

EL ANÁLISIS DE INFORMACIÓN EN TAXONOMÍA NUMÉRICA: APLICACIÓN AL ESTUDIO DE LA TRIBU *BRASSICEAE* (*CRUCIFERAE*)

por

J. ESTEBAN HERNÁNDEZ BERMEJO &
MARGARITA CLEMENTE MUÑOZ*

Resumen

HERNÁNDEZ BERMEJO, J. E. & M. CLEMENTE MUÑOZ (1985). El análisis de información en taxonomía numérica: aplicación al estudio de la tribu *Brassiceae* (*Cruciferae*). *Anales Jard. Bot. Madrid* 41(2):313-331.

Se aplica el análisis de la información a los resultados obtenidos en trabajos previos sobre taxonomía numérica de la tribu *Brassiceae* en base a caracteres procedentes de las piezas estériles de la flor. Se desarrollan las aplicaciones de la función de Shannon en: *a)* el análisis de la variabilidad intragenérica y comparación del grado de evolución y diversificación de los géneros; *b)* la elección de caracteres con mayor capacidad de diagnóstico tanto para la diferenciación de los géneros como para la distinción de las especies de cada uno de ellos; *c)* análisis de la interdependencia entre caracteres estudiados, consideraciones y método para determinar el grado de redundancia marginal de la información aportada por cada carácter respecto a los demás.

Con la aplicación de estos métodos a los caracteres y tribu estudiados resulta posible reivindicar caracteres como «nectarios medianos», «color de pétalos» y «posición de sépalos respecto al eje floral» para la diferenciación intergenérica, o de «simetría de la corola», «pilosidad de sépalos» y «presencia de sépalos cuculados» para la intragenérica. Se extraen asimismo conclusiones sobre la interdependencia entre caracteres, la capacidad de diagnóstico de cada uno de ellos dentro de cada género estudiado, el nivel de heterogeneidad interna de algunos de ellos, y sobre el grado de definición y constancia de caracteres florales en géneros como *Crambe*, *Hutera* y *Moricandia*.

Abstract

HERNÁNDEZ BERMEJO, J. E. & M. CLEMENTE MUÑOZ (1985). Information analysis in numerical taxonomy; application to the study of the tribe *Brassiceae* (*Cruciferae*). *Anales Jard. Bot. Madrid* 41(2):313-331 (In Spanish).

In this paper, information analysis is applied to a series of results obtained in previous papers on numerical taxonomy of the tribe *Brassiceae* based on characters derived from the sterile floral parts. Shannon's function is applied to: *a)* the analysis of the intragenetic variability and the comparison of the degree of evolution and diversification of the genera; *b)* the selection of the characters with a greater diagnostic value both for the differentiation of the genera and for the discrimination of each one of their species and *c)* the analysis of the interrelationship between the characters studied, considerations and method for the determination of the degree of marginal redundancy of the information furnished by each character with respect to the others.

(*) Cátedra de Botánica Agrícola. E.T.S. Ingenieros Agrónomos. Córdoba.

By applying these methods to the characters and tribe studied, it is possible to revalidate characters such as the «median nectaries», «colour of the petals» and «position of the petals with respect to the floral axis» for intergeneric differentiation or «symmetry of the corolla», «hairiness of the sepals» and «presence of hooded sepals» for intrageneric differentiation.

Conclusions are also obtained on the interrelationship between the characters, their diagnostic value within each genus studied, the level of internal heterogeneity of some of these characters and the degree of definition and constancy of the floral characters in genera such as *Crambe*, *Hutera* and *Moricandia*.

INTRODUCCIÓN

La teoría de la información y especialmente las aplicaciones de la función de Shannon (SHANNON, 1948) han sido objeto de numerosos trabajos en el campo de la investigación biológica. Sus relaciones con la estadística fueron exploradas por KULLBACK (1959) y a partir de ellas, utilizadas en el análisis numérico de diversos problemas biológicos. Margalef desarrolló las aplicaciones de la teoría de la información al estudio de la heterogeneidad espacial de comunidades (MARGALEF, 1957) y a la medida de la diversidad ecológica (MARGALEF, 1958). También se utilizó el cálculo de la información para el estudio de la interdependencia existente entre dos o más especies, así como entre aquéllas y los factores ecológicos que condicionan su distribución (GODRON, 1968). Un resumen de las aplicaciones de la Teoría de la Información en el campo de la Ecología puede encontrarse en LEGENDRE & LEGENDRE (1979). Casos concretos de aplicación a problemas de ecología terrestre en la Península hay publicados también, por ejemplo, en GONZÁLEZ BERNÁLDEZ & *al.* (1968) que utilizan un estadístico desarrollado por LANCE & WILLIAMS (1966) inspirado en la expresión de Shannon para el estudio numérico de comunidades de pastizal del occidente ibérico, o en HERNÁNDEZ BERMEJO & SAINZ OLLERO (1978) que aplican el cálculo de la «información mutua» desarrollado por GODRON (1968) para analizar las relaciones entre *Fagus sylvatica* L. y algunos factores ecológicos que condicionan su distribución en el extremo meridional de su área.

También en taxonomía numérica ha sido empleada la expresión de Shannon y así ESTABROOK (1967) plantea un desarrollo teórico de sus aplicaciones en la medida de la información contenida por los caracteres utilizados en un estudio taxonómico, en el cálculo de la interdependencia entre dichos caracteres y en la comparación entre diferentes sistemas de clasificación a fin de conocer el más válido o eficaz de ellos. Posteriormente HAWKSWORTH & *al.* (1968) aplican estos criterios en el análisis de caracteres y sistemas de clasificación de las especies del género *Arcuthobium* y BAUM (1974) lo hace revisando la taxonomía del género *Avena*. Más recientemente DUNCAM & ESTABROOK (1976) proponen nuevas aplicaciones metodológicas en la evaluación de sistemas de clasificación.

Tomando como punto de partida los trabajos de taxonomía numérica en *Brassicaceae* (*Cruciferae*) relativos a nectarios florales (CLEMENTE MUÑOZ & HERNÁNDEZ BERMEJO, 1978a), corola (CLEMENTE MUÑOZ & HERNÁNDEZ BERMEJO, 1978b) y cáliz (CLEMENTE MUÑOZ & HERNÁNDEZ BERMEJO, 1980a), así como la clasificación jerárquica conseguida a partir de los citados

caracteres florales (CLEMENTE MUÑOZ & HERNÁNDEZ BERMEJO, 1980b), se utilizan ahora los métodos desarrollados a partir de la teoría de la información para: a) valorar la información (heterogeneidad) conseguida por cada uno de los caracteres analizados, tanto sobre el conjunto de las especies estudiadas como sobre las de cada uno de los géneros de la tribu, b) evaluar las interrelaciones existentes entre dichos caracteres, c) posibilitar la elaboración de claves taxonómicas con criterios de máxima operatividad, al poder elegir el carácter con más contenido de información para diferenciar las especies de un género, éstos entre sí o de forma global las especies del conjunto de la tribu, y d) comparar el nivel de diversificación conseguido por cada uno de los géneros estudiados, consiguiendo de esta forma nuevos elementos para enjuiciar las pautas de evolución de la tribu*.

La matriz de datos sobre la que se inicia el análisis es la misma que sirvió de base a la clasificación jerárquica ya publicada (CLEMENTE MUÑOZ & HERNÁNDEZ BERMEJO, 1980b). Los doce caracteres estudiados producían sobre los ciento cuarenta y cinco táxones de *Brassicaceae* (cuarenta y un géneros) una serie de agrupaciones c_1 c_2 ... c_{12} con diferente número de clases según las posibilidades particulares de cada carácter. El elemento a_{ij} de la matriz** representaba la clase o grupo en que resultaba incluida la especie i de acuerdo con la partición o agrupamiento correspondiente al carácter j . En los resultados de las tablas, los caracteres estudiados se abrevian con las siguientes siglas:

Carácter	Siglas	N.º de clases que distingue el carácter
Morfometría de pétalos	mp	19
Morfometría de sépalos	ms	17
Sépalos cuculados	cu	6
Pilosidad de sépalos	p	5
Nerviación de pétalos	n	5
Nectarios medianos	nm	9
Color de pétalos	cp	8
Simetría de la corola	sm	2
Posición de los sépalos respecto al eje floral	ef	5
Color de sépalos	cs	7
Escotadura de pétalos	e	2
Alas en la zona de transición limbo/uña del pétalo	a	6

(*) Un resumen y adelanto de este trabajo fue remitido por los autores al XIII Congreso Internacional de Botánica, celebrado en Sidney.

(**) Se omite la inclusión de dicha matriz; puede consultarse en CLEMENTE MUÑOZ & HERNÁNDEZ BERMEJO (1980b).

Como se ha indicado, nos proponemos analizar el valor taxonómico de dichos caracteres para el conjunto de la tribu y para cada uno de los géneros politípicos considerados, estimando además en cada uno de ellos el grado de diversificación conseguido. Utilizaremos a este fin la expresión de Shannon que expresa en bits (unidades de elección binaria) la información o dicho de otro modo, la entropía provocada por un criterio de partición (aquí caracteres taxonómicos) sobre un conjunto de elementos (en nuestro caso táxones, a los niveles de especie y subespecie) agrupables en un cierto número de clases (en este caso, cada una de las posibles expresiones del carácter).

MÉTODO

Aunque parte del desarrollo teórico que fundamenta las aplicaciones de la teoría de la información a la taxonomía pueden encontrarse en ESTABROOK (1967), HAWKSWORTH & *al.* (1968), DUNCAN & ESTABROOK (1976) e incluso en aplicaciones a otras disciplinas biológicas (GODRON, 1968; LEGENDRE & LEGENDRE, 1979), resumiremos el método de cálculo señalando las variaciones al mismo por nosotros sugeridas y utilizadas.

Se estudia un conjunto de **S** especies (especies más subespecies) pertenecientes a **1, 2, 3 ... i ... n** géneros, cada uno de los cuales incluye **s_i** especies.

Evidentemente: $S = \sum_{i=1}^n s_i$

Se consideran los caracteres taxonómicos **1, 2 ... j ... m** que consiguen otras tantas particiones sobre las **S** especies. El número de clases **r** utilizado (o distinguido) por cada carácter **j** depende de la naturaleza de éste. Así el carácter **j** distinguirá las clases **j₁, j₂ ... j_k ... j_r***. Llamaremos **s(j_k)** al número de especies que son clasificadas como pertenecientes a la clase **j_k** y como **s_i(j_k)** al de especies del género **i** que son clasificadas dentro de la citada clase.

También es evidente que $s(j_k) = \sum_{i=1}^n s_i(j_k)$.

Definiremos ahora las dos siguientes probabilidades:

La probabilidad de que una especie elegida al azar pertenezca a la clase **j_k** tras estudiarla con arreglo al carácter **j**, es:

$$P_k = \frac{s(j_k)}{S} \quad (1)$$

La probabilidad de que una especie del género **i** pertenezca a la clase **j_k** será:

$$P_k^i = \frac{s_i(j_k)}{s_i} \quad (2)$$

(*) Aunque obviamente **r** es función de **j**, escribiremos **j_r** y no **j_{rj}**) para no complicar aún más la notación.

Aplicando con (1) la función de Shannon calculamos la entropía que consigue el carácter j sobre las S especies:

$$H_j = -\sum_{k=1}^r p_k \log_2 p_k \tag{3}$$

Los sucesivos valores $H_1, H_2 \dots H_j \dots H_m$ nos permiten comparar la entropía producida por cada uno de los m caracteres sobre las S especies y elegir los más adecuados para la diferenciación y descripción de cada uno de los táxones, así como para la elaboración de claves para su determinación.

Aplicando con (2) la función de Shannon, analizamos el mismo efecto de forma intragenérica:

$$H_j^i = -\sum_{k=1}^r p_k^i \log_2 p_k^i$$

Así calculamos la heterogeneidad producida por el carácter j sobre las s_i especies del género i . Los caracteres óptimos para la diferenciación intragenérica son de esta forma detectados.

A la hora de representar los valores H_j^i en función de la variable r y a fin de conocer, por ej., cuáles son los mejores caracteres para diferenciar las especies del género i entre sí, habrá que tener en cuenta la existencia de unos valores máximos que serán:

$$H_{\text{máx}} = \log_2 r, \quad \text{cuando } r < s_i \tag{4}$$

$$H_{\text{máx}} = \log_2 s_i, \quad \text{cuando } s_i < r \tag{5}$$

El valor (4) se deduce al considerar el caso en que todas las clases sean igualmente probables. La expresión de Shannon alcanza entonces su máximo valor correspondiente a una heterogeneidad o entropía máximas. Pero no siempre hay tantas especies en un género como clases distinguidas por el carácter j . En este caso el valor máximo de H es el de (5).

Proponemos a partir del cálculo de los sucesivos valores $H_j^1, H_j^2 \dots H_j^i \dots H_j^m$ el cálculo de la entropía media \bar{H}_j^i conseguida por el carácter j entre los géneros estudiados. También hemos utilizado la relación \bar{H}_j^i/\bar{H}_j^i para estimar de alguna forma el valor discriminante intergenérico del carácter j ; ésta nos compara la entropía total producida sobre las S especies con la que resulta absorbida —en valor promediado— por la distinción de especies intragenéricamente.

Para el estudio de la interdependencia entre caracteres, ESTABROOK (1967) utiliza como estimador de dicha interdependencia el valor de una distancia entre dos caracteres j y l , definida como:

$$d(j, l) = \frac{H(j/l) + H(l/j)}{H(j \cup l)} \tag{6} (*)$$

(*) Aunque otros autores prefieren escribir $H(j, l)$ y $R(j, l)$ elegimos la notación $H(j \cup l)$ y $H(j \cap l)$ respectivamente.

o el de la semejanza, transformación de (6) como:

$$\text{sem}(j, l) = \sqrt{1 - d(j, l)^2} \quad (7)$$

Si $1, 2, \dots, k \dots r$ son las clases distinguidas por el carácter j y $1, 2 \dots k' \dots r'$ las utilizadas por el carácter l , las expresiones (6) y (7) se calculan en base a que:

$$H(j/l) = \sum_{k'=1}^{r'} H(j/l_{k'}) \cdot p(l_{k'})$$

$$H(j/l_{k'}) = \sum_{k=1}^r p(j_k/l_{k'}) \cdot \log_2 p(j_k/l_{k'})$$

$$p(j_k/l_{k'}) = \frac{P(j_k l_{k'})}{p(l_{k'})}$$

Cumpléndose naturalmente que:

$$H(j) - H(j/l) = H(l) - H(l/j) = H(j \cap l)$$

y

$$H(j \cup l) = H(j/l) + H(l/j) + H(j \cap l)$$

Parecidos algoritmos y razonamientos inspiraron a GODRON (1968) algunas aplicaciones en el campo de la ecología vegetal para el cálculo de la información mutua entre especies de una serie de inventarios o comunidades, o entre aquellas y determinados factores ambientales.

Según el coeficiente d propuesto por Estabrook, el valor máximo se alcanza cuando los dos caracteres j y l son absolutamente disyuntivos, es decir cuando $H(j \cap l) = 0$. El valor mínimo $d = 0$ implica por el contrario que $H(j/l) = H(l/j) = 0$, es decir que j y l son el mismo carácter. Entre estos valores d mide el grado de interdependencia, considerándose a j y l como caracteres tanto más correlacionados cuando mayor sea la información común (o redundancia) de ambos, esto es $H(j \cap l)$ versus $H(j/l)$ y $H(l/j)$.

Cuando el número de clases de los dos caracteres comparados es muy similar ($r = r'$) no existe ninguna limitación para la aplicación de los coeficientes d o sem de Estabrook, alcanzándose el valor $d = 0$ cuando $\forall (k, k')$ se cumple que $p(j_k/l_{k'}) = 1$ y entonces $\log_2 p(j_k/l_{k'}) = 0$, lo que significa que todas las especies incluidas en la clase j_k pertenecen también a la clase $l_{k'}$ para todo par de clases (k, k') . (8)

Sin embargo, hemos detectado en la aplicación de estos coeficientes, que cuando $r \neq r'$ (esto es frecuente y a veces de forma muy señalada; véanse por ejemplo en nuestro caso los caracteres «morfología de pétalos» y «simetría de la corola», con $r = 19$ y $r' = 2$, respectivamente), surgen importantes limi-

taciones al no cumplirse el razonamiento anterior (8). Se imposibilita, en consecuencia, el valor mínimo $d=0$ y se distorsionan los valores de d y sem enmascarándose incluso en ocasiones la posible dependencia de los caracteres con reducido número de clases.

Por esta razón hemos calculado y preferimos los coeficientes:

$$q = \frac{H(j \cap l)}{H(j)} \times 100 \tag{9}$$

$$q' = \frac{H(j \cap l)}{H(l)} \times 100 \tag{10}$$

inspirados en el cálculo de coeficientes de semejanza referidas a totales marginales, que ya fueron utilizados con éxito, ante parecidas dificultades, por HERNÁNDEZ BERMEJO & SAINZ OLLERO (1984) en ciertas aplicaciones del análisis de semejanza a la fitogeografía.

Estos coeficientes permiten la comparación de caracteres con marcadas diferencias en contenido de información y la detección de cuál de los dos caracteres resulta más superfluo al manifestarse la redundancia $H(j \cap l)$ en términos relativos a la información conseguida por cada carácter $H(j)$ y $H(l)$.

Una aportación metodológica, también aquí sugerida, consiste en el cálculo de los H_j^i valores a fin de analizar el grado de diversificación conseguido por cada género i a través de sus especies. Para ello calculamos el valor medio \bar{H}_j^i , es decir el que resulta en cada género al promediar los m caracteres: $\bar{H}_j^i = (H_1^i + H_2^i + \dots + H_j^i + \dots + H_m^i) / m$. La representación en coordenadas cartesianas de \bar{H}_j^i versus s permite interpretar el nivel de heterogeneidad o diversificación alcanzado por cada género en función de los caracteres analizados en el estudio y del número de especies que lo integran.

RESULTADOS

Aplicando el método descrito al conjunto de géneros analizados, se obtienen en primer lugar los valores de la matriz H_j^i (cuadro 1). Estos corresponden a la información que cada carácter j produce sobre cada género i . La fila \bar{H}_j incluye los contenidos de información de cada carácter considerando globalmente todas las especies estudiadas sin atender a su agrupación en géneros. Por el contrario, \bar{H}_j^i es el valor medio de los valores $H_1^i, H_2^i, \dots, H_j^i, \dots, H_m^i$. La columna \bar{H}_j^i recoge la heterogeneidad media de cada género en función de los doce caracteres. Los mayores valores corresponden a *Brassica*, *Diplotaxis* y *Erucastrum*.

En la figura 1 se representan los valores medios \bar{H}_j^i de información contenida en cada género (en base a los j caracteres) versus el número de especies de cada uno de ellos. La curva de regresión representada ($\bar{H}_j^i = 0,0563 + 0,9572 \log s_i$) se ajusta a los valores observados con $r = 0,9666$. En ella se aprecia cómo a pesar de corresponder los géneros con mayor

CUADRO 1

Caracteres n.º clases	(f) (r)	mp 19	ms 17	cu 6	p 5	n 5	nm 9	cp 8	sm 2	ef 5	cs 7	e 2	a 6	\bar{H}_j^i
Géneros	Núm. espec. estud.													
<i>Brassica</i>	28	2,48	2,42	2,25	1,79	1,78	2,08	1,22	0,59	1,07	1,37	0,37	0,37	1,48
<i>Cakile</i>	3	0,92	0,92	0,92	0,92	0,00	0,92	0,00	0,00	0,92	0,00	0,00	0,00	0,46
<i>Coringia</i>	2	0,00	1,00	1,00	0,00	1,00	0,00	1,00	0,00	0,00	1,00	0,00	0,00	0,42
<i>Grambe</i>	12	2,35	1,96	0,41	0,41	0,55	0,92	0,00	0,92	0,00	1,65	0,00	0,00	0,85
<i>Diploaxis</i>	18	2,05	2,71	1,54	2,11	1,39	1,86	1,23	0,99	0,00	1,23	0,00	0,00	1,26
<i>Eruca</i>	3	0,00	0,92	0,00	1,58	0,00	0,00	0,00	0,92	0,92	0,92	0,92	0,00	0,51
<i>Erucaria</i>	4	1,50	1,50	1,00	0,81	0,81	0,00	0,00	1,00	0,81	1,50	0,00	0,81	0,81
<i>Erucastrum</i>	14	1,96	2,49	1,73	1,66	0,95	1,48	0,75	0,94	0,75	0,75	0,00	0,00	1,12
<i>Hirschfeldia</i>	2	1,00	1,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,25
<i>Hutera</i>	2	0,00	1,00	0,00	0,00	0,00	0,00	0,00	1,00	1,00	0,00	0,00	0,00	0,25
<i>Hutera</i> (*)	9	1,75	1,91	1,30	0,95	0,54	0,81	0,00	0,81	1,55	0,95	0,00	0,00	0,88
<i>Moricandia</i>	7	0,99	2,24	1,38	0,00	0,00	0,00	0,00	0,00	0,86	1,67	0,99	0,59	0,73
<i>Raphanus</i>	4	0,81	2,00	1,00	1,50	0,00	0,81	0,00	1,00	0,81	1,00	0,81	0,00	0,81
<i>Rapistrum</i>	2	1,00	1,00	1,00	1,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,42
<i>Rhynchosinapis</i>	6	1,92	1,58	0,65	1,00	0,65	0,92	0,00	0,65	0,65	1,00	0,00	0,00	0,75
<i>Sinapidendron</i>	3	1,58	0,00	0,92	0,92	1,58	1,58	0,92	0,92	0,00	0,00	0,00	0,00	0,70
<i>Sinapis</i>	6	1,79	1,00	1,79	1,92	1,00	0,92	0,00	1,00	0,00	1,00	0,00	0,00	0,87
<i>Trachystoma</i>	3	0,92	1,58	0,92	0,00	0,92	1,58	0,00	0,00	1,58	0,00	0,00	0,00	0,63
<i>Vella</i>	2	0,00	1,00	1,00	0,00	1,00	1,00	0,00	1,00	0,00	0,00	0,00	0,00	0,42
\bar{H}_j		1,18	1,46	0,97	0,92	0,76	0,78	0,28	0,61	0,52	0,73	0,17	0,27	
\bar{H}_j		3,34	3,53	2,11	2,05	1,97	2,85	2,37	0,98	1,91	1,93	0,48	0,52	
\bar{H}_j/\bar{H}_j		2,83	2,41	2,17	2,22	2,61	3,64	8,34	1,61	8,69	2,66	2,82	1,94	

Información producida por los caracteres sobre los géneros poliploídicos estudiados.
 (*) *Hutera* sensu Gómez Campo 1977 (este valor no ha sido considerado en los promedios).

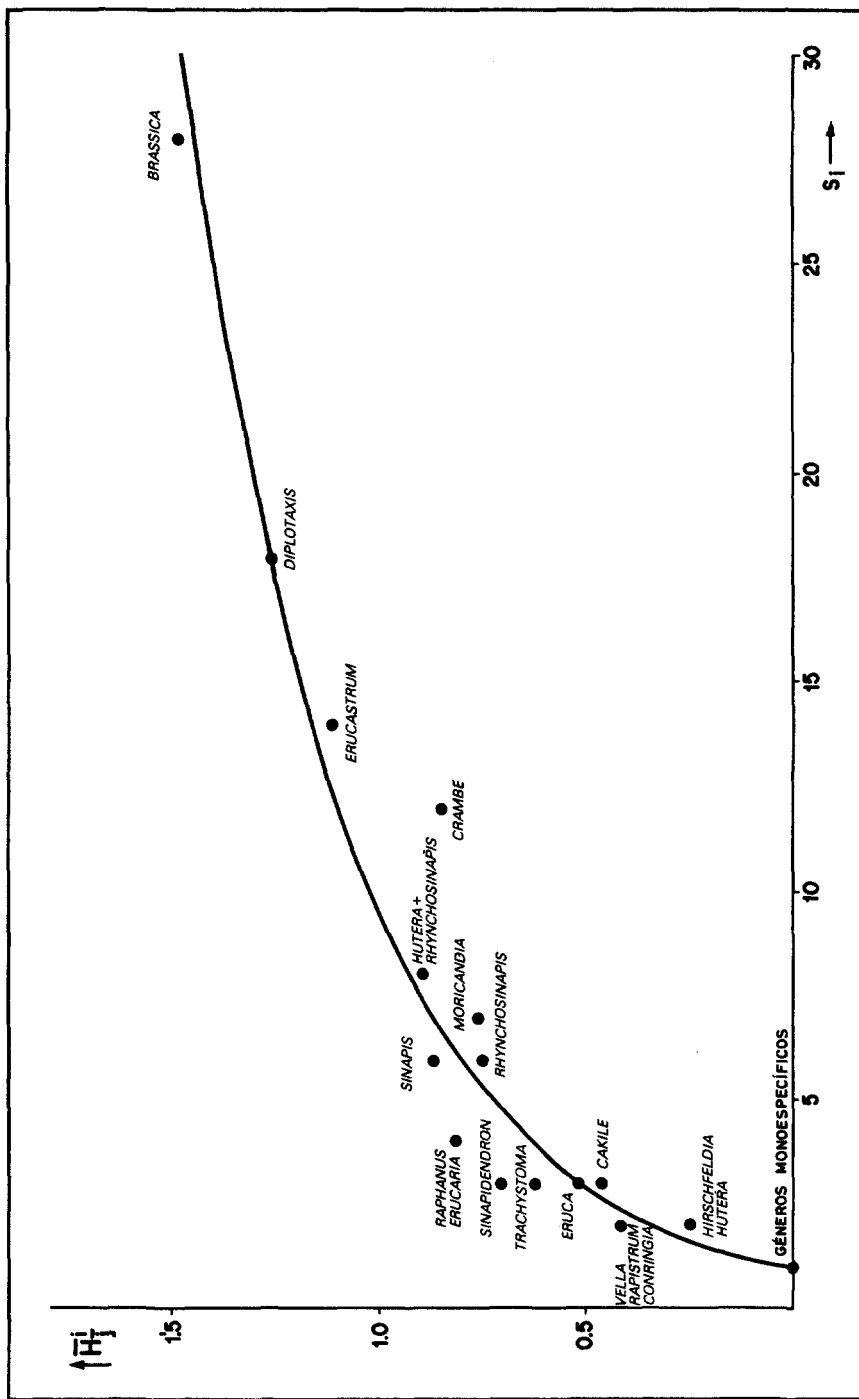


Fig. 1.—Información (heterogeneidad) media (\bar{H}_i) manifestada por cada género en función de los caracteres estudiados (S_1). La curva representada es la de regresión $\bar{H}_i = 0,0563 + 0,9572 \log S_1$ ($r = 0,99666$).

número de especies a los de mayor contenido de información, hay otros que se sitúan por encima de los valores esperados por la curva, lo que aumenta el valor diagnóstico de la figura.

El cuadro 2 recoge los coeficientes de distancia y semejanza **d** y **sem** estimadores —según ESTABROOK (1967)— de la interdependencia entre caracteres. Los valores de la redundancia relativa a los valores marginales **q** y **q'** aparecen en el cuadro 3. En el cuadro 4 se calculan los valores medios de los cuatro coeficientes anteriores, para cada uno de los caracteres taxonómicos considerados.

CUADRO 2

	mp	ms	cu	p	n	nm	cp	sm	ef	cs	e	a
mp	—	.598	.378	.330	.527	.546	.528	.241	.550	.440	.224	.379
ms	.802	—	.396	.385	.386	.542	.546	.182	.555	.454	.239	.285
cu	.926	.918	—	.212	.225	.388	.360	.102	.216	.267	.224	.153
p	.944	.923	.977	—	.231	.288	.405	.139	.214	.303	.178	.244
n	.850	.923	.974	.973	—	.472	.497	.117	.261	.302	.156	.192
nm	.837	.840	.922	.958	.881	—	.593	.175	.482	.382	.223	.244
cp	.849	.838	.933	.914	.868	.805	—	.189	.579	.483	.259	.348
sm	.971	.983	.995	.990	.993	.985	.982	—	.214	.249	.012	.207
ef	.835	.832	.976	.977	.965	.876	.815	.977	—	.423	.290	.389
cs	.898	.891	.964	.953	.953	.924	.876	.968	.906	—	.246	.342
e	.975	.971	.975	.984	.988	.975	.966	.999	.957	.969	—	.164
a	.925	.958	.988	.970	.981	.970	.937	.978	.921	.940	.986	—

Interrelación entre pares de caracteres. Valores de la semejanza (**sem.**) —valores por encima de la diagonal— y distancia (**d**) —valores por debajo de la diagonal—.

$$d(j, l) = \frac{H(j/l) + H(l/j)}{H(j \cup l)}; \quad sem(j, l) = \sqrt{1 - d^2(j, l)}$$

CUADRO 3

	mp	ms	cu	p	n	nm	cp	sm	ef	cs	e	a
mp	—	34,1	11,3	8,6	16,9	30,1	22,5	3,7	22,3	14,6	2,8	8,0
ms	28,3	—	12,1	11,3	11,2	24,5	23,4	2,1	22,2	15,1	3,2	4,6
cu	17,8	20,3	—	4,4	4,9	17,0	13,4	0,8	4,4	6,7	3,0	1,5
p	14,0	19,6	4,5	—	5,2	9,7	17,1	1,4	4,4	8,6	1,9	3,7
n	24,8	20,1	5,2	5,4	—	26,1	25,7	1,0	6,6	8,8	1,5	2,3
nm	35,3	35,2	12,6	7,0	18,1	—	29,9	2,0	18,4	11,8	2,9	3,5
cp	31,4	34,7	11,9	14,7	21,3	35,7	—	2,5	28,1	19,9	4,0	7,2
sm	12,5	7,6	1,7	3,0	2,1	5,9	6,2	—	6,6	9,0	0,0	3,2
ef	38,9	41,1	4,9	4,7	6,8	27,4	35,1	3,4	—	17,2	5,1	9,3
cs	25,5	28,1	7,4	9,3	9,1	17,5	24,9	4,6	17,2	—	3,7	7,2
e	19,7	23,7	13,4	8,2	6,2	17,1	19,8	0,2	20,5	14,9	—	2,8
a	51,4	31,0	5,9	14,5	8,7	19,0	32,7	6,1	34,0	26,5	2,6	—

Valores de los coeficientes $q = 100 \times H(j \cap l) / H(j)$ —por debajo de la diagonal— y $q' = 100 \times H(j \cap l) / H(l)$ —por encima de la diagonal—.

Podemos utilizar la diferencia o el coeficiente entre \bar{H}_j y \bar{H}_j^i es decir entre la entropía que cada carácter j produce sobre el conjunto de las S especies y la entropía media absorbida intragenéricamente, para dar una idea de la entropía debida o producida por la diferenciación intergenérica. Así, la figura 2, podrá mostrarnos cómo los caracteres **cp** (color de pétalos), **nm** (nectarios

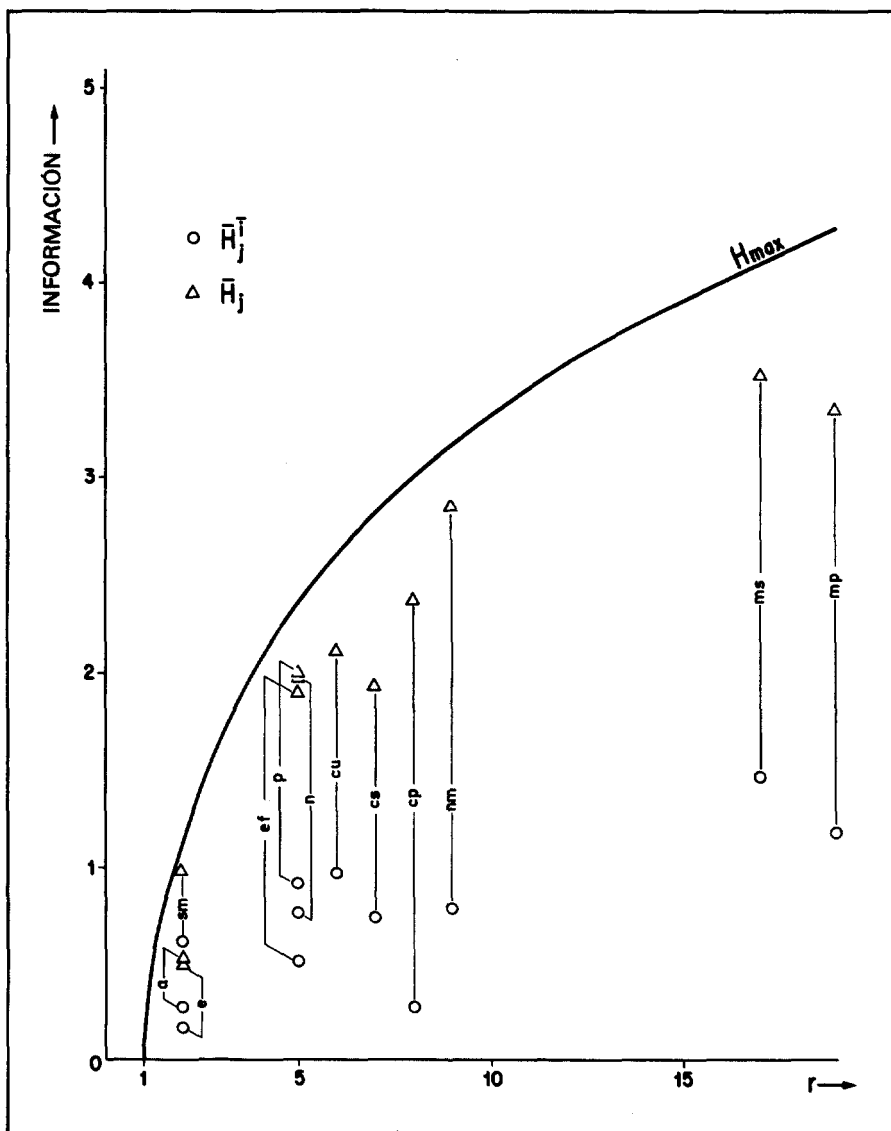


Fig. 2.—Representación de los valores \bar{H}_j^i y \bar{H}_j versus r , bajo la curva de H_{max} . Las diferencias $\bar{H}_j - \bar{H}_j^i$ pueden dar idea de la capacidad de diagnóstico intergenérico de cada carácter j .

medianos), o **ef** (posición de sépalos respecto al eje floral) son mejores caracteres respecto a la distinción entre géneros que **cs** (color de sépalos), **n** (nerviación) o **a** (alas limbo/uña del pétalo).

CUADRO 4

	mp	ms	cu	p	n	nm	cp	sm	ef	cs	e	a
$\bar{d}(j, l)$.892	.898	.959	.960	.941	.907	.889	.984	.912	.931	.977	.959
$\overline{sem}(j, l)$.431	.415	.266	.266	.306	.394	.435	.166	.379	.353	.201	.268
Valor medio												
$\bar{q} = \frac{H(j \cap l)}{H(j)} \forall j$	27,6	26,3	9,6	9,5	11,9	21,0	24,4	4,8	21,3	26,8	13,3	21,1
$\bar{q}' = \frac{H(j \cap l)}{H(l)} \forall l$	15,9	14,9	7,2	6,9	10,0	16,0	17,6	2,9	13,1	11,1	2,6	4,8

Valores medios de la interdependencia de cada carácter con los restantes, expresados según coeficientes de distancia $d(j, l)$, de semejanza $sem(j, l)$ y de redundancia relativa a los valores marginales según q y q' .

CUADRO 5

	Brassica	Cakile	Conringia	Crambe	Diplolaxis	Eruca	Erucaria	Erucastrum	Hirschfeldia	Hutera	Moricandia	Raphanus	Rapistrum	Rhynchosinapis	Sinapis	Sinapidendron	Trachystoma	Vella	Diferenciación intergenérica
Sépalos cuculados	3	1	2	1	1	0	1	2	0	0	2	1	2	1	2	1	1	2	1
Pilosidad sépalos	2	1	0	1	3	2	1	2	2	0	0	2	2	1	3	1	0	0	1
Simetría corola	2	0	0	2	2	2	2	2	0	2	0	2	0	2	2	2	0	2	0
Nerviación pétalos	2	0	2	2	1	0	1	1	0	0	0	0	2	1	1	2	1	2	1
Color sépalos	1	0	2	2	1	1	2	1	0	0	2	1	0	1	1	0	0	0	1
Morfometría sépalos	1	1	2	1	2	1	2	1	2	2	2	2	2	1	1	0	2	2	1
Morfometría pétalos	1	1	0	1	1	0	2	1	2	0	1	1	2	2	2	2	1	0	1
Nectarios medianos	1	1	0	1	1	0	0	1	0	0	1	0	1	1	2	2	2	2	2
Posición sépalos-eje flor	1	1	0	0	0	1	1	1	0	2	1	1	0	1	0	0	2	0	2
Color pétalos	1	0	2	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0	3
Escotadura pétalos	1	0	0	0	0	2	0	0	0	0	2	2	0	0	0	0	0	0	0
Alas en la transición limbo/uña del pétalo	1	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0

Interpretación de la figura 2. Se valora la capacidad de diagnóstico de los caracteres dentro del conjunto de especies de cada género (3: elevada; 2: aceptable; 1: escasa; 0: nula).

Una vez conseguida la diferenciación intergenérica, puede plantearse el que suele ser más serio problema taxonómico: la diferenciación de especies de un mismo género. La figura 3 analiza el valor de diagnóstico, es decir la información o heterogeneidad producida por los doce caracteres dentro de cada uno de los géneros politípicos estudiados, siempre en función del número de clases utilizado por cada carácter, pues a igualdad de información conseguida será preferible el que menos clases distinga, es decir el de más sencilla observación. Los valores de la información máxima posible (H_{\max} y H'_{\max}) ayudan también a la elección de los caracteres con mayor capacidad de diagnóstico. Un resumen de la interpretación de la figura 3 se presenta en el cuadro 5 con el que ya se puede proceder de forma inmediata a la elección ordenada de los mejores caracteres, tanto para las especies de cada género, como para la elaboración de claves de determinación.

DISCUSIÓN Y CONCLUSIONES

El conjunto de aplicaciones de la teoría de la información a la taxonomía numérica resulta muy útil tanto en los aspectos ya ensayados por otros autores (estudio de la interdependencia entre caracteres, comparación entre sistemas de clasificación) como en los aquí propuestos (análisis de la variabilidad intragenérica, cálculo de la información redundante marginal entre caracteres). En el caso de la tribu *Brassicaceae* y manejando caracteres taxonómicos procedentes en exclusiva de las piezas estériles de la flor, estos métodos han permitido además elegir los caracteres más eficaces, tanto para la diferenciación de táxones entre las especies de un mismo género, como incluso intergenéricamente. La elaboración de claves resulta así rápida, y de máxima operatividad.

Una aplicación perfecta de estos métodos requeriría haber considerado la totalidad de los táxones integrantes del conjunto estudiado. Esto, obviamente, no siempre es posible, en especial cuando se manejan familias, tribus o géneros muy numerosos, geográfica y ecológicamente dispersos. En nuestro caso los ciento cuarenta y cinco táxones estudiados significan un porcentaje muy elevado y representativo de la tribu (formada por unas doscientas treinta especies). De los cincuenta y dos géneros que la constituyen, sólo ocho no han podido ser considerados. La mayor parte de los géneros monotípicos están incluidos. De los politípicos el % de especies estudiadas están en muchos casos entre el 75 y 100% (por ej. en *Diplotaxis*, *Sinapis*, *Eruca*, *Erucastrum*, *Moricandia*, *Raphanus* y *Trachystoma*). Tan sólo en cuatro géneros el porcentaje es inferior al 50% (*Enarthrocarpus*, *Conringia*, *Pseudoerucaria* y *Sinapidendron*). En definitiva, los resultados pueden ser interpretados como muy representativos con sólo escasas restricciones.

La figura 1 nos permite observar la tendencia general de los géneros de la tribu respecto a su grado de variabilidad. Los pares de valores (H'_j , s_j) se ajustan muy aceptablemente a la curva de la figura. Como era de esperar, esa variabilidad (estimada en base a la heterogeneidad o información media producida por los doce caracteres florales estudiados) tiende a estabilizarse al

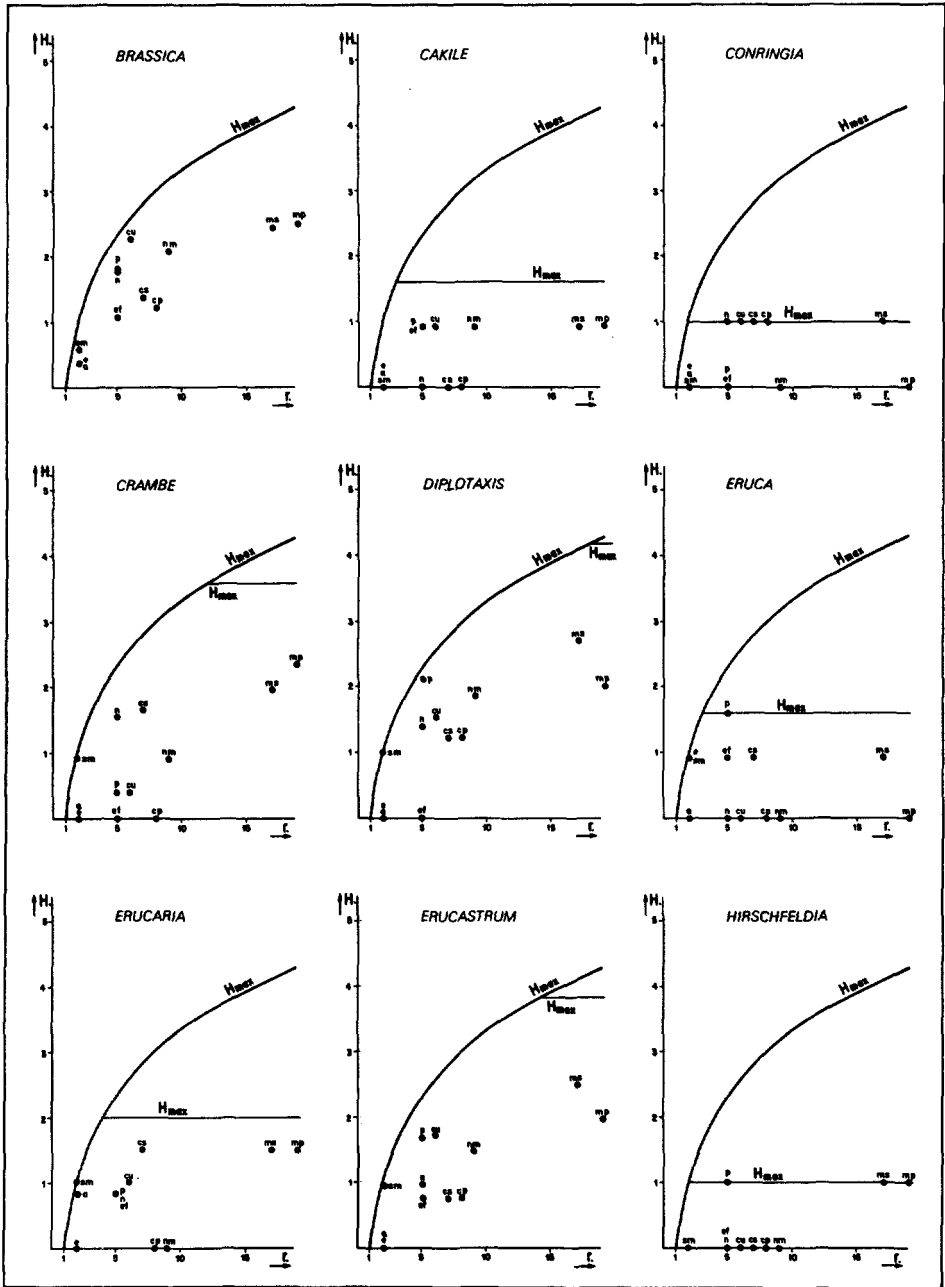
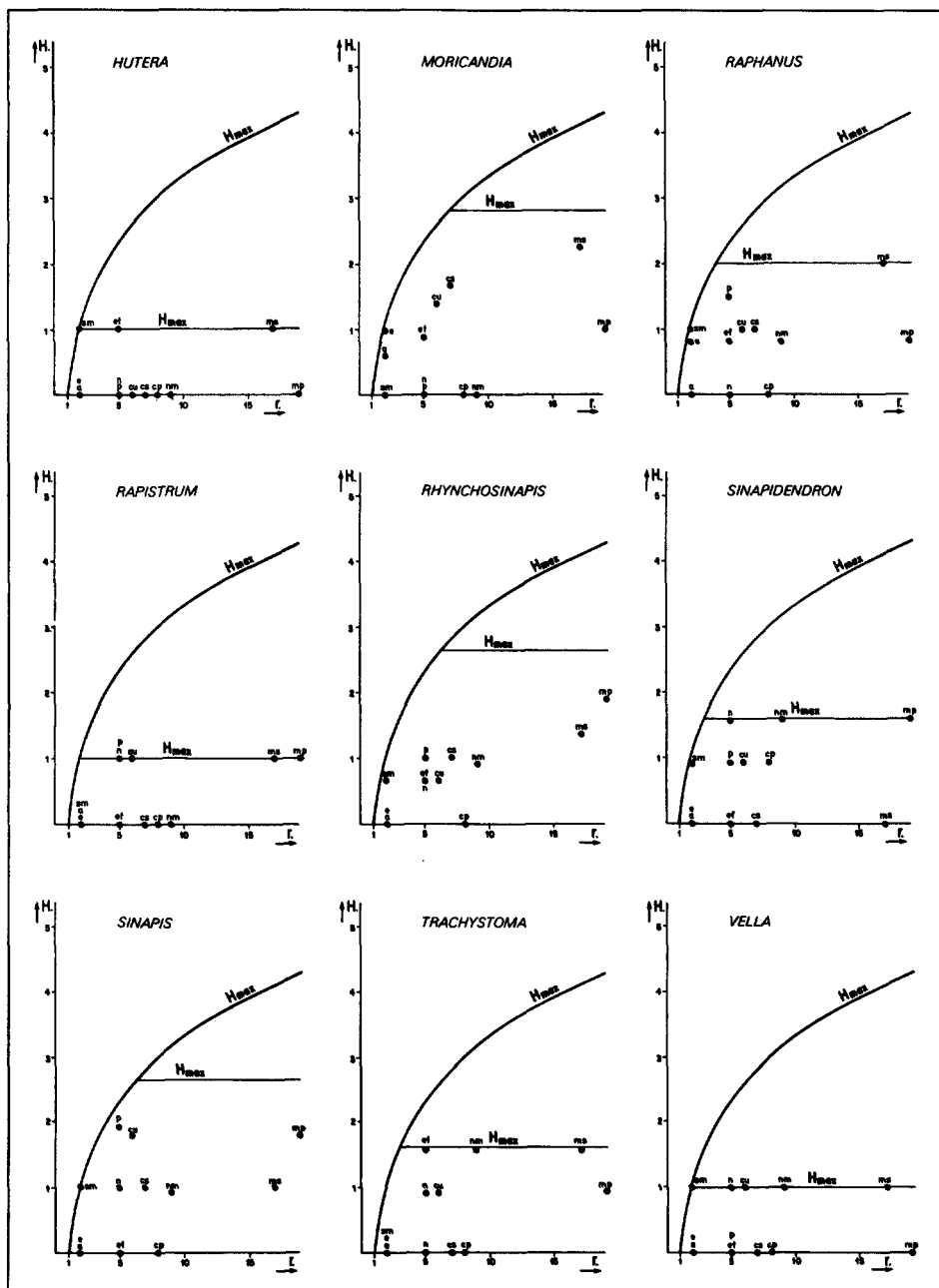


Fig. 3.—Capacidad de diagnóstico de los caracteres en función del número de clases por ellos distinguidos. Representaciones separadas para cada uno de los géneros politépicos estudiados.



aumentar el número de especies constitutivas del género. El techo podría estar muy próximo al que alcanza *Brassica* (1,48 bits respecto a 28 especies), puesto que la información media de los caracteres para el total de la tribu se sitúa en torno a los 2,02 bits (realmente estas cifras corresponden a bits/especie). Algunos géneros con muchas menos especies que *Brassica*, como *Sinapis* o *Hutera* (sensu GÓMEZ-CAMPO, 1977) presentan sólo un 40% menos de variabilidad, con cuatro veces menor número de especies. El orden de colocación de los géneros habla en cualquier caso de un aumento progresivo de la variabilidad parejo (pero progresivamente menor) al del número de especies. Sorprenden, no obstante, algunas distorsiones a esta norma. Así *Crambe* y en menor medida *Moricandia* manifiestan una heterogeneidad mucho menor que la esperada a partir de la tendencia general. Esto puede significar un menor grado de diversificación real que podría sugerirnos una especiación más reciente; o lo contrario, la adopción de una vía evolutiva de estatigénesis, es decir, de envejecimiento del phylum y constancia de caracteres por mayor compromiso evolutivo, lo que en nuestro caso implicaría una solución floral muy definitiva y la adopción de estrategias relativamente estables en cuanto al sistema, momento y vector de polinización. Nos inclinamos por esta interpretación que confirmaría anteriores resultados al respecto (CLEMENTE MUÑOZ & HERNÁNDEZ BERMEJO, 1980b) en los que *Crambe* y *Moricandia* eran ya interpretados como dos de los géneros de la tribu con flor más constante y definida.

Conviene matizar, no obstante, que la mayor o menor variabilidad registrada en un género puede también ser función de la mayor o menor minuciosidad de los autores del mismo y de sus especies, o de un grado de conocimiento y definición aún imperfecto del citado taxon. Este puede ser el caso de *Sinapidendron*, en el que tal vez cabría replantearse su consistencia y límites.

La elección de los caracteres que producen mayor heterogeneidad y por lo tanto mayor capacidad de diagnóstico es otra de las aplicaciones citadas y aquí comprobadas del análisis de información. Una primera cuestión puede ser la búsqueda de caracteres con mejor diagnosis intergenérica. En este sentido, la interpretación de la figura 2 resulta de utilidad. La diferencia de valores $\bar{H}_i - \bar{H}_j$ nos informa en cierta medida de esta cualidad. A igualdad de este valor entre dos o más caracteres es evidente que será preferible el que utilice un menor número de clases. Con estos criterios se deduce que el carácter «color de pétalos» es el mejor para el reconocimiento de los géneros de la tribu, siguiéndole en interés «posición de sépalos respecto al eje floral» y «nectarios medianos». En contraste con esta conclusión resulta notable el poco uso que las claves dicotómicas publicadas hacen de los mencionados caracteres. *Flora Europaea* (TUTIN & al., 1964) por ejemplo, utiliza el color de los pétalos de forma muy limitada en la diferenciación de los géneros de *Brassicaceae* y para nada los otros dos caracteres citados. Sin embargo hay que recordar que SCHULZ (1936) sí empleó el carácter «nectarios medianos», justamente para la distinción de las subtribus de *Brassicaceae*.

El método es todavía más inmediato y contundente a la hora de buscar los caracteres óptimos para la diferenciación interespecifica dentro de cada uno de los géneros politípicos estudiados. Los gráficos de la figura 3, interpretados

y resumidos en el cuadro 5 permiten elegir los caracteres más adecuados en cada género. Así, si tuviéramos que definir las especies de *Brassica*, el mejor carácter sería, con diferencia, el de «sépalos cuculados» que, junto con «pilosidad de sépalos», «simetría de la corola» y «nerviación de pétalos», permite diferenciar bastante bien las especies de este nutrido género. Ningún carácter sobresale especialmente para *Cakile*. *Conringia* puede diferenciarse por «sépalos cuculados», «nerviación de pétalos» y «color de pétalos y sépalos». *Crambe*, por «simetría de la corola», «nerviación de pétalos» y «color de sépalos». *Diplotaxis*, fundamentalmente por «pilosidad de sépalos», así como también por «simetría de la corola», y «morfometría de sépalos». *Eruca*, por «pilosidad de sépalos», «simetría de la corola», y «sépalos cuculados». *Erucaria*, por «simetría de la corola», «color de sépalos» y «morfometría de sépalos y pétalos». *Erucastrum*, por «sépalos cuculados», «pilosidad de sépalos» y «simetría de la corola». *Hirschfeldia*, por «pilosidad de sépalos» y «morfometría de pétalos y sépalos». *Hutera*, por «simetría de la corola», «morfometría de pétalos y sépalos», y «posición de sépalos respecto al eje floral». *Moricandia*, fundamentalmente por «sépalos cuculados» y «morfometría y color de sépalos». *Raphanus*, por «pilosidad de sépalos», «simetría de la corola», «morfometría de sépalos» y «escotadura de pétalos». *Sinapis*, especialmente por la «pilosidad de sépalos», así como por «simetría de la corola», «morfometría de pétalos» y «sépalos cuculados». *Sinapidendron*, por «simetría de la corola», «nerviación de pétalos», «morfometría de pétalos» y «nectarios medianos». *Trachystoma*, por «morfometría de sépalos», «nectarios medianos» y «posición de sépalos respecto al eje floral». Y *Vella*, finalmente, por «sépalos cuculados», «nectarios medianos», «simetría de la corola», «nerviación de pétalos» y «morfometría de sépalos».

Un primer test de la capacidad de diagnosis global de los caracteres contrastados podríamos obtenerlo con el sumatorio por filas del cuadro 3. Los peores caracteres para el análisis interespecífico serían «color de pétalos», «nectarios medianos» y «posición de sépalos respecto al eje floral», precisamente los tres mejores para el diagnóstico intergenérico. Lo contrario ocurre con «simetría de la corola», «sépalos cuculados», «pilosidad de sépalos» y las dos «morfometrías» de pétalos y sépalos, buenos diferenciadores a nivel intragenérico.

Queda todavía por analizar la interdependencia entre caracteres. Los coeficientes de distancia (**d**) y semejanza (**sem**) utilizados por ESTABROOK (1968) resultaron siempre superiores a $d=0,800$ ($sem \simeq 0,600$). La mayor parte de los valores se sitúan incluso por encima de $d>0,950$ lo que indica —a pesar de las dificultades que pueda haber para establecer unos niveles objetivos de significación— una generalizada independencia que habla a favor de su correcta utilización en trabajos anteriores (CLEMENTE MUÑOZ & HERNÁNDEZ BERMEJO 1978a, 1978b, 1980a, 1980b). Los valores registrados entre «color de pétalos» y «nectarios medianos» así como entre el par «morfometría de pétalos» y «morfometría de sépalos» son los que presentan una distancia (**d**) menor, como resultado posiblemente de que ciertos géneros presentan unos modelos de flor muy definidos respecto a ellos (*Crambe*, *Moricandia*, *Hutera*, ...), por lo que sus especies han podido provocar cierto grado de correlación.

La utilización de los coeficientes de redundancia q y q' hace patente algunos casos de interdependencia que los coeficientes d y sem de Estabrook enmascaraban por causa de la marcada diferencia entre el número de clases de los dos caracteres comparados. Esto es particularmente notorio en el caso del par «morfometría de pétalos» y «alas transición limbo/uña», cuya distancia es $d=0,925$ pero que presenta sin embargo un valor $q=51,4\%$, es decir, que más de la mitad de la información del segundo de los caracteres citados está contenida en el primero.

En el cuadro 4 se calculan los valores medios de la interdependencia de cada carácter con los restantes, según valores d y sem . También, los valores medios de q informan del grado de inclusión o redundancia media de los restantes caracteres respecto al considerado y los de q' , del % de información del carácter que aparece incluido —en valor medio— dentro de los once restantes. El carácter más independiente y a la vez menos redundante parece ser «simetría de la corola» pese a tratarse de un carácter con bajo contenido de información. «Morfometría de pétalos», «morfometría de sépalos» y «color de sépalos» son los caracteres que incluyen mayor contenido de información redundante con otros caracteres (del 26 al 28%). «Color de pétalos» presenta el valor q' más alto, como consecuencia de la relativa redundancia que mantiene con otros caracteres como «nectarios medianos», «nerviación de pétalos», «posición de sépalos respecto al eje floral» y «alas transición limbo/uña». Esto puede estar provocado por la existencia de modelos de flor muy definidos como el ya citado de las especies de *Moricandia*, también presente en otros géneros como *Pseudoerucaria* y *Foleyola*, consistente en la asociación de pigmentación autociánica en los pétalos (color morado) con la presencia simultánea y constante de nerviación bronquidódroma, presencia de alas en la base del limbo de pétalo y cálices cerrados con cierto nivel de concrecencia entre los sépalos. Este modelo de flor parece muy eficaz en ambientes con cierto grado de xericidad y su valor adaptativo fue ya discutido en anteriores trabajos (CLEMENTE MUÑOZ & HERNÁNDEZ BERMEJO, 1978a, 1978b, 1980a, 1980b).

REFERENCIAS BIBLIOGRÁFICAS

- BAUM, B. R. (1974). Classification of the oat species (*Avena*, Poaceae) using various taximetric methods and an information —theoretic model. *Canad. J. Bot.* 52:2241-2262.
- CLEMENTE MUÑOZ, M. & J. E. HERNÁNDEZ-BERMEJO (1978a). El aparato nectarígeno de la tribu Brassiceae (Cruciferae). *Anales Inst. Bot. Cavanilles* 35:279-296.
- CLEMENTE MUÑOZ, M. & J. E. HERNÁNDEZ-BERMEJO (1978b). La corola en la tribu Brassiceae (Cruciferae). *Anales Inst. Bot. Cavanilles* 35:297-334.
- CLEMENTE MUÑOZ, M. & J. E. HERNÁNDEZ-BERMEJO (1980a). El cáliz de la tribu Brassiceae (Cruciferae). *Anales Inst. Bot. Cavanilles* 36:77-96.
- CLEMENTE MUÑOZ, M. & J. E. HERNÁNDEZ-BERMEJO (1980b). Clasificación jerárquica de las Brassiceae según caracteres de las piezas estériles de su flor. *Anales Jard. Bot. Madrid* 36:97-113.
- DUNCAN, T. & G. F. ESTABROOK (1976). An operational method for evaluating classifications. *Syst. Bot.* 1:373-382.
- ESTABROOK, G. F. (1967). An information theory model for character analysis. *Taxon* 16:86-97.
- GÓMEZ-CAMPO, C. (1977). Studies on Cruciferae: II. New names for *Rhynchosinapis* species under *Hutera*. *Anales Inst. Bot. Cavanilles* 34:147-149.

- GONZÁLEZ BERNÁLDEZ, F., P. MONTSERRAT RECORDER & A. GIL CRIADO (1968). Elaboración automática de datos fitosociológicos. *Bol. Soc. Esp. Hist. Nat.* 66:151-176.
- GODRON, M. (1968). Quelques applications de la notion de fréquence en écologie végétale. *Oecol. Pl.* 3:185-212.
- HAWKSWORTH, F. G., G. F. ESTABROOK & D. J. ROGERS (1968). Application of information theory model for character analysis in the genus *Arcentobium* (Viscaceae). *Taxon* 17:605-619.
- HERNÁNDEZ-BERMEJO, J. E. y M. CLEMENTE MUÑOZ (1981). Information analysis in numerical taxonomy: an application in the study of the Brassiceae (Cruciferae) tribe. *Abstracts XIII International Botanical Congress*. Sydney: 331.
- HERNÁNDEZ-BERMEJO, J. E. & H. SAINZ OLLERO (1978). *Ecología de los hayedos meridionales ibéricos: El macizo de Ayllón*. Serie Recursos Naturales. Ministerio de Agricultura, Madrid.
- HERNÁNDEZ-BERMEJO, J. E. & H. SAINZ OLLERO (1984). El análisis de semejanza aplicado al estudio de barreras y fronteras fitogeográficas: su aplicación a la corología y endemoflora ibéricas. *Anales Jard. Bot. Madrid* 41: 421-432.
- KULLBACK, S. (1959). *Information theory and statistics*. Wiley, New York.
- LANCE, G. N. & W. T. WILLIAMS (1966). Computer programmes for hierarchical polythetic classifications. *Computer J.* 9: 60-64.
- LEGENDRE, L. & P. LEGENDRE (1979). *Écologie numérique: 1. Le traitement multiple des données écologiques*. Masson, Paris.
- MARGALEF, R. (1957). La teoría de la información en Ecología. *Mem. Real Acad. Ci. Barcelona* 32:39-97.
- MARGALEF, R. (1958). Information theory in ecology. *Genet. Syst.* 3:36-71.
- SCHULZ, O. E. (1936). Cruciferae. In: A. Engler & al. (2nd ed.), *Die natürlichen Pflanzenfamilien*. 17b:227-658. Wilhelm Engelmann, Leipzig.
- SHANNON, C. E. (1948). A Mathematical theory of communications. *Bell System Tech. J.* 27: 379-423, 623-656.
- TUTIN, T. G. & al. (Eds.) (1964). *Flora Europea. Vol. 1. Lycopodiaceae to Platanaceae*. Cambridge.

Accepted para publicación: 6-II-84