

Anidamiento y los análisis Rq y Qr en PAM's

J. Andrés Christen

CIMAT, Guanajuato, Mexico jac@cimat.mx

Jorge Soberón

U. Kansas, USA jsoberon@ku.edu

Resumen

Usando una aproximación continua para la función divisoria en matrices anidadas, investigamos la forma resultante en las gráficas Rq y Qr de Arita, H.T., Christen, J.A., Rodríguez, P. y Soberón, J. (2008), "Range size and species richness: Mathematical constraints y biological implications", *American Naturalist*, **172**(4), 519-532. Este análisis nos permite identificar matrices anidadas dentro de matrices de presencia-ausencia (de especies, o PAM's, por sus siglas en inglés), sin tener que recurrir a métodos computacionalmente intensivos como reordenamientos óptimos, termomapas, etc. Introducimos el uso de notación matricial para mejorar y clarificar presentaciónes anteriores.

KEYWORDS: PAM's; biogeografía; ecología.

1. Introducción

La biogeografía es el estudio de la biodiversidad en el espacio y en el tiempo. Una herramienta fundamental de la biogeografía son las matrices de presencia-ausencia de especies o PAM's por sus siglas en inglés. La región geográfica en estudio, típicamente a escala continental, es dividida en cuadrantes y, dentro de un "ensamble" de especies (e. g. un taxón, como los mamíferos), se hace un conteo completo de las especies que habitan (1) y no habitan (0) cada cuadrante de la región. Si enumeramos las especies identificándolas con los reglones de una matriz, y

enumeramos los sitios o cuadrantes con las columnas de dicha matriz, lo que obtenemos es precisamente una PAM (en la región, a la escala y en relación al ensamble seleccionados). El estudio de las relaciones y propiedades y descripción de dichas PAM's constituye una actividad sustantiva de la biogeografía.

Los análisis "Rq" y "Qr" de Arita, Christen, Soberón y Rodríguez (2008) han demostrado ser una herramienta útil para analizar PAM's. En este artículo ampliamos dicha investigación al elaborar sobre la relación que tienen las gráficas Rq y Qr de Arita *et al.* (2008) con el anidamiento en PAM's (Ulrich, Almeida-Neto y Gotelli, 2009), introduciendo a su vez una notación matricial más compacta. En la siguiente sección introduciremos nuestra notación matricial lo que son los análisis de "Rq" y "Qr". Más adelante describiremos la relación de dichos análisis con el fenómeno biogeográfico de anidamiento.

2. Los análisis Rq y Qr

Una matriz de presencia y ausencia (de especies) es definida como un matriz de dimensiones $S \times N \Delta = (\delta(i, j))$ donde

 $\delta(i,j) = \begin{cases} 1 & \text{si la especie } i \text{ está presente en el sitio } j \\ 0 & \text{en otro caso,} \end{cases}$

para i = 1, 2, ..., S especies y j = 1, 2, ..., N sitios, y para cada especie i existe un sitio j tal que $\delta(i, j) = 1$, y viceversa, para cada j existe un sitio i tal que $\delta(i, j) = 1$. Esto es, una PAM está formada por S renglones que representan especies y N columnas que representan sitios. Sea $\mathbf{1}_k^T = (1, 1, ..., 1)$ un vector columna de unos de dimensión k. El vector $S \times 1$ de "áreas de distribución" (el número de sitios que ocupa cada especie) \boldsymbol{n} y el vector $N \times 1$ de "diversidad de especies" (el número de especies que habitan cada sitio) \boldsymbol{s} pueden ser calculados como

$$\boldsymbol{n} = \boldsymbol{\Delta} \mathbf{1}_N$$
 and $\boldsymbol{s} = \boldsymbol{\Delta}^T \mathbf{1}_S$.

Entonces si $\boldsymbol{n} = (n_i)$, n_i es el "área de distribución" de la especie i y si $\boldsymbol{s} = (s_j)$, s_j es la "diversidad de especies" del sitio j. (Note que una PAM es una matriz de 0's y 1's tal que ambos \boldsymbol{n} y \boldsymbol{s} tienen todas sus entradas diferentes de cero). Las "áreas de distribución proporcional" y "las diversidades proporcionales de especies" son $\boldsymbol{n}^* = \frac{1}{N}\boldsymbol{n} = (n_i^*)$ y $\boldsymbol{s}^* = \frac{1}{S}\boldsymbol{s} = (s_j^*)$, respectivamente.

El vector $S \times 1$ de "volúmenes del campo de riquezas" $\mathbf{D} = (D_i)$ y el vector $N \times 1$ de "volúmenes del campo de distribuciones" son definidos

 como

$$oldsymbol{D} = oldsymbol{\Delta} oldsymbol{s} = oldsymbol{\Delta} oldsymbol{\Delta}^T oldsymbol{1}_S \;\; ext{and} \;\; oldsymbol{R} = oldsymbol{\Delta}^T oldsymbol{n} = oldsymbol{\Delta}^T oldsymbol{\Delta} oldsymbol{1}_N.$$

Los "volúmenes proporcionales del campo de riquezas" y los "volúmenes proporcionales del campo de distribuciones" (Graves y Rahbek, 2005) son $\frac{1}{S}\boldsymbol{D}$ y $\frac{1}{N}\boldsymbol{R}$, respectivamente.

El número de unos en la PAM Δ lo llamamos el "peso" f y la "densidad" de la matriz Δ es $f^* = \frac{f}{NS}$ (esto viene del término "fill" usado en inglés en biogeografía, nosotros preferimos usar el término más común de "densidad" importándolo de la física, que a su vez es el término usado para las matrices de adyacencias, similares a las PAM's, en la literatura de teoría de grafos, e. g. Coleman y Moré, 1983). La mal llamada "diversidad beta" de Whittaker de la PAM es $\beta = \frac{NS}{f} = 1/f^*$ (esto es, el inverso de la densidad de la matriz; note que β no es una diversidad, es comúnmente vista como una razón de cambio o, más bien, una proporción entre el promedio de diversidades locales con la diversidad global). Note que $f = \mathbf{1}_N^T \Delta^T \mathbf{1}_S = \mathbf{1}_N^T \mathbf{s} = \mathbf{1}_S^T \Delta \mathbf{1}_N = \mathbf{1}_S^T \mathbf{n}$. Estas equivalencias dan pie a la interpretación dual de la beta de Whittaker en la literatura de biogeografía (eg. Whittaker, 1972, y Routledge, 1977).

Los análisis Rq y Qr de Arita *et al.* (2008) se basan en las gráficas de $\frac{n_i}{N}$ vs. $\frac{D_i}{Sn_i}$ y $\frac{s_j}{S}$ vs. $\frac{R_j}{s_jN}$, respectivamente. Vea por ejemplo las figuras 6(A) y 6(B) de Arita *et al.* (2008).

Finalmente, la matriz $N\times N$ de varianza-covarianza definida en la forma usual es

$$\boldsymbol{C}_{si} = \frac{1}{S} \boldsymbol{\Delta}^T \boldsymbol{\Delta} - (\boldsymbol{s}^*) (\boldsymbol{s}^*)^T.$$

El vector $N \times 1$ de covarianzas promedio para cada sitio $\boldsymbol{\tau} = (\tau_j)$ se define como

$$\boldsymbol{\tau} = \frac{1}{N} \boldsymbol{C}_{si} \boldsymbol{1}_N = \frac{1}{NS} (\boldsymbol{\Delta}^T \boldsymbol{\Delta} \boldsymbol{1}_N - (\boldsymbol{s}^*) \boldsymbol{s}^T \boldsymbol{1}_N)$$

y por lo tanto $\boldsymbol{\tau} = \frac{1}{NS}\boldsymbol{R} - \frac{f}{NS}\boldsymbol{s}^* = \frac{1}{NS}\boldsymbol{R} - \beta^{-1}\boldsymbol{s}^*$. Equivalentemente el vector $S \times 1$ de covarianzas promedio por cada especie $\boldsymbol{\rho} = (\rho_i)$ se define como

$$\boldsymbol{\rho} = \frac{1}{S} \boldsymbol{C}_{sp} \boldsymbol{1}_{S} = \frac{1}{SN} (\boldsymbol{\Delta} \boldsymbol{\Delta}^{T} \boldsymbol{1}_{S} - (\boldsymbol{n}^{*}) \boldsymbol{n}^{T} \boldsymbol{1}_{S})$$

$$\operatorname{con} \boldsymbol{C}_{sp} = \frac{1}{N} \boldsymbol{\Delta} \boldsymbol{\Delta}^{T} - (\boldsymbol{n}^{*}) (\boldsymbol{n}^{*})^{T}.$$

Es entonces claro que $\boldsymbol{\rho} = \frac{1}{NS}\boldsymbol{D} - \beta^{-1}\boldsymbol{n}^*$. $\boldsymbol{\tau} = \frac{1}{NS}\boldsymbol{R} - \beta^{-1}\boldsymbol{s}^*$ y $\boldsymbol{\rho} = \frac{1}{NS}\boldsymbol{D} - \beta^{-1}\boldsymbol{n}^*$ son formas generales de las ecuaciones (9) y (5) de Arita *et al.* (2008). Vea también Soberón y Ceballos (2009).



Figura 1: Mapeo de la PAM (izq.) al cuadrado unitario (dcha.), los 1's se representan con rectángulos negros y los 0's con rectángulos blancos. La PAM está perfectamente anidada y esto resulta en una "frontera" (función no creciente) que separa los espacios negros y blancos.

3. Anidamiento

Es sencillo encontrar un mapeo de cualquier PAM al cuadrado unitario $[0,1] \times [0,1]$ identificando cada $\delta(i,j)$ con un pequeño rectángulo (negro si $\delta(i,j) = 1$ y blanco si $\delta(i,j) = 0$, vea por ejemplo Almeida-Neto, Guimaraes y Lewinsohn, 2007). Esto es simplemente una imagen en blanco y negro de la matriz Δ . PAM's perfectamente anidadas son PAM's que, bajo un reordenamiento (no único) de renglones y columnas, existe una función $\gamma(x) : [0, x_0] \rightarrow [0, y_0]$ que separa los rectángulos negros (los 1's) de los blancos (los 0's) en dicho mapeo al $[0, 1] \times [0, 1]$. Asumiremos aquí que $\gamma(0) = y_0$ y $\gamma(x_0) = 0$ y que γ es decreciente, vea la Figura 1. A γ la llamamos "función divisoria" de la PAM anidada. Para un panorama general sobre anidamiento en biogeografía vea Ulrich, Almeida-Neto y Gotelli (2009).

Considerando muchos sitios y muchas especies en una matriz perfectamente anidada, uno puede asumir, como aproximación, que $\gamma(x)$ es suave (i. e. diferenciable); vea la curva continua en la Figura 2(a). Una vez que conocemos S, $N \neq y = \gamma(x)$ la PAM (perfectamente anidada) queda definida. Consecuentemente, en este caso, podemos desarrollar el análisis Rq y Qr como en la sección anterior usando la función $\gamma(x)$ como sigue:

1. El equivalente de índice para renglones i que representan las especies es ahora la variable y (eje y) y el equivalente para el índice de columna que representa sitios es ahora la variable x (eje x).



Figura 2: (a-c) PAM perfectamente anidada (aquí ponemos los 1's en gris) con S = 50 especies y N = 125 sitios, la línea continua representa una aproximación suave a la función divisoria $\gamma(x)$ ($x_0 = 1$, y $y_0 = 1$). Es claro que la diversidad proporcional de especies en cada sitio es $s(x) = \gamma(x)$ y que las áreas de distribución proporcionales para la cada especie es $n(y) = \gamma^{-1}(y)$. El volumen proporcional del campo de riquezas es aproximado con la función D(y) y es el área (sumando diversidades de especies a lo largo de sus sitios de hábitat) bajo $\gamma(x)$ desde 0 hasta la línea vertical azul que se encuentra en $\gamma^{-1}(y)$; mostramos las áreas correspondientes para (a) y = 0.2, (b) y = 0.3 y (c) y = 0.5. Más aun, siendo que dicha área se adelgaza conforme crece y, debemos de tener que D(y) es una función decreciente. (d) Análisis Rq correspondiente para esta PAM graficado junto con su correspondiente aproximación, (D(y)/n(y), n(y)) (en rojo). Se prueba también que D(y)/n(y) es una función creciente y dado que n(y) es decreciente la gráfica Rq es una función decreciente en el cuadrado unitario (vea texto).

- 2. Las diversidades proporcionales de especies (s_j^*) s son sencillamente $s(x) = \gamma(x)$ y las áreas de distribución proporcional (n_i^*) s son $n(y) = \gamma^{-1}(y)$.
- 3. Los términos $\frac{D_i}{S}$'s son la suma de las diversidades proporcionales de especies sobre sus áreas de hábitat. En este caso de PAM's perfectamente anidadas, los sitios de presencia son contiguos, desde la columna 1 hasta su frontera o división donde se pasa del 1 al 0. Esto significa que los términos $\frac{D_i}{S}$'s pueden ser aproximados con

$$D(y) = \int_0^{\gamma^{-1}(y)} \gamma(x) dx,$$

vea Figura 2(a-c). Usando el cambio de variable $u = \gamma(x)$ y recurriendo a integración por partes llegamos a la expresión $D(y) = y\gamma^{-1}(y) + \int_{y}^{y_0} \gamma^{-1}(u) du = yn(y) + \int_{y}^{y_0} n(u) du$. En la misma tabla de razonamientos, los términos $\frac{R_i}{NS}$'s pueden ser aproximados con

$$R(x) = \int_0^{\gamma(x)} \gamma^{-1}(y) dy,$$

lo cual es igual a $R(x) = x s(x) + \int_x^{x_0} n(u) du.$

4. La gráfica Rq es (aproximadamente) representada con la función paramétrica $(D(y)/n(y), n(y)), y \in [0, 1]$ (note que $\frac{D_i}{Sn_i} = \frac{D_i}{NSn_i^*} \approx \frac{D(y)}{n(y)}; y = i/S$). A su vez, la gráfica Qr es (aproximadamente) representada por la función paramétrica en el cuadrado unitario con $(R(x)/s(x), s(x)), x \in [0, 1]$.

La siguiente pregunta nos es ahora aparente, ¿cómo resulta la gráfica Rq (o Qr) en el caso de (sub)matrices anidadas? Note que si tenemos dos funciones $h: I_1 \to I_2$ y $g: I_1 \to I_3$ (para $I_j \subset \mathcal{R}$) la función paramétrica de $(g(y), h(y)), y \in I_1$, es lo mismo (visto como subconjunto de $I_3 \times I_2$) que la gráfica de $h(g^{-1}(z)), z \in g(I_1)$ (esto es, los puntos $(z, h(g^{-1}(z)))$, desde luego suponiendo que $g^{-1}(z)$ existe. Ya que $D(0) = \int_0^{y_0} n(u) du =$ f^* (la densidad de la PAM) tenemos que $\frac{D(0)}{n(0)} = f^*/x_0$ y lím $_{y \to y_0} \frac{D(y)}{n(y)} =$ y_0 (más aún, vemos que $f^* < x_0 y_0$ y por lo tanto $f^*/x_0 < y_0$). Al calcular la derivada correspondiente tenemos que

$$\frac{d}{dy}\frac{D(y)}{n(y)} = -\frac{1}{\gamma'(\gamma^{-1}(y))}\frac{\int_{y}^{y_0}\gamma^{-1}(u)du}{\{\gamma^{-1}(y)\}^2}.$$

Dado que $\gamma(x)$ y su inversa son positivas y decrecientes, la integral en la expresión es positiva pero $\gamma'(x)$ es negativa, por ende

$$\frac{d}{dy}\frac{D(y)}{n(y)} > 0. \tag{1}$$

Usando sendos argumentos para el análisis Qr arribamos al siguiente teorema.

Teorema 1 La gráfica Rq para una PAM anidada con función divisoria decreciente $\gamma : [0, x_0] \rightarrow [0, y_0] \ (\gamma([0, x_0]) = [0, y_0])$ es la gráfica de la función decreciente $R : [f^*/x_0, y_0] \rightarrow [0, x_0]$ definida como R(z) = $n(g_1^{-1}(z)), \text{ donde } g_1(y) = \frac{D(y)}{n(y)}.$ La correspondiente gráfica Qr es la gráfica de la función $Q : [f^*/y_0, x_0] \rightarrow [0, y_0]$ definida como Q(z) = $s(g_2^{-1}(z)), \text{ donde } g_2(x) = \frac{R(x)}{s(x)}.$

Demostración: De (1) vemos que g_1 es creciente y entonces R existe (un argumento similar garantiza la existencia de Q). Más aun, dado que ambas g_1 y g_2 son funciones crecientes y dado que n y s son funciones decrecientes, R y Q son decrecientes. •

3.1. Niveles de Isocovarianza y Anidamiento

La estructura de isocovarianza de la gráficas Rq y Qr fue estudiada por Arita *et al.* (2008) y más en detalle por Soberón y Ceballos (2009); en esta oportunidad estudiamos su relación con matrices anidadas. De la relación $\boldsymbol{\rho} = \frac{1}{NS} \boldsymbol{D} - \beta^{-1} \boldsymbol{n}^*$ es inmediato que

$$n_i^* = \frac{\rho_i}{\frac{D_i}{NSn_i^*} - f^*}.$$

Si para un conjunto de especies (un "ensamble") su covarianza promedio permanece constante, esto es $\rho_i = \rho$, entonces los puntos correspondientes a estas especies en su gráfica Qr resultan estar acomodados a lo largo de la hipérbola

$$n_i^* = \frac{\rho}{\frac{D_i}{NSn_i^*} - f^*}.$$

Este conjunto de puntos de la gráfica Qr son $(z_i = \frac{D_i}{NSn_i^*}, n_i^* = \frac{\rho}{z_i - f^*})$. Si ordenamos los n_i^* 's en orden decreciente, entonces los z_i 's deben de resultar en orden creciente y entonces estos puntos deben de corresponder a una matriz anidada con función divisoria $\gamma^{-1}(y) = n_{[yS]}^*$. Lo mismo se sigue para cualquier otra forma funcional $n_i = h(z_i)$ en la cual $h(\cdot)$ sea decreciente. Las curvas de isocovarianza son un ejemplo de lo anterior.

3.2. Ejemplo

Suponga que nuestra PAM Δ tiene una submatriz, un subconjunto de especies (un ensamble) y un subconjunto de sitios (una región), que es perfectamente anidada. Las especies del ensamble rara vez tiene sitios de hábitat fuera de la región, pero muchas especies fuera del ensamble comparten la región con este. Una PAM de este tipo se presenta en la Figura 3(a). Desde luego que la dificultad primordial en el análisis de anidamiento de PAM's es encontrar un reordenamiento correcto de sitios y especies, en principio un problema de complejidad O(S!N!). Existen algoritmos para analizar grados de anidamiento en PAM's, contándose con una batería sustancial de "índices de anidamiento", sin embargo ninguno es capaz de recuperar una submatriz anidada dentro de un PAM con una configuración arbitraria (vea Ulrich, Almeida-Neto y Gotelli, 2009). Ahora reordenamos aleatoriamente la PAM en la Figura 3(a) para obtener la Figura 3(b); la submatriz anidada queda completamente diluida. Lo primero que se nos puede ocurrir en un análisis de anidamiento es reordenar la matriz en secuencia creciente de áreas de distribución (renglones) y diversidad de especies (columnas), para obtener la Figura 3(c). Aún así la submatriz anidada no puede ser decantada correctamente.

El análisis Rq de esta matriz se presenta en la Figura 4(a). Este análisis es *invariante a permutaciónes de columnas y renglones* y por lo tanto es idéntico al ser aplicado a cualquiera de las matrices presentadas en Figura 3(a-c). Note como en la Figura 4(a) existe una subconjunto de puntos que forman una curva decreciente a la derecha del cúmulo principal de puntos, los cuales hemos encerrado en un rectángulo gris. Como hemos explicado, esta es la huella característica de una matriz anidada. Si seleccionamos dichos puntos, esto nos lleva a la matriz en la Figura 4(b). Al tomar esta matriz, tomamos los puntos restantes de la matriz desordenada de la Figura 3(b) y concatenamos dichas matrices obtenemos una versión permutada de la original. Esta matriz se presenta en la Figura 4(d) y casi toda la submatriz anidada es recuperada (compare las Figuras 3(a) y 3(d)).

4. Discusión

Hemos demostrado que existe una relación importante entre el análisis Rq y Qr de Arita *et al.* (2008) y el anidamiento en PAM's, correspondiéndose con arreglos de puntos en orden decreciente (a contraposición de nubes de puntos) en las gráficas Rq y Qr. Sin duda esto



Figura 3: (a) PAM simulada con S = 400 y N = 5000 conteniendo una submatriz de 50 especies y 200 sitios. (b) Permutación aleatoria de dicha matriz. (c) Reordenamiento de la matriz en (b) en orden decreciente de áreas de distribución y diversidad de especies; la submatriz anidada ya no es aparente. (d) Al seleccionar la curva decreciente de puntos en la gráfica Rq de la Figure 4(a), reordenando de nuevo, y concatenando el resto de los puntos recobramos casi la totalidad de la matriz anidada.



Figura 4: (a) Gráfica de Rq para la PAM presentada en la Figura 3. Se hace una selección de puntos (rectángulo en gris) siguiendo el patrón decreciente de puntos a la derecha del cúmulo principal, para obtener la matriz anidada que se presenta en (b).

coadyuva en la fuerza interpretativa de la gráficas Rq y Qr, y seguramente contribuirá al desarrollo de análisis más detallados de matrices de presencia y ausencia de especies. La complejidad algorítmica de estos análisis es relativamente baja a comparación de otros análisis de anidamiento (e. g. búsquedas de reordenamientos óptimos) y nuestro análisis de anidamiento viene simplemente libre de costo computacional obteniéndose como subproducto de las gráficas Rq y Qr. En la medida en que más ensambles se depuren y estén disponibles, se requerirán analizar PAM's de carácter global (S > 5,000 y $N > 10^6$, vea Soberón y Ceballos, 2009) y la complejidad asociada a los métodos de análisis emergerá como un factor a tomar en cuenta, sino como de hecho un elemento prohibitivo. En contraposición, el análisis Rq y Qr depende de manipulaciones simples de los vectores n, s, D, R. El tamaño conjunto de estos es 2(S+N), siendo que el tamaño de la PAM Δ es NS, el cual para PAM grandes es, con mucho, mayor al anterior. Esto redunda en el hecho de que, para llevar a cabo el análisis Rq y Qr, no es ni siguiera necesario mantener Δ en memoria; esta puede ser leída una vez para obtener n, s y una segunda vez para generar D, R; siendo entonces la descrita operación de orden 2NS, lineal con el tamaño de la matriz.

Aunque todavía sería necesario un conjunto sustancial de análisis Rq y Qr en PAM's de ensambles reales para confirmar el uso e importancia de la perspectiva de anidamiento en la interpretación de las gráficas Rq y Qr en contextos de biogeografía y ecología, nuestro trabajo recién demuestra como podemos contribuir al cuerpo teórico en el estudio de patrones espaciales en biodiversidad al definir formalmente un núcleo pequeño de conceptos ("diversidad" beta, los volúmenes del campo de riqueza y del campo de distribuciones, anidamiento, etc.) y consecuentemente estudiar sus interrelaciones y propiedades matemáticas. En Ecología es práctica común proponer índices diversos para conceptos vagos. Esto repercute en una proliferación de medidas a veces hasta contradictorias que obstaculizan, en vez de promover, nuestro entendimiento y poder interpretativo. Esperamos que nuestra perspectiva, esbozada aquí, albergue una alternativa fructífera.

Referencias

- Almeida-Neto, M., P. R. Guimaraes, y T. M. Lewinsohn (2007), "On nestedness analyses: rethinking matrix temperature and anti-nestedness", *Oikos*, **116**, 716–722.
- [2] Arita, H.T., Christen, J.A., Rodríguez, P. y Soberón, J. (2008), "Range size and species richness: Mathematical constraints and biological implications", *American Naturalist*, **172**(4), 519-532.
- [3] Coleman, T. F. y Moré, J. J. (1983), "Estimation of sparse Jacobian matrices and graph coloring Problems", SIAM Journal on Numerical Analysis, 20(1), 187–209.
- [4] Graves, G y Rahbek, C. (2005), "Source pool geometry and assembly of continental avifaunas", Proceedings of National Academy of Science, 102, 7871-7876.
- [5] Soberón, J., y G. Ceballos (2009), "Biological Signature and Mathematical Artifacts in Global Biodiversity Patterns: case of Mammals of World", *American Naturalist* (submitted).
- [6] Routledge, R.D. (1977), "On Whittaker's components of diversity", *Ecology*, 58, 1120–1127.
- [7] Ulrich W., Almeida-Neto, M. y Gotelli, N.J. (2009), "A consumer's guide to nestedness analysis", *Oikos*, **118**, 3–17.
- [8] Whittaker, R.H. (1972), "Evolution and measurements of species diversity", *Taxon*, 21, 213–251.