

**MARTITRACKS: A GEOMETRICAL APPROACH FOR IDENTIFYING
GEOGRAPHICAL PATTERNS OF DISTRIBUTION**

SUSY ECHEVERRÍA LONDOÑO

UNIVERSIDAD INDUSTRIAL DE SANTANDER

FACULTAD DE CIENCIAS

ESCUELA DE BIOLOGÍA

BUCARAMANGA

2010

**MARTITRACKS: A GEOMETRICAL APPROACH FOR IDENTIFYING
GEOGRAPHICAL PATTERNS OF DISTRIBUTION**

SUSY ECHEVERRÍA LONDOÑO

**Trabajo de investigación presentado como requisito para optar el título de
Bióloga**

DIRECTOR:

Daniel Rafael Miranda Esquivel

Doctor en Ciencias Naturales

UNIVERSIDAD INDUSTRIAL DE SANTANDER

FACULTAD DE CIENCIAS

ESCUELA DE BIOLOGÍA

BUCARAMANGA

2010

Contents

1	INTRODUCTION	10
2	RESULTS	12
2.1	The general framework	12
2.2	First step: Minimum spanning trees (MST)	13
2.3	Second step: Spatial congruence among species	13
2.3.1	Spatial congruence between MST	13
2.3.2	Spatial congruence among MST	15
2.4	Empirical analyses	17
2.4.1	Panbiogeographical analysis of the genus <i>Bomarea</i> (Alstroemeriaceae)	17
2.4.2	Panbiogeographical analysis for genera from the North-Andes Cordillera	17
3	DISCUSSION	18
4	MATERIALS AND METHODS	19
4.1	Availability and Requirements	19
5	ACKNOWLEDGMENTS	19
6	REFERENCES	20

List of Figures

1	Figure 1. MartiTracks' framework	23
2	Figure 2. Basic units of congruence.	24
3	Figure 3. Distance from a point to a segment.	24
4	Figure 4. First condition of congruence.	25
5	Figure 5. Decision rules of congruence.	26
6	Figure 6. Comparison between segments.	27
7	Figure 7. Index of congruence.	28
8	Figure 8. Default search strategy of MartiTracks.	29
9	Figure 9. Panbiogeographical analysis of the genus <i>Bomarea</i> (Alstroemeriaceae).	30
10	Figure 10. Panbiogeographical analysis for genera from the North-Andes Cordillera	31

RESUMEN

TÍTULO: MARTITRACKS: UN ACERCAMIENTO GEOMÉTRICO PARA IDENTIFICAR PATRONES DE DISTRIBUCIÓN GEOGRÁFICA ¹

AUTOR: SUSY ECHEVERRÍA LONDOÑO ²

PALABRAS CLAVE: BIOGEOGRAFÍA HISTÓRICA, PANBIOGEOGRAFÍA, PATRONES GEOGRÁFICOS, GEOMETRÍA

DESCRIPCIÓN

Los datos geográficos han incrementado exponencialmente, por esta razón, es necesario el uso de herramientas biogeográficas capaces de manipular bases de datos de este tipo. La panbiogeografía es una disciplina de la biogeografía, la cual tiene características útiles para el manejo de la complejidad de los datos distribucionales. A pesar de esto, no existen suficientes métodos cuantitativos para el análisis de gran cantidad de datos geográficos.

Este trabajo describe el programa Martitracks, un acercamiento geométrico, como una alternativa al análisis panbiogeográfico tradicional. El algoritmo de MartiTracks ha sido diseñado para definir patrones generales de distribución a partir de datos geográficos de las especies. Este algoritmo está dividido en dos fases, la primera calcula los árboles de tendido mínimo para cada especie, los cuales son equivalentes a los trazos individuales; la segunda fase delimita los patrones de distribución de acuerdo al nivel de congruencia fijada por el usuario. El archivo de salida de Martitracks, es un archivo de tipo KML (Keyhole markup language), el cual es compatible con cualquier programa de sistemas de información geográfica como GoogleEarth, Qgis, etc.

El algoritmo de Martitracks es un acercamiento viable para realizar un análisis panbiogeográfico. Este algoritmo permite eliminar la ambigüedad, y la subjetividad inherente del análisis panbiogeográfico de reconstrucción manual con una gran cantidad de datos. Además, Martitracks es una herramienta útil para delimitar patrones de distribución general de especies, reduciendo la complejidad y el tiempo necesario para la manipulación de gran cantidad de puntos geográficos.

¹Proyecto de Investigación

²Facultad de Ciencias, Escuela de Biología; Director: Daniel Rafael Miranda Esquivel, Doctor en Ciencias Naturales

ABSTRACT

TITLE: MARTITRACKS: A GEOMETRICAL APPROACH FOR IDENTIFYING GEOGRAPHICAL PATTERNS OF DISTRIBUTION ¹

AUTHOR: SUSY ECHEVERRÍA LONDOÑO ²

KEY WORDS: HISTORICAL BIOGEOGRAPHY, PANBIOGEOGRAPHY, GEOGRAPHICAL PATTERNS, GEOMETRY

DESCRIPTION

The geographic data of species have increased exponentially. For this reason, it is necessary to use a biogeographic tool to be able to deal with such geographical datasets. Panbiogeography is a biogeographic approach, which have useful characteristics to manage the complexity of distributional data; however, there are not enough quantitative methods to handle large amounts of geographical data.

We describe the program MartiTracks, a geometric-based approach, as an alternative to traditional panbiogeographic analysis. MartiTracks' algorithm has been designed to define general distributional patterns from geographical data of species. This algorithm is divided into two main components, a minimum spanning tree for each species is computed to create individual tracks in a panbiogeographic context; and, the spatial congruence among minimum spanning tree's segments is defined using congruence parameters in a geometric context. Finally, MartiTracks' algorithm delimits general distributional patterns according to the level of congruence set by the user. The output is a keyhole markup language file, which is compatible with any geographic information system program like GoogleEarth, Qgis, etc.

MartiTracks' algorithm is a feasible quantitative approach for a panbiogeographic analysis. This algorithm allows removal of the ambiguity, and the subjective factor included in a manual panbiogeographic analysis with overcrowded points. Likewise, MartiTracks is a useful tool for determining general distribution patterns of species reducing both the complexity, and time needed for managing large datasets.

¹Thesis project

²Facultad de Ciencias, Escuela de Biología; Director: Daniel Rafael Miranda Esquivel, Doctor en Ciencias Naturales

1 INTRODUCTION

The geographic distribution of species has been considered a substantial source of analysis in biology. This data is regarded meaningful for documenting and conserving biodiversity[1]. Given the exponential growing of distributional data [2, 3], the necessity of procedures and tools to facilitate the management of distributional points has increased.

Biogeographic approaches have tools, which can manage and analyze this kind of data in a suitable and easy way. For example, Ecological biogeography takes specific richness patterns of distribution of life forms for its analysis; Aerography is based on chronological patterns while Macroecology is supported by structural and functional patterns of ecological systems; and finally, Evolutionary or Historical biogeography takes distributional patterns of distribution in order to understand evolutionary patterns and processes[4, 5].

In historical biogeography, the panbiogeographic approach is considered an important tool for the primary management of distributional data [5]. Panbiogeography delimits general distributional patterns taking into account only the geographical data of the species. The method was developed by Croizat [6, 7, 8], who criticized Darwin's biogeographic ideas about natural selection and the means of dispersal in the geographic distribution [9]. The panbiogeographic approach is based on four biogeographic assumptions [4]:

- ”1. Distribution patterns constitute an empirical database for biogeographical analysis;
2. Distribution patterns provide information about where, when, and how animals and plants evolve;

3. The spatial and temporal component of these distribution patterns can be graphically represented;

4. Testable hypotheses about historical relationship between the evolution of distributions and earth history can be derived from geographic correlations between distribution graphs as geological/geomorphic features.”

The method of the panbiogeographic approach, which delimits distributional pattern of multiple species, is called track analysis; and it is based on five elements: individual tracks, generalized tracks, nodes, main massing, and baselines [10, 4, 11, 12]. An individual track is made up of lines drawn on a map, on which different localities or distribution points of a particular taxon or group of taxa are connected, such the sum of the segment lengths that connects all the distribution points is the smallest possible. In graph theory, an individual track is a minimum spanning tree (hereafter MST) [4, 13, 14]. Generalized tracks or standard tracks are considered repetitive patterns because these summarize distributions of diverse individual taxa [15]. These are lines on a map resulting from the overlapping of the individual tracks. Generalized tracks are interpreted as distributional patterns of an ancestral biota that had been fragmented by tectonic or climatic events [16]. Nodes are areas or localities where two or more generalized tracks overlap. These are complex areas or tectonic and biotic convergence zones [14, 16, 10, 13]. The last two elements (main massing and baselines) are considered supplementary information; therefore, we do not consider these elements in the tracks analysis.

Since the panbiogeography approach focuses on the spatial or geographical component, as a fundamental pre-condition to any analysis of the patterns and processes of evolutionary change [4, 17, 11], this approach is a rational cost efficient way to depict a pattern for the data and reduce the initial complexity that we find in this kind of data sets [18]. Nevertheless, there are few quantitative approaches, like Craw’s

compatibility track analysis [18, 19], or Page's method based on graph theory [14], which has not been implemented, for mapping generalized biogeographic tracks.

On account of individual and generalized tracks are lines in a geometrical context, and the congruence among individual tracks is a geometrical property, in this work, we describe a new program called MartiTracks, which contains a new algorithm to perform a panbiogeography track analysis based on a geometric approach. The algorithm includes geometrical functions and processes, which makes a geometric approach a feasible quantitative alternative to the traditional track analysis. Likewise, this approach is a unique and very useful technique to capture any distributional pattern/structure in studies that use geographical data.

2 RESULTS

2.1 The general framework

For a new MartiTracks project, distributional points (latitude, and longitude data) of a particular set of taxa must be defined. A typical MartiTracks input file is a text file, which has the following structure: Taxon-name, latitude, and longitude data. These data points are used for building an individual track for each species. The spatial congruence among the individual tracks is then evaluated through the congruence algorithm in order to determine whether there are generalized tracks representing the general patterns of distribution of the species compared. The individual tracks of each species and the generalized patterns of distribution are represented into a Keyhole Markup Language (kml) file that can be visualized using any Geographic Information System (GIS) program like GoogleEarth, or Qgis. The structure of MartiTracks' framework is shown in Figure 1.

2.2 First step: Minimum spanning trees (MST)

As most of the biogeographic programs like Croizat[19], or Trazos2004[20], MartiTracks creates in the first stage a minimum spanning tree (MST) as the individual track for each species. This stage integrates a minimum euclidian distance parameter, that we called cut value, to reduce overcrowded points. When several points are found at the same place or are close enough to be considered the same sampling point, these points are reduced to a single point. Therefore, this parameter allows reducing the quantity of the initial redundancy in the data sets, speeding up the MSTs calculation.

2.3 Second step: Spatial congruence among species

2.3.1 Spatial congruence between MST

Once the individual tracks are defined, the panbiogeographic method determines the spatial congruence among the individual tracks, to delimitate generalized tracks or general patterns of distribution.

The geometric approach of MartiTracks considers each MST's segment or edge as the basic unit of congruence among species. Therefore, given an individual track or MST as $T = (V(T), E(T))$, a segment s_i belonging to the MST a is defined as the edge that connects two endpoint vertices $s_{i,a}=(v_{i,a}, e_{i,a})$ Figure 2.

The core in the geometrical approach is the function that counts the shortest distance from a point to a segment, this function was developed by Paul Bourke and could be found at <http://local.wasp.uwa.edu.au/~pbourke/geometry/pointline/>.

Given a segment (P1-P2) and a point P3 in the figure 3, the distance from the point P3 to the segment P1-P2, is defined as the distance between the intersecting point P, which result of the perpendicular extension of P3 towards the segment

P1-P2, to the point P3. If the intersecting point does not fall within the segment, the function will take the shorter distance from the point P3 to an endpoint of the segment(P1-P2).

Given two segments $s_{i,a}=(v_{i,a}, e_{i,a})$, and $s_{j,b}=(v_{j,b}, e_{j,b})$ belonging to the species a and b , respectively, we consider as the first condition, that these two segments are congruent if any of the vertices ($v_{i,a}$) in the segment $s_{i,a}$ has an intersecting point (P_a) on the segment b , or if any of the vertices ($v_{j,b}$) in the segment $s_{j,b}$ has an intersecting point (P_b) on the segment a Figure 4.

As congruence also depends on the euclidian distances between segments and points, the maximum, and minimum distances between segments are calculated in order to define two decision rules of congruence. Using these rules, two segments are congruent if the minimum, and the maximum distance between segments do not exceed the limit of the specific values.

For the first rule, $s_{i,a}=(v_{i,a}, e_{i,a})$, and $s_{j,b}=(v_{j,b}, e_{j,b})$ belonging to the species a and b , respectively; and $dmin$ as the minimum distance, $dmax$ as the maximum distance, $lmin$ as the boundary of the minimum distance, and $lmax$ as the boundary of the maximum distance. Two segments are congruent, if the first congruence condition is fulfilled (Figure 3); and if $(0 \leq dmin \leq lmin, \text{ and } 0 \leq dmax \leq lmax)$ is true, see figure 5a.

The second rule is defined by the maximum distance inside the spatial range. Given two segments $s_{i,a}=(v_{i,a}, e_{i,a})$, and $s_{j,b}=(v_{j,b}, e_{j,b})$ belonging to the species a and b , respectively; and $dmax\text{-line}$ as the maximum distance inside of the segment, and $lmax\text{-line}$ as the boundary of the maximum distance inside of the line segment. Two segments are congruent, if both $v_{i,a}$ have intersecting points on $e_{j,b}$ or if both $v_{j,b}$ have intersecting points on $e_{i,a}$ (Figure 3); and if $(0 \leq dmax\text{-line} \leq lmax\text{-line})$ is true. See figure 5b.

Finally, if two segments are found to be congruent, their points will be connected through a new MST. Then, each segment of a species a is compared with the other segments of the species b until the whole MST of the species a has been compared. The same procedure is carried out from b to a . The congruence between two species could be null and therefore no tracks or new MSTs would be created. Figure 6.

2.3.2 Spatial congruence among MST

The spatial congruence of a MST with the remaining MSTs is the criterion to delimitate whether there is a generalized track, if a species does not have any level of congruence with the remain species no generalized tracks or a new MST is generated.

Once each species is contrasted with the remaining species and there are some levels of congruence, a generalized track is created. When the analysis is completed, there is a pool of degenerate tracks (i.e there have several generalized tracks with the same answer). For delimitating the repetitive pattern of distribution among these degenerate tracks, MartiTracks quantifies the similarity among tracks by means of a index of congruence (IC). This index is calculated taking into account the length of the congruent segments between two MST: a, b (either individual, or generalized track), and the total length of the MST a or b .

It is important to emphasize that this is an asymmetrical index, because if we calculate the IC from the MST of the species a to the MST of the species b , this value will be different when the index is calculated from the MST of the species b to the MST of the species b . As both have the same congruent segments but not the same length for the divisor, which could be either the total MST_a length or total MST_b length. For resolving this asymmetrical characteristic, the congruence algorithm takes the highest value of IC. Figure 7.

IC_{ab} = congruent segments length of the MST_{ab} /total MST_a length

IC_{ba} = congruent segments length of the MST_{ba} /total MST_b length.

The index of congruence (IC) is designed to measure the similarity among tracks, and determines if two tracks can be considered the same element. Given i as the higher value between IC_{ab} , and IC_{bc} , and min-IC as the minimum IC value for considering two MSTs as the same element. If ($i \geq \text{min-IC}$) the geographical points of the MST of the species a , and b will be jointed and will be part of the same MST.

The values of the cut value, l_{min} , l_{max} , and $l_{\text{max-line}}$, and min-IC are defined according to the congruence level of the user's requirements. The default values which are used by Martitracks, were obtained following the empirical results of the panbiogeographical analysis of the genus *Bomarea*, which are described in the empirical analyses section.

It is important to take into account that the value of each parameter of congruence depends on the value of the other parameters. In like manner, there is a constrain rule for these values, hence cut value < l_{min} < l_{max} < $l_{\text{max-line}}$.

In summary, the geometrical algorithm of MartiTracks uses the cut value, l_{max} , l_{min} , and $l_{\text{max-line}}$ of the decision rules, and the index of congruence IC for decreasing both the complexity, and time of analysis, and for delimitating repetitive patterns of distribution among geographical points of species. These parameters can be use in different levels of the analysis. Thus, we can use the IC for reducing the number of individual tracks at the beginning, then calculate the congruent segments among the species, delimit generalized tracks using the decision rules and, finally reduce the similar generalized tracks by the IC at the end. The figure 8 shown the default search strategy of MartiTracks using the congruence parameters in different levels of the analysis.

2.4 Empirical analyses

2.4.1 Panbiogeographical analysis of the genus *Bomarea* (Alstroemeriaceae)

An empirical analysis was developed with 2340 records belonging to 61 species of the genus *Bomarea*, which were obtained from the Global Biodiversity Information Facility GBIF (<http://www.gbif.org/datasets/resources/> 24/07/2010). We used three different parameters sets of congruence in order to calculate general distributional patterns with different levels of congruence. The general distributional patterns obtained by one of the sets are shown in figure 9.

To run the program, we used a PC-compatible computer with an Intel Core 2 Quad Q6600 at 2.40 GHz and 4GB of RAM, running Ubuntu 9.04 64 bits. The panbiogeographical analyses of the genus required from 30 to 60 seconds.

These results were compared with the results from a previous panbiogeographic work of *Bomarea*, which was made by Alzate *et al.*, 2008 [21]. In contrast to our analysis, Alzate *et al.*, included 2205 records belonging to 101 species of the genus *Bomarea*. Although there is a difference between the number of species evaluated in both analyses, we found similar patterns of distribution between our results, and Alzate's *et al.* results. Figure 9.

2.4.2 Panbiogeographical analysis for genera from the North-Andes Cordillera

We analyzed 100118 georeferences belonging to 1031 genera, which are distributed across the North Andes, with the aim of evaluating Martitracks efficiency to deal with large datasets. This dataset was obtained from the Global Biodiversity Information Facility GBIF (<http://www.gbif.org/datasets/resources/> 26/06/2009), and was not filtered for errors on distributions or taxonomy, therefore this mimics an exploratory analysis to evaluate a huge data set. Four parameter sets were employed to visual-

ize general patterns of distribution with different levels of congruence.

Depending on the parameters used, the analyses of North-Andes data generated several patterns, from three to 27 generalized tracks. In figure 10 we show the three general patterns found with one of the sets evaluated. The analyses required from 15 to 30 minutes. These results prove the high ability of MartiTracks program for reducing the complexity of the data and to find a common pattern of distribution with large datasets, in a reasonable time.

3 DISCUSSION

As the amount of geographical information is rapidly growing, the necessity of bioinformatic tools able to deal with this kind of data, has increased. For panbiogeographical analyses, MartiTracks is a feasible quantitative alternative to the traditional track analysis (as Manual reconstruction or Craw's compatibility track analysis); As a consequence, the ambiguity and the subjective factor, produced when overcrowded geographical points are evaluated [22, 23], is eliminated from the analyses. Another significant advantage of MartiTracks is that the geometrical approach eliminates the high amount of time needed for analyzing large datasets as shown in North-Andes analysis. Thus, a single computer could easily deal with datasets involving thousands of geographical records. Finally, by setting different distance parameters, which define the level of congruence, the users can explore several levels of resolution for analyzing their datasets. Hence, the users can explore several settings according to their requirements.

4 MATERIALS AND METHODS

MartiTracks was designed using Freepascal language under the Unix Operative System, Linux - Ubuntu 10.04 64 bits. Compiled versions of the program for Windows and Linux platforms, along with the source code are freely available under a GNU General Public license GPL 2.1 at <http://tux.uis.edu.co/labsist/martitracks> and <http://code.google.com/p/martitracks>

4.1 Availability and Requirements

Project name: MartiTracks

Project home page:

<http://tux.uis.edu.co/labsist/martitracks>

<http://code.google.com/p/martitracks>

Bug tracking <http://code.google.com/p/martitracks>

Operating system(s): Plataform independent

(But 64 bits OS is recommended for medium/huge datasets)

Programming language: Freepascal

License: GNU GPL 2.1

Any restrictions to non-academics: none

5 ACKNOWLEDGMENTS

The authors would like to thank Paul Bourke for information, and code source for the function of minimum distance between a point and a line, and Salvador Arias for his suggestions and ideas to improve this article.

6 REFERENCES

References

- [1] Richardson DM, Whittaker RJ (2010) Conservation biogeography –foundations, concepts and challenges. *Diversity and Distributions* 16: 313-320.
- [2] Crovello TJ (1981) Quantitative Biogeography: An Overview. *Taxon* 30: 563–575.
- [3] Soria-Auza R, Kessler M (2008) The influence of sampling intensity on the perception of the spatial distribution of tropical diversity and endemism. A case study of ferns from Bolivia. *Diversity and Distributions* 14: 123-130.
- [4] Craw RC, Grehan JR, Heads MJ (1999) *Panbiogeography: Tracking the History of Life*. Oxford University Press, 229-229 pp.
- [5] Morrone JJ (2009) *Evolutionary Biogeography: An integrative approach with case studies*. Columbia University Press.
- [6] Croizat L (1952) *Manual of Phytogeography*. The Hague: Junk.
- [7] Croizat L (1958) *Panbiogeography*. Caracas: Published by the author.
- [8] Croizat L (1964) *Space, Time, Form: The Biological Synthesis*. Caracas: Published by the author.
- [9] Craw RC (1987) Panbiogeography and structuralist biology. *Rivista di Biologia - Biology Forum* 80: 226-229.
- [10] Crisci JV, Morrone JJ (1992) Panbiogeografía y biogeografía cladística: paradigmas actuales de la biogeografía histórica. *Ciencias* 6: 87-97.

- [11] Grehan JR (2001) Introducción a la Biogeografía en Latinoamérica: Teorías, Conceptos, Métodos y Aplicaciones, Mexico: Las Prensas de Ciencias, Universidad Nacional Autónoma de México, chapter Panbiogeografía y la geografía de la vida. pp. 181-195.
- [12] Espinos-Organista D, Morrone JJ, Llorente-Bousquets J, Villela-Flores O (2002) Patrones de homología espacial, Las Prensas de Ciencias, Universidad Nacional Autónoma de México, chapter 4. pp. 53-64.
- [13] Morrone JJ (2004) Homología Biogeográfica: Las Coordinadas Espaciales de la Vida, volume 37 of *Cuadernos del Instituto de Biología*. Mexico: Instituto de Biología, Universidad Nacional Autónoma de México, 199 pp.
- [14] Page RDM (1987) Graphs and generalized tracks: quantifying Croizat's panbiogeography. *Systematic Zoology* 36: 1-17.
- [15] Michaux B (1989) Generalized tracks and geology. *Systematic Zoology* 38: 390–398.
- [16] Craw RC (1988) Continuing the synthesis between Panbiogeography, Phylogenetic systematics and Geology as illustrated by empirical studies on the biogeography of New Zealand and the Chatham Islands. *Systematic Zoology* 37: 291-310.
- [17] Grehan JR (1994) The Beginning and End of Dispersal: The Representation of Panbiogeography. *Journal of Biogeography* 21: 451–462.
- [18] Craw RC (1989) Quantitative panbiogeography: Introduction to methods. *New Zealand Journal of Zoology* 16: 485-494.
- [19] Cavalcanti MJ (2009) Croizat: A software package for quantitative analysis in Panbiogeography. *Biogeografia* 4: 4-6.

- [20] Rojas-Parra CA (2007) Una herramienta automatizada para realizar análisis panbiogeográficos. *Biogeografia* 1: 31-33.
- [21] Alzate F, Quijano-Abril MA, Morrone JJ (2008) Panbiogeographical analysis of the genus *Bomarea* (Alstroemeriaceae). *Journal of Biogeography* 35: 1250-1257.
- [22] Franco-Rosselli P (2001) Estudios panbiogeograficos en Colombia, Mexico: Las Prensas de Ciencias, Universidad Nacional Autonoma de Mexico. pp. 221-224.
- [23] Liria J (2008) Geographic information systems and spatial analysis: a combined method for panbiogeographic studies. *Revista Mexicana de Biodiversidad* 79: 281-284.

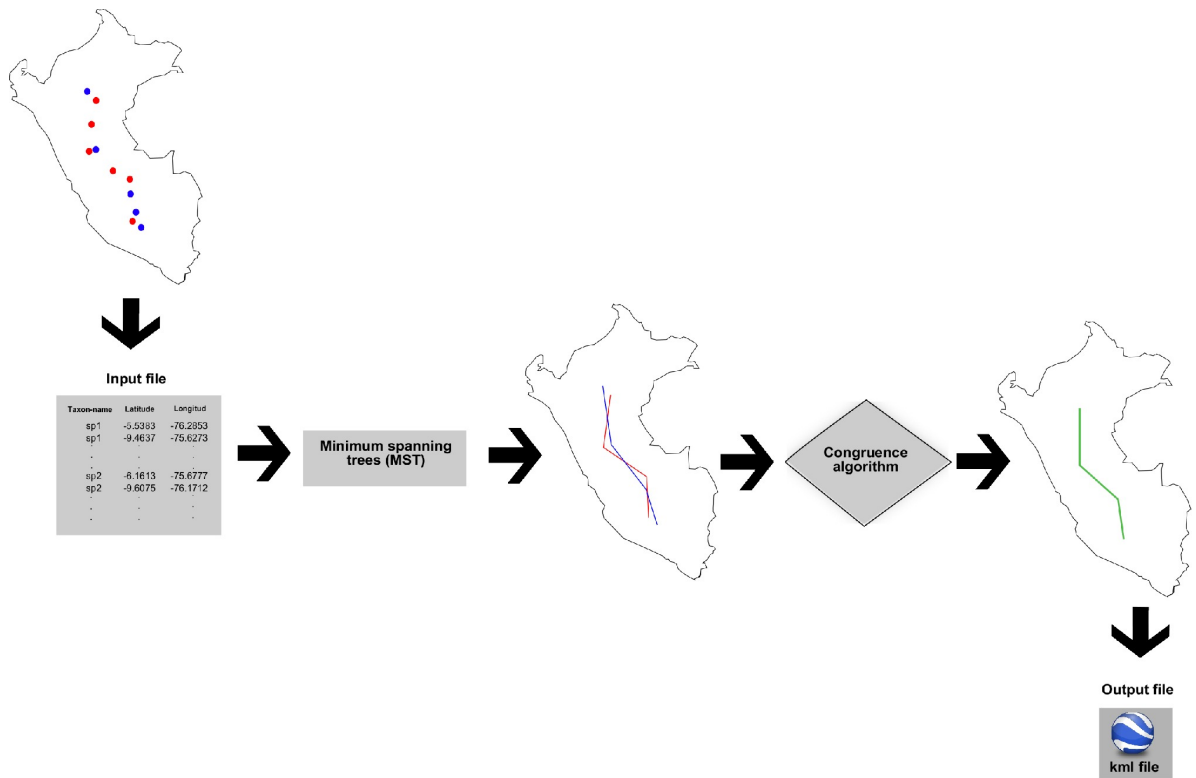


Figure 1: **MartiTracks' framework.** The user specifies the input file with distributional data of species (latitude-longitude). These geographic points are used to calculate a minimum spanning tree (MST) for each species. Then the MSTs are analyzed by the congruence algorithm in order to delimit general patterns of distribution. The output file is a kml file.

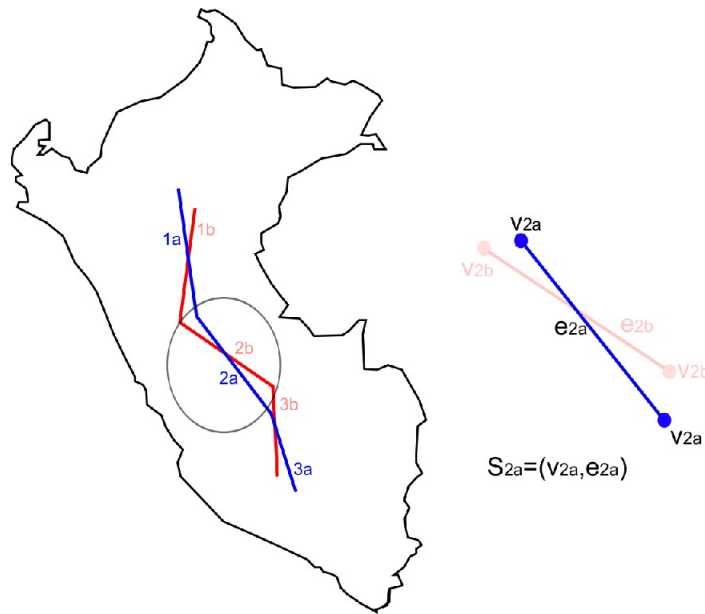
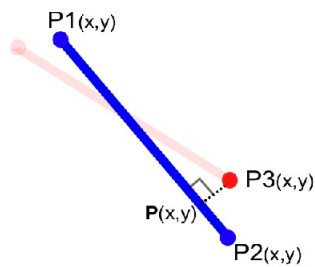


Figure 2: **Basic units of congruence.** The MSTs' segments are the basic units of congruence among species. Each segment s_i belonging to the MST a is defined as an edge $e_{i,a}$ that connects two endpoint vertices $v_{i,a}$.



$$u = \frac{(P3x - P1x)(P2x - P1x) + (P3y - P1y)(P2y - P1y)}{(P2x - P1x)^2 + (P2y - P1y)^2}$$

$$Px = P1x + u(P2x - P1x) \quad Py = P1y + u(P2y - P1y)$$

$$d(P3, P) = (Px - P3x)^2 + (Py - P3y)^2$$

Figure 3: **Distance from a point to a segment.** The distance from the point (P3) to the segment (P1-P2) is calculated by the distance between the intersecting point (P), resulting from the perpendicular extension of P3 towards the segment (P1-P2), to the point P3. For further information <http://local.wasp.uwa.edu.au/~pbourke/geometry/pointline/>.

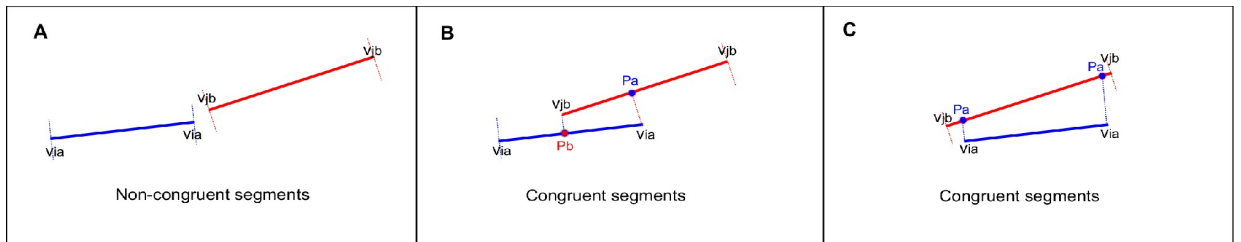


Figure 4: **First condition of congruence.** MartiTracks considers two segments $s_{i,a}$, and $s_{j,b}$ as congruent, if any of the vertices ($v_{i,a}$) has an intersecting point (P_a) on the segment b , or if any of the vertices ($v_{j,b}$) has an intersecting point (P_b) on the segment a

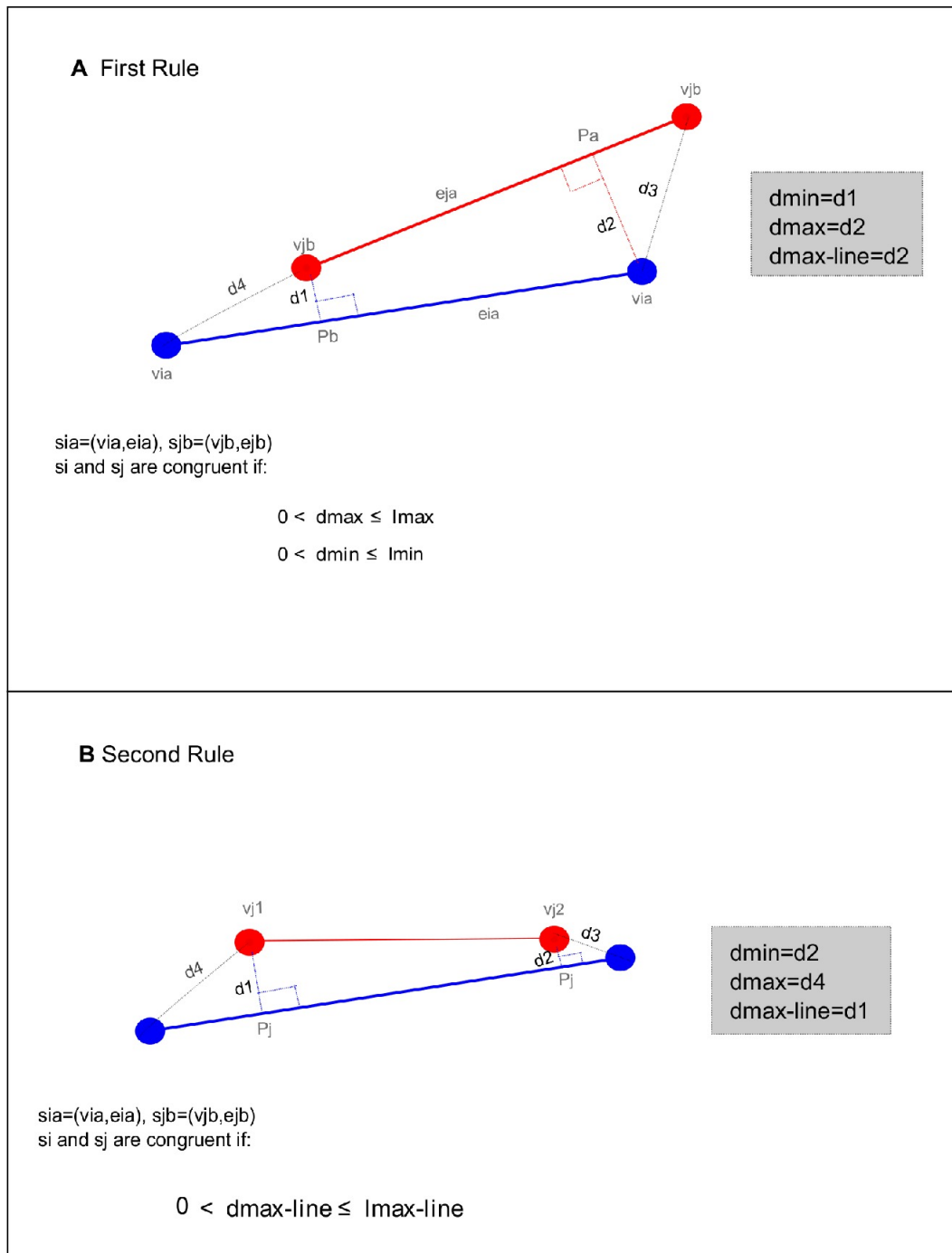


Figure 5: **Decision rules of congruence.** MartiTracks takes the minimum, and the maximum distance between segments to define decision rules of congruence. Given two segments $s_{i,a}=(v_{i,a}, e_{i,a})$, and $s_{j,b}=(v_{j,b}, e_{j,b})$ belonging to the species a and b , respectively; two segments are congruent if: A. The first condition of congruence is fulfilled see figure 2, and if $(0 \leq d_{min} \leq l_{min}, \text{ and } 0 \leq d_{max} \leq l_{max})$ is true. B. If both $v_{i,a}$ have intersecting points on $e_{j,b}$ or if both $v_{j,b}$ have intersecting points on $e_{i,a}$ (Figure 3); and if $(0 \leq d_{max-line} \leq l_{max-line})$ is true.

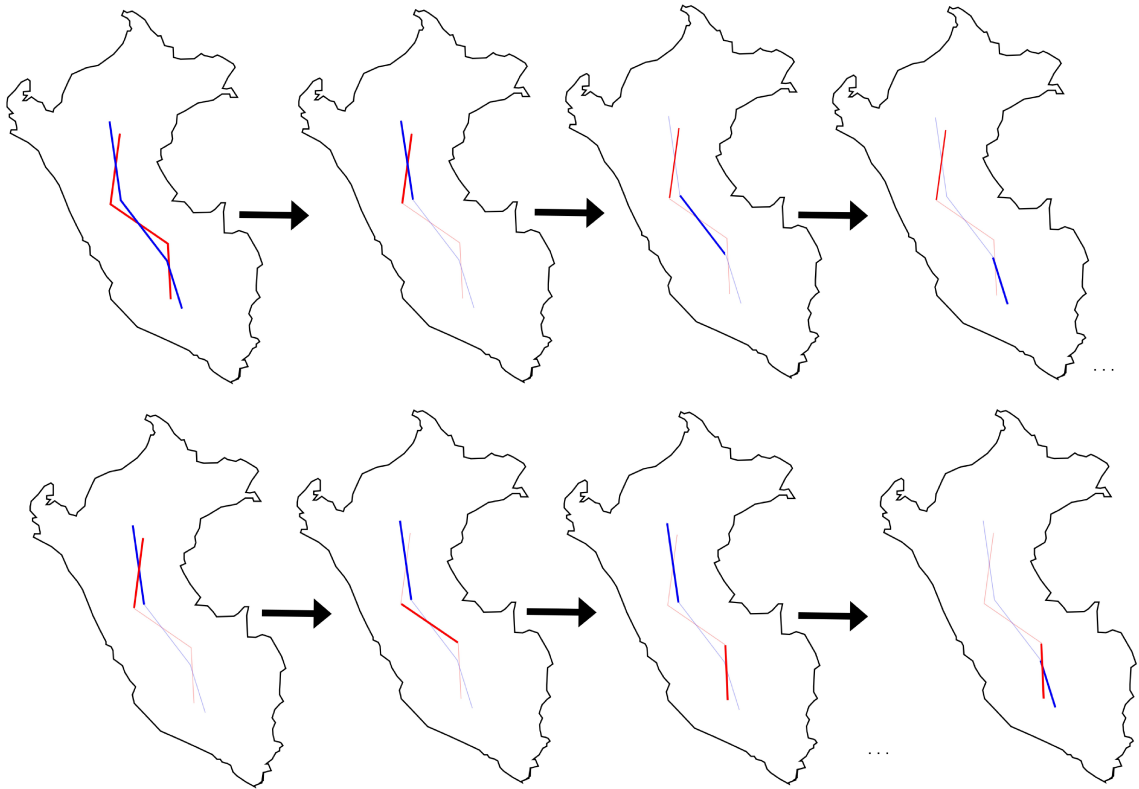
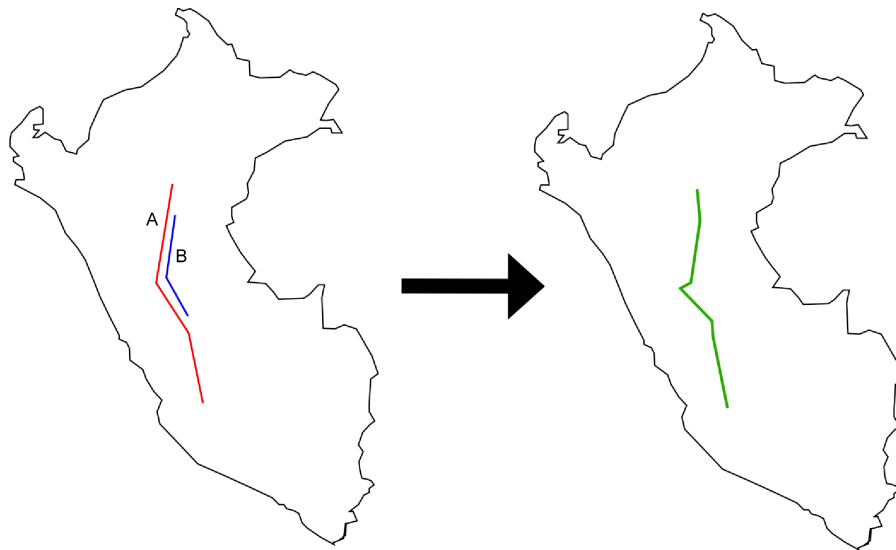


Figure 6: **Comparison between segments.** Each segment of the species a is compared with the other segments of the species b until the whole MST_a has been analyzed. The same procedure is carried out from b to a .



	Number of segments	Number of congruent segments	Total MST length	Congruent segments length	IC
A	4	3	7.85	5.27	0.67
B	2	2	3.65	3.65	1

Figure 7: **Index of congruence.** MartiTracks calculates the similarity among degenerate tracks by the index of congruence (IC). Given two tracks a, b (either individual, or generalized tracks), the index of congruence IC_{ab} from a to b is the length of the congruent segments from a to b divided by the total length of the MST_a . In the same way as the track a is calculated the index of congruence IC_{ba} from b to a . MartiTracks takes the higher value IC_{ab} , and IC_{bc}

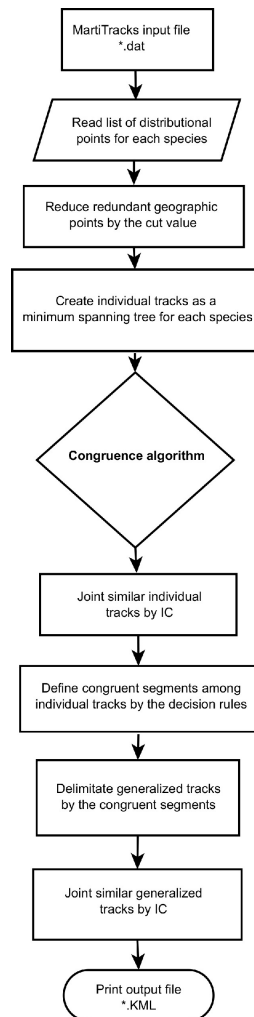


Figure 8: **Default search strategy of MartiTracks.** At the beginning of the analysis the user must define the values of the congruence parameters (cut value, I_{min} , I_{max} , and $I_{max-line}$ for the decision rules; and $min-IC$). These parameters are used in different levels of MartiTracks analysis. Thus, the cut value is utilized for reducing the redundancy of the initial data; then the index of congruence among individual tracks is calculated for jointing similar tracks; I_{min} , I_{max} , and $I_{max-line}$ values are used in the decision rules to define congruent segments among individual tracks; finally, the index of congruence is again calculated, but for generalized tracks to find the general patterns of distribution among the species.

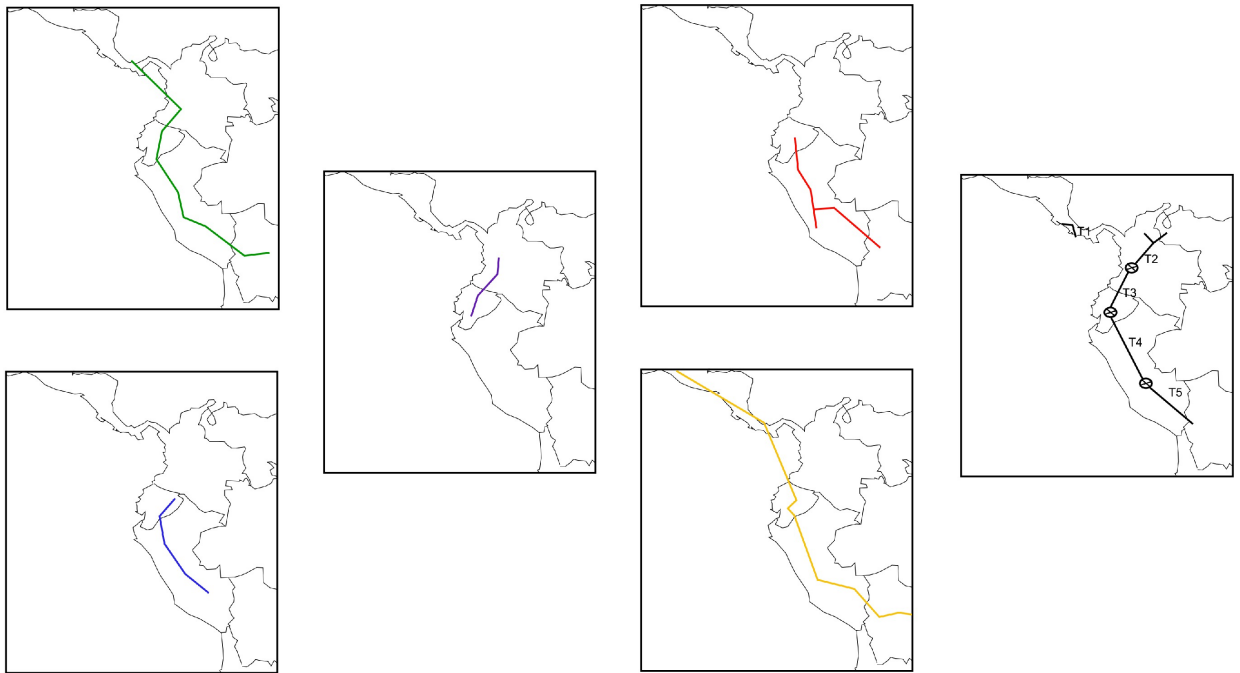


Figure 9: **Panbiogeographical analysis of the genus *Bomarea* (Alstroemeriaceae)**. General patterns of distribution obtained with: cut value=2, $I_{min}=2.5$, $I_{max}=3$. $I_{max-line}=4$, and $min-IC=0.8$ in MartiTracks, and Alzate *et al*'s generalized tracks of 101 species of the genus *Bomarea*

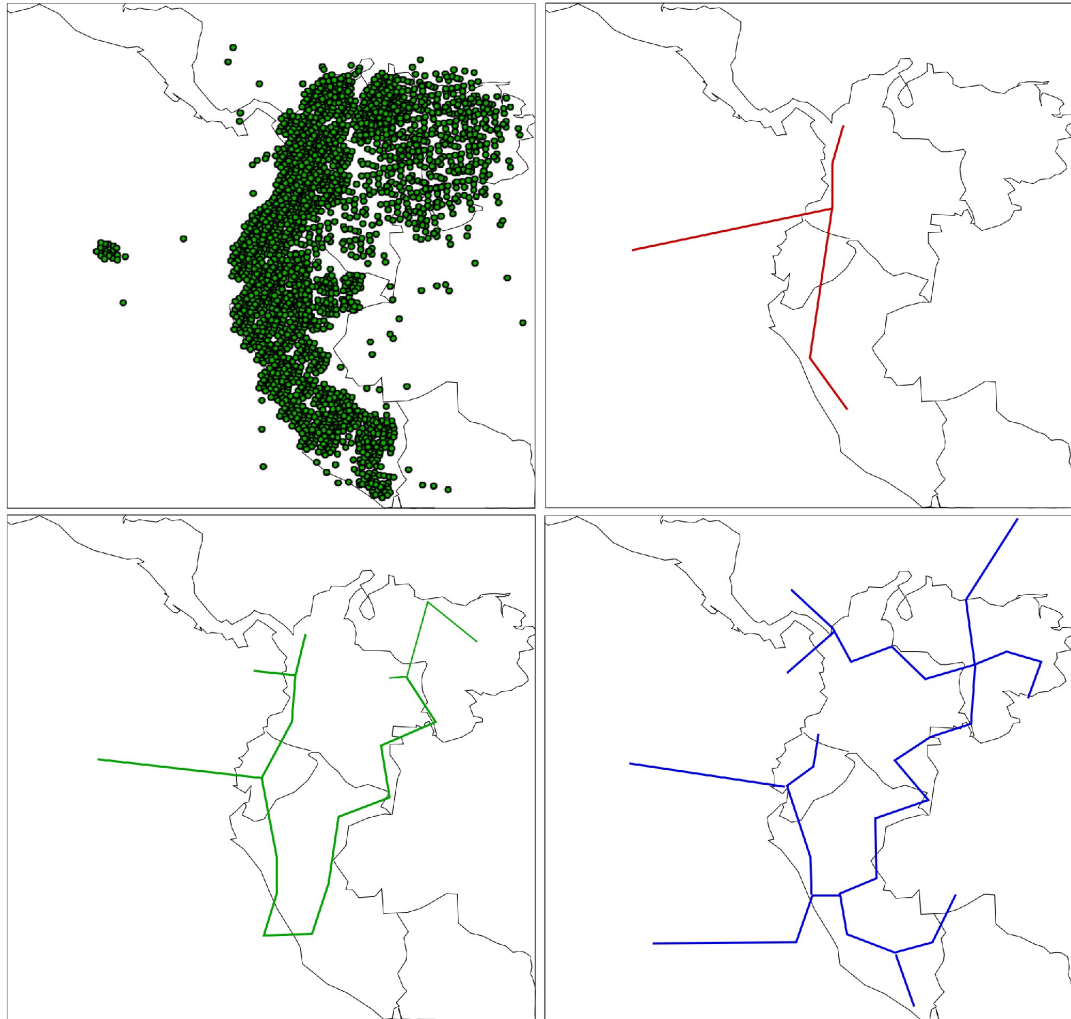


Figure 10: **Panbiogeographical analysis for genera from the North-Andes Cordillera.** 100118 georeferences of 1013 genera distributed across the North Andes, and general patterns of distribution obtained with: cut value=2, $I_{min}=12$, $I_{max}=10$, $I_{max-line}=16$, and $min-IC=0.6$ in MartiTracks.