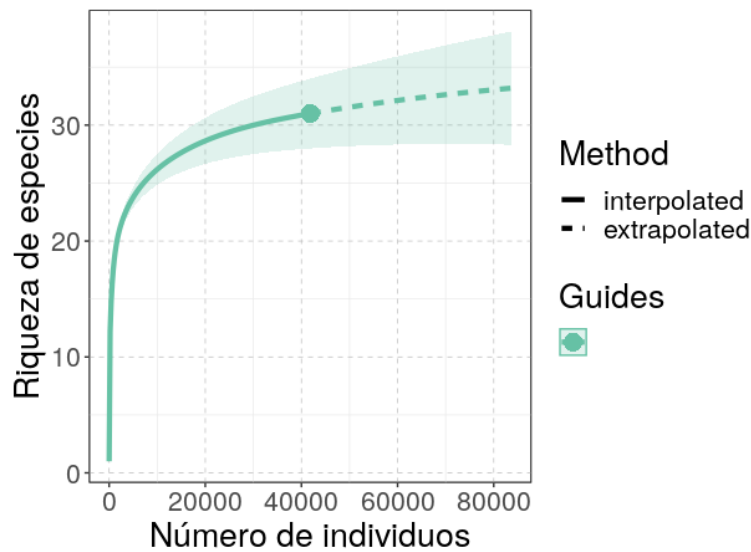


Figure S7: Curva de acumulación de especies hacia el doble de la abundancia encontrada en BCI. Se estima que la riqueza aumentaría en cuatro especies en consecuencia de duplicar el censo, o replicarlo en otra zona de Barro Colorado. El intervalo para un 95% de confianza se representa con los márgenes verde pálido.


```

#' ### Abundancia por especie
sort(colSums(mc_rubic))
summary(colSums(mc_rubic)) # Resumen estadístico

#' ### Riqueza numérica de toda la "comunidad"
specnumber(colSums(mc_rubic))

#' ### Abundancia de toda la comunidad
sum(colSums(mc_rubic))

#' ### Una tabla para el manuscrito, es necesario asignarle nombre
#' Para esto, usaré la colección "tidyverse"
abun_sp <- censo_rubic %>%
  group_by(Latin) %>%
  count() %>%
  arrange(desc(n))
abun_sp

#' ### Un gráfico para el manuscrito
#' Gráfico de mosaicos de la abundancia por especie por cuadros
abun_sp_q <- crear_grafico_mosaico_de_mc(mc_rubic, tam_rotulo = 7)
abun_sp_q

```

aed5_correlacion_variables_ambientales.R

```

#' ---
#' title: "Análisis exploratorio de datos. Correlaciones entre variables ambientales"
#' author: "JR"
#' date: "25 de octubre, 2020"
#' output: github_document
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ### Cargar paquetes
library(tidyverse)
library(sf)
library(ez)
library(psych)
library(vegan)
library(graphics)

#' ### Cargar datos
load('biodata/matriz_ambiental.Rdata')
load('biodata/Rubiaceae.Rdata')

```

```

#' ### Una correlación simple
cor(bci_env_grid$pendiente_media, bci_env_grid$geomorf_vertiente_pct)
plot(bci_env_grid$pendiente_media, bci_env_grid$geomorf_vertiente_pct)
cor.test(bci_env_grid$pendiente_media, bci_env_grid$geomorf_vertiente_pct)

#' ### Generar objeto de columnas numéricas
#' El objeto que generaré, denominado 'env_num', no tendrá las columnas 'id' y las de coordenada
env_num <- bci_env_grid %>%
  dplyr::select_if(is.numeric) %>%
  dplyr::select(-id, -matches('^U.*')) %>%
  st_drop_geometry %>%
  mutate(
    riqueza_mifam = specnumber(mc_rubic),
    abundancia_mifam = rowSums(mc_rubic)) %>%
  rename_all(gsub, pattern = '_pct$', replacement = '') %>%
  rename_all(gsub, pattern = '_| ', replacement = '\\n')
env_num %>% tibble

#' ### Panel de correlaciones con herramientas del paquete 'graphics' y 'psych'
cor(env_num)
ncol(env_num)
pairs(env_num[,sample(1:33, 15)]) # paquete graphics
env_num[,sample(1:33, 15)] %>% pairs.panels #paquete psych

#' ### Panel de correlaciones con 'ez'
#'
#' ##### Todas las variables (se empasta). Comentado, sólo mostrado para fines didácticos
# p_cor_todos <- env_num %>%
#   ezCor(r_size_lims = c(4,8), label_size = 4)
# p_cor_todos

#' ##### Sólo suelo (elementos y pH), abundancia/riqueza
p_cor_suelo_ar <- env_num %>%
  dplyr::select(matches('^([A-T,Z]|abundancia|riqueza|^pH$'), ignore.case = F)) %>%
  ezCor(r_size_lims = c(4,8), label_size = 3)
p_cor_suelo_ar

#' ##### Sólo heterogeneidad, geomorfología, abundancia/riqueza
p_cor_geomorf_ar <- env_num %>%
  dplyr::select(-matches('^([A-T,Z]|pH$'), ignore.case = F)) %>%
  ezCor(r_size_lims = c(4,8), label_size = 3)
p_cor_geomorf_ar

#' ##### Matriz de comunidad
p_cor_mc_rubic <- mc_rubic %>%
  rename_all(gsub, pattern = '_| ', replacement = '\\n') %>%
  ezCor(r_size_lims = c(4,8), label_size = 2)

```

mapas_abundancia_riqueza_rubiaceae.R

25

```

mapa_cuadros_abun_global <- mapView(
  bci_env_grid,
  layer.name = 'abundancia',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T, zoom = 14,
  col.regions = azul,
  zcol = 'abundancia_global') %>%
addStaticLabels(label = bci_env_grid$abundancia_global, textsize = "7pt") %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 15)
mapa_cuadros_abun_global
mapa_cuadros_abun_global %>% mapshot(file = 'mapa_cuadros_abun_global.png')

#' ### Mapa de cuadros, simbología por riqueza global
mapa_cuadros_riq_global <- mapView(
  bci_env_grid,
  layer.name = 'riqueza',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T, zoom = 14,
  col.regions = rojo,
  zcol = 'riqueza_global') %>%
addStaticLabels(label = bci_env_grid$riqueza_global, textsize = "7pt") %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 15)
mapa_cuadros_riq_global
mapa_cuadros_riq_global %>% mapshot(file = 'mapa_cuadros_riq_global.png')

#' ### Mapa de cuadros, simbología por abundancia de mi familia
mapa_cuadros_abun_rubic <- mapView(
  bci_env_grid %>% mutate(abun = rowSums(mc_rubic)),
  layer.name = 'abundancia',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T, zoom = 14,
  col.regions = azul,
  zcol = 'abun') %>%
addStaticLabels(label = rowSums(mc_rubic), textsize = "7.3pt") %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 16)

```

```

mapa_cuadros_abun_rubic
mapa_cuadros_abun_rubic %>% mapshot(file = 'mapa_cuadros_abun_rubic.png')

#' ### Mapa de cuadros, simbología por riqueza de mi familia
mapa_cuadros_riq_rubic <- mapView(
  bci_env_grid %>% mutate(riq = specnumber(mc_rubic)),
  layer.name = 'riqueza',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T, zoom = 14,
  col.regions = verde,
  zcol = 'riq') %>%
  addStaticLabels(label = specnumber(mc_rubic), textsize = '11') %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15.9)
mapa_cuadros_riq_rubic
mapa_cuadros_riq_rubic %>% mapshot(file = 'mapa_cuadros_riq_rubic.png')

```

intento_correlacion_densidad.R

```

bci_env_density <- bci_env_grid %>%
  mutate(area= st_area(geometry), densidad_indiv = abundancia_global/area)
env_num_density <- bci_env_grid_density %>%
  dplyr::select(-id, -matches('^U.*')) %>%
  st_drop_geometry %>%
  mutate(
    riqueza_rubic = specnumber(mc_rubic),
    abundancia_rubic = rowSums(mc_rubic)) %>%
  rename_all(gsub, pattern = '_pct$', replacement = '') %>%
  rename_all(gsub, pattern = '_| ', replacement = '\\n')
##
p_cor_env_num_density_suelo_spear <- env_num_density %>%
  dplyr::select(matches('^[A-T,Z]|abundancia|densidad|riqueza|^pH$', ignore.case = F)) %>%
  ezCorM(r_size_lims = c(4,8), label_size = 3, method = 'spearman')
p_cor_env_num_density_suelo_spear

png(
  filename = 'p_cor_env_density_suelo_spear.png',
  width = 1920, height = 1080, res = 125
)
p_cor_env_num_density_suelo_spear
dev.off()
##dplyr::select_if(is.numeric) %>%

```

medicion_asociacion.R

```
#' ---
#' title: "Medición de asociación. Modo Q aplicado a mi familia asignada"
#' author: "JR"
#' date: "9 de noviembre, 2020"
#' output: github_document
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ## Preámbulo

#' ### Cargar paquetes
library(vegan)
library(adespatial)
library(broom)
library(tidyverse)
library(sf)
library(cluster)
library(gclus)
source('biodata/funciones.R')

#' ### Cargar datos
#'
load('biodata/matriz_ambiental.Rdata')
load('biodata/Rubiaceae.Rdata')
#'
#' ## Modo Q: matrices de disimilaridad entre objetos
#'
#' ### Modo Q para datos cuantitativos de especies (abundancia). Datos de mi familia asignada
#'
#' Aplicado a mi familia asignada de BCI, en la forma de matriz de distancia euclídea, utilizando
#'
rubic_d_hel <- dist.ldc(mc_rubic, "hellinger", silent = T)
rubic_d_hel %>% tidy # Para evitar desbordar la consola
#'
#' Para interpretar esta matriz, es necesario representarla gráficamente. En la representación e
#'
coldiss(rubic_d_hel, diag = T)
#'
#' Mejorable el gráfico, quizá este es más explícito:
#'
coldissgg(rubic_d_hel, ordered = T, nc = 4, fsz = 0)
#'
#' Con valores de distancia sobreimpresos (se empastan un poco)
#'
coldissgg(rubic_d_hel, ordered = T, nc = 4, fsz = 1.5)
```

```

#’
#’ Puedes guardar el gráfico usando el botón ‘Export’ de la pestaña ‘Plots’
#’
#’ Una forma alterna de guardar el gráfico es mediante funciones de R. La calidad de gráficos ex
#’
png(
  filename = 'matriz_disimilaridad_hellinger.png',
  width = 2400, height = 1200, pointsize = 35
)
coldiss(rubic_d_hel, diag = T,)
dev.off()
#’
#’ MUY IMPORTANTE. La última función, ‘dev.off()’, es necesaria para cerrar el dispositivo. Si n
#’
#’ ### Modo Q para datos binarios (presencia/ausencia)
#’
#’ Habitualmente, sólo dispones de datos de presencia/ausencia. En tales casos, existe un conjun
#’
#’ Un error común consiste en referirse a los índices de Jaccard y de Sorensen "a secas", sin es
#’
#’ Si alguna vez te enfrentas a textos donde no se especifica qué tipo de métrica se usa, te sug
#’
#’ Recalco: es imprescindible declarar qué tipo de métrica estás usando. Ejemplos de redacci
#’
#’ - Correcto: "índice de disimilaridad de Jaccard", "índice de similaridad de Sorensen"
#’
#’ - Incorrecto: "índice de Jaccard", "índice de Sorensen".
#’
#’ A continuación, muestro cómo calcular la distancia de Jaccard ( $D_{J}$ ) en un ú
#’
rubic_jac <- vegdist(mc_rubic, method = 'jac', binary = T)
rubic_jac %>% tidy # Mostrando sólo las primeras 10 combinaciones en modo data.frame
#’
#’ El argumento ‘binary=T’ en ‘vegdist’ "ordena" que se realice primero ‘decostand(mc_apcyn_meli
#’
#’ En esta matriz de disimilaridad, al igual que en la anterior, un valor pequeño (rosa) signifi
#’
coldiss(rubic_jac, diag = T)
#’
#’ La distancia de Jaccard ( $D_{J}$ ) se puede expresar como "la proporción de especies
#’
#’ Como la distancia de Jaccard ( $D_{J}$ ) es el complemento a 1 de la similaridad de J
#’
(1 - rubic_jac) %>% tidy %>% rename(similaridad=distance) #Similaridad
#’
#’ Dado que este resultado muestra la similaridad, podemos leerlo como "el sitio 1 y el 2 compar
#’

```



```

#’ La fórmula de la similaridad de Jaccard es  $S_{J} = a/(a+b+c)$ , donde a es el número
#’
#’ Para obtener las variables a, b y c, usaré La función ‘betadiver’ del paquete ‘vegan’
#’
mi_fam_abc <- betadiver(mc_apcyn_melic_saptc)
mi_fam_abc %>%
  map(tidy) %>%
  map(slice, 1) %>%
  map_df(I, .id = 'tipo') %>%
  dplyr::select(tipo, n_especies=distance)
#’
#’ Puedes notar que ambos sitios comparten 11 especies (a), que el sitio 2 no tiene especies
#’
round(11/12*100,2) #Porcentaje de especies compartidas = similaridad
#’
#’ Con ‘betadiver’ también puedes calcular índices de similaridad. Por ejemplo, el Jaccard se calcula como
#’
betadiver(mc_rubic, method = 'j') %>% tidy
#’
#’ No obstante, usaremos esta función en los análisis de diversidad beta más adelante.
#’
#’ Además de la distancia de Jaccard, otra distancia muy utilizada es la de Sorensen o Bray-Curtis
#’
rubic_sor <- vegdist(mc_rubic, method = 'bray', binary = T)
rubic_sor %>% tidy
coldiss(rubic_sor, diag = T)
#’
#’ ### Modo Q para datos cuantitativos, NO de abundancia de especies (variables ambientales)
#’
#’ En este ejemplo, usaré sólo variables de suelo, todas cuantitativas, puedes combinar con otras
#’
env_suelo_punt_z <- bci_env_grid %>%
  st_drop_geometry() %>%
  dplyr::select(matches('^A-T,Z|~pH$'), ignore.case = F) %>%
  scale()
env_suelo_punt_z_d <- dist(env_suelo_punt_z)
env_suelo_punt_z_d %>% tidy
coldiss(env_suelo_punt_z_d, diag = T)
#’
#’ Al y ph
#’ ### Modo Q para datos cualitativos y cuantitativos (mixtos), NO de abundancia de especies (variables ambientales)
#’
#’ En este ejemplo, usaré las siguientes variables mixtas (funciona igualmente para datos cualitativos)
#’
#’ - ‘heterogeneidad_ambiental’. Índice cuantitativo calculado como la diversidad de Simpson a nivel de sitio
#’
#’ - ‘habitat’. Tipo de hábitat. Asume los siguientes valores posibles: *OldHigh*, *OldLow* y *OldMed*

```

```

#'
#' - 'quebrada'. Informa sobre si hay o no quebrada. Los valores posibles son *Yes* o *No*.
#'
env_mix <- bci_env_grid %>%
  st_drop_geometry() %>%
  dplyr::select(heterogeneidad_ambiental, habitat, quebrada)
env_mix_d <- daisy(x = env_mix, metric = 'gower')
env_mix_d %>% as.dist %>% tidy
env_mix_d %>% coldiss(diag = T)
#'

```

medicion_asociacion_especies.R

```

#' ---
#' title: "Medición de asociación. Modo R aplicado a mi familia asignada"
#' author: "JR"
#' date: "3 de noviembre, 2020"
#' output: github_document
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ## Preámbulo

#' ### Cargar paquetes
library(vegan)
library(adespatial)
library(broom)
library(tidyverse)
library(sf)
library(gclus)
source('biodata/funciones.R')

#' ### Cargar datos
#'
load('biodata/matriz_ambiental.Rdata')
load('biodata/Rubiaceae.Rdata')
#'
#' ## Modo R: matrices de dependencia entre variables (índice de correlación)
#'
#' ### Modo R para datos cuantitativos de especies (abundancia)
#'
#' En este caso, las variables usaré los valores de abundancias de especies como variables. Es d
#'
#' Aunque se podría usar el índice de correlación como métrica de la dependencia (tal como mostr
#'

```

```

#' Por tal razón, es recomendable aplicar la transformación *Chi* a la matriz de comunidad trans
#'
#' Primero, sustituyo el caracter de espacio por un <enter> en los nombres de las especies (cara
#'
mi_fam_t <- mc_rubic %>%
  rename_all(gsub, pattern = ' ', replacement = '\n') %>%
  t()
mi_fam_t %>% tibble
#'
#' Segundo, transformo la matriz transpuesta usando estandarización *Chi*.
#'
mi_fam_t_chi <- decostand(mi_fam_t, "chi.square")
mi_fam_t_chi %>% tibble
#'
#' Tercero, calculo la distancia euclídea.
#'
mi_fam_t_chi_d %>% tidy
#'
#' Finalmente, creo el "mapa de calor".
#'
coldiss(mi_fam_t_chi_d, diag = TRUE)
#'
#' En el mapa de calor **ordenado** (el de la derecha), se identifica al menos un patrón de depe
#'
#' ### Modo R para datos binarios (presencia/ausencia)
#'
#' Arriba usé la distancia de Jaccard para evaluar asociación entre sitios. Dicha métrica tambié
#'
mi_fam_t_jac <- vegdist(mi_fam_t, "jaccard", binary = TRUE)
mi_fam_t_jac %>% tidy
coldiss(mi_fam_t_jac, diag = TRUE)

#' ### Modo R para datos cuantitativos, NO de abundancia de especies (variables ambientales)
#' En modo R evaluás asociación entre descriptores, es decir, entre variables. La métrica común
#'
#' En este ejemplo, mostraré la correlación entre variables de suelo y la abundancia y riqueza g
#'
env_num <- bci_env_grid %>%
  dplyr::select_if(is.numeric) %>%
  dplyr::select(-id, -matches('^U.*')) %>%
  st_drop_geometry %>%
  mutate(
    riqueza_mifam = specnumber(mc_rubic),
    abundancia_mifam = rowSums(mc_rubic) %>%
    rename_all(gsub, pattern = '_pct$', replacement = '') %>%
    rename_all(gsub, pattern = '_| ', replacement = '\n')
  )
env_num %>% tibble

```

```

p_cor_suelo_ar <- env_num %>%
  dplyr::select(matches('^[A-T,Z]|abundancia|riqueza|^pH$', ignore.case = F)) %>%
  ezCorM(r_size_lims = c(4,8), label_size = 3, method = 'pearson')
p_cor_suelo_ar

p_cor_suelo_ar_spearman <- env_num %>%
  dplyr::select(matches('^[A-T,Z]|abundancia|riqueza|^pH$', ignore.case = F)) %>%
  ezCorM(r_size_lims = c(4,8), label_size = 3, method = 'spearman')
p_cor_suelo_ar_spearman

png(
  filename = 'matriz_correlacion_suelo_abun_riq_rubic_spearman.png',
  width = 1920, height = 1080, res = 125
)
p_cor_suelo_ar_spearman
dev.off() #NO OLVIDAR ESTA IMPORTANTE SENTENCIA

p_cor_geomorf_ar <- env_num %>%
  dplyr::select(-matches('^[A-T,Z]|pH$', ignore.case = F)) %>%
  ezCorM(r_size_lims = c(4,8), label_size = 3, method = 'pearson')
p_cor_geomorf_ar

p_cor_geomorf_ar_spearman <- env_num %>%
  dplyr::select(-matches('^[A-T,Z]|pH$', ignore.case = F)) %>%
  ezCorM(r_size_lims = c(4,8), label_size = 3, method = 'spearman')
p_cor_geomorf_ar_spearman

png(
  filename = 'panel_cor_geomorf_abun_riq_spear.png',
  width = 1920, height = 1080, res = 110
)
p_cor_geomorf_ar_spearman
dev.off() #NO OLVIDAR ESTA IMPORTANTE SENTENCIA

```

a_agrupamiento1.R

```

#' ---
#' title: "Análisis de agrupamiento (cluster analysis). <br> Parte 1: agrupamiento jerárquico"
#' author: "JR"
#' date: "11 de noviembre, 2020"
#' output: github_document
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

```

```

#' ## Preámbulo

#' ### Cargar paquetes
library(vegan)
library(magrittr)
library(broom)
source('biodata/funciones.R')

#' ### Cargar datos
#'
load('biodata/Rubiaceae.Rdata')
mi_fam <- mc_rubic
#'

#' ## Características de las técnicas de agrupamiento
#'
#' Las técnicas de agrupamiento se clasifican según los algoritmos que emplean y el orden de eje
#'
#' - Secuenciales o simultáneos.
#' - Por aglomeración o por división. En referencias en español encontrarás "aglomerativos" y "d
#' - Monotéticos o politéticos.
#' - Jerárquicos o no jerárquicos.
#' - Probabilísticos o no probabilísticos.
#' - Restringidos o no restringidos.
#'

#' ## Agrupamiento jerárquico
#'
#' El agrupamiento jerárquico (AJ) es una técnica de agrupamiento secuencial que consiste en la
#'
#' Dentro del AJ es frecuente usar un enfoque aglomerativo, lo cual implica aplicar algoritmos s
#'
#' El AJ aglomerativo dispone de varios algoritmos de resolución del agrupamiento por pares, que
#'
#' Normalmente, en el análisis de agrupamiento nos interesa agrupar sitios en función de sus des
#'
#' Dado que los cuadros en BCI están autocorrelacionados espacialmente, violamos el supuesto de
#'

#' ### Agrupamiento "aglomerativo" por enlace simple
#'
#' Este método utiliza, como criterio de enlace para agrupar sucesivamente pares de objetos, la
#'
#' Para aplicar este método, debes transformar la matriz de comunidad utilizando alguno de los m
#'
mi_fam_norm <- decostand(mi_fam, "normalize")
mi_fam_norm_d <- vegdist(mi_fam_norm, "euc")
mi_fam_norm_d %>% tidy
#'

#' Es importante, para garantizar consistencia a lo largo del agrupamiento, asignar los nombres

```

```

#’
attr(mi_fam_norm_d, "labels") <- rownames(mi_fam)
#’
#’ Posteriormente, el agrupamiento jerárquico lo realizaré con la función ‘hclust’ del paquete ‘
#’
(cl_single <- hclust(mi_fam_norm_d, method = ‘single’))
#’
#’ Finalmente, el dendrograma a continuación:
plot(cl_single, labels = rownames(mi_fam), hang = -1,
     main = "Sitios de BCI según composición de especies de Rubiaceae\nEnlace simple a partir de
     xlab = ‘Sitios’, ylab = ‘Altura’)
#’
#’ ### Agrupamiento "aglomerativo" por enlace completo
#’
#’ En este caso, el criterio de enlace para agrupar sucesivamente pares de objetos es la menor s
#’
(cl_complete <- hclust(mi_fam_norm_d, method = ‘complete’))
plot(cl_complete, labels = rownames(mi_fam), hang = -1,
     main = "Sitios de BCI según composición de especies de Rubiaceae\nEnlace completo a partir
     xlab = ‘Sitios’, ylab = ‘Altura’)
#’
#’ ### Agrupamiento "aglomerativo" por enlace promedio
#’
#’ En este caso, el criterio de enlace para agrupar sucesivamente pares de objetos es el promedi
#’
#’ Así, dependiendo de si se media o centroide, o si se ponderan o no las distancias originales,
#’
#’ Sólo crearé el dendrograma del método UPGMA.
#’
(cl_upgma <- hclust(mi_fam_norm_d, method = ‘average’))
plot(cl_upgma, labels = rownames(mi_fam), hang = -1,
     main = "Sitios de BCI según composición de especies de Rubiaceae\nUPGMA a partir de matriz
     xlab = ‘Sitios’, ylab = ‘Altura’)
#’
#’ ### Agrupamiento por el método de Ward de varianza mínima
#’
#’ Se basa en los mismos supuestos y criterios de la regresión lineal por mínimos cuadrados, sim
#’
(cl_ward <- hclust(mi_fam_norm_d, method = ‘ward.D2’))
plot(cl_ward, labels = rownames(mi_fam), hang = -1,
     main = "Sitios de BCI según composición de especies de Rubiaceae\nMétodo de Ward a partir d
     xlab = ‘Sitios’, ylab = ‘Altura’)

```

```
#' ---
#' title: "Análisis de agrupamiento (cluster analysis). <br> Parte 2: Interpretación y comparación de dendrogramas"
#' author: "JR"
#' date: "11 de noviembre, 2020"
#' output: github_document
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ## Preámbulo
#'
#' ### Cargar paquetes
#'
library(vegan)
library(tidyverse)
library(broom)
library(cluster)
library(gclus)
library(pvclust)
library(sf)
source('biodata/funciones.R')
#'
#' ### Cargar datos
#'
load('biodata/Rubiaceae.Rdata')
mi_fam <- mc_rubic
load('biodata/matriz_ambiental.Rdata')
mi_fam %>% tibble
bci_env_grid %>% tibble
#'
#' ### Generar matriz de distancias de cuerdas
#'
mi_fam_norm <- decostand(mi_fam, "normalize")
mi_fam_norm_d <- vegdist(mi_fam_norm, "euc")
mi_fam_norm_d %>% tidy
#'
#' ## Interpretación visual de dendrogramas
#'
# [En el script anterior](aa_analisis_de_agrupamiento_1_jerarquico.md) realicé los dendrogramas
#
# La primera evaluación de los dendrogramas NO debe venir de la mano de sofisticados análisis n
#
# Evita concentrar tu vista en grupos extremadamente pequeños; comienza analizando el árbol des
#
# Anota tus impresiones, para que las compares con los resultados que posteriormente obtendrás;
#
```



```

# Para la exploración visual, generaré los objetos de cluster dentro de una lista:
#
lista_cl <- list(
  cl_single = hclust(mi_fam_norm_d, method = 'single'),
  cl_complete = hclust(mi_fam_norm_d, method = 'complete'),
  cl_upgma = hclust(mi_fam_norm_d, method = 'average'),
  cl_ward = hclust(mi_fam_norm_d, method = 'ward.D2')
)
#
# Un plot en panel 2x2 ayuda a visualizarlos todos de manera conjunta. En tu caso, observa y co
#
par(mfrow = c(2,2))
invisible(map(names(lista_cl), function(x) plot(lista_cl[[x]], main = x, hang = -1)))
par(mfrow = c(1,1))
#
# En mi caso, exceptuando el dendrograma generado por medio del enlace simple, detecto al menos
#
# - Un grupo pequeño, compuesto por los sitios 1, 42, 12, 21, 11, 2 y 16.
# - Un "grupo" heterogéneo y grande, conformado por 25, 31,..., 26,..., 35,..., 34,...,32, 17,..
#
# Además de los grupos anteriores, detecto elementos que no forman grupos, es decir, sitios que
#
# ## Elegir método y número de clústers
#
# Existen varios criterios para elegir un dendrograma idóneo, como por ejemplo, los gráficos ti
#
# ### Seleccionar método de agrupamiento por correlación cofenética
#
# La correlación cofenética implica conocer la distancia cofenética, y esta última se entiende m
#
# Usando la lista de objetos de clústers, calcularé la correlación cofenética dentro de un 'map
#
map_df(lista_cl, function(x) {
  coph_d <- cophenetic(x)
  corr <- cor(mi_fam_norm_d, coph_d)
  return(corr)
})
#
# Habrás notado que, tanto UPGMA como enlace completo, tienen valores altos de correlación cofe
#
# ### Elegir número de clústers
#
# Elegiré UPGMA como método de agrupamiento y determinaré cuántos grupos son idóneos de acuerdo
#
# El objetivo de la función 'calcular anchuras siluetas' está implícito en su nombre, y requiere
#
# 1. Las anchuras promedio para cada partición, excepto para las particiones 'i=1' y 'i=50', po

```

```

#’
#’ 2. Número óptimo de grupos. Haré los cálculos para UPGMA y Ward, y luego explico en qué consi
#’
#’ Para UPGMA:
#’
anch_sil_upgma <- calcular_anchuras_siluetas(
  mc_orig = mi_fam,
  distancias = mi_fam_norm_d,
  cluster = lista_cl$cl_upgma)
anch_sil_upgma
#’
#’ El objeto ‘anchuras_siluetas’ de la lista ‘anch_sil_upgma’ te muestra un vector con los prome
#’
#’ Igualmente, el objeto ‘n_grupos_optimo’ te indica cuál es el número óptimo de clústers a crea
#’
#’ Haré el gráfico de dendrograma, aunque nota que en este caso primero reordenaré los sitios co
#’
u_dend_reord <- reorder.hclust(lista_cl$cl_upgma, mi_fam_norm_d)
plot(u_dend_reord, hang = -1)
rect.hclust(
  tree = u_dend_reord,
  k = anch_sil_upgma$n_grupos_optimo)
#’
#’ Ahora compararé el dendrograma con el mapa de calor en un mismo gráfico, colocando los dendro
#’
heatmap(
  as.matrix(mi_fam_norm_d),
  Rowv = as.dendrogram(u_dend_reord),
  symm = TRUE,
  margin = c(3, 3),
  col = rev(cm.colors(4))
)
#’
#’ En general, hay dos grupos, uno grande y otro pequeño, y parece haber un tercero en el mapa d
#’
#’ Mostraré el resultado para Ward:
#’
anch_sil_ward <- calcular_anchuras_siluetas(
  mc_orig = mi_fam,
  distancias = mi_fam_norm_d,
  cluster = lista_cl$cl_ward)
anch_sil_ward
#’
#’ En este caso, el valor máximo, que ocupa la posición número 2, no se diferencia mucho del de
#’
w_dend_reord <- reorder.hclust(lista_cl$cl_ward, mi_fam_norm_d)
plot(w_dend_reord, hang = -1)

```

```

rect.hclust(
  tree = w_dend_reord,
  k = anch_sil_ward$n_grupos_optimo)
plot(w_dend_reord, hang = -1)
rect.hclust(
  tree = w_dend_reord,
  k = anch_sil_ward$n_grupos_optimo + 1)
#'
#' Comparando el dendrograma con el mapa de calor. Verificar si el número de grupos hace sentido
#'
heatmap(
  as.matrix(mi_fam_norm_d),
  Rowv = as.dendrogram(w_dend_reord),
  symm = TRUE,
  margin = c(3, 3),
  col = rev(cm.colors(4))
)
#'
#' Nótese que este dendrograma hace más sentido que el sugerido por UPGMA. En cualquier casos, c
#'
#' ### Evaluación mediante remuestreo por *bootstrap* multiescalar
#'
#' Con suerte, un agrupamiento aplicado a datos muestrales reflejará los patrones naturales de o
#'
#' Sin embargo, los datos de BCI también tienen sesgo, pues se usa un DAP de corte para decidir
#'
#' No obstante, aun con todas sus bondades, los datos censales carecen de una fortaleza: no refl
#'
#' Por estas razones, los análisis de agrupamientos realizados hasta este punto, reflejan tanto
#'
#' El método de remuestreo por *bootstrap* multiescalar está implementado en el paquete 'pvclust'
#'
#' La función 'pvclust' devolverá un dendrograma enriquecido, que incluirá los valores de AU y BP
#'
#' - Rectángulos de borde azul, para todos aquellos grupos que resulten con valores de AU>0.91 e
#'
#' - Líneas inferiores rojas, que resaltan aquellos grupos (o subgrupos) que obtuvieron AU>0.95.
#'
#' Ten presente que, al realizar remuestreo por *bootstrap* multiescalar, cada corrida puede arr
#'
#' #### UPGMA
#'
cl_pvclust_upgma <-
  pvclust(t(mi_fam_norm),
    method.hclust = "average",
    method.dist = "euc",
    iseed = 91, # Resultado reproducible

```

```

        parallel = TRUE)
# Añadir los valores de p
plot(cl_pvclust_upgma, hang = -1)
# Añadir rectángulos a los grupos significativos
lines(cl_pvclust_upgma)
pvrect(cl_pvclust_upgma, alpha = 0.91, border = 4)
#
#' #### Ward
#'
cl_pvclust_ward <-
  pvclust(t(mi_fam_norm),
    method.hclust = "ward.D2",
    method.dist = "euc",
    iseed = 191, # Resultado reproducible
    parallel = TRUE)
# Añadir los valores de p
plot(cl_pvclust_ward, hang = -1)
# Añadir rectángulos a los grupos significativos
lines(cl_pvclust_ward)
pvrect(cl_pvclust_ward, alpha = 0.91, border = 4)
#
#' ### Recapitulando los grupos de sitios.
#'
#' #### Patrones comunes y dispares
#'
#' Detecto algunos patrones consistentes en cuanto a grupos de sitios según composición de las e
#'
#' - Tanto en UPGMA como en Ward, detecté al menos dos o tres grandes grupos. Con el primer méto
#'
#' - En el caso específico del dendrograma Ward, 'pvclust' atomizó los sitios en demasiados grup
#'
#' #### ¿Cómo declaro los grupos de sitios?
#'
#' Para conservar las clasificaciones de grupos de sitios anteriores, crearé un vector con el id
#'
#' UPGMA:
(grupos_upgma_k2 <- as.factor(cutree(lista_cl$cl_upgma, k = 2)))
#'
#' En este caso, los sitios 1 y 2 pertenecen al grupo 1, los sitios 3 al 6 pertenecen al grupo 2
#'
table(grupos_upgma_k2)
#'
#' Nota lo desiguales que son estos grupos, un efecto esperado dado el alto grado de autocorrela
#'
#' Ward:
#'
(grupos_ward_k2 <- as.factor(cutree(lista_cl$cl_ward, k = 2)))

```

```

table(grupos_ward_k2)
#’
#’ Guardaré estos vectores en archivos para reutilizarlos en *scripts* posteriores:
#’
saveRDS(grupos_upgma_k2, 'grupos_upgma_k2.RDS')
saveRDS(grupos_ward_k3, 'grupos_ward_k3.RDS')
#’
#’ Evita usar este, y cualquier otro procedimiento, de manera mecánica. En tu caso, quizá tengas
#’
#’ En el próximo *script*, aprenderás a comparar este resultado con las variables ambientales. T

```

a_agrupamiento3_variabl_ambientls.R

```

#’ ---
#’ title: "Análisis de agrupamiento (cluster analysis). <br> Parte 3: Grupos (clústers), variabl
#’ author: "JR"
#’ date: "15 de noviembre, 2020"
#’ output: github_document
#’ ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#’ ## Preámbulo
#’
#’ ### Cargar paquetes
#’
library(mapview)
library(tidyverse)
library(sf)
library(RColorBrewer)
source('biodata/funciones.R')
#’
#’ ### Cargar datos
#’
load('biodata/Rubiaceae.Rdata')
load('biodata/matriz_ambiental.Rdata')
grupos_complete_k2 <- readRDS('grupos_complete_k2.RDS')
table(grupos_complete_k2) #Importante, tener en cuenta los desiguales tamaños de los grupos
grupos_ward_k2 <- readRDS('grupos_ward_k2.RDS')
table(grupos_ward_k2)
#’
#’ ### Paletas
#’
rojo <- colorRampPalette(brewer.pal(8, "Reds"))
rojo_inv <- colorRampPalette(rev(brewer.pal(8, "Reds")))

```

```

colores_grupos <- brewer.pal(8, "Accent")
#'
#' ## Explorar efectos
#'
#' ### Pruebas de igualdad de promedios de las variables entre 2 grupos
#'
#' Para evaluar homogeneidad de promedios usaré las pruebas *t* (medias), basada en la distribuc
#'
#' Primero crearé un objeto que permita realizar tanto las pruebas como los diagramas de cajas.
#'
(m_amb_complete_k2 <- bci_env_grid %>%
  select_if(is.numeric) %>% select(-id) %>%
  mutate(grupos_complete_k2) %>%
  st_drop_geometry() %>%
  pivot_longer(-grupos_complete_k2, names_to = "variable", values_to = "valor"))
#'
#' para ward_k2 por mi
#'
(m_amb_ward_k2 <- bci_env_grid %>%
  select_if(is.numeric) %>% select(-id) %>%
  mutate(grupos_ward_k2) %>%
  st_drop_geometry %>%
  pivot_longer(-grupos_ward_k2, names_to = "variable", values_to = "valor"))
#'
#' A continuación, las pruebas:
#'
m_amb_complete_k2 %>%
  group_by(variable) %>%
  summarise(
    p_valor_t = t.test(valor ~ grupos_complete_k2)$p.value,
    p_valor_w = wilcox.test(valor ~ grupos_complete_k2, exact = F)$p.value) %>%
  arrange(p_valor_t) %>%
  print(n=Inf)
#'
#' para ward_k2 por mi
#'
m_amb_ward_k2 %>%
  group_by(variable) %>%
  summarise(
    p_valor_t = t.test(valor ~ grupos_ward_k2)$p.value,
    p_valor_w = wilcox.test(valor ~ grupos_ward_k2, exact = F)$p.value) %>%
  arrange(p_valor_t) %>%
  print(n=Inf)
#'
#' Interesa observar las variables que obtuvieron valores de  $p < 0.01$ . Reitero que, en mi caso, mi
#'
#' Por otra parte, este es un buen momento para "revisitar" tus análisis exploratorios de datos

```

```

#'
#' Los diagramas de caja:
#'
m_amb_complete_k2 %>%
  group_by(variable) %>%
  ggplot() + aes(x = grupos_complete_k2, y = valor, fill = grupos_complete_k2) +
  geom_boxplot() +
  scale_fill_brewer(palette = 'Accent') +
  theme_bw() +
  theme(legend.position="none") +
  facet_wrap(~ variable, scales = 'free_y')
#'
#' para ward_k2 por mi
m_amb_ward_k2 %>%
  group_by(variable) %>%
  ggplot() + aes(x = grupos_ward_k2, y = valor, fill = grupos_ward_k2) +
  geom_boxplot() +
  scale_fill_brewer(palette = 'Accent') +
  theme_bw() +
  theme(legend.position = "none") +
  facet_wrap(~variable, scales = 'free_y')
#'
#' Mapas:
#'
mapa_complete_k2 <- mapView(
  bci_env_grid %>% mutate(grupos_complete_k2),
  layer.name = 'Grupos (2) complete',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T,
  col.regions = colores_grupos[1:2],
  zcol = 'grupos_complete_k2') %>%
  addStaticLabels(label = bci_env_grid$id) %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_complete_k2
mapa_complete_k2 %>% mapshot(
  file = 'mapa_complete_k2.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
#' para ward_k2 por mi
#'
mapa_ward_k2 <- mapView(
  bci_env_grid %>% mutate(grupos_ward_k2),

```

```

layer.name = 'Grupos (2) ward',
alpha.regions = 0.6,
map.types = 'OpenTopoMap',
legend = T,
col.regions = colores_grupos[1:2],
zcol = 'grupos_ward_k2') %>%
addStaticLabels(label = bci_env_grid$id) %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 15)
mapa_ward_k2
mapa_ward_k2 %>% mapshot(
  file = 'mapa_ward_k2.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
#' Mapa de una de las variables donde se presentó efecto de su promedio ( $p < 0.01$ ), en este caso,
#'
mapa_fosforo <- mapView(
  bci_env_grid,
  layer.name = 'Fósforo',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T,
  col.regions = rojo,
  zcol = 'P') %>%
addStaticLabels(label = bci_env_grid$id) %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 15)
mapa_fosforo
mapa_fosforo %>% mapshot(
  file = 'mapa_fosforo.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
#' ### Pruebas de igualdad de promedios de las variables entre 3 grupos o más
#'
#' Objeto común:
#'
(m_amb_ward_k3 <- bci_env_grid %>%
  select_if(is.numeric) %>% select(-id) %>%
  mutate(grupos_ward_k3) %>%
  st_drop_geometry() %>%
  pivot_longer(-grupos_ward_k3, names_to = "variable", values_to = "valor"))

```



```

#'
#' Pruebas, en este caso ANOVA (evalúa homogeneidad de medias; no se cumplen muchos de los supue
#'
m_amb_ward_k3 %>%
  group_by(variable) %>%
  summarise(
    p_valor_a = oneway.test(valor ~ grupos_ward_k3)$p.value,
    p_valor_k = kruskal.test(valor ~ grupos_ward_k3)$p.value) %>%
  arrange(p_valor_k) %>%
  print(n=Inf)
#'
#' Gráficos:
#'
m_amb_ward_k3 %>%
  group_by(variable) %>%
  ggplot() + aes(x = grupos_ward_k3, y = valor, fill = grupos_ward_k3) +
  geom_boxplot() +
  scale_fill_brewer(palette = 'Accent') +
  theme_bw() +
  theme(legend.position="none") +
  facet_wrap(~ variable, scales = 'free_y')
#'
#' Mapas:
#'
mapa_ward_k3 <- mapView(
  bci_env_grid %>% mutate(grupos_ward_k3),
  layer.name = 'Grupos (3) Ward',
  alpha.regions = 0.6,
  map.types = 'OpenTopoMap',
  legend = T,
  col.regions = colores_grupos[1:3],
  zcol = 'grupos_ward_k3') %>%
  addStaticLabels(label = bci_env_grid$id) %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_ward_k3
mapa_ward_k3 %>% mapshot(
  file = 'mapa_ward_k3.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
#' Mapa de una de las variables donde se presentó efecto de su promedio ( $p < 0.01$ ), en este caso,
#'
mapa_ph <- mapView(
  bci_env_grid,

```

```

layer.name = 'pH',
alpha.regions = 0.6,
map.types = 'OpenTopoMap',
legend = T,
col.regions = rojo_inv,
zcol = 'pH') %>%
addStaticLabels(label = bci_env_grid$id) %>%
leaflet::setView(
  lng = -79.85136,
  lat = 9.15097,
  zoom = 15)
mapa_ph
mapa_ph %>% mapshot(
  file = 'mapa_ph.png',
  remove_controls = c("zoomControl", "layersControl", "homeButton")
)
#'
```

a_agrupamiento4_sps_ind_pref.R

```

#' ---
#' title: "Análisis de agrupamiento (cluster analysis). <br> Parte 4: Especies indicadoras, espe
#' author: "JR"
#' date: "15 de noviembre, 2020"
#' output: github_document
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ## Preámbulo
#'
#' ### Cargar paquetes
#'
library(indicspecies)
source('biodata/funciones.R')
#'
#' ### Cargar datos
#'
load('biodata/Rubiaceae.Rdata')
mi_fam <- mc_rubic
grupos_complete_k2 <- readRDS('grupos_complete_k2.RDS')
table(grupos_complete_k2)
grupos_ward_k2 <- readRDS('grupos_ward_k2.RDS')
table(grupos_ward_k2)
#'
#' ## Análisis de especies indicadoras mediante IndVal
```

```

#'
#' ### Complete*
#'
iva_complete_k2 <- multipatt(
  x = mi_fam,
  cluster = grupos_complete_k2,
  func = 'IndVal.g',
  max.order = 1,
  control = how(nperm = 999))
summary(iva_complete_k2, indvalcomp = TRUE)
sort(colSums(mi_fam))
(p_complete_adj <- p.adjust(iva_complete_k2$sign$p.value))
(iva_complete_boot <- strassoc(
  X = mi_fam,
  cluster = grupos_complete_k2,
  func = "IndVal.g",
  nboot = 1000))
#'
#' Ward*
#'
iva_ward_k2 <- multipatt(
  x = mi_fam,
  cluster = grupos_ward_k2,
  func = 'IndVal.g',
  max.order = 1,
  control = how(nperm = 999))
summary(iva_ward_k2, indvalcomp = TRUE, alpha = 0.01)
colSums(mi_fam)
(p_ward_adj <- p.adjust(iva_ward_k2$sign$p.value))
(iva_ward_boot <- strassoc(
  X = mi_fam,
  cluster = grupos_ward_k2,
  func = "IndVal.g",
  nboot = 1000))
#'
#' ## Análisis de especies con preferencia por hábitat mediante el coeficiente de correlación bi
#'
#' ### UPGMA
#'
phi_complete_k2 <- multipatt(
  mi_fam,
  grupos_complete_k2,
  func = "r.g",
  max.order = 1,
  control = how(nperm = 999))
summary(phi_complete_k2)
colSums(mi_fam)

```

```

(phi_complete_boot <- strassoc(
  X = mi_fam,
  cluster = grupos_complete_k2,
  func = "r.g",
  nboot = 1000))
#'
#' Ward
#'
phi_ward_k2 <- multipatt(
  mi_fam,
  grupos_ward_k2,
  func = "r.g",
  max.order = 1,
  control = how(nperm = 999))
summary(phi_ward_k2)
colSums(mi_fam)
(phi_ward_boot <- strassoc(
  X = mi_fam,
  cluster = grupos_ward_k2,
  func = "r.g",
  nboot = 1000))

```

t_ordenacion_1.R

```

#' ---
#' title: "Técnicas de ordenación. <br> Parte 1: Ordenación no restringida. <br> PCA, CA y PCoA"
#' author: "JR"
#' date: "21 de noviembre, 2020"
#' output: github_document
#'
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ## Preámbulo
#'
#' ### Cargar paquetes
#'
library(vegan)
library(tidyverse)
library(sf)
library(mapview)
source('biodata/funciones.R')
#'
#' ### Cargar datos

```

```

#'
load('biodata/Rubiaceae.Rdata')
load('biodata/matriz_ambiental.Rdata')
mi_fam <- mc_rubic
(colnames(mi_fam) <- make.cepnames(colnames(mi_fam)))
(df_equivencias <- data.frame(
  nombre_original = colnames(mc_rubic),
  colnames(mi_fam)))
bci_env_grid %>% tibble
grupos_complete_k2 <- readRDS('grupos_complete_k2.RDS')
table(grupos_complete_k2)
grupos_ward_k2 <- readRDS('grupos_ward_k2.RDS')
table(grupos_ward_k2)
#'
#' ## Ordenación
#'
#' La ordenación se basa en los mismos principios que la medición de asociación (similaridad) y
#'
#' A diferencia del análisis de agrupamiento, o como complemento de éste, el análisis de ordenación
#'
#' El análisis de ordenación puede ser no restringido (o simple) y restringido (o 'canónico'). E
#'
#' Las principales técnicas de ordenación no restringida son análisis de componentes principales
#'
#' ### Análisis de componentes principales (PCA)
#'
#' Es el método tradicional basado en vectores propios que comúnmente se aplica a datos cuantitativos
#'
#' #### PCA aplicado a datos ambientales
#'
#' Para aplicar PCA a datos ambientales, es necesario que todas las variables sean numéricas y
#'
#' Dado que se requiere que las variables de entrada sean exclusivamente numéricas, el primer paso es
#'
#' ¡IMPORTANTE! Haré esta demostración sólo con las variables de suelo, **pero puedes (y debes)
#'
#' A partir de los datos de suelo, la función 'rda', de 'vegan' realizará los siguientes pasos:
#'
env_suelo <- bci_env_grid %>%
  st_drop_geometry %>%
  dplyr::select(matches('~[A-T,Z]|~pH$', ignore.case = F))
env_suelo %>% tibble
env_suelo_pca <- rda(env_suelo, scale = TRUE)
env_suelo_pca
summary(env_suelo_pca)
#'
#' para variables geomorfológicas

```

```

#'
env_geomorf <- bci_env_grid %>%
  st_drop_geometry %>%
  select_if(is.numeric) %>%
  select(-id) %>%
  dplyr::select(-matches('^([A-T,Z,U]|~pH$)', ignore.case = F)) %>%
  rename_all(gsub, pattern = '_pct$', replacement = '') %>%
  rename_all(gsub, pattern = '_| ', replacement = '\\n')
env_geomorf %>% tibble
env_geomorf_pca <- rda(env_geomorf, scale = TRUE)
summary(env_geomorf_pca)
#'
#'  Cobre y fosforo? idk
#'
env_fosforo_cobre <- bci_env_grid %>%
  st_drop_geometry %>%
  dplyr::select(matches('^P$|Cu$', ignore.case = F))
env_fosforo_cobre %>% tibble
env_fosforo_cobre_pca <- rda(env_fosforo_cobre, scale = TRUE)
summary(env_fosforo_cobre_pca)
#'
#' Para agilizar la producción de scripts analíticos de referencia, trasladaré las explicaciones
#'
#' En el vídeo asociado, explico el significado de:
#'
#' - Inercia, *Inertia*
#' - Valores propios, autovalores, *Eigenvalues*
#' - Escalamiento, *Scaling*
#' - Puntuaciones de "especies", *Species scores*
#' - Puntuaciones de "sitios", *Site scores*
#'
screepplot(env_suelo_pca, bstick = TRUE)
#'
#' biplot Usando función 'cleanplot.pca'
#'
par(mfrow = c(1, 2))
cleanplot.pca(env_suelo_pca, scaling = 1, mar.percent = 0.08, cex.char1 = 0.7)
cleanplot.pca(env_suelo_pca, scaling = 2, mar.percent = 0.04, cex.char1 = 0.7)
par(mfrow = c(1, 1))
#'
#' #' screepplot para variables geomorfológicas
#'
screepplot(env_geomorf_pca, bstick = TRUE)
#'
par(mfrow = c(1, 2))
cleanplot.pca(env_geomorf_pca, scaling = 1, mar.percent = 0.08, cex.char1 = 0.7)
cleanplot.pca(env_geomorf_pca, scaling = 2, mar.percent = 0.04, cex.char1 = 0.7)

```

```

par(mfrow = c(1, 1))
#'
#' #' Comparar distribución de los sitios en biplots con distribución real en el mapa:
#'
#' ### Generar mapa de cuadros sin simbología
#'
mapa_cuadros <- mapView(
  bci_env_grid,
  col.regions = 'grey80',
  alpha.regions = 0.3,
  map.types = 'OpenTopoMap',
  legend = F, zoom = 14,
  zcol = 'id') %>% addStaticLabels() %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
mapa_cuadros
#'
#' Comparar con resultados de un análisis de agrupamiento del mismo conjunto de datos. Primero a
#'
(env_agrupamiento <- hclust(dist(scale(env_suelo)), 'ward.D'))
(env_grupos <- cutree(env_agrupamiento, k = 3))
(mi_cluster <- factor(env_grupos))
(mi_cluster_l <- levels(mi_cluster))
(mi_cluster_l_seq <- 1:length(mi_cluster_l))
#'
#' Observa que estoy generando un agrupamiento basado en los datos de suelo. No estoy comparando
#'
#' Luego calculo las puntuaciones de los sitios para usarlas luego como coordenadas de los puntos
#'
(puntuaciones <- scores(env_suelo_pca, display = 'wa', scaling = 1))
#'
#' Luego creo el gráfico base, coloco los puntos sobre el gráfico usando las puntuaciones, les c
#'
grafico_base <- plot(
  env_suelo_pca,
  display = "wa",
  scaling = 1,
  type = "n",
  main = "PCA y grupos"
)
abline(v = 0, lty = "dotted")
abline(h = 0, lty = "dotted")
for (i in mi_cluster_l_seq) {
  points(puntuaciones[mi_cluster == i, ],
    pch = (14 + i),

```

```

        cex = 2,
        col = i + 1)
}
text(puntuaciones, row.names(env_suelo), cex = 1, pos = 3)
legend(
  "topleft", # Otras alternativas: "bottomleft", "bottomright" y "topleft"
  paste("Grupo", c(mi_cluster_l_seq)),
  pch = 14 + c(mi_cluster_l_seq),
  col = 1 + c(mi_cluster_l_seq),
  pt.cex = 2
)
#'
#' Es razonable que el análisis cluster y el biplot muestren patrones consistentes, puesto que s
#'
#' Si hago lo mismo, pero usando mi análisis de agrupamiento anterior (*scripts* "aa_analisis_de
#'
par(mfrow = c(1, 2))
(mi_cluster_anterior <- grupos_complete_k2)
#(mi_cluster_anterior <- grupos_ward_k2)
(mi_cluster_anterior_l <- levels(mi_cluster_anterior))
(mi_cluster_anterior_l_seq <- 1:length(mi_cluster_anterior_l))
grafico_base <- plot(
  env_suelo_pca,
  display = "wa",
  scaling = 1,
  type = "n",
  main = "PCA y grupos Complete"
)
abline(v = 0, lty = "dotted")
abline(h = 0, lty = "dotted")
for (i in mi_cluster_anterior_l_seq) {
  points(puntuaciones[mi_cluster_anterior == i, ],
        pch = (14 + i),
        cex = 2,
        col = i + 1)
}
text(puntuaciones, row.names(env_suelo), cex = 1, pos = 3)
legend(
  "topright", # Otras alternativas: "bottomleft", "bottomright" y "topleft"
  paste("Grupo", c(mi_cluster_anterior_l_seq)),
  pch = 14 + c(mi_cluster_anterior_l_seq),
  col = 1 + c(mi_cluster_anterior_l_seq),
  pt.cex = 2
)
#'
(mi_cluster_anterior <- grupos_ward_k2)
(mi_cluster_anterior_l <- levels(mi_cluster_anterior))

```



```

(mi_cluster_anterior_l_seq <- 1:length(mi_cluster_anterior_l))
grafico_base <- plot(
  env_suelo_pca,
  display = "wa",
  scaling = 1,
  type = "n",
  main = "PCA y grupos Ward"
)
abline(v = 0, lty = "dotted")
abline(h = 0, lty = "dotted")
for (i in mi_cluster_anterior_l_seq) {
  points(puntuaciones[mi_cluster_anterior == i, ],
    pch = (14 + i),
    cex = 2,
    col = i + 1)
}
text(puntuaciones, row.names(env_suelo), cex = 1, pos = 3)
legend(
  "topright", # Otras alternativas: "bottomleft", "bottomright" y "topleft"
  paste("Grupo", c(mi_cluster_anterior_l_seq)),
  pch = 14 + c(mi_cluster_anterior_l_seq),
  col = 1 + c(mi_cluster_anterior_l_seq),
  pt.cex = 2
)
par(mfrow = c(1, 1))
#'
#' Esto podría significar que las tendencias/patrones de mi matriz de comunidad (cuadros de 1 Ha
#'
#' #### PCA aplicado a datos de comunidad transformados
#'
mi_fam_hel <- decostand(mi_fam, method = 'hellinger')
mi_fam_hel %>% tibble
mi_fam_hel_pca <- rda(mi_fam_hel)
summary(mi_fam_hel_pca)
screeplot(
  mi_fam_hel_pca,
  bstick = TRUE,
  npcs = length(mi_fam_hel_pca$CA$eig)
)
mi_fam_hel_pca_sc1 <- scores(mi_fam_hel_pca,
  display = "species", scaling = 1)
mi_fam_hel_pca_sc2 <- scores(mi_fam_hel_pca,
  display = "species", scaling = 2)
par(mfrow = c(1, 2))
cleanplot.pca(mi_fam_hel_pca, scaling = 1, mar.percent = 0.06, cex.char1 = 0.7)
cleanplot.pca(mi_fam_hel_pca, scaling = 2, mar.percent = 0.06, cex.char1 = 0.7)
par(mfrow = c(1, 1))

```

```

#'
#' Si intentáramos realizar el PCA a datos de comunidad no transformados, no recogeríamos apropi
#'
#' Las especies que contribuyen mucho a los ejes 1 y 2 del PCA (aquellas cuyos vectores sobresal
#'
#' Evaluaré el ajuste del PCA de datos de comunidad a datos ambientales, mediante la función 'en
#'
biplot(
  mi_fam_hel_pca,
  main = "PCA, escalamiento 2, ajuste a variables ambientales")
<- envfit(mi_fam_hel_pca, env_suelo, scaling = 2))
plot(mi_fam_hel_pca_envfit, p.max = 0.05 , col = 3)
#'
#' Comento los resultados en el vídeo asociado. También probaré ajuste con todas las numéricas d
#'
#' NOTA: te recomiendo probar otros métodos de selección de variables, como por ejemplo, usando
#'
env_num <- bci_env_grid %>%
  select_if(is.numeric) %>%
  select(-id) %>%
  st_drop_geometry
(mi_fam_hel_pca_envfit_num <- envfit(mi_fam_hel_pca, env_num, scaling = 2))
biplot(
  mi_fam_hel_pca,
  main = "PCA, escalamiento 2, ajuste a variables ambientales")
plot(mi_fam_hel_pca_envfit_num, p.max = 0.05 , col = 3)
biplot(
  mi_fam_hel_pca,
  main = "PCA, escalamiento 2, ajuste a variables ambientales")
plot(mi_fam_hel_pca_envfit_num, p.max = 0.1 , col = 3)
#'
#' Comento los resultados en el vídeo asociado.
#'
#' ¿Cuándo o a qué datos aplicar PCA?
#'
#' - PCA no es especialmente sensible a datos muy desviados de la normalidad.
#' - Como toda técnica, PCA tiene limitaciones.
#' - Las variables deben ser dimensionalmente homogéneas (unidades comparables o adimensionales)
#' - No usar en matriz transpuestas (no hace sentido la covarianza entre objetos).
#' - Es posible usar PCA con dato de presencia/ausencia, en cuyo caso, la matriz de comunidad de
#' - Las relaciones entre variables se miden por ángulos, no por proximidad de las puntas de los
#'
#' ### Análisis de correspondencia (CA)
#'
mi_fam_ca <- cca(mi_fam)
summary(mi_fam_ca)
summary(mi_fam_ca, scaling = 1)

```

```

#'
#' Screeplot
#'
screeplot(mi_fam_ca, bstick = TRUE, npcs = length(mi_fam_ca$CA$eig))
#'
#' Biplots
#'
par(mfrow = c(1, 2))
plot(mi_fam_ca,
     scaling = 1,
     main = "Análisis de correspondencia, escalamiento 1"
)
plot(mi_fam_ca,
     scaling = 2, # Por defecto scaling=2, lo escribo sólo para fines didácticos
     main = "Análisis de correspondencia, escalamiento 2")
par(mfrow = c(1, 1))
#'
#' Excluyendo especie *Thevetia ahouai*, abreviada como *Thevahou*. #' excluyendo Amaioua corymbosa
#'
mi_fam_ca <- cca(mi_fam[, -grep('Amaicory', colnames(mi_fam))])
summary(mi_fam_ca)
summary(mi_fam_ca, scaling = 1)
screeplot(mi_fam_ca, bstick = TRUE, npcs = length(mi_fam_ca$CA$eig))
par(mfrow = c(1, 2))
plot(mi_fam_ca,
     scaling = 1,
     main = "CA, escalamiento 1, sin Amaioua corymbosa"
)
plot(mi_fam_ca,
     scaling = 2,
     main = "CA, escalamiento 2, sin Amaioua corymbosa")
par(mfrow = c(1, 1))
#'
#' Análisis de coordenadas principales (PCoA)
#'
#' Las técnicas de ordenación anteriores preservan la distancia euclídea entre los objetos. Si m
#'
#' La función que realiza el PCoA en '{vegan}' es 'cmdscale' (de *Classical (Metric) Multidimens
#'
mi_fam_d_jacc <- vegdist(mi_fam, method = 'jaccard') # En realidad, 'bray' es la opción por defe
mi_fam_d_jacc_pcoa <- cmdscale(
  mi_fam_d_jacc,
  k = (nrow(mi_fam) - 1),
  add = T,
  eig = TRUE)
round(mi_fam_d_jacc_pcoa$eig, 2)
round(sum(mi_fam_d_jacc_pcoa$eig[mi_fam_d_jacc_pcoa$eig<0]),2)

```

```

round(sum(mi_fam_d_jacc_pcoa$eig[mi_fam_d_jacc_pcoa$eig>=0]),2)
ordiplot(scores(mi_fam_d_jacc_pcoa, choices = c(1, 2)),
          type = "t",
          main = "PCoA de distancias Jaccard con promedios ponderados de especies")
abline(h = 0, lty = 3)
abline(v = 0, lty = 3)
mi_fam_d_jacc_pcoa_wa <- wascores(mi_fam_d_jacc_pcoa$points[, 1:2], mi_fam)
text(
  mi_fam_d_jacc_pcoa_wa,
  rownames(mi_fam_d_jacc_pcoa_wa),
  cex = 0.7, col = "red")
(mi_fam_d_jacc_pcoa_env <- envfit(mi_fam_d_jacc_pcoa, env_num))
plot(mi_fam_d_jacc_pcoa_env, p.max = 0.05, col = 3)

```

t_ordenacion_2.R

```

#' ---
#' title: "Técnicas de ordenación. <br> Parte 2: Ordenación restringida o 'canónica'. <br> RDA,
#' author: "JR"
#' date: "21 de noviembre, 2020"
#' output: github_document
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ## Preámbulo
#'
#' ### Cargar paquetes
#'
library(vegan)
library(tidyverse)
library(sf)
source('biodata/funciones.R')
#'
#' ### Cargar datos
#'
load('biodata/Rubiaceae.Rdata')
load('biodata/matriz_ambiental.Rdata')
mi_fam <- mc_rubic
(colnames(mi_fam) <- make.cepnames(colnames(mi_fam)))
(df_equivencias <- data.frame(
  nombre_original = colnames(mc_rubic),
  colnames(mi_fam)))
bci_env_grid %>% tibble
#'

```

```

#' ## Ordenación restringida
#'
#' Con este conjunto de técnicas de ordenación, las tendencias detectadas en un conjunto de datos
#'
#' Las principales técnicas de ordenación restringida son análisis de redundancia o RDA (siglas)
#'
#' ### Análisis de redundancia (RDA)
#'
#' En el análisis de ordenación simple o no restringida, la matriz de datos (de comunidad o ambiental)
#'
#' RDA combina la regresión y el análisis de componentes principales. El RDA es una regresión lineal
#'
#' - Cada variable de respuesta (e.g. matriz de comunidad) es ajustada (por regresión) a la matriz de
#'
#' - Se prueba significancia estadística de la relación entre las variables de respuesta y explicativas
#'
#' - Si la prueba resulta significativa, se realiza un PCA de la matriz de valores ajustados. Es decir, se
#'
#' - Luego se usa la matriz de vectores propios canónicos para calcular dos tipos de puntuaciones:
#'
#'     - Usando la matriz de valores ajustados para obtener una ordenación en el espacio de las variables
#'
#'     - Usando la matriz de valores centrados de las variables de respuesta para obtener una ordenación
#'
#' Mostraré a continuación ejemplos aplicados a mi familia asignada. Para agilizar, trasladaré los datos
#'
#' Nota (de nuevo): recuerda que los datos de BCI están autocorrelacionados espacialmente, por lo que
#'
#' #### Ejemplo usando la matriz ambiental de variables suelo:
#'
mi_fam_hel <- decostand(mi_fam, method = 'hellinger')
mi_fam_hel %>% tibble
env_suelo <- bci_env_grid %>%
  st_drop_geometry %>%
  dplyr::select(matches('^[A-T,Z]|~pH$', ignore.case = F))
env_suelo %>% tibble
mi_fam_hel_rda_suelo <- rda(mi_fam_hel ~ ., env_suelo)
summary(mi_fam_hel_rda_suelo)
#'
#' ¿Qué partes del resumen debes mirar?
#'
#' - La varianza, particionada, de la cual se muestra tanto la inercia (absoluta) como la proporción
#'
#' - Los valores propios (*eigenvalues*) y su contribución a la varianza. Existen *eigenvalues* y
#'
#' - Las puntuaciones de sitios (*site scores*) y de especies (*species scores*) tienen el mismo
#'

```

```

#’ - Las puntuaciones restringidas de sitio (en ‘{vegan}’ se usa la etiqueta *Site constraints (
#’
#’ - Finalmente, el resumen presenta las puntuaciones para el *biplot*, que más adelante usaré e
#’
#’ Tal como comenté arriba, es importante calcular un  $R^2$  insesgado, siendo este más ú
#’
RsquareAdj(mi_fam_hel_rda_suelo)$adj.r.squared
#’
#’ Normalmente, el  $R^2$  insesgado es mucho más bajo que el sesgado, porque se relativiz
#’
#’ Otro diagnóstico importante es la determinación de los factores de inflación de la varianza (
#’
vif.cca(mi_fam_hel_rda_suelo)
#’
#’ Variables con valores por encima de 10 deben ser examinadas y, desde una posición conservador
#’
#’ Finalmente, la representación del modelo se realiza en un *triplot*, que es un gráfico enriqu
#’
#’ Escalamiento 1:
#’
plot(mi_fam_hel_rda_suelo,
      scaling = 1,
      display = c("sp", "lc", "cn"),
      main = "Triplot de RDA especies ~ var. suelo, escalamiento 1"
)
mi_fam_hel_rda_suelo_sc1 <-
  scores(mi_fam_hel_rda_suelo,
         choices = 1:2,
         scaling = 1,
         display = "sp"
  )
arrows(0, 0,
       mi_fam_hel_rda_suelo_sc1[, 1] * 0.9,
       mi_fam_hel_rda_suelo_sc1[, 2] * 0.9,
       length = 0,
       lty = 1,
       col = "red"
)
#’
#’ Paso la interpretación de este gráfico y la del siguiente al video correspondiente.
#’
#’ Escalamiento 2
#’
plot(mi_fam_hel_rda_suelo,
      scaling = 2,
      main = "Triplot de RDA especies ~ var. suelo, escalamiento 2"
)

```

```

mi_fam_hel_rda_suelo_sc2 <-
  display = c("sp", "lc", "cn"),
  scores(mi_fam_hel_rda_suelo,
    scaling = 2,
    choices = 1:2,
    display = "sp"
  )
arrows(0, 0,
  mi_fam_hel_rda_suelo_sc2[, 1] * 0.9,
  mi_fam_hel_rda_suelo_sc2[, 2] * 0.9,
  length = 0,
  lty = 1,
  col = "red"
)
#'
#' ##### Ejemplo usando las matriz ambiental con variables seleccionadas
#'
#' El RDA anterior mostró que las variables de suelo son útiles para predecir la matriz de comun
#'
#' - Hay mucha colinealidad entre ellas.
#'
#' - No se probó mejorar el modelo añadiendo otras variables, además de las de suelo.
#'
#' Crearé una matriz ambiental con las variables que resultaron significativas en el ajuste *pos
#'
#''# Agregué mis "variables significativas" en rubiaceae
#'
env_selec <- bci_env_grid %>%
  select(
    abundancia_global,
    elevacion_media,
    geomorf_interfluvio_pct,
    geomorf_llanura_pct,
    geomorf_valle_pct,
    UTM.NS,
    Al, B, Ca, Cu, Fe, K, Mg, Mn, P, Zn, N, N.min., pH) %>%
  st_drop_geometry
mi_fam_hel_rda_selec <- rda(mi_fam_hel ~ ., env_selec)
#'
vif.cca(mi_fam_hel_rda_selec)
#'
#' Haré el plot del escalamiento 2 para comprobar gráficamente asociación entre variables sin la
#'
plot(mi_fam_hel_rda_selec,
  scaling = 2,
  display = c("sp", "lc", "cn"),
  main = "Triplot de RDA especies ~ var. selec, escalamiento 2"
)

```

```

)
#’
#’ Tal como comenté arriba, variables con valores VIF por encima de 10, deben ser examinadas. En
#’
#’#’ Quité Ca para ver que pasaba, igual no hay tanto solapamiento en rubic
#’
env_selec2 <- bci_env_grid %>%
  select(
    abundancia_global,
    elevacion_media,
    geomorf_interfluvio_pct,
    geomorf_llanura_pct,
    geomorf_valle_pct,
    UTM.NS,
    Al, B, Cu, Fe, Mg, K, Mn, P, Zn, N, N.min., pH) %>%
  st_drop_geometry
mi_fam_hel_rda_selec2 <- rda(mi_fam_hel ~ ., env_selec2)
vif.cca(mi_fam_hel_rda_selec2)
plot(mi_fam_hel_rda_selec2,
      scaling = 2,
      display = c("sp", "lc", "cn"),
      main = "Triplot de RDA especies ~ var. selec2, escalamiento 2"
)
#’
#’ Nota que las posiciones rotaron, no así la asociación entre la mayoría de las variables. ‘B’
#’
#’#’ Ahora quito "UTM.NS" y potasio. El vif.cca sigue bajando de a poquito
#’
env_selec3 <- bci_env_grid %>%
  select(
    abundancia_global,
    elevacion_media,
    geomorf_interfluvio_pct,
    geomorf_llanura_pct,
    geomorf_valle_pct,
    Al, B, Cu, Fe, Mg, Mn, P, Zn, N, N.min., pH) %>%
  st_drop_geometry
mi_fam_hel_rda_selec3 <- rda(mi_fam_hel ~ ., env_selec3)
vif.cca(mi_fam_hel_rda_selec3)
plot(mi_fam_hel_rda_selec3,
      scaling = 2,
      display = c("sp", "lc", "cn"),
      main = "Triplot de RDA especies ~ var. selec3, escalamiento 2"
)
#’
#’ Finalmente, la coordenada ‘UTM.EW’ tiene un alto valor VIF, por lo que es preferible excluirla
#’

```



```

#' #' Ahora quito magnesio, boro y valle, aunque no quería, es eso o elevacion media
#'
env_selec4 <- bci_env_grid %>%
  select(
    abundancia_global,
    elevacion_media,
    geomorf_interfluvio_pct,
    geomorf_llanura_pct,
    Al, Cu, Fe, Mn, P, Zn, N, N.min., pH) %>%
  st_drop_geometry
mi_fam_hel_rda_selec4 <- rda(mi_fam_hel ~ ., env_selec4)
vif.cca(mi_fam_hel_rda_selec4)
plot(mi_fam_hel_rda_selec4,
     scaling = 2,
     display = c("sp", "lc", "cn"),
     main = "Triplot de RDA especies ~ var. selec4, escalamiento 2"
)
#
#' Las variables 'N.min.' y 'Zn' ahora presentan valores VIF aceptables. Habrás notado que es to
#
summary(mi_fam_hel_rda_selec4)
RsquareAdj(mi_fam_hel_rda_selec4)$adj.r.squared
#
#' Triplot, pero ahora con las flechas para las especies.
#
#' Escalamiento 1:
#
plot(mi_fam_hel_rda_selec4,
     scaling = 1,
     display = c("sp", "lc", "cn"),
     main = "Triplot de RDA especies ~ var. selec4, escalamiento 1"
)
mi_fam_hel_rda_selec4_sc1 <-
  scores(mi_fam_hel_rda_selec4,
        choices = 1:2,
        scaling = 1,
        display = "sp"
  )
arrows(0, 0,
      mi_fam_hel_rda_selec4_sc1[, 1] * 0.9,
      mi_fam_hel_rda_selec4_sc1[, 2] * 0.9,
      length = 0,
      lty = 1,
      col = "red"
)
#
#' Escalamiento 2

```

```

#’
plot(mi_fam_hel_rda_selec4,
     scaling = 2,
     display = c("sp", "lc", "cn"),
     main = "Triplot de RDA especies ~ var. selec4, escalamiento 2"
)
mi_fam_hel_rda_selec4_sc2 <-
  scores(mi_fam_hel_rda_selec4,
        scaling = 2,
        choices = 1:2,
        display = "sp"
  )
arrows(0, 0,
       mi_fam_hel_rda_selec4_sc2[, 1] * 0.9,
       mi_fam_hel_rda_selec4_sc2[, 2] * 0.9,
       length = 0,
       lty = 1,
       col = "red"
)
#’
#’ ### Análisis de correspondencia canónica (CCA)
#’
#’ #### Ejemplo usando las matriz ambiental con variables seleccionadas:
#’
mi_fam_cca_selec4 <- cca(mi_fam ~ ., env_selec4)
summary(mi_fam_cca_selec4)
RsquareAdj(mi_fam_cca_selec4)
#’
#’ Escalamiento 1
#’
plot(mi_fam_cca_selec4,
     scaling = 1,
     display = c("sp", "lc", "cn"),
     main = "Triplot de CCA especies ~ var. selec4, escalamiento 1"
)
#’
#’ Escalamiento 2
#’
plot(mi_fam_cca_selec4,
     scaling = 2,
     display = c("sp", "lc", "cn"),
     main = "Triplot de CCA especies ~ var. selec4, escalamiento 2"
)
#’
#’ Excluyendo especies con abundancia menor a 100 individuos
#’
colSums(mi_fam) %>% sort

```

```

mi_fam_no_raras <- mi_fam %>% select_if(colSums(.) > 100)
intersect(colnames(mi_fam), colnames(mi_fam_no_raras))
setdiff(colnames(mi_fam), colnames(mi_fam_no_raras))
mi_fam_no_raras_cca_selec4 <- cca(mi_fam_no_raras ~ ., env_selec4)
summary(mi_fam_no_raras_cca_selec4)
RsquareAdj(mi_fam_no_raras_cca_selec4)
#'
#' Escalamiento 1
#'
plot(mi_fam_no_raras_cca_selec4,
     scaling = 1,
     display = c("sp", "lc", "cn"),
     main = "Triplot de CCA especies no raras ~ var. selec4, escalamiento 1"
)
#'
#' Escalamiento 2
#'
plot(mi_fam_no_raras_cca_selec4,
     scaling = 2,
     display = c("sp", "lc", "cn"),
     main = "Triplot de CCA especies no raras ~ var. selec4, escalamiento 2"
)

```

diversidad_1_alpha.R

```

#' ---
#' title: "Análisis de diversidad. <br> Parte 1: Diversidad alpha"
#' author: "JR"
#' date: "29 de noviembre, 2020"
#' output: github_document
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ## Preámbulo
#'
#' ### Cargar paquetes
#'
library(vegan)
library(adespatial)
library(plyr)
library(RColorBrewer)
library(tidyverse)
library(sf)
library(SpadeR)





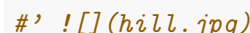
```

```

library(iNEXT)
source('biodata/funciones.R')
#
#' ### Cargar datos
#
load('biodata/Rubiaceae.Rdata')
load('biodata/matriz_ambiental.Rdata')
mi_fam <- mc_rubic
bci_env_grid %>% tibble
grupos_complete_k2 <- readRDS('grupos_complete_k2.RDS')
table(grupos_complete_k2)
grupos_ward_k2 <- readRDS('grupos_ward_k2.RDS')
table(grupos_ward_k2)
#
#' ## Diversidad alpha
#
#' **La tentación de medir la diversidad mediante un único número, ha atrapado a muchos investigadores.
#
#' **La riqueza de especies, una cifra de uso común en estudios ecológicos, es un elemento muy sensible.
#
#' El término "diversidad biológica" se asume que fue acuñado a principios de los 80, y se refiere a la
#
#' biodiversidad es de factura más reciente, y es una contracción de "biológica-diversidad".
#
#' La definición de biodiversidad del PNUMA es: "variabilidad entre organismos vivos" de todo tipo,
#
#' Hubbell (2001) ofrece una definición más adaptada a la práctica actual y mucho más restringida.
#
#' En estas dos últimas acepciones (Hubbell y Magurran), la diversidad biológica puede dividirse en
#
#' ### La diversidad de especies como un único número
#
#' Usaré la notación *q* para designar el número de especies o riqueza de especies, y *n* para el número de
#
#' #### Riqueza de especies y rarefacción
#
#' Magurran (2004) distingue entre "densidad de especies", que equivale al número de especies, y "riqueza de especies",
#
#' Para asegurar la comparabilidad entre sitios, se han propuesto distintos métodos. Uno es la rarefacción.
#
#' 
#
#' #### Componentes de la diversidad de especies basada en abundancia: riqueza y equidad
#
#' Asumamos que un sitio de muestreo es una variable cualitativa, y cada especie es un "estado".
#
#' 

```

```

#’
#’ Desde el punto de vista ecológico, la H tiene dos propiedades importantes:
#’
#’ - Crece al aumentar la riqueza de especies q.
#’ - Crece con la uniformidad (=equidad o equitabilidad, es decir, qué tan bien repartida se enc
#’
#’ Para una q dada, H asume su valor máximo ( $H_{\max}$ ) cuando todas las especies es
#’
#’ 
#’
#’ Un ejemplo ilustra lo anterior:
#’
#’ - Cinco especies, 55 individuos, abundancias desiguales:
#’
foo1 <- c(25, 16, 9, 4, 1)
diversity(foo1)
#’
#’ - Cinco especies, 55 individuos, abundancias homogéneas
#’
foo2 <- c(11, 11, 11, 11, 11)
diversity(foo2)
log(5)
#’
#’ Otros índices miden exclusivamente la homogeneidad de la abundancia, a lo que se denomina equi
#’
#’ Uno de los más usados es la equidad de Pielou o J (1966), que es la razón entre la entr
#’
#’ 
#’
#’ Teóricamente, este índice sólo debería aumentar cuando aumenta la homogeneidad de la abunda
#’
#’ Todos los índices de equidad se relacionan con la forma de los modelos de abundancia de esp
#’
#’ Otra medida común en trabajos de ecología numérica, pero que es inversa o contraria a la equi
#’
#’ 
#’
#’ Este valor aumenta con la dominancia, por lo que realmente no mide diversidad, sino más bien
#’
#’ La riqueza de especies (q), la entropía de Shannon (H), la equidad de Pielou (J) y el i
#’
#’ 
#’
#’ donde a es el orden de la medida de entropía (a=0,1,2...*), la cual cuantifica la importan
#’
#’ 
#’

```

```

#' Las tres primeras entropías de Renyi ( $H_a$ ), donde  $a=0, 1$  y  $2$ , y los correspondientes
#'
#' Por otra parte, los tres primeros números de diversidad de Hill tienen significados que conocemos
#'
#'  <br>
#' > Según Borcard et al., 2018.
#'
#' Bajo esta notación:
#'
#' - La equidad de Pielou (denominada también equidad de Shannon) equivale a  $J=H_1/H_0$ 
#'
#' Usando la misma aproximación, Hill propuso también otras ratios que, a diferencia de lo que ocurre con
#'
#' -  $E_1=N_1/N_0$  a la cual el propio Hill denominó como su versión de la equidad
#'
#' -  $E_2=N_2/N_0$ , a la cual Hill denominó como su versión de la equidad
#'
#' Los números de diversidad y las ratios de Hill son menos sensibles a las matrices de comunidades
#'
#' **Estas afirmaciones tienen implicaciones muy importantes desde el punto de vista ecológico,
#'
#' **Índices, entropías, equidades, ratios**
#'
(indices <- alpha_div(mi_fam))
pairs(indices,
      lower.panel = panel.smooth,
      upper.panel = panel.cor,
      diag.panel = panel.hist,
      main = "Pearson Correlation Matrix")
indices_env <- bind_cols(
  indices,
  bci_env_grid %>%
    select_if(is.numeric) %>%
    st_drop_geometry %>%
    select(-id, -UTM.EW,
          -elevacion_media,
          -pendiente_media,
          -curvatura_perfil_media,
          -curvatura_tangencial_media,
          -heterogeneidad_ambiental) %>%
    select(-matches('^geom.*pct$')))
indices_env %>% tibble
ezCorM(indices_env, r_size_lims = c(2,4), label_size = 2)
#'
#' **Modelos de abundancia de especies**
#'
mi_fam_mae <- radfit(mi_fam)

```

```

plot(mi_fam_mae)
# '
# ' **Rarefacción**
# '
# ' Riqueza por sitios
# '
riqueza <- specnumber(mi_fam)
riqueza %>% sort
# '
# ' Sitios con riqueza mínima y máxima
# '
riqueza[riqueza == min(riqueza)]
riqueza[riqueza == max(riqueza)]
range(riqueza)
# '
# ' Abundancia por sitios
# '
abundancia <- rowSums(mi_fam)
abundancia %>% sort
# '
# ' Sitios con abundancias mínima y máxima
# '
abundancia[abundancia == min(abundancia)]
abundancia[abundancia == max(abundancia)]
(rango_abun <- range(abundancia))
# '
# ' Abundancia en el sitio más pobre
# '
abundancia[riqueza == min(riqueza)]
# '
# ' Abundancia en el sitio más rico
# '
abundancia[riqueza == max(riqueza)]
# '
# ' Riqueza en el sitio con menor abundancia
# '
riqueza[abundancia == min(abundancia)]
# '
# ' Riqueza en el sitio con mayor abundancia
# '
riqueza[abundancia == max(abundancia)]
# '
# ' Rarefacción a la abundancia más pequeña encontrada
# '
riqueza_menor_abun <- rarefy(mi_fam, sample = rango_abun[1])
sort(riqueza)
sort(round(riqueza_menor_abun))

```

```

rarecurve(
  mi_fam,
  step = 1,
  sample = rango_abun[1],
  xlab = "Número de individuos (tamaño de muestra)",
  ylab = "Especies",
  label = TRUE,
  col = "blue", cex = 0.4
)
#'
#' ### Riqueza de especies, estimación y comparación, "completitud de muestra" (existe en el diccionario)
#'
#' Aproximación básica:
#'
specpool(mi_fam)
specpool(mi_fam)[[1]]/specpool(mi_fam)*100
#'
#' Lista comprensiva de métodos (incluyendo recientes):
#'
#' - **Enfoques asintóticos. Estiman la riqueza de especies**:
#'   - Paramétricos:
#'     - Modelo homogéneo (estándar y MLE)
#'   - No paramétricos:
#'     - Chao1 y Chao1-bc
#'     - iChao1
#'     - Basados en "cobertura" o "completitud de muestra". ACE para datos de abundancia
#'     - Estimadores Jackknife (de primer y segundo órdenes)
#' - **Enfoques no asintóticos. Se utilizan para hacer rarefacción y extrapolación**:
#'   - Basados en tamaño de la muestra
#'   - Basados en "cobertura" o "completitud de muestra"
#'
#' ##### Enfoques asintóticos y no asintóticos aplicados a la matriz de comunidad combinada
#'
#' Generar la matriz de comunidad combinada, en la que todos los sitios forman uno.
#'
mi_fam_combinada <- colSums(mi_fam)
mi_fam_combinada %>% sort
mi_fam_combinada_chao <- estimacion_riqueza_chao(
  mc = mi_fam_combinada,
  n_raras = 18)
mi_fam_combinada_chao$asintoticos_estimacion
mi_fam_combinada_chao$no_asintoticos_rarefaccion_extrapolacion
mi_fam_combinada_chao$no_asintoticos_rarefaccion_extrapolacion_grafico
#'
#' ##### Enfoques asintóticos y no asintóticos aplicados a una matriz de comunidad agrupada
#'
#' # Generar matriz de comunidad agrupada según el método de Ward (tres grupos), procedente de

```



```

#'
mi_fam_k2 <- mi_fam %>%
  mutate(g=grupos_complete_k2) %>%
  group_by(g) %>%
  summarise_all(sum) %>%
  select(-g) %>%
  data.frame
mi_fam_k2 %>% rowSums %>% sort
mi_fam_k2_chao <- estimacion_riqueza_chao(
  mc = mi_fam_k2,
  n_raras = 13)
mi_fam_k2_chao$asintoticos_estimacion
mi_fam_k2_chao$no_asintoticos_rarefaccion_extrapolacion
mi_fam_k2_chao$no_asintoticos_rarefaccion_extrapolacion_grafico

```

diversidad_2_beta.R

```

#' ---
#' title: "Análisis de diversidad. <br> Parte 2: Diversidad beta"
#' author: "JR"
#' date: "2 de diciembre, 2020"
#' output: github_document
#' ---

knitr::opts_chunk$set(fig.width=12, fig.height=8)

#' ## Preámbulo
#'
#' ### Cargar paquetes
#'
library(vegan)
library(adespatial)
library(plyr)
library(tidyverse)
library(sf)
library(vegetarian)
library(mapview)
source('biodata/funciones.R')
#'
#' ### Cargar datos
#'
load('biodata/Rubiaceae.Rdata')
load('biodata/matriz_ambiental.Rdata')
mi_fam <- mc_rubic
bci_env_grid %>% tibble

```

```

#'
#' ## Diversidad beta
#'
#' En la concepción de Whittaker, la diversidad alfa es la local o de sitio, la diversidad beta
#'
#' La diversidad alfa la medimos usando métricas que ponderan comúnmente la riqueza y la equidad
#'
#' ### Diversidad beta con un único número
#'
beta_multiplicativa <- calcular_beta_multiplicativa(
  mc = mi_fam,
  orden = 0:20)
beta_multiplicativa
#'
#' Notar que, para esta matriz de comunidad, la diversidad beta multiplicativa disminuye en esca
#'
#' ### Contribución de especies a la diversidad beta (SCBD, *species contribution to beta divers
#'
determinar_contrib_local_y_especie(
  mc = mi_fam,
  alpha = 0.05,
  nperm = 9999,
  metodo = 'hellinger')
#'
mapa_cuadros <- mapView(
  bci_env_grid,
  col.regions = 'grey80',
  alpha.regions = 0.3,
  map.types = 'OpenTopoMap',
  legend = F, zoom = 14,
  zcol = 'id') %>% addStaticLabels() %>%
  leaflet::setView(
    lng = -79.85136,
    lat = 9.15097,
    zoom = 15)
#'
#' Utilizar el mapa de cuadros para identificar aquellos que contribuyen más a la diversidad bet
#'
mapa_cuadros

```

Referencias

Appelhans, T., Detsch, F., Reudenbach, C., & Woellauer, S. (2019). *Mapview: Interactive viewing of spatial data in r*. Retrieved from <https://CRAN.R-project.org/package=mapview>

Baillie, I., Elsenbeer, H., Barthold, F., Grimm, R., & Stallard, R. (2006). *Semi-detailed soil survey*

of barro colorado island, panama.

Borcard, D., Gillet, F., & Legendre, P. (2018). *Numerical Ecology with R. Second Edition* (pp. 52–66). <https://doi.org/10.1007/978-3-319-71404-2>

Cáceres, M. D., & Legendre, P. (2009). Associations between species and groups of sites: Indices and statistical inference. *Ecology*, 90(12), 3566–3574.

Chao, A., Ma, K. H., Hsieh, T. C., & Chiu, C.-H. (2016). *SpadeR: Species-richness prediction and diversity estimation with r*. Retrieved from <https://CRAN.R-project.org/package=SpadeR>

Condit, R., Chisholm, R. A., & Hubbell, S. P. (2012). Thirty years of forest census at Barro Colorado and the importance of immigration in maintaining diversity. *PLOS ONE*, 7(11), 1–6. <https://doi.org/10.1371/journal.pone.0049826>

Condit, R., Pérez, R., Lao, S., Aguilar, S., & Hubbell, S. P. (2017). Demographic trends and climate over 35 years in the Barro Colorado 50 ha plot. *Forest Ecosystems*, 4(1), 17. <https://doi.org/10.1186/s40663-017-0103-1>

Condit, R., Pitman, N., Leigh, E., Chave, J., Terborgh, J., Foster, R., ... Hubbell, S. (2002). Beta-diversity in tropical forest trees. *Science (New York, N.Y.)*, 295, 666–669. <https://doi.org/10.1126/science.1066854>

Davis, A. P., Govaerts, R., Bridson, D. M., Ruhsam, M., Moat, J., & Brummitt, N. A. (2009). A global assessment of distribution, diversity, endemism, and taxonomic effort in the rubiaceae1. *Annals of the Missouri Botanical Garden*, 96(1), 68–78.

De Cáceres, M., & Legendre, P. (2009). Associations between species and groups of sites: Indices and statistical inference. In *Ecology*. Retrieved from <http://sites.google.com/site/miqueldecaceres/>

Dufrene, M., & Legendre, P. (1997). Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecological Monographs*, 67, 345–366. <https://doi.org/10.2307/2963459>

Grandtner, M., & Chevrette, J. (2013). *Dictionary of trees, volume 2: South america: Nomenclature, taxonomy and ecology*. Retrieved from <https://books.google.com.do/books?id=XALRl1qzcLMC>

Hsieh, T. C., Ma, K. H., & Chao, A. (2020). *INEXT: Interpolation and extrapolation for species diversity*. Retrieved from http://chao.stat.nthu.edu.tw/wordpress/software_download/

Hubbell, S. P., & Foster, R. B. (1983). Diversity of canopy trees in a neotropical forest and implications for conservation. In T. Whitmore, A. Chadwick, & A. Sutton (Eds.), *Tropical rain forest: Ecology and management* (pp. 25–41). Oxford: The British Ecological Society.

Hubbell, S. P., Condit, R., Foster, R. B., Grubb, P. J., Thomas, C. D., Hassell, M. P., & May, R. M. (1990). Presence and absence of density dependence in a neotropical tree community. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 330(1257), 269–281. <https://doi.org/10.1098/rstb.1990.0198>

Hubbell, S., Condit, R., & Foster, R. (2021). Forest Census Plot on Barro Colorado Island. Retrieved March 23, 2021, from <http://ctfs.si.edu/webatlas/datasets/bci/>

Jansen, S., Robbrecht, E., Beeckman, H., & Smets, E. (2000). Aluminium accumulation in rubiaceae: An additional character for the delimitation of the subfamily rubioideae? *IAWA Journal*, 21. <https://doi.org/10.1163/22941932-90000245>

Legendre, P., & Gallagher, E. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia*, 129, 271–280. <https://doi.org/10.1007/s004420100716>

Martínez Batlle, J. R. (2020). biogeografia-master/scripts-de-analisis-BCI: Long coding sessions (Version v0.0.0.9000). <https://doi.org/10.5281/zenodo.4402362>

McAlister, R. L., Kolterman, D. A., & Pollard, A. J. (2015). Nickel hyperaccumulation in populations of *Psychotria grandis* (rubiacae) from serpentine and non-serpentine soils of puerto rico.

Australian Journal of Botany, 63(2), 85–91.

Moreno, C. E. (2001). *Manual de métodos para medir la biodiversidad*. Universidad Veracruzana.

Néda, Z., Horvat, S., Toháti, H. M., Derzsi, A., & Balogh, A. (2008). A spatially explicit model for tropical tree diversity patterns. *arXiv E-Prints*, arXiv:0803.3704.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ... Wagner, H. (2019). *Vegan: Community ecology package*. Retrieved from <https://CRAN.R-project.org/package=vegan>

Patino, L., Velbel, M., Price, J., & Wade, J. (2003). Trace element mobility during spheroidal weathering of basalts and andesites in hawaii and guatemala. *Chemical Geology*, 202. <https://doi.org/10.1016/j.chemgeo.2003.01.002>

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1), 439–446. <https://doi.org/10.32614/RJ-2018-009>

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

Ricotta, C. (2004). A parametric diversity measure combining the relative abundances and taxonomic distinctiveness of species. *Diversity and Distributions*, 10(2), 143–146. <https://doi.org/https://doi.org/10.1111/j.1366-9516.2004.00069.x>

Torres-Leite, F., Cavatte, P. C., Garbin, M. L., Hollunder, R. K., Ferreira-Santos, K., Capetine, T. B., ... Carrijo, T. T. (2019). Surviving in the shadows: Light responses of co-occurring rubiaceae species within a tropical forest understory. *Flora*, 261, 151487. <https://doi.org/https://doi.org/10.1016/j.flora.2019.151487>

Volkov, I., Banavar, J. R., Hubbell, S. P., & Maritan, A. (2003). Neutral theory and relative species abundance in ecology. *Nature*, 424(6952), 1035–1037. <https://doi.org/10.1038/nature01883>

Wickham, H. (2017). *Tidyverse: Easily install and load the 'tidyverse'*. Retrieved from <https://CRAN.R-project.org/package=tidyverse>