Quim. Nova, Vol. 35, No. 11, 2146-2152, 2012

Artigo

CHEMOTAXONOMY OF THREE GENERA OF THE Annonaceae FAMILY USING SELF-ORGANIZING MAPS AND ^{13}C NMR DATA OF DITERPENES[#]

Luciana Scotti, Josean Fechine Tavares e Marcelo Sobral da Silva
Departamento de Ciências Farmacêuticas, Centro de Ciências da Saúde, Universidade Federal da Paraíba, Campus I, 58051-970 João Pessoa – PB, Brasil
Emanuela Viana Falcão, Luana de Morais e Silva, Gabriela Cristina da Silva Soares e Marcus Tullius Scotti*
Departamento de Engenharia e Meio Ambiente, Universidade Federal da Paraíba, Campus IV, 58297-000 Rio Tinto – PB, Brasil

Recebido em 4/6/12; aceito em 17/7/12; publicado na web em 26/10/12

The Annonaceae family is distributed throughout Neotropical regions of the world. In Brazil, it covers nearly all natural formations particularly *Annona*, *Xylopia* and *Polyalthia* and is characterized chemically by the production of sources of terpenoids (mainly diterpenes), alkaloids, steroids, polyphenols and, flavonoids. Studies from ^{13}C NMR data of diterpenes related with their botanical occurrence were used to generate self-organizing maps (SOM). Results corroborate those in the literature obtained from morphological and molecular data for three genera and the model can be used to project other diterpenes. Therefore, the model produced can predict which genera are likely to contain a compound.

Keywords: chemotaxonomy; self-organizing maps; ^{13}C NMR.

INTRODUCTION

The Annonaceae family is composed of about 2000 species in 129 genera found throughout Neotropical regions of the world. In Brazil, the family encompasses around 29 genera with approximately 260 species occurring in all natural formations predominantly *Xylopia, Annona* (genus native) and *Polyalthia* (genus input). This family is known to produce many edible fruits, and many of its plants are commonly used in folk medicine. The Annonaceae and all woody plants of the *Magnoliales* has very rich chemical characteristics and are recognized sources of terpenoids (mainly diterpenes), alkaloids (in large amounts, especially the core isoquinoline derivatives), steroids, polyphenols and, flavonoids.[1]

Xylopia comprises about 150 species.[2] Some fruits from this genus are used for culinary purposes as condiments while others serve as a source of fiber for rope manufacture. The timber produced is light, durable and displays medical properties.[3] Several species have broad applications, particularly in folk medicine as vermifuges and antimicrobial agents.[4] Numerous chemical compounds have been isolated from *Xylopia*, including biologically active acetogenins,[5] kauranes and labdanes (diterpene types), sesquiterpenes, alkaloids and, flavonoids. Notably, diterpenes exhibit metabolic characteristics of the *Xylopia* genus.[6]

The genus *Annona* has about 120 species (*A. squamosa*, *A. cherimola* Mill., *A. reticulate* L., *A. muricata* L., *A. dioica* among others) found in Central and South America, Africa, Asia and Australia. Some have been investigated for their chemical compounds and pharmacological activities. Many of the species are used in traditional medicines for the treatment of a variety of diseases. Several annonacaeous species have been found to contain acetogenins, a class of natural compounds with a wide variety of biological activities.[7-12]

Polyalthia contains several species: *P. angustissima*, *P. chrysotricha*, *P. elmeri*, *P. glabra*, *P. hirtifolia*, *P. hookeriana,* among others. It is found natively in India and Sri Lanka and was introduced into gardens in many tropical countries around the world. These plants produce a great diversity of substances that could be of therapeutic significance in many areas of medicine. These have shown marked antimicrobial, anti-inflammatory, cytotoxic, immunosuppressive, antibacterial, antifungal and other pharmacological properties.[13-15]

Phylogenetic studies have contributed to the knowledge of evolution relationships among organisms. However, to date few studies have investigated Annonaceae phylogeny, and those conducted have tended to focus on the relationship among sister families or phylogenetic studies on a genus level.[16-19] Concerning the three genera of the present study, *Polyalthia* is nearest to the tree's basal, while *Xylopia* and *Annona* have closed relationship (Figures 1and 2).[16-19] These previous studies were based on molecular data sets and morphological analyses. However, chemistry data can also contribute to phylogenetic analysis. Secondary metabolites have specific botanic origin, which can be used as a chemotaxonomic marker and thus, represent a bridge in phylogeny between genetics and morphology.[20] In addition, studies of natural chemical products have made a significant contribution to the bioprospection/development of new drugs, as well as to associating compounds, properties and botanic occurrence.[21]

The use of chemical data for classification has been employed in several studies and secondary metabolites have been proposed and used by Gottlieb and co-workers.[22-28] Chemotaxonomic studies have been applied at several levels using different classes of secondary metabolites: superorders of angiosperms;[20,24,25] families such as Asteraceae,[29] Meliaceae,[27] Apocynaceae,[28] Lamiaceae;[30] tribes of Asteraceae.[23,31-35]

ANNs (Artificial Neural Networks) are a method or, more precisely, a set of methods, used extensively since the 1990s. Since ANNs are not restricted to linear correlations and can also take into account non-linear data correlations, they can be efficiently applied for modeling, prediction and classification. The book by Zupan and Gasteiger remains the main text on this area.[36] Self-organizing maps (SOMs) are widely used by ANNs for pattern recognition and classification as proposed by Kohonen who called his algorithm a "self-organizing network".[37,38] This procedure can map multivariate data onto a two dimensional grid, grouping similar patterns near each other. Therefore, Kohonen learning is best suited for mapping of data. In this projection, the similarity relationship between objects

*e-mail: mtscotti@ccae.ufpb.br
[#]Artigo em homenagem ao Prof. Otto R. Gottlieb (31/8/1920-19/6/2011)

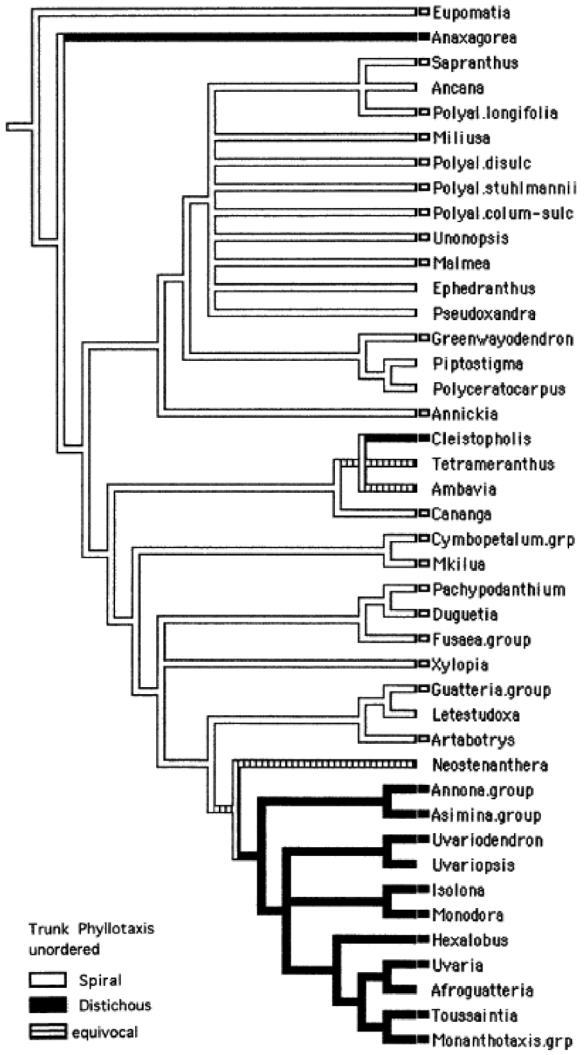


Figure 1. *The rbcL consensus tree of Doyle and co-workers Reproduzida da ref. 19, com permissão do autor*

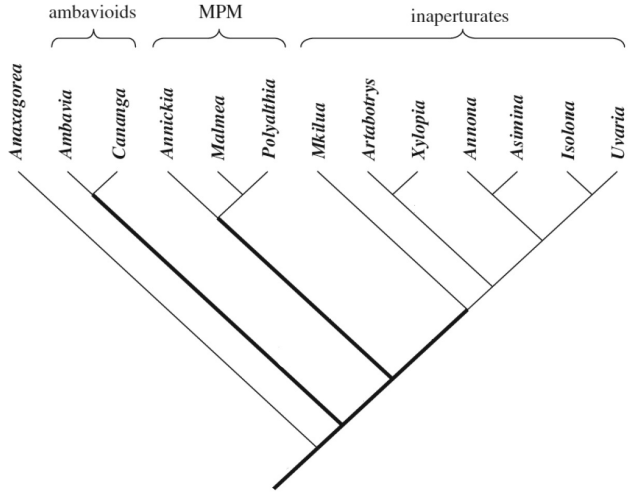


Figure 2. *Annonaceae relationships found in the strict consensuses of 2000 and 78 most-parsimonious trees found from the molecular combined and total combined analyses (morphology + molecular data). MPM = malmeoid-piptostigmoid-miliusoid clade. Reproduzida da ref. 17, com permissão de John Wiley and Sons*

is conserved.[38] Thus, in principle, Kohonen networks can be used for clustering of objects. It is important to note that the training of these networks (SOMs) is unsupervised, but there are some algorithms that combine the original SOM with supervised methods.[36]

In the original SOM procedure, the investigated property is not used during the training process. Each neuron in the grid is associated with a weight, and similar patterns stimulate neurons with similar weight, so that similar patterns are mapped near one another.[36] In chemistry for example, there are several applications of ANNs such as in HPLC, reactivity, and the classification of olive oils.[36] Another very useful application is in the prediction and classification of spectra such as infrared,[39] mass,[40] and nuclear magnetic resonance[41-43], including some QSAR studies.[44] In natural products chemistry, there are a few studies available showing applications of ANNs, such as the classification of Asteraceae tribes,[31,32,35] and the prediction of skeletal types.[45,46]

^{13}C NMR (Nuclear Magnetic Resonance) data yield rich information about the molecular structure and are sufficiently sensitive to detect small differences in the molecule.[47] These differences are measured by the variation in chemical shifts, and the values can be used to associate the chemical structure with the respective botanical occurrence. This association can help to understand the influence of the chemical environment of secondary metabolite as a molecular descriptor, and, for this reason, ^{13}C NMR data may be useful in chemotaxonomic studies.

In the present study, ^{13}C NMR chemical shift values of 20 carbons of skeletal structures of 137 diterpenes, selected from the literature[48] and Self-Organizing Maps (SOMs), were used to perform chemotaxonomic studies of three genera of the Annonaceae family, namely: *Annona*, *Xylopia* and *Polyalthia* and to compare with previous studies using morphological and molecular data.

EXPERIMENTAL AND CHEMIOINFORMATIC STUDIES

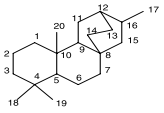
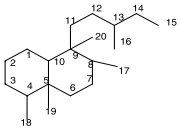
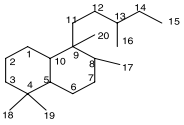
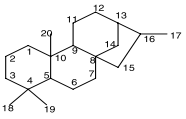
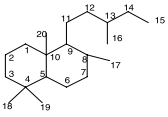
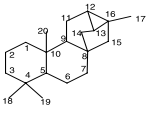
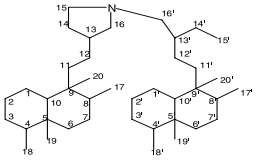
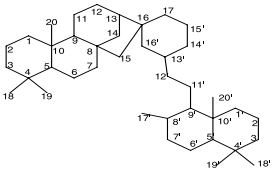
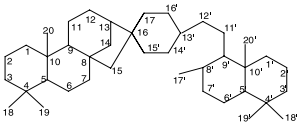
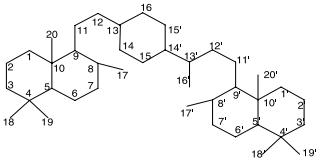
We selected 118 diterpenes from the literature (Table 1) together with their ^{13}C NMR chemical shifts and respective botanical occurrence in three genera: *Annona*, *Polyalthia* and *Xylopia*, of the Annonaceae tribe. The respective skeletal types are listed in Table 2. The ^{13}C NMR chemical shifts of the diterpenes were introduced as input data as showed in Table 3. Each diterpene can appear n times within a delimited taxon, in this case a genus (*Annona*, *Polyalthia*, *Xylopia*). The number of occurrences for a taxon was defined by counting how many times a compound appeared in a given species belonging to that taxon (genus). The 118 diterpenes have 169 botanical occurrences. Therefore, the input data constitutes a matrix of 169 samples and 20 variables where each variable corresponds to a chemical shift. The chemical shifts are sorted by sequence number of the diterpene skeletal atoms as shown in Table 2. The samples are labeled according to botanical occurrence in the three genera used in this study, namely Annonaceae: *Annona*, *Polyalthia* and *Xylopia.*

A Kohonen-ANN was trained using the Matlab 6.5 computing environment by Mathworks and SOM Toolbox 2.0.[49] SOM toolbox is a set of Matlab functions that can be used to develop and implement SOM neural networks, and which contains functions for creation, visualization and analysis of self-organizing maps. A SOM grid with square geometry 13 x 5 in size was created and trained. The training was conducted through the Batch-training algorithm. In this algorithm, the whole dataset is presented to the network before any adjustment is made. In each training step, the dataset is partitioned according to the regions of the map weight vectors. Within the algorithm, the new weight vector is based on simple averages and there is no learning rate.[38] This feature allows missing values to be ignored by the net. The number of epochs is automatically chosen by

Table 1. Diterpenes from Annonaceae used in this study

ID	Substance name	Skeleton number[a]
1	Trachyloban-18-oate methyl,*ent*:	6
2	Kauran-16-a-17-dihydroxy	4
3	Kauran-19-oic acid, 16-a-17-dihydroxy, *ent*:	4
4	Kauran-16-a-17-dihydroxy: methyl ester	4
5	Kauran-19-oate, methyl: 16-b-17-dihydroxy: (-):	4
6	Kauran-16-b-17-dihydroxy	4
7	Kauran-16-b-hydroxy-17-acetoxy-19-oate: methyl, (-):	4
8	Kauran-16-b-17-diacetoxy-19-oate: methyl, (-):	4
9	Kauran-17-b-hydroxy-19-oate	4
10	Kauran-17-a-hydroxy-19-oate	4
11	Kaur-16-en-19-oic acid, (-):	4
12	Kauran-19-oic acid, 16-b-17-dihydroxy	4
13	Kauran-16-17-19-triol, 16-b: *ent*:	4
14	Kauran-19-oic acid, 16-b-*ent*: 17-hydroxy:	4
15	Kauran-19-oic acid, 16-a-*ent*: 17-hydroxy:	4
16	Kaur-16-en-19-ol, *ent*: (-):	4
17	Kauran-19-oic acid, 16-b-17-dihydroxy: *ent*:	4
18	Kauran-19-oic acid, 17-acetoxy-16-b: *ent*:	4
19	Kauran-17-oic acid, 19-formyl: *ent*:	4
20	Kauran-19-al, 16-b-17-dihydroxy: *ent*:	4
21	Kauran-19-al,16-b-*ent*: 17-hydroxy :	4
22	Kauran-19-al, 16-b-hydroxy-17-acetoxy:*ent*:	4
23	Kauran-17-oic acid, 19-nor-4-a-hydroxy: *ent*:	4
24	Annosquamosin B	4
25	Annoglabasin A	4
26	Annoglabasin B	4
27	Kauran-19-oic acid, 16-17-diacetoxy: *ent*:	4
28	Kauran-17-oic acid, 19-carbomethoxy: *ent*:	4
29	Labda-8-trans-13-dien-15-oic acid, 18-carboxy: methylester (4S,9R,10R)	4
30	Kaur-16-en-19-oate	4
31	Trachyloban-18-b-oic acid, 7-b-acetoxy:	4
32	Trachyloban-18-b-oic acid, 7-oxo	4
33	Trachyloban-19-a-oic acid, 7-a-hydroxy:	4
34	Kauran-17-oic acid, 16S: (-):	4
35	Kauran-19-oic acid, 16-a-hydroxy: (-)	4
36	Annoglabasin C	4
37	Annoglabasin D	4
38	Annoglabasin E	4
39	Annoglabasin F	4
40	Kauran-19-oic acid, 16-a-methoxy: *ent*:	4
41	Kauran-17-19-dimethyl ester, 16-a-hydro:*ent*:	4
42	Cleroda-3-13(14)-dien-15-16-olide, 16-a-hydroxy:	4
43	Cleroda-3-13(14)-dien-15-16-olide, 16-a-acetoxy:	4
44	Cleroda-3-13(14)-E-dien-15-methyl ester, 16-oxo:	4
45	Kolava-3-13-trans-dien-15-oic acid	4
46	Kolava-3-11-trans-dien-13-one, 14-15-bisnor:	4
47	Kolava-3-13-cis-dien-16-15-olide-2-one, (16R):	4
48	Kolava-3-13-cis-dien-16-15-olide-2-one, (16S):	4
49	Kolava-2-13-cis-dien-16-15-olide-3-al, (4-2)-abeo: (16R):	4
50	Kolava-2-13-cis-dien-16-15-olide-3-al, (4-2)-abeo: (16S):	4
51	Isoozate dimer-13'-oxo-methyl, 13'-nor: *ent*:	4
52	Isoozate dimer-13-epoxy-methyl, *ent*:	4
53	Isoozate dimer-*ent*-methyl, *ent*:	4
54	Polyalthiaditerpeno 4	4
55	Kolava-3-13-cis-dien-16-al-15-oic acid	4
56	Kolava-3-12-cis-dien-16-al, 15-oic acid	4
57	Polyalthialdoic acid	4
58	Cleroda-3-13-dien-15-16-olide, 16-hydroxy:	4
59	Cleroda-4(18)-13-dien-15-16-olide, 16-hydroxy:	5
60	Halima-5(10)-13-dien-15-16-olide, 16-hydroxy: *ent*:	4
61	Cleroda-3-trans-13-dien-15-oic acid	4
62	Cleroda-4(18)-13-trans-dien-15-oic acid	4
63	Halima-5(10)-trans-13-dien-15-oic acid, *ent*:	6
64	Halima-1(10)-trans-13-dien-15-oic acid,*ent*:	6
65	Cleroda-3-trans-13-dien-15-oic acid, 16-oxo:	6
66	Cleroda-4(18)-trans-13-dien-15-oic acid, 16-oxo:	4
67	Halima-5(10)- trans-13-dien-15-oic acid, 16-oxo: *ent*:	4
68	Cleroda-3,13-E-dien-16-15-olide	4
69	Cleroda-4(18)-13-dien-16-15-olide	4
70	Halima-5(10)-13-dien-16-15-olide,*ent*:	4
71	Halima-1(10)-13-dien-16-15-olide,*ent*:	4
72	Cleroda-4(18)-cis-13(14)-dien-15-16-olide, 3-b-16-a-dihydroxy:	4
73	Cleroda-cis-13(14)-en-15-16-olide, 4-b-16-a-dihydroxy:	4
74	Cleroda-4(18)-cis-13(14)-dien-15-16-olide, 16-a-hydroxy:	4
75	Cleroda-4(18)-trans-13(14)-dien-15-oic acid	2
76	Cleroda-4(18)-cis-13(14)-dien-15-16-olide	2
77	Cleroda-3-13(14)-dien-15-16-olide, 16-b-hydroxy:	2
78	Kolavenic acid	2
79	Kolava-3-13-cis-dien-16-15-olide, 16-(R)-hydroxy:	2
80	Kolava-3-13-cis-dien-16-15-olide, 16-(R)-acetoxy:	2
81	Labd-8(17)-13(16)-14-trien-18-methyl ester, *ent*:	2
82	Labd-8(17)-14-dien-13-16-diol-18-methyl ester, *ent*:	2
83	Trachyloban-3-b-acetoxy:*ent*:	2
84	Atisan-16-a-ol, *ent*:	2
85	Atisan-16-a-18-diol, *ent*:	10
86	Atisan-16-a-ol-18-oic, *ent*:	10
87	Atisan-4-b-16-a-diol, 18-nor: *ent*:	10
88	Atisan-4-b-hydroperoxid-16-a-ol, 18-nor: *ent*:	10
89	Atisan-4(18)-en-16-a-ol, 19-nor: *ent*:	7
90	Adduct 1a	2
91	Adduct 2a	2
92	Adduct 3a	2
93	Adduct 4a	2
94	Trachyloban-19-oic acid, 15-oxo: (-):	2
95	Trachyloban-19-oic acid, 15-a-hydroxy:*ent*:	3
96	Kaur-16-en-17-al-19-oic acid, (-):	8
97	Acutifloric acid	8
98	Frutoic acid	8
99	Acutifloric acid, 7-b-acetoxy:	6
100	Trachyloban-19-b-oic acid, 7-a-hydroxy: (-):	1
101	Atisan-4-b-16-a-diol, 19-nor: *ent*:	1
102	Atisan-4-a-16-a-diol, 18-nor: *ent*:	1
103	Atisan-16-a-ol, 18-nor: *ent*:	1
104	Atisan-16-a-ol, 19-nor: *ent*:	1
105	Atisan-4(19)-en-16-a-ol, 18-nor: *ent*:	4
106	Kaur-16-en-18-nor: 4-(R)-hydroperoxide:*ent*:	4
107	Kauran-methyl-16S-17-nor-15-oxo-16-(1-pyrazoline)-19-oate,*ent*:	4
108	Kauran-methyl-17-nor-15-oxo-16-cyclopropyl-19-oate,*ent*:	5
109	Labdan-8-13-dien-15-oate-methyl:	5
110	Labdan-8-13-dien	4
111	Kauran-19-oic acid, 15-oxo	4
112	Kaur-16-en-19-oic acid, 15-oxo:*ent*:	4
113	Trachyloban-19-a-oic acid	4
114	Trachyloban-19-oate methyl,*ent*:	4
115	Kaur-16-en-19-oic acid, 7-oxo: *ent*: (-):	4
116	Kolavenic acid, 2-oxo:	6
117	Kaur-16-en-18-oate: methyl	4
118	Labd-8(17)-13(16)-14-trien-18-oic acid, *ent*:	5

Table 2. Basic skeleton types of the diterpenoids isolated from the Annonaceae used in this study. The number in the structures corresponding to biosynthetic order

Type	Structure	Number
Artisane		1
Clerodane		2
Halimane		3
Kaurane		4
Labdane		5
Trachylobane		6
Bis-clerodane-imide		7
Kaurane-labdane Type 1		8
Kaurane-labdane Type 2		9
Labdane-labdane		10

the Toolbox, i.e., the neural network is trained until its convergence to minimal error.

All SOMs were generated with the same topology: for the local lattice structure, the rectangular grid was used, while a sheet was used to indicate the global map shape, using the Gaussian neighborhood function. The literature shows that the determination of the size of the SOMs is an empirical process.[38] Initially, a heuristic formula of $m = 5(n)^{0.5}$ is used for total number of map units, where n is the number of samples. The ratio of side lengths is based on the two biggest eigenvalues of the covariance matrix of the given data. Some different maps sizes were prepared, based on the initial map, generated as described earlier. The SOM toolbox automatically labels the map based on the previously labeled data. The label with most instances is added to the map unit. In the case of a match, the first encountered label is used. A hit is a sample which has the same label as the map unit where it is located. For each map, 10 cross-validations were performed, splitting 10% of the data by dividing into training and test sets, which consisted of approximately 80% (137 samples) and 20% (32 samples) of total samples, respectively (Table 4).

Table 3. Chemical shifts of ^{13}C NMR used as input vector in of the 10 diterpenes

	Number of the substances									
C	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	38.4	42	41.1	42.1	41.8	41.4	41.2	41.5	41.6	42
2	17.3	18.2	19.8	19.1	19	18.7	18.9	19	19	19
3	37.7	42	38.7	38.1	38	42	38	37.9	38	38
4	47.4	33.4	43.9	43.9	43.7	33.2	43.5	43.7	44.1	44.6
5	51.7	56.1	57	56.9	56.8	56.1	56.7	56.7	56.6	56.9
6	22.4	20.5	22.9	22.2	21.5	20	21.5	21.6	22.1	22.4
7	38.3	37.2	42.7	40.7	40.6	38.2	40.5	40.6	40.7	40.8
8	40.8	44.6	44.9	44.7	43.4	43.5	43.6	43.2	43.6	43.7
9	53.2	56.7	56.3	55.8	56	56.9	55.9	55.8	55.3	56.4
10	36.8	39.4	40	39.5	39.4	39.3	39.3	39.3	39.3	39.8
11	19.5	18.3	18.9	18.6	18.9	18.6	18.9	18.9	18.3	18.8
12	20.5	26.3	26.8	26.2	26.5	26.7	26.4	26.8	26	28.9
13	24.2	45.5	45.8	45.4	40.5	52.6	41.7	41.7	37.1	37.1
14	33.4	40.4	37.8	37.3	37.9	40	37.8	37	38.1	40.3
15	50.3	53.4	53.9	53.2	52.3	56.1	52.3	50.4	43.6	44.9
16	22.9	81.6	81.6	81.9	79.8	79.7	78.4	86.9	43.2	42.3
17	20.5	66.2	66.4	66.4	69.7	69.7	71.6	66	64.1	67.4
18	179.4	33.4	29.3	28.7	28.6	33.6	28.6	28.6	31.3	31.3
19	16.5	21.5	180.1	177.9	177.9	21.5	177.9	177.9	177.9	177.9
20	14.9	17.7	15.4	15.4	15.1	17.6	15.4	15.1	18.7	18.7

Table 4. Summary of the ten different training and test sets

	Train set		Test set		Total	
	Train	% Total	Test	% Total	Total	% Total
Annona	48	80.0	12	20.0	60	100
Polyalthia	34	79.1	9	20.9	43	100
Xylopia	55	83.3	11	16.7	66	100
Total	137	81.1	32	18.9	169	100

RESULTS AND DISCUSSION

The results are summarized in Table 5. Training and test sets have overall higher values of matches: 90.9 and 86.3%, respectively. The *Polyalthia* genus showed the highest match values of botanical occurrence for both training (100%) and test sets (98.9%), while *Xylopia* exhibited the lowest match results: 79.8% for training and 70% for test set.

The SOM (Figure 3) show a clear separation among the botanical occurrences of the three genera. Three distinct regions are evidenced: northwest of the map with black squares (*Annona* region), gray

squares (*Xylopia* region) and, light gray squares (*Polyalthia* region). The ^{13}C chemical shifts and SOM are able to clearly distinguish diterpenes from *Annona* (top of map), and from *Polyalthia* (bottom of SOM). Diterpenes of *Xylopia* are situated in the middle of the map, between *Annona* and *Polyalthia* regions, and share some neurons with diterpenes mainly from *Annona* (top of map) and *Polyalthia* (bottom region), thus explaining the poorer match results of this genus.

Table 5. Summary of the training and test match results (%)

	Annona	*Polyalthia*	*Xylopia*	Total
Train set 1	97.9	100.0	80.0	91.2
Train set 2	93.8	100.0	83.6	91.2
Train set 3	93.8	100.0	83.6	91.2
Train set 4	100.0	100.0	81.8	92.7
Train set 5	100.0	100.0	80.0	92.0
Train set 6	100.0	100.0	76.4	90.5
Train set 7	93.8	100.0	83.6	91.2
Train set 8	97.9	100.0	76.4	89.8
Train set 9	100.0	100.0	74.5	89.8
Train set 10	93.8	100.0	78.2	89.1
Average	97.1	100.0	79.8	90.9
Standard deviation	3.0	0.0	3.4	1.1
Test set 1	83.3	100.0	90.9	90.6
Test set 2	100.0	100.0	63.6	87.5
Test set 3	100.0	100.0	72.7	90.6
Test set 4	100.0	88.9	72.7	87.5
Test set 5	83.3	100.0	45.5	75.0
Test set 6	91.7	100.0	63.6	84.4
Test set 7	91.7	100.0	63.6	84.4
Test set 8	83.3	100.0	72.7	84.4
Test set 9	83.3	100.0	81.8	87.5
Test set 10	100.0	100.0	72.7	90.6
Average	91.7	98.9	70.0	86.3
Standard deviation	7.9	3.5	12.2	4.7

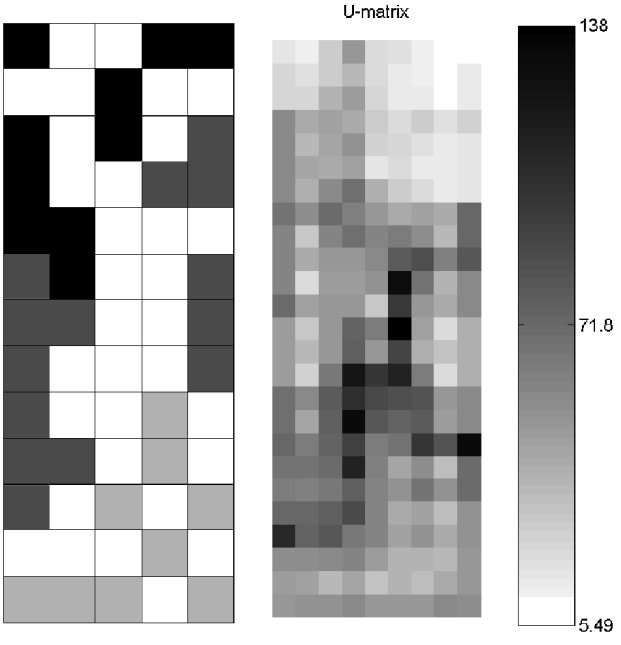


Figure 3. *Kohonen map obtained using 13C NMR data. Annona, Xylopia and Polyalthiaregions in black, gray light gray colors, respectively. U-matrix, on right, that shows distances between neighboring map unit: high values of indicate a cluster border, uniform areas of low values indicate clusters themselves*

Figure 3 shows the U-matrix of the SOM, which visualizes distances between neighboring map units, and thus shows the cluster structure of the map: high values of the U-matrix indicate a cluster border while uniform areas of low values indicate clusters themselves. Dark squares divide the SOM into two main parts: northwest (*Annona* region) and southeast (*Polyalthia* region). Analysing Table 5, Figures 3 and 4 (mainly), it is evident that the chemistry of *Xylopia* is more similar to *Annona* genus regarding the diterpenes structures.

A A(16) X(9)			A A(12) X(1)	A A(5)
		A A(1)		
A A(6)		A A(7)		X X(4) A(3)
A A(3)			X X(6)	X X(3)
A A(4)	A A(1)			
X X(1)	A A(2) X(2)			X X(6)
X X(2)	X X(3)			X X(2)
X X(5)				X X(11)
X X(2)			Po Po(1)	
X X(3)	X X(2)		Po Po(5)	
X X(3)		Po Po(3)		Po Po(5)
			Po Po(8)	
Po Po(9)	Po Po(1)	Po Po(6)		Po Po(5) X(1)

Figure 4. *SOM with frequency of diterpenes and respective botanical occurrence*

These results corroborate previous phylogenetic studies, in which these two genera (*Annona* and *Xylopia)* were more closely related and *Polyalthia* farther from these in cladograms using mainly molecular and morphological data. Thus, for these three genera, the chemistry regarding the diterpenes is closely related to the molecular and morphological data used in the previous studies on phylogeny. The carbons responsible for the division in botanical occurrence of the diterpenes of the genera used in this study can be seen in Figure 5, which shows the weight that each descriptor has in this Kohonen map. Generally, the most representative descriptors are those that have two main characteristics: the greatest weights in the genus predominant region; a considerable difference between the highest and lowest descriptors' (^{13}C NMR chemical shifts) values.

The following carbons have highest ^{13}C chemical shift values for the *Polyalthia* region: 2, 3, 4, 6, 10, 11, 13, 14 and 20. For example, compound **57** (Figure 6) polyalthialdoic acid, a clerodane which has a double bond between carbons 13 and 14. On the other hand,

Annona diterpene carbons have higher chemical shift values mainly for atoms 1, 7, 9, 17 and 19, for example Annoglabasin A and B, compounds **26** and **27** respectively (Figure 6), two diterpene kauranes that have ester and carboxyl groups on carbon 17, plus aldehyde and acetoxyl groups on carbon 19, respectively. These higher chemical shifts are less representative for diterpenes from the *Xylopia* genus but remain significant. Some diterpenes from this genus have higher chemical shift values for carbons 15 and 18, such as those found in intrachyloban-19-oic acid.

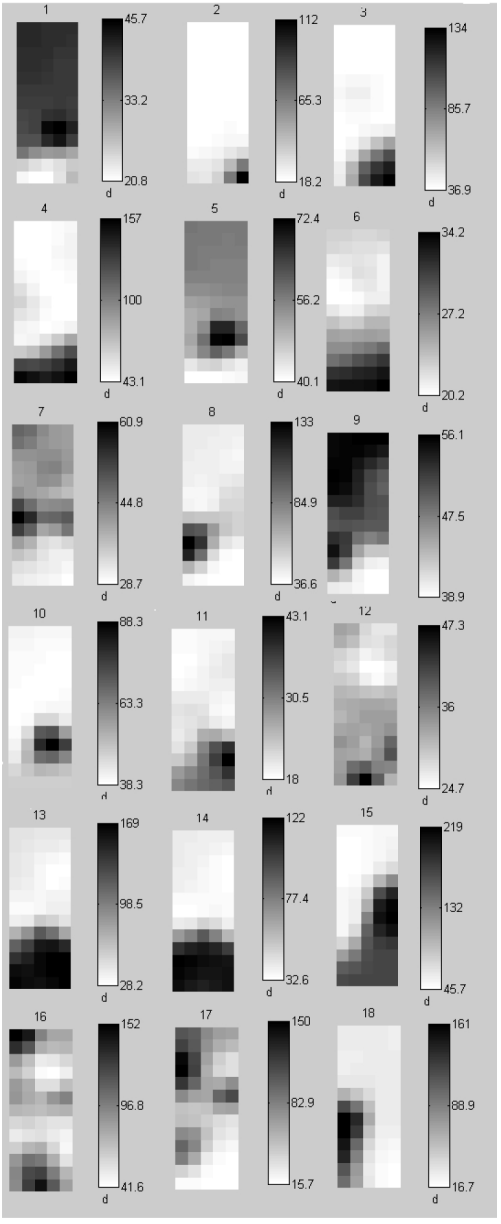


Figure 5. *The weight maps of the Kohonen map obtained of ^{13}C NMR shifts*

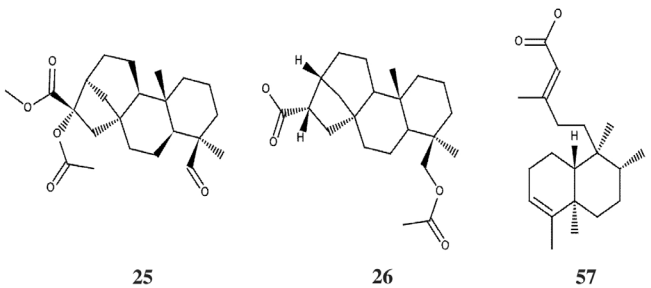


Figure 6. *Structures of Annoglabasin A (**25**) and B (**26**) and polyalthialdoic acid (**57**)*

CONCLUSIONS

SOMs and ^{13}C NMR data of skeletal carbons of diterpenes produce clusters according to their botanical occurrence. Moreover, their similarities corroborate previous studies using morphological and molecular data and can predict the botanical occurrence of similar compounds. Some ^{13}C chemical shift values are specific for skeletal carbons of diterpenes from these three genera of Annonaceae. Therefore, ^{13}C NMR can be used as a molecular descriptor, including in QSAR studies, enabling the methodology to find new structures with potential biological activities.

ACKNOWLEDGEMENTS

The authors would like to thank the Brazilian National Research Council (CNPq) for financial support.

REFERENCES

1. Chatrou, L. W.; Rainer, H.; Maas, P. J. M. In *Flowering Plants of Neotropics;* Smith, N.; Mori, S. A.; Henderson, A.; Stevenson D. W.; Heald, S. V., eds.; Botanical Garden: New York, 2004, p. 18-20.
2. Brummitt, R. K.; Powel, C. E.; *Authors of Plant Names,* Royal Botanic Gardens: Kew, 1988.
3. Silva, J. B.; Rocha, A. B.; *Rev. Cienc. Farm.* **1981***, 3*, 33.
4. Yiadom, B. K.; Fiagbe, N. I. Y.; Ayim, J. S. K.; *J. Nat. Prod.* **1977**, *40*, 543.
5. Alfonso, D.; Colman, S. T.; Zhao, G. X.; Shi, G.; Ye, Q.; Schwedler, J. T.; Mclaughlin, J. L.; *Tetrahedron* **1996**, *52,* 4215.
6. Vilegas, W.; Felício, J. D.; Roque, N. F.; Gottlieb, H. E.; *Phytochemistry* **1991**, *30*, 1869.
7. Santos, A. S.; Andrade, E. H. A.; Zoghbi, M. G. B.; Maia, J. G. S.; *Flavour Frag. J.* **1998**, *13*, 148.
8. Costa, E. V.; Pinheiro, M. L. B.; Xavier, C. M.; Silva, J. R. A.; Amaral, A. C. F.; Souza, A. D. L.; Barison, A.; Campos, F. R.; Ferreira, A. G.; Machado, G. M. C.; Leon, L. L. P.*; J. Nat. Prod.* **2006**, *69*, 292.
9. Joly, A. B.; *Botânica*: *Introdução à taxonomia vegetal,* 13ª ed., Companhia Editora Nacional: São Paulo, 2002.
10. Leboeuf, M.; Cavé, A.; Bhaumik, P. K.; Mukherjee, B.; Mukherjee, R.; *Phytochemistry* **1982**, *21*, 2783.
11. Manica, I.; Icuma, I. M.; Junqueira, K. P.; Oliveira, M. A. S.; Cunha, M. M.; Oliveira Jr., M. E.; Junqueira, N. T. V.; Alves, R. T.; *Frutas Anonáceas: Ata ou Pinha, Atemoia, Cherimólia e Graviola - Tecnologia de Produção, Pós-colheita, Mercado,* Cinco Continentes: Porto Alegre, 2003.
12. Pontes, A. F.; Barbosa, M. R. V.; Maas, P. J. M.; *Acta Bot. Bras.* **2004**, *18*, 281.
13. Chen, C. Y.; Chang, F. R.; Shih, Y. C.; Hsieh, T. J.; Chia, Y. C.; Tseng, H. Y.; Chen, H. C.; Chen, S. J.; Hsu, M. C.; Wu, Y. C.; *J. Nat. Prod*. **2000**, *63*, 1475.
14. Ravikumar, Y. S.; Mahadevan, K. M.; Kumaraswamy, M. N.; Vaidya, V. P.; Manjunatha, H.; Kumar, V.; Satyanarayana, N. D.; *Environ. Toxicol. Pharmacol.* **2008**, *26*, 142.
15. Verma, M.; Singh, S. K.; Bhushan, S.; Sharma, V. K.; Datt, P.; Kapahi, B. K.; Saxena, A. K.; *Chem. Biol. Interact.* **2008**, *171*, 45.
16. Johnson, D. M.; *Syst. Bot.* **2003**, *28*, 503.
17. Sauquet, H.; Doyle, J. A.; Scharaschkin, T.; Borsch, T.; Hilu, K. W.; Chatrou, L. W.; Le Thomas A.; *Bot. J. Linn. Soc***. 2003**, *142*, 125.
18. Scharaschkin, T.; Doyle, J. A.; *Syst. Bot.* **2005**, *30,* 712.

19. Doyle, J. A.; Bygrave, P.; Le Thomas, A. In *Pollen and spores: morphology and biology*; Harley, M. M.; Morton, C. M.; Blackmore, S., eds.; Royal Botanic Gardens: Kew, 2000.
20. Emerenciano, V. P.; Kaplan, M. A.; Gottlieb, O. R.; *Biochem. Syst. Ecol*. **1985**, *13*, 145.
21. Queiroz, E. F.; Wolfender, J. L.; Hostettmann, K.; *Curr. Drug Targets* **2009**, *10*, 202.
22. Gottlieb, O. R.; *Phytochemistry* **1989**, *28*, 2545.
23. Emerenciano, V. P.; Ferreira, Z. S.; Kaplan, M. A.; Gottlieb, O. R.; *Phytochemistry* **1987**, *26*, 3103.
24. Da Silva, M. F. G. F.; Gottlieb, O. R.; *Biochem. Syst. Ecol*. **1987**, *15*, 85.
25. Kaplan, M. A.; Gottlieb, O. R.; *Biochem. Syst. Ecol*. **1982**, *10*, 329.
26. Emerenciano, V. P.; Kaplan, M. A.; Gottlieb, O. R.; Bonfanti, M. R. D. M.; Ferreira, Z. S.; Comegno, L. M. A.; *Biochem. Syst. Ecol*. **1986**, *14*, 585.
27. Da Silva, M. F. G. F.; Gottlieb, O. R.; Dreyer, D. L.; *Biochem. Syst. Ecol*. **1984**, *12*, 299.
28. Bolzani, V. S.; Da Silva, M. F. G. F.; Rocha, A. I.; Gottlieb, O. R.; Dreyer, D. L.; *Biochem. Syst. Ecol*. **1984**, *12*, 159.
29. Emerenciano, V. P.; Militão, J. S. L. T.; Campos, C. C.; Romoff, P.; Kaplan, M. A.; Zambon, M.; Brant, A. J. C.; *Biochem. Syst. Ecol*. **2001**, *29*, 947.
30. Alvarenga, S. A. V.; Gastmans, J. P.; Rodrigues, G. V.; Moreno, P. R. H.; Emerenciano, V. P.; *Phytochemistry* **2001**, *56*, 583.
31. Da Costa, F. B.; Terfloth, L.; Gasteiger, J.; *Phytochemistry* **2005**, *66*, 345.
32. Hristozov, D.; Da Costa, F. B.; Gasteiger, J.; *J. Chem. Inf. Model.* **2007**, *47*, 9.
33. Calabria, L. M.; Emerenciano, V. P.; Ferreira, M. J. P.; Scotti, M. T.; Mabry, T. J.; *Nat. Prod. Commun.* **2007**, *2*, 277.
34. Emerenciano, V. P.; Cabrol-Bass, D.; Ferreira, M. J. P.; Alvarenga, S. A.; Brant, A. J. C.; Scotti, M. T.; Barbosa, K. O.; *Nat. Prod. Commun.* **2006**, *1*, 495.
35. Scotti, M. T.; Emerenciano, V. P.; Ferreira, M. J. P.; Scotti, L.; Stefani, R.; Da Silva, M. S.; Mendonça, F. J. B.; *Molecules* **2012**, *17*, 4684.
36. Zupan, E.; Gasteiger, J.; *Neural Networks for Chemists- An Introduction*, *VCH Verlagsgesellschaft*: Weinheim, 1993.
37. Kohonen, T.; *Biol. Cybern.* **1982**, *43*, 59.
38. Kohonen, T.; *Self-Organizing Maps*, 1st ed., Springer: Berlin, 2001.
39. Cleva, C.; Cachet, C.; Cabrol-Bass, D.; *Analusi*s **1999**, *27*, 81.
40. Eghbaldar, A.; Forrest, T. P.; Cabrol-Bass, D.; *Anal. Chim. Acta* **1998**, *359*, 283.
41. Binev, Y.; Aires, S. J.; *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 940.
42. Ivanciuc, O.; Rabine, J. P.; Cabrol-Bass, D.; Panaye, A.; Doucet, J. P.; *J. Chem. Inf. Model.* **1997**, *37*, 587.
43. Meiler, J.; Maier. W.; Meusinger, R.; *J. Magn. Reson.* **2002**, *157*, 242.
44. Fernandes, M. B.; Scotti, M. T.; Ferreira, M. J. P.; Emerenciano, V. P.; *Eur. J. Med. Chem.* **2008**, *43*, 2197.
45. Fraser, L. A.; Mulholland, D. A.; Fraser, D. D.; *Phytochem. Anal.* **1997**, *8,* 301.
46. Emerenciano, V. P.; Alvarenga, S. A.; Scotti, M. T.; Ferreira, M. J.; Stefani, R.; Nuzillard, J. M.; *Anal. Chim. Acta* **2006**, *579*, 217.
47. Da Costa, F. B.; Scotti, M. T. In *Revisões em Processos e Técnicas Avançadas de Isolamento e Determinação Estrutural de Ativos de Plantas;* De Souza, G. H. B.; De Mello, J. C. P.; Lopes, N. P., eds.; Ed. UFOP: Ouro Preto, 2012, p. 73-118.
48. Andrade, N. C.; Cunha, E. V. L.; Silva, M. S.; Agra, M. F.; Barbosa-Filho, J. M. In *Recent Research Developments in Phytochemistry*; Pandalai, S. G., ed.; Research Signpost: Kerala, 2003.
49. Vesanto, J.; Himberg, J.; Alhoniemi, E.; Parhankangas, J.; *SOM Toolbox for Matlab,* available at http://www.cis.hut.fi/projects/somtoolbox, accessed May 2011.