Botanical Sciences 104 (3): 755-769. 2026
DOI: 10.17129/botsci.3787

Received: August 13, 2025, Accepted: March 11, 2026
On line first: June 8, 2026

Taxonomy and Floristics / Taxonomía y Florística

G-patterns: Biogeographic classification based on the distribution of assemblages

Miguel Murguía-Romero[1]*, José Luis Villaseñor[2]

[1] Unidad de Informática para la Biodiversidad, Instituto de Biología, Universidad Nacional Autónoma de México, Mexico City, Mexico.
[2] Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México, Mexico City, Mexico.
*Corresponding author: miguel.murguia@ib.unam.mx

Abstract

Background: The overlapping distribution of two or more taxa creates a biogeographic pattern. Biogeographic classifications can be directly based on this concept.

Hypotheses: Each coincident species distribution is postulated as a biogeographic pattern.

Studied species: Asteraceae species from Mexico.

Study site: Mexico

Methods: Biogeographic patterns (G-patterns) of the Asteraceae of Mexico were identified using a data matrix of 3,195 species from 32 states. A new methodology is proposed and applied in which a G-pattern is defined as the set of geographic units in which two or more species are distributed completely and exclusively: each species is distributed in each geographic unit of the G-pattern and in no other geographic unit of the analysis area.

Results: A total of 244 G-patterns were identified, with 94% comprising contiguous geographic units and only 6 % being disjointed. Just a third of the species are not involved in a G-pattern. The 32 states form a G-pattern (holoregion) defined by six species, while 22 states form a G-pattern (monoregion) independently. A linear relationship was observed between the logarithm of the number of species and the number of G-patterns per state ($R^2 = 0.68$).

Conclusions: G-patterns allow biogeographic areas to be proposed based on logical rather than arithmetic or statistical criteria; they help identify or disprove biogeographic classes.

Keywords: Asteraceae, descriptive analysis, holoregions, logical methods, Mexico, monoregions

Resumen:

Antecedentes: La distribución coincidente de dos o más taxones determina un patrón biogeográfico. La construcción de clasificaciones biogeográficas puede fundamentarse directamente en este concepto.

Hipótesis: Cada una de las distribuciones coincidentes de dos o más especies se postula como un patrón biogeográfico.

Especies de estudio: Asteraceae de México.

Sitio: México

Métodos: Se identificaron los patrones biogeográficos (patrones G) de las Asteraceae en México a partir de una matriz de datos con 3,195 especies distribuidas en 32 entidades federativas. Se propone y aplica una nueva metodología en la que se define como patrón G el conjunto de unidades geográficas en las que se distribuyen dos o más especies de forma completa y exclusiva: cada especie se distribuye en cada unidad geográfica del patrón G y en ninguna otra unidad geográfica del área de análisis.

Resultados: Se obtuvieron 244 patrones G, 94 % de los cuales están constituidos por unidades geográficas contiguas y solo 6 % por unidades geográficas disjuntas. Solo la tercera parte de las especies participa en un patrón G. Los 32 estados conforman un patrón G (holorregión) definido por seis especies, mientras que 22 estados conforman por sí mismos un patrón G (monorregión). Se encontró una relación lineal entre el logaritmo del número de especies y el número de patrones G por estado ($R^2 = 0.68$).

Conclusiones: Los patrones G permiten proponer áreas biogeográficas directamente, basándose en criterios lógicos y no en criterios aritméticos o estadísticos; además, permiten la identificación o refutación de clases biogeográficas.

Palabras clave: Asteraceae, análisis descriptivo, métodos lógicos, México, holorregiones, monorregiones.

Biogeographic classifications originate from biogeographic analyses. Their meanings vary depending on the biogeographic school or approach used and the method employed for construction. For instance, in ecological biogeography, classification is purely descriptive, whereas in historical biogeography, it also serves as a well-supported evolutionary hypothesis. Methods in historical biogeography often involve a more logical (deductive) component when building their taxonomic and standardized cladograms, compared to ecological biogeography methods for creating dendrograms, which incorporate elements based on arithmetic averages (statistical inferences).

Level of abstraction of biogeographic classifications. Biogeographic classification explains how species are distributed across different geographic areas. From an ecological biogeography perspective, a common approach to creating biogeographic classifications is to compare the similarity between each pair of Operational Geographical Units (OGUs) based on the species they share. To determine these similarities, an incidence matrix (showing the presence and absence of taxa) is typically created (Murguía & Villaseñor 2000), which is then converted into a similarity matrix using an index such as the Jaccard or Simpson coefficients. Afterwards, a clustering algorithm is applied to the similarity matrix to produce a dendrogram, in which a specified similarity threshold identifies biogeographic regions.

From the perspective of historical biogeography, similarity is used as evidence to establish historical relationships (homologies) between species and their distribution. Among the most widely used methods are the parsimony analysis of endemisms (PAE) (Rosen 1988) or the Brooks parsimony analysis (BPA, Brooks 1990, Brooks & Van Veller 2003). The biogeographic classification obtained represents a hypothesis of the historical relationships among the three fundamental elements that explain a taxon's evolutionary history: time, space, and form. In other words, it helps to understand the relationship between species and their geographic distribution over time.

The meaning of biogeographic classifications varies by biogeographic school. For example, in ecological biogeography, the classification (the area dendrogram) is primarily descriptive, illustrating the different regions identified. Conversely, in historical biogeography, the area cladograms function as hypotheses that explain distribution patterns across time and space.

The structure of knowledge about biogeographic processes is examined at various levels of abstraction, from simple observation to description and ultimately to hypothesis generation, with each level building on the previous one. Based on this, dendrograms represent a lower level of abstraction than area cladograms, since the former are at the descriptive stage, while the latter are at the hypothesis-generating stage.

Logical bases versus inferential bases in biogeographic classification. Since it is always possible to generate a dendrogram of areas regardless of the data structure, another difference between the two schools (ecological or historical biogeography) in their approach to biogeographic classification concerns the types of assumptions they implicitly make in their methods. For instance, in ecological biogeography, producing a dendrogram is unavoidable because, regardless of the data structure, it is always feasible to create one for areas. Only in rare circumstances, when the similarities are highly homogeneous, will it be impossible to produce such a dendrogram. The creation of the dendrogram is a deliberate result and does not depend on whether the data are consistent. An analogy can be drawn from calculating the average of, say, five real numbers: regardless of how different or similar these numbers are, an average can be computed continuously. The dendrogram is then analyzed to determine the cut-off threshold (level) for defining the final classes (regions or biogeographic units).

In contrast, historical biogeography constructs the cladogram of areas using different strategies. For example, the PAE (Rosen 1988, Morrone & Crisci 1995) results from selecting relationships between areas based on logical criteria, specifically parsimony, concerning a priori hypothesis defined by biogeographic processes (areas of species distribution). The construction of the area cladogram is guided by the criteria of maximizing the number of synapomorphies (shared unique areas) and minimizing the number of homoplasies (areas with incongruent or redundant distributions). They constitute a logical-deductive procedure rather than a statistical-inferential one, as in the construction of dendrograms. The area cladogram is constructed according to logical rules, while the area dendrogram is formed from averages. Nonetheless, both methods tend to produce similar results—identifying G-patterns.

Morrone & Crisci (1995) describe three steps for conducting a PAE analysis in historical biogeography: 1) Recognizing spatial homology, 2) Identifying areas of endemism, and 3) Developing hypotheses about relationships. They propose their method as an alternative to panbiogeographic techniques, acting as a first step in identifying common traits that reflect ancestral biotas and spatial homologies.

The panbiogeographic method (Croizat 1982), considered a strategy in historical biogeography, does not include preformed hypotheses about the processes that create its structures, such as tracks, nodes, or baselines (Page 1987). However, constructing generalized tracks by considering the overlap of multiple individual tracks can lead to the formation of forced biogeographic patterns; that is, combining a sufficiently large number of individual tracks can identify a generalized track. Simberloff *et al*. (1981) suggested that the concept of generalized tracks has not been tested and might be a statistical artifact. This issue was addressed by Page (1987), who proposed a method to test similarities between tracks.

One of the strengths of the panbiogeographic method, as already noted, is that it does not rely on a priori hypotheses about the processes that form its constructs. Instead, inferences about the processes that create the observed patterns are developed afterward. Individual tracks serve as the basis for pattern identification; they are constructed using a minimum spanning tree, a rule applied when no other information is available (Page 1987). This encourages us to explore alternative methods to the traditional approach to detecting biogeographical patterns; for example, is there another approach that aligns with the philosophy behind the concurrent distribution of multiple species?

An alternative approach is to identify biogeographic patterns before constructing individual tracks by applying the minimum spanning tree transformation. The coincident distribution of taxa observed in localities or area units where their presence is recorded, rather than from subsequent constructs such as individual tracks, offers an alternative method for detecting biogeographic patterns. Detecting these patterns, as part of the initial step proposed by Morrone & Crisci (1995), with minimal transformations, reduces the likelihood that they are merely statistical artifacts, as defined by Simberloff *et al*. (1981).

This paper introduces a method for creating biogeographic classifications based on the overlapping distributions of species across a set of OGUs. The concurrent presence of two or more taxa is a specific event that forms a biogeographic pattern; therefore, in a biogeographic analysis involving multiple species, there are likely to be several such overlapping distributions. Each of these simultaneous distributions is also regarded as a biogeographic pattern, regardless of whether their causes or explanations are aligned or consistent. The proposed method includes logical elements for identifying biogeographic patterns based on the overlapping distributions of species within operational geographic units (OGUs). The approach does not rely on prior assumptions and aims to provide descriptive insights rather than definitive hypotheses, although some hypotheses may emerge from the analysis.

Materials and methods

Based on a matrix of species presence and absence in a set of OGUs (where rows are species and columns are OGUs), a coincident distribution—called the G-pattern—is the group of OGUs where a specific set of species exactly occurs. This set of species is known as a *spatial species assemblage*. It helps identify a biogeographic pattern, but it is not an initial hypothesis about the processes or causes that create it.

The Q(n) histogram. The graph of taxon frequencies by number of OGUs (histogram Q(n)) is explained below to help clarify the method and introduce a nomenclature that facilitates interpretation of the results of this newly proposed construct. To illustrate the process, a data matrix of 3,195 species from the Asteraceae family and their distribution (presence) across the 32 federal entities (states) of Mexico is used. The species frequency was determined by the number of states in Mexico in which they were present, referred to below as OGUs.

G-patterns. As previously mentioned, the overlapping distributions of multiple taxa create a biogeographic pattern, regardless of whether they correlate. A set of OGUs showing a coincident distribution is referred to as a G-pattern.

The taxa of species assemblages with coincident distribution belong to only one of the classes Q(1), Q(2), ..., Q(m), where m is the total number of OGUs in the analysis area. Each species assemblage defines a set of OGUs,

and each taxon within the assemblage is present in all of its OGUs but not in others. Each of these groups of OGUs is considered a biogeographic pattern, and to distinguish them from other biogeographic patterns characterized by different methods, they are referred to as G-patterns. Regarding G-patterns, the following can be emphasized:

1) A biogeographic pattern G_{ij} includes a specific number of OGUs, indicated by *i*; the second subscript j is added to distinguish each pattern from other possible patterns within the same class Q(i). Therefore, there can be up to m patterns of type G_{1j}, each containing only one OGU: $G_{1,1}$, $G_{1,2}$, $G_{1,3}$, ... G_{1m}.

2) To establish the presence of a G_{1j} pattern (i.e., a pattern of class Q(1) in OGU_j), it is necessary to have more than one taxon that is exclusively found in that OGU.

3) The presence of the G_{1j} pattern for each OGU shows evidence that, within that OGU, there is a biogeographic class not shared by other OGUs.

4) There is only one G-pattern (or possibly none) that contains *m* OGUs: G_{m1}. This pattern includes all the OGUs in the analysis region. The presence of the G_{m1} pattern indicates the existence of a biogeographic unit encompassing the entire study area. Conversely, its absence suggests some biogeographic heterogeneity within the whole analysis territory.

5) Generally, each G_{ij} pattern indicates the presence of a biogeographic class to which the OGUs it comprises belong, and which is not shared with other OGUs in the study region. Table 1 summarizes the rules for interpreting the G-patterns.

The set of OGUs where taxa not belonging to any G-pattern are distributed is called 'No G-pattern'. The importance of this group lies in its absence, which indicates that all taxa belong to some G-pattern. At the same time, the number of OGUs it contains serves as an inverse indicator of how consistently taxa are distributed across G-patterns.

Null model for G-patterns. The ways to select *r* elements from *n* are calculated as:

$$C_r^n = \frac{n!}{r!\,(n-r)!}$$

The possible combinations of 1, 2, 3, and so on species in 32 OGUs are shown in Table 2. For example, translating the values in Table 2, for a set of 32 OGUs, there are 201,376 ways to form groups of 5 OGUs [Q(5)] with the same matching species. Suppose only two species are recorded in Q(5). In that case, the probability that these two species occupy the same five OGUs is 1/201,376 = 0.000005, an extremely rare event that biogeographically indicates an equally rare pattern. In the example matrix, 81 species (Q(5)) comprise only 122 distinct patterns.

G-patterns of the Mexican Asteraceae species. As an empirical exercise, G-patterns were identified for the Asteraceae species in Mexico. For this purpose, a presence/absence matrix of 3,195 Asteraceae species across the 32 states of Mexico was used. To verify the completeness of the data matrix, the nonparametric $Chao_2$ richness estimator (Colwell & Coddington 1994) was applied, resulting in an estimate of 3,629 species. This figure indicates a completeness level of 88 %, a high percentage given the taxonomic scope and geographic coverage of the matrix.

For each identified G-pattern, the number of states comprising each pattern and the recorded species in each state were counted (Table 3). They were also analyzed to determine whether they formed disjunct or non-disjunct patterns (Table S1). In the first case, it was confirmed that all the federative entities that make up a G-pattern are connected by their borders. A web interface was provided for quick visualization of the detected G-patterns (www.sipmx.net).

Biogeographical classifications. After obtaining the list of identified G-patterns, various combinations were created to find sets that included all 32 federal entities without redundancy. This ensures that each class within the defined biogeographic classification is exclusive from the others. Using a heuristic criterion, the combination that

Table 1. Rules for the interpretation of G-patterns.

1. Definition of G-pattern A G-pattern is a set of OGUs in which more than one taxon is distributed coincidentally; each taxon is distributed in each of the OGUs in the G-pattern and is not distributed in any other OGU in the analysis region.
2. Descriptive significance of G-patterns A G-pattern is considered a form of biogeographic description, but not as a priori hypothesis of the processes or causes of such a pattern.
3. Each G-pattern defines a biogeographic class Each G-pattern indicates the presence of a biogeographic class in the OGUs that comprise it and that is not shared with other OGUs not included in such G-pattern.
4. G-patterns of a single OGU (monoregions) The presence of the G_{1j} pattern corresponding to each OGU indicates evidence that in such a OGU there is a biogeographic class not shared by other OGUs.
5. Pattern-G comprising all the OGUs in the study region (holoregion) There is only one possible pattern (or it may not exist) that contains *m* OGUs: G_{m1}. This is the pattern made up of all the OGUs in the analysis region.
6. Significance of the G_{m1} pattern The presence of the G_{m1} pattern indicates evidence of a geographic unity of the entire study region.
7. Significance of the absence of the G_{m1} pattern The absence of the pattern formed by all the OGUs (i.e., when the G_{m1} pattern does not exist) may suggest the existence of biogeographic processes in a wider region than the analysis area.
8. Independence of each G-pattern Each G_{ij} pattern is biogeographic evidence regardless of whether it is congruent or not with other G-patterns.
9. Hierarchy of G-patterns G-patterns can be hierarchized to build a tree based on containment relationships.
10. OGUs that do not belong to any G-pattern OGUs that do not belong to any biogeographic pattern are indicators that there is no evidence that they constitute by themselves (G_{1j}) or that they are part of a biogeographic pattern (and of course, also of possible poor sampling).
11. No-G-pattern and species that do not belong to any G-pattern The UGOs of species that do not belong to any G-pattern are found in the No-G-pattern; their absence indicates strong congruence in the geographic distribution of species within the UGOs. The number of species and the number of UGOs that comprise it are also (inversely) indicative of congruence.
12. Biogeographic cohesion and complexity The proportion of taxa participating in G-patterns is a measure of the biogeographic cohesion of the group in the study area. In contrast, the number of G-patterns identified, as well as the proportion of disjunct *vs* non-disjunct G-patterns, are measures of the degree of complexity of the biogeographic patterns.

Table 2. Possible combinations of taking r elements in 32 cases (for example, the 32 states of Mexico to build up G-patterns).

r	Combinations of r in 32
1	32
2	496
3	4,960
4	35,960
5	201,376
6	906,192
7	3,365,856
8	10,518,300
9	28,048,800
10	64,512,240
11	129,024,480
12	225,792,840
13	347,373,600
14	471,435,600
15	565,722,720
16	601,080,390
17	565,722,720
18	471,435,600
19	347,373,600
20	225,792,840
21	129,024,480
22	64,512,240
23	28,048,800
24	10,518,300
25	3,365,856
26	906,192
27	201,376
28	35,960
29	4,960
30	496
31	32
32	1

maximized the sum of incidences in the presence/absence matrix for each pattern was selected to visually depict a possible classification among multiple options. A comparison is made between the results of the G-patterns and a recent classification of Mexican states based on Magnoliophyta species using an ecological biogeography (phenetic) approach.

Comparison between approaches. To perform an initial analysis comparing G-patterns with ecological biogeography methods, the number of species and the number of G-patterns shared between a specific state and others were measured. The latter methods have shown consistency with other biogeographic classification approaches; they can, on average, reveal the biogeographic structure that links species to their distributions. Therefore, a positive correlation is expected: the greater the number of shared species, the more shared G-patterns should be present.

Table 3. Number of G-patterns by number of states and their characteristics, analyzing the Asteraceae species in the federal entities (states) of Mexico.

Number of states	Number of G-patterns	Possible number of patterns	Species in G-patterns	Total species	Species in G-patterns / Total species
1	22	32	710	714	99.4 %
2	39	496	560	588	95.2 %
3	47	4,960	357	400	89.3 %
4	32	35,960	182	256	71.1 %
5	32	201,376	122	201	60.7 %
6	20	906,192	52	124	41.9 %
7	11	3,365,856	30	111	27.0 %
8	8	10,518,300	24	90	26.7 %
9	3	28,048,800	6	67	9.0 %
10	5	64,512,240	10	65	15.4 %
11	2	129,024,480	4	54	7.4 %
12	2	225,792,840	4	50	8.0 %
13	1	347,373,600	2	44	4.5 %
14	0	471,435,600		32	0 %
15	0	565,722,720		38	0 %
16	0	601,080,390		29	0 %
17	0	565,722,720		31	0 %
18	0	471,435,600		30	0 %
19	0	347,373,600		28	0 %
20	0	225,792,840		24	0 %
21	0	129,024,480		30	0 %
22	2	64,512,240	5	35	14.3 %
23	2	28,048,800	6	38	15.8 %
24	3	10,518,300	7	22	31.8 %
25	2	3,365,856	4	20	20.0 %
26	3	906,192	8	21	38.1 %
27	4	201,376	9	18	50.0 %
28	1	35,960	2	8	25.0 %
29	0	4,960		11	0 %
30	1	496	2	6	33.3 %
31	1	32	2	3	66.7 %
32	1	1	6	6	100 %
TOTAL	244		2,114	3,194	66.2 %

Results

The Q(n) histogram. (Figure 1) shows the distribution histogram of the analyzed matrix by class Q(n). The graph displays the distribution of species by OGUs, revealing a typical inverted J shape, indicating that many species have limited distributions. In contrast, a few have a broad distribution. The first bar represents the number of species found in only one OGU, specifically 714. The second bar shows the number of species found in only two OGUs (588 species), and so on. The second-to-last bar indicates the number of species occurring in 31 OGUs (3 species), and the

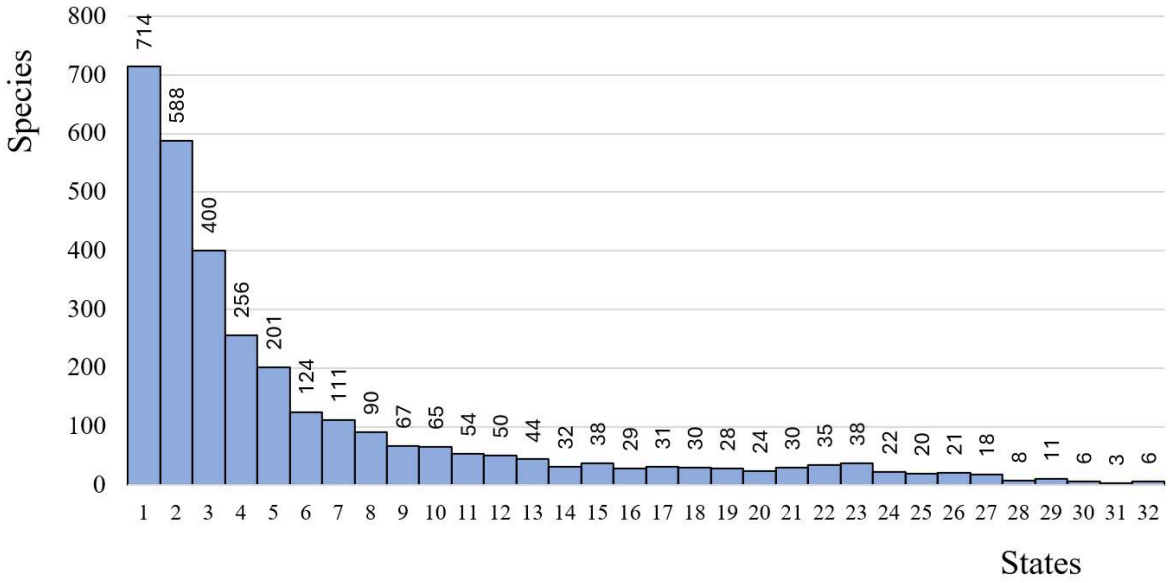

Figure 1. Species frequency by number of OGUs. Each class (horizontal axis) shows the number of OGUs (states) where each species occurs. For example, the first bar indicates the number of species found in a single state (class Q(1)).

last bar represents the species present in all 32 OGUs, encompassing the entire study region (6 species). Each group of species belongs to one of the classes Q(1), Q(2), …, Q(32), depending on whether species in that group are found in only one OGU, two OGUs, or all 32 OGUs.

G-Patterns. A total of 244 G-patterns were identified (Table 3), including 2,114 species, which represent 66 % of the 3,195 Asteraceae species in Mexico. The remaining species do not exhibit any discernible pattern; their presence in specific states is unique, occurring in a single pattern that no other species shares. In Mexico's Asteraceae, the G_{m1} pattern was supported with six species occurring across all federal entities. These species are *Artemisia ludoviciana* Nutt., *Bidens pilosa* L., *Erigeron bonariensis* L., *Erigeron canadensis* L., *Sonchus oleraceus* L., and *Symphyotrichum expansum* (Poepp. ex Spreng.) G.L.Nesom.

All states are part of some G-pattern. Durango appears in the highest number of G-patterns (82), while Yucatán is included in the fewest, with only 13 patterns (Table 4). In contrast, not all entities form a G-pattern on their own (G_{1j} patterns); the 10 entities that do not match a G(1) pattern are ROO, YUC, AGU, TLA, GUA, TAB, CAM, ZAC, NAY, and CMX (state abbreviations follow the standard ISO 3166-2), which may indicate there is no evidence to define a distinct biogeographic unit within their territories. Only 14 patterns are disjoint, while the remaining 230 have all OGUs as contiguous.

For each of the 244 G-patterns identified in the study, the number of states and species that comprise them is provided in Table S1. The patterns with the most species belong to classes G_{1j} and G_{2j}, with 714 and 588 species, respectively. Collectively, these species account for 41 % of the total Asteraceae species in Mexico, limited to 61 G-patterns. For illustration, Figure 2A displays eight Asteraceae species patterns formed by eight federal entities; all of these patterns demonstrate a continuous distribution, and Figure 2B depicts G-pattern 221 in the SIP-MX system (www.sipmx.net).

The "non-pattern G" consists of the geographic units where taxa that do not belong to any G-pattern are found. In Mexico's Asteraceae, the "non-pattern G" includes 32 states and accounts for 1,080 species out of 3,195, meaning that only about a third (33.8 %) do not belong to any G-pattern. Conversely, when considering the total number of incidences in the presence/absence matrix (states by species), the percentage of incidences occurring in some G-pattern is 34.0 % (6,562 out of 19,282 incidences).

Table 4. Total number of Asteraceae G-patterns in Mexico by state. The numbers in the columns indicate the number of states in a pattern. For example, the state of Coahuila (COA) participates in 69 patterns, five of which are made up of two states (G_{2j} patterns). All states belong to Pattern G_{32}.

		Number of G-patterns by the number of states that encompass them																					
Mexican state	**Number of G-patterns**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**30**	**31**	**32**
AGU	34		1	1	2	2	1	2	2	1	2	1	1		1	2	3	1	3	4	1	1	1	1
BCN	18	1	2	2	2	4	2	2												1			1	1
BCS	15	1	1	1	1	3	1	2											1	2	1			1
CAM	16		1	2	1	2	2				1				1			1	1	1		1	1	1
CHP	43	1	3	7	5	4	4	2	1		1				1		2	1	3	4	1	1	1	1
CHH	75	1	3	8	9	11	6	5	4	3	3	1	2		1	2	3	2	3	4	1	1	1	1
COA	69	1	5	7	9	10	6	4	3	3	3	1	1		1	2	2	1	2	4	1	1	1	1
COL	23	1	1	2	2	1			1				1		1	1		2	3	3	1	1	1	1
CMX	19			1								1		1	1	1	3	1	2	4	1	1	1	1
DUR	82	1	4	8	9	12	8	6	5	3	3	1	2		2	2	3	2	3	4	1	1	1	1
GUA	38		2	2	2	4	1	1	2		3		1	1	1	2	3	2	3	4	1	1	1	1
GRO	62	1	5	6	9	8	5	3	1		1	1	1	1	2	2	3	2	3	4	1	1	1	1
HID	48	1	2	4	2	4	5	2	3	1	1	1	1	1	2	2	3	2	3	4	1	1	1	1
JAL	72	1	5	10	8	9	5	4	3		3	1	2	1	2	2	3	2	3	4	1	1	1	1
MEX	47	1	2	4	4	5	4	2	1		1	1	1	1	2	2	3	2	3	4	1	1	1	1
MIC	57	1	3	6	7	7	4	3	2		1	1	1	1	2	2	3	2	3	4	1	1	1	1
MOR	34	1			3	4	3	1				1		1	2	2	3	2	3	4	1	1	1	1
NAY	53		3	8	3	5	4	4	3	1	1	1	1		2	2	3	2	2	4	1	1	1	1
NLE	60	1	2	6	5	7	7	3	4	3	3	1	1		1	2	3	1	2	4	1	1	1	1
OAX	59	1	4	7	8	5	6	3	1		2	1	1		2	2	3	2	3	4	1	1	1	1
PUE	53	1	3	5	8	4	5	3	1		1	1		1	2	2	3	2	3	4	1	1	1	1
QUE	47	1	3	4	1	5	3	2	3	1	1	1	1	1	2	2	3	2	3	4	1	1	1	1
ROO	14			3	1	2	1				1				1			1	1			1	1	1
SLP	67	1	5	6	3	7	8	3	5	3	4		1	1	2	2	3	2	3	4	1	1	1	1
SIN	54	1	3	6	4	6	5	5	3		1	1	1		1	1	3	2	3	4	1	1	1	1
SON	56	1	3	6	4	9	3	3	3	2	3	1	1			2	2	2	3	4	1	1	1	1
TAB	17		1	1	1	2	2	1			1				1			1	1	1	1	1	1	1
TAM	60	1	3	4	3	7	6	3	5	3	3	1	1		2	2	3	2	3	4	1	1	1	1
TLA	18											1		1	1	1	3	1	2	4	1	1	1	1
VER	60	1	4	7	6	5	7	3	3		2	1		1	2	2	3	2	3	4	1	1	1	1
YUC	13		1	2	1	1	1				1				1			1	1			1	1	1
ZAC	62		3	5	5	5	5	5	5	3	3	1	2		2	2	3	2	3	4	1	1	1	1

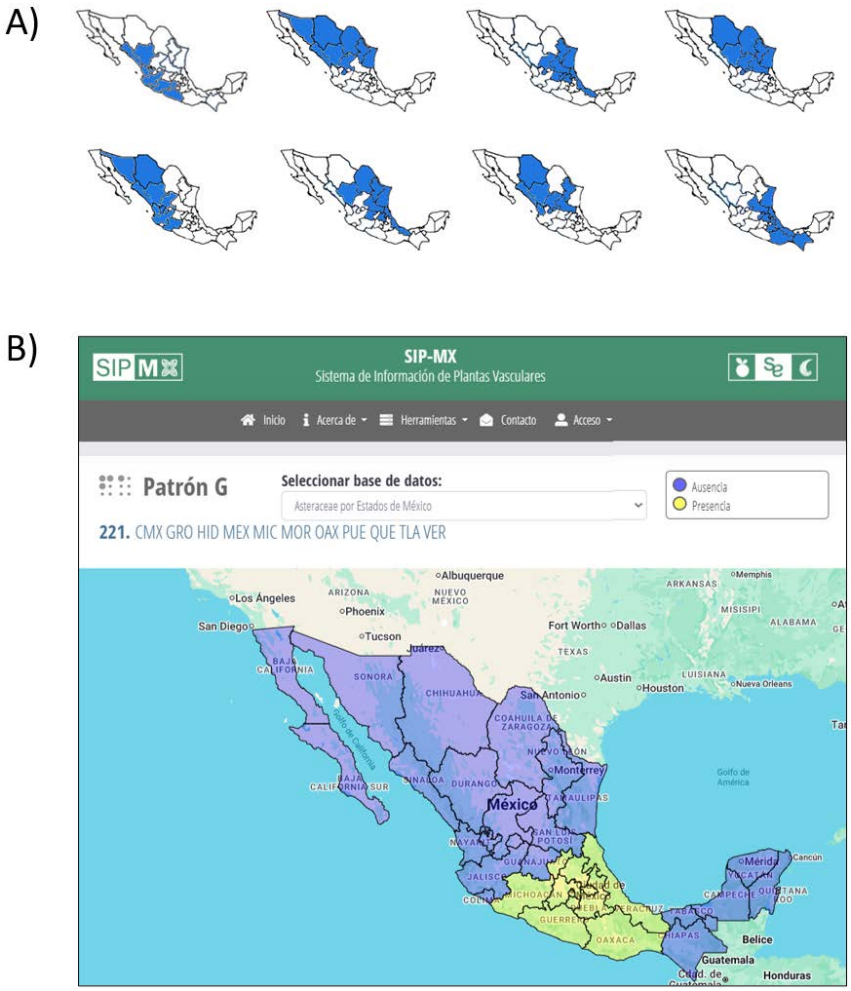

Figure 2. Samples of G-Patterns of the Asteraceae of Mexico. A) Geographic distribution of the states that form a pattern consisting of 8 OGUs (Patterns-G_{8j}). B) G-Pattern 221 (Pattern-$G_{11,2}$) as shown in the SIPMX (www.sipmx.net) web system.

Another useful metric in biogeographic analysis is the average number of species per disjunct versus non-disjunct G-patterns. In the case of the Compositae of Mexico, species that define the disjunct G-patterns total 35, while those in the non-disjunct patterns total 2,079. Therefore, the average number of species per G-pattern is 2.5, calculated as 35/14, and for non-disjunct patterns, it is 9.04, calculated as 2,079/230. This indicates that contiguous states form G-patterns more often than non-contiguous ones, and that non-disjunct patterns are supported by more species than disjunct ones- on average, about 3.6 times more (9.04/2.05). From a biogeographic perspective, these metrics can support or challenge arguments for vicariance versus dispersal phenomena.

There is a linear relationship between the logarithm of the number of species and the number of G-patterns by states (Figure 4B; $R^2 = 0.68$). The graph in Figure 4B shows that states BCN and BCS are the furthest from the trend line. This is because both states form G_2 patterns and have the highest average number of species per G-pattern, and both states form G_1 patterns. In other words, both entities exhibit high biogeographic cohesion, meaning they are associated with a low number of states and share a large number of species exclusively among themselves. Therefore, a helpful tool is the graph plotting the number of G-patterns per geographic unit against the total number of species they contain. If this relationship is logarithmic, it allows us to identify geographic units that deviate from this pattern and draw conclusions about their specific biogeographic aspects, as exemplified here with the states of BCN and BCS.

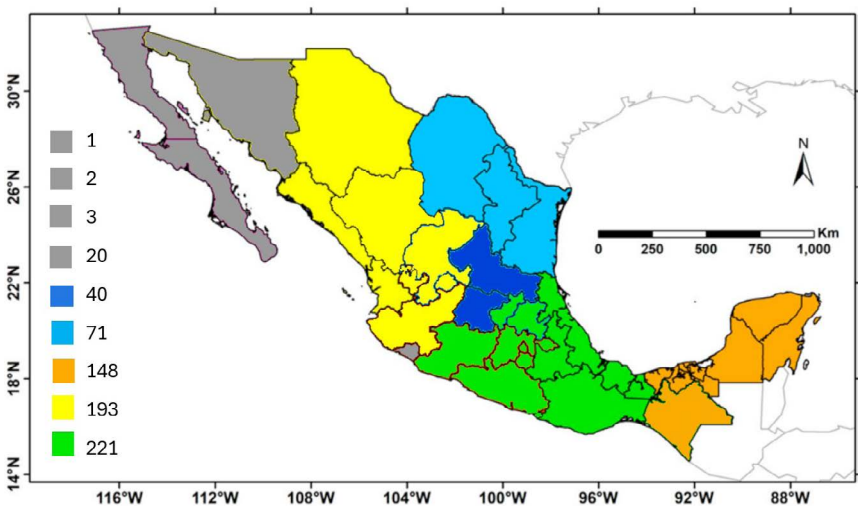

Figure 3. Biogeographic classification based on G-patterns shown in the SIP MX internet portal. The map displays the distribution of patterns from classification number 28 in Table S2, which includes 249 species. Monoregions 1, 2, 3, and 20 are shaded in the same color to improve visualization.

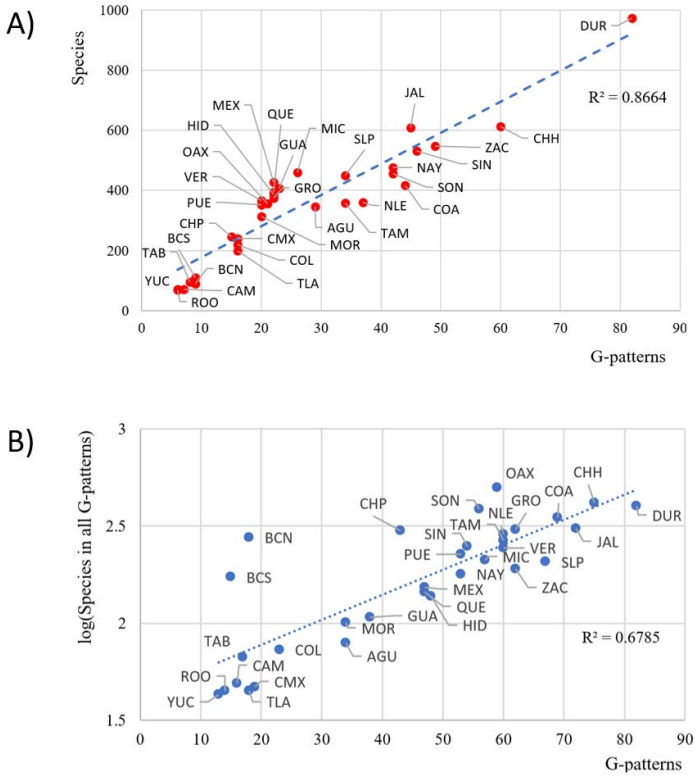

Figure 4. Linear relationships between G-patterns and the number of species. A) Number of species recorded in the state of Durango (DUR) *versus* the number of G-patterns shared with Durango. B) log (Number of species) *versus* the number of G-patterns.

Biogeographic regionalization. Each G-pattern creates its own biogeographic pattern; however, constructing a hierarchical regionalization to develop a biogeographic classification that covers an entire territory with non-overlapping classes is always beneficial. Multiple biogeographic regionalizations can be derived from G-patterns. In the study of Asteraceae from Mexico, 244 patterns were identified (Table S1). From these, a subset can be selected that includes

all 32 federal entities without overlapping. By exploring all possible combinations of the 244 G-patterns, potential biogeographical regionalizations were generated (Table S2), ensuring that each includes all federal entities and that no G-pattern shares any federal entity with another G-pattern. For example, classification 5 consists of 8 G-patterns, in which 192 species are exclusively distributed in one of these eight patterns and not found in any of the other seven patterns.

A metric for prioritizing geographic classifications based on G-patterns is the total number of incidences that include the G-patterns comprising the classification. This value is part of the OGU incidence matrix by taxa, so it can be interpreted as the extent to which each classification contributes to explaining the biogeographic configuration. The higher this value, the greater the explanatory power of the classification. Therefore, in the exercise on Asteraceae in the states of Mexico (Table S2), the classification with the highest number of occurrences (28) includes patterns G 1, 2, 3, 20, 46, 71, 148, 193, 221 (Figure 3), which are: 1: BCN; 2: BCS; 3: COL; 20: SON; 46: GUA, SLP; 71: COA, NLE, TAM; 148: CAM, CHP, ROO, TAB, YUC; 193: AGU, CHH, DUR, JAL, NAY, SIN, ZAC; and 221: CMX, GRO, HID, MEX, MIC, MOR, OAX, PUE, QUE, TLA, VER (see Table S2). Four patterns (1, 2, 3, 20) are monoregions.

To heuristically evaluate the G-patterns obtained here, a comparison can be made with the classification proposed by Villaseñor & Ortiz (2025), which groups the federative entities into six floristic categories based on the presence of native vascular plant species (Table 5). Of the six classes proposed by these authors, only two are represented in our G-patterns (I. Baja California Peninsula and III. Northwestern)—their region II. The Yucatán Peninsula includes the states CAM, ROO, TAB, and YUC; this classification differs from our G-pattern in that whenever TAB is included, CHP is included as well. Therefore, based on the G-patterns obtained, it can be proposed that this region consists of either CAM, CHP, ROO, TAB, YUC, or a combination of ROO, TAB, and YUC, but not CAM, ROO, TAB, and YUC together.

Regarding Region IV. Central-Northeastern states (AGU, COA, GUA, HID, QUE, NLE, SLP, TAM, ZAC) exhibit G-patterns that more closely connect AGU and ZAC to DUR. Among the 35 patterns involving AGU, DUR appears in 30, and for ZAC, it appears in 50 of 63. In fact, these two states share the most G-patterns with DUR.

Region V, or West-central V (COL, GRO, JAL, MEX, CMX, MIC, MOR, TLA), differs from the G-patterns because it does not include NAY, which shares the most G-patterns with JAL (47 out of 54). Similarly, the GRO state is linked to more G-patterns, including those of OAX and PUE, which are not part of this V region.

Regarding Region VI, Southeastern (CHP, OAX, PUE, VER), as previously mentioned, the GRO state is linked with more G-patterns, including OAX and PUE. Additionally, VER is more strongly associated with HID and PUE than with OGUs from Region IV.

The observed differences are likely due to the comparison method used. However, the exact definition of OGUs (states of Mexico) varies between the two data sets: the G-patterns based on Mexican Asteraceae species and the regionalization by Villaseñor & Ortíz (2025), which was performed using all vascular plant species in Mexico. While Asteraceae can serve as a good proxy for flowering plants (*e.g.,* Villaseñor *et al*. 2007, del-Val *et al*. 2015), there may be significant differences in the distribution patterns between the states of both plant groups.

Numerical comparison. The numerical comparison was performed using Durango as a case study; this state participates in the most G-patterns. There is a strong linear correlation ($R^2 = 0.87$) between the number of G-patterns and the number of species shared with other federal entities (Figure 4a). Notably, the five points closest to DUR in the graph in Figure 4a represent entities involved in the G193 pattern (Figure 3), which includes seven entities, among them DUR.

Discussion

Several G-pattern metrics for the Asteraceae of Mexico by state show biogeographic consistency, supporting this method as a potential way to define regionalizations. These metrics include a high proportion of non-disjunct

Table 5. Comparison of the classification of the states of Mexico using UPGMA (Villaseñor & Ortiz, 2025) with the G-patterns (this study).

Floristic group	States	G-Pattern
I. Baja California peninsula	BCN, BCS	24
II. Yucatán peninsula	CAM, ROO, TAB, YUC	No detected
III. Northwestern	CHH, DUR, NAY, SIN, SON	161
IV. Central-Northeastern	AGU, COA, GUA, HID, NLE, QUE, SLP, TAM, ZAC	No detected
V. West-central	COL, GRO, JAL, MEX, CMX, MIC, MOR, TLA	No detected
VI. Southeastern	CHP, OAX, PUE, VER	No detected

patterns compared to disjunct ones. Additionally, there is a correlation between the number of G-patterns per state and the number of species involved in each. Another metric is the high percentage of species participating in G-patterns, along with the large proportion of occurrences in the presence/absence matrix that are included in G-patterns. Therefore, using G-patterns allows proposing biogeographic regions composed of a single OGU (monoregions) as well as areas encompassing all the OGUs under study (holoregions). These features are either absent or difficult to identify using other biogeographic classification methods, such as the ecological biogeography approaches.

The entity with the most G-patterns is DUR, appearing in 82 patterns. YUC and ROO are involved in the fewest patterns (13 and 14, respectively) among the entities, and they appear in most of the patterns in which they are recorded.

Considering G-patterns with fewer than 20 entities, CAM, ROO, and YUC participate in disjunct patterns, except for those formed by CAM-ROO-YUC and CAM-YUC, which support the idea of recognizing the Yucatan Peninsula as a distinct biogeographic region (*e.g.,* Villaseñor & Ortiz 2025). It is also noted that none of the three entities form a G-pattern, meaning that none of them contains more than one exclusive species.

The exercise of numerical analysis, which compares the number of shared G-patterns with the number of shared species, yielded a high correlation, supporting the congruence between the two methods. Most importantly, since these methods have different foundations, it demonstrates that biogeographic patterns can be revealed through G-patterns.

Here, a comparison is made with a phenetic (ecological biogeography) classification, mainly because it uses the same geographical units (federal entities of Mexico). However, possible future comparisons with biogeographical classifications built with other types of methods, such as the PAE, would be in the same terms, that is, verify if there are G-patterns that recover their results, and always taking into account that the classification arising from the PAE can be interpreted as a synthesis of those detected as G-patterns.

No biogeographic classification can precisely represent the distribution areas of all species it is based on. In any case, the areas included in such a classification may correspond to patterns near the distribution areas of each species. They will only reflect the distribution of some of them. This highlights the complexity of species distribution within a group; therefore, it is essential to recognize that biogeographic classifications are merely heuristic models that simplify complexity for human understanding and make it easier to study species distribution as a whole. Identifying cases where a given classification does not accurately depict species distribution is highly valuable both scientifically and heuristically, as it reveals a model's strengths and weaknesses, enabling critique and advancing scientific knowledge.

G-patterns provide a straightforward way to examine how species are distributed across geographic areas; they also help us understand the complexity of these patterns and offer an objective method to identify fundamental trends, which can be used to assess the successes and limitations of biogeographic classifications.

It is crucial to understand that any biogeographic classification will be accurate only for a subset of the species considered; the remaining species will be classified only partially. This principle applies to any method of biogeographic regionalization, whether for ecological biogeography (as indicated by the similarity percentages that never reach 100 %) or for historical biogeography (as shown by its consistency index, which also does not reach 100 %, reflecting the number of homoplasies in the cladogram). The G-patterns obtained in this study highlight inconsistencies in both types of analysis.

G-patterns enable the direct proposal of biogeographic areas based on logical criteria rather than arithmetic ones (*e.g.,* using set operations rather than averages or statistical measures). This makes them easier to interpret and discuss with colleagues. By reporting all possible G-patterns for a region (like Mexico in this case), the complexity of geographic regionalization is addressed, potentially revealing full or partial nestings.

Supplementary material

Supplemental data for this article can be accessed here: https://doi.org/10.17129/botsci.3787

Acknowledgments

Bernardo Serrano-Estrada developed and programmed the module for visualizing G-patterns in the SIPMX web system. We are deeply grateful to the anonymous reviewer, whose comments substantially improved the final version of the manuscript.

Literature cited

Brooks DR. 1990. Parsimony analysis in historical biogeography and coevolution: methodological and theoretical update. *Systematic Zoology* 39: 14-30. DOI: https://doi.org/10.2307/2992205

Brooks DR, van Veller MGP. 2003. Critique of parsimony analysis of endemicity as a method of historical biogeography. *Journal of Biogeography* 30: 819-825. DOI: https://doi.org/10.1046/j.1365-2699.2003.00848.x

Colwell RK, Coddington JA. 1994. Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 345: 101-118. DOI: https://doi.org/10.1098/rstb.1994.0091

Croizat L. 1982. Vicariance/Vicariism, Panbiogeography, "Vicariance Biogeography," Etc.: A Clarification. *Systematic Zoology* 31:291–304. DOI: https://doi.org/10.2307/2413236

Del-Val E, Balvanera P, Castellarini F, Espinosa-García FJ, Murguía M, Pacheco C. 2015. Identifying areas of high invasion risk: a general model and an application to Mexico. *Revista Mexicana de Biodiversidad* 86: 208-216. DOI: https://doi.org/10.7550/rmb.44743

Morrone JJ, Crisci JV. 1995. Historical biogeography: introduction to methods. *Annual review of ecology and systematics* 26: 373-401.

Murguía M, Villaseñor JL. 2000. Estimating the quality of the records used in quantitative biogeography with presence-absence matrices. *Annales Botanici Fennici* 37: 289-296.

Page RDM. 1987. Graphs and Generalized Tracks: Quantifying Croizat's Panbiogeography. *Systematic Zoology* 36:1-17. DOI: https://doi.org/10.2307/2413304

Rosen BR. 1988. From fossils to earth history: applied historical biogeography. *In*: Myers AA, Giller P. eds. *Analytical Biogeography: An Integrated Approach to the Study of Animal and Plant Distributions*. Chapman & Hall, London, pp. 437-481. ISBN 978-0-4124-0050-6

Simberloff D, Heck KL, McCoy ED, Connor EF. 1981. There have been no statistical tests of cladistic biogeographical hypotheses. *In*: Nelson G, Rosen DE, eds. *Vicariance Biogeography: A Critique*. New York, Columbia University Press, pp. 40-72. ISBN: 9780231048088

Villaseñor JL, Maeda P, Rosell JA, Ortiz E. 2007. Plant families as predictors of plant biodiversity in Mexico. *Diversity and Distributions* 13: 871-876. DOI: https://doi.org/10.1111/j.1472-4642.2007.00385.x

Villaseñor JL, Ortiz E. 2025. Floristic richness comparison among the Mexican states. *Revista Mexicana de Biodiversidad* 96: e965505. DOI: https://doi.org/10.22201/ib.20078706e.2025.96.5505

Associate editor: Gilberto Ocampo

Author contributions: MMR conceived the idea and codified the Prolog programs for the analysis; JLV compiled the source database; MMR and JVL performed the study, wrote the manuscript, and approved the final version.

Supporting Agencies: Not applicable.

Conflict of interests: The authors declare no conflicts of interest, whether financial or personal, regarding the information, data presentation, and results of this article.