Journal of Medicinal Plants Research Vol. 6(7), pp. 1207-1218, 23 February, 2012
Available online at http://www.academicjournals.org/JMPR
DOI: 10.5897/JMPR11.942
ISSN 1996-0875 ©2012 Academic Journals

Full Length Research Paper

Self-organizing maps as a good tool for classification of subfamily *Astereoideae*

Mauro Vicentini Correia[1], Harold Hilarion Fokoue[1], Marcelo José Pena Ferreira[1], Luciana Scotti[5]*, Marcus Tullius Scotti[4], Sandra Aparecida Vestri Alvarenga[2], Gilberto do Vale Rodrigues[3] and Vicente de Paulo Emerenciano[2]

[1]Instituto de Química, Universidade de São Paulo, Caixa Postal 26077, São Paulo, SP, Brazil, 05513-970.
[2]Faculdade de Engenharia de Guaratinguetá, UNESP, Guaratinguetá, SP, Brazil, 12516-410.
[3]Departamento de Química, Instituto de Ciências Exatas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil.
[4]Center of Applied Sciences and Education, Federal University of Paraiba, Campus IV, Rio Tinto 58297-000, Brazil.
[5]LTF Laboratory of Pharmaceutical Technology, Federal University of Paraíba, Brazil.

Accepted 22 November, 2011

Artificial neural network (ANN) is defined as computational models with structures derived from the simplified concept of the brain in which a number of nodes are interconnected in a network-like structure. The most used ANNs architecture for pattern recognition and classification is the self-organizing map (SOM). SOM is a powerful visualization tool as it is able to reduce dimensions of projections and displays similarities among objects and was successfully used in several applications with chemistry database. In this work, we used SOM as good methodology of classification of a database containing various types of compounds from the *Asteroideae* subfamily (*Asteraceae*). The Kohonen neural network was trained using Matlab version 6.5 with the package Somtoolbox 2.0. Some chemical evolutionary descriptors and the numbers of occurrences of 12 chemical classes in different taxa of the subfamily were used as variables. The study shows that SOM applied to chemical data can contribute to differentiate genera, tribes, and branches of subfamily, as well as to tribal and subfamily classifications of *Asteroideae,* exhibiting a high hit percentage comparable to that of an expert performance, and in agreement with the previous tribe classification proposed by Funk.

Key words: *Asteraceae, Asteroideae*, self-organizing maps, secondary metabolites.

INTRODUCTION

There are several classifications of plants at higher hierarchical levels based only on botanical data (Takhtajan, 1973; Dahlgren, 1980; Cronquist, 1988; Bremer and Anderberg, 1994), but recently new classification was performed using molecular and morphological data (Funk et al., 2005, 2009). However, several authors have discussed the classification of plants, mainly angiosperms, using chemical data of the species (Gershenzon and Mabry, 1983; Waterman and Gray, 1987). The methods for chemical classifications include generally either phenetics (Sneath and Sokal, 1973) or phyletics using either macromolecules (Jansen et al., 1990) or secondary metabolites (Calabria et al., 2007). Generally, phonetic methods employ chemical

*Corresponding author. E-mail: luciana.scotti@gmail.com.

Abbreviations: ANN, Artificial neural network; **SOM,** self-organizing map; **SOM Kohonen map,** Kohonen self–organizing feature map; **O,** oxidation state; **S,** skeletal specialization; **OS,** oxidation step (**OS** for ring **A** and as **OSB** for ring **B**).

characters involving “chemical indexes” computed specifically for each compound class (Richardson and Young, 1982; Emerenciano et al., 2001, 2006). At lower hierarchical levels, the classification is more difficult and several misunderstandings between authors are frequently published for the large families of angiosperms, such as the *Asteraceae* (Bentham, 1873; Carlquist, 1976; Wagenitz, 1976; Cronquist, 1977; Jansen et al., 1990; Bremer, 1996; Bayer and Starr, 1998; Funk et al., 2005, 2009). At subfamily and tribal level, the problems of classification in *Asteraceae* become more complex as, for example, in the tribe *Heliantheae* of the subfamily *Asteroideae,* where the genera were grouped differently into subtribes by four authors (Stuessy, 1977; Robinson, 1981; Karis, 1993; Baldwin et al., 2002). The subfamily *Asteroideae* is well characterized by morphological and molecular characters (Bremer and Anderberg, 1994) and nearly worldwide distributed. There are comprised of ca. 15, 500 species (over 60% of the species in the family) placed in ca. 1229 genera (over 70%) and 20 tribes (approximately 60%), and is the largest subfamily of Asteraceae (Pelser and Watson, 2009). Also it has a nearly worldwide distribution. An area optimization analysis on a supertree (=metatree) of the family suggests a sub-Saharan origin of *Asteroideae* (Funk et al., 2005).

In the middle of the last century, artificial neural networks (ANN) were introduced as a scientific tool, and are now routinely used by industry and universities. Various software are now available (Massart et al., 1988; Statsoft, 2000; Mathworks, 2004), and the book of Zupan and Gasteiger (Zupan and Gasteiger, 1993) is the principal text in the area for chemists. The self-organizing maps (SOM) are the most used ANN architecture for pattern recognition and classification. This procedure can map multivariate data onto a two dimensional grid, grouping similar patterns near each other (Kohonen, 1982, 2001; Zupan and Gasteiger, 1993).

Botanists and chemists work at the classification of plants at several hierarchical levels. If the classification of a taxon is natural, in the botanical sense of the term, the prediction in related species that have not been studied of the presence or absence of certain chemical compounds can be surmised (Taketo et al., 2010). The aim of this study is to compare the chemical data with the Funk classification (Funk et al., 2009) for the *Asteroideae* subfamily, using traditional methods and artificial neural networks (ANN).

MATERIALS AND METHODS

Our database at São Paulo University now contains about 36.000 occurrences (a compound X can appear n times within a delimited taxon, such as genus, subtribe, tribe, subfamily or family. We define number of occurrences for a superior taxon, counting how many times a compound appears in determined species belonging to that taxon) of compounds isolated from *Asteraceae,* totaling approximately 10.000 compounds divided into 12 classes (Figure 1) containing 548 skeletal types. The twelve classes of compound and their codes were presented in Table 1. Also, each type of different skeleton has received a code according to their class. The skeletal type is obtain by removing all functional groups of secondary metabolites (e.g terpenoids), or partially, maintaining a common core (e.g flavonoids). The Figure 2 shows some examples of flavonoids and sesquiterpenes classes, with their skeletons.

The database arose through an extensive inspection of chemical abstracts from 1960 to 2008. The data were exported to EXCEL from the chemical database obtained from our own system (SISTEMAT) (Gastmans et al., 1990). For each compound present in the database were calculated the skeletal specialization (S) and the oxidation state (O) (Gottlieb, 1989). The oxidation step (OS) of a compound was calculated from his oxidation state (O) (Emerenciano et al., 1998). For Flavonoids, because of their mixed biogenetic origin, two values of the oxidation state were calculated, one for the ring A and other for the ring B, and from them were calculated their oxidation step values (as OS for ring A and as OSB for ring B) (Emerenciano et al., 1998).

According to the Gottlieb (1989), the oxidation state (O) of a compound (per Carbon) is determined by counting, for each carbon of the compound, -1 for each bond to H and +1 for each bond to a Heteroatom, these counts are added and divided by number of C-atom of the compound, and the skeletal specialization (S) of compound with respect to the general skeletal precursor of its biosynthetic class is determined by counting the number of bonds (to C) broken and the numbers of bonds (to C, or to heteroatom if this involves formation of a new cycle) formed for each carbon of the compound; the total counts obtained are then divided by the numbers of C-atoms in the compound. Emerenciano (1998) defined oxidation step as the subtraction of the oxidation state of a compound with biogenetic compound consider as the precursor dividing the result by two.

From the overall database, we extracted all data concerning the occurrence of chemical classes for species of *Asteroideae* (around 12.000 occurrences).

For this study in all data sets, one sample corresponds to one occurrence and the S, OS, OSB, Class and Skeletal Type are independent variables. The determined taxon is considered as the dependent variable. The Table 2 shows the abbreviation of the studied tribes. The analyses were performed using the software MATLAB 6.5 (Mathworks, 2004) and the package SOM Toolbox 2.0 for Matlab (SOM Toolbox 2.0 for Matlab, 2011).

The training was conducted through the Batch–training algorithm. In this, 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. After this, the weights are calculated, as stated by Equation (1):

$$m_i(t+1) = \frac{\sum_{j=1}^{n} h_{ic}(t) X_j}{\delta \sum_{j=1}^{n} h_{ic}(t)} \quad (1)$$

Where xj is data set partitioned according to the Voronoi regions of the map weight vector. The new weight vector is the average of the data samples, where the weight of each data sample is the

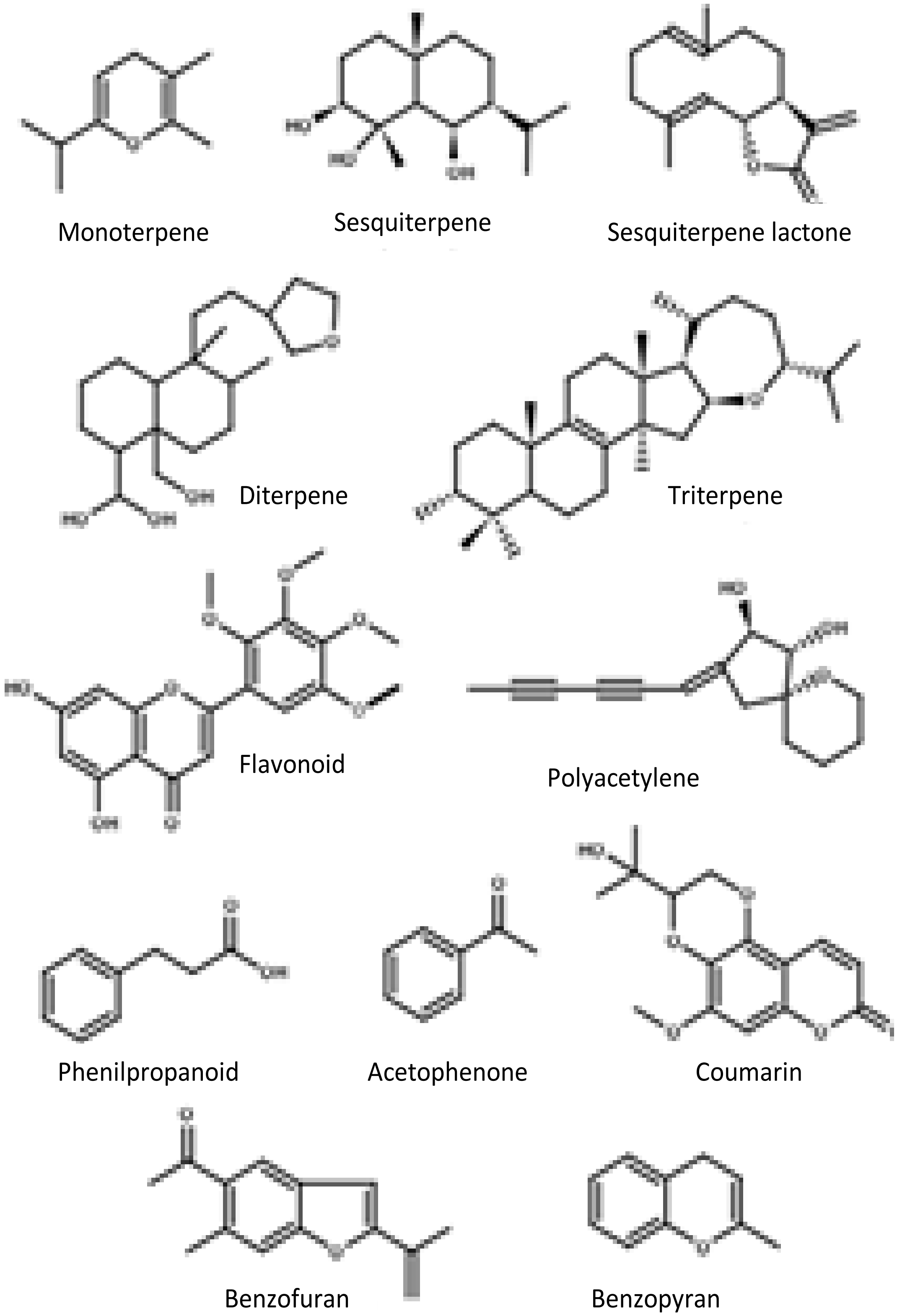

Figure 1. Examples of the twelve classes of compounds studied.

Table 1. The twelve classes of compounds and their respective codes.

Class	Code
Monoterpene	1
Sesquiterpene	3
Sesquiterpene Lactone	4
Diterpene	5
Triterpene	7
Coumarin	8
Flavonoid	9
Polyacetylene	10
Benzofuran	11
Benzopyran	12
Acetophenone	13
Phenilpropanoid	16

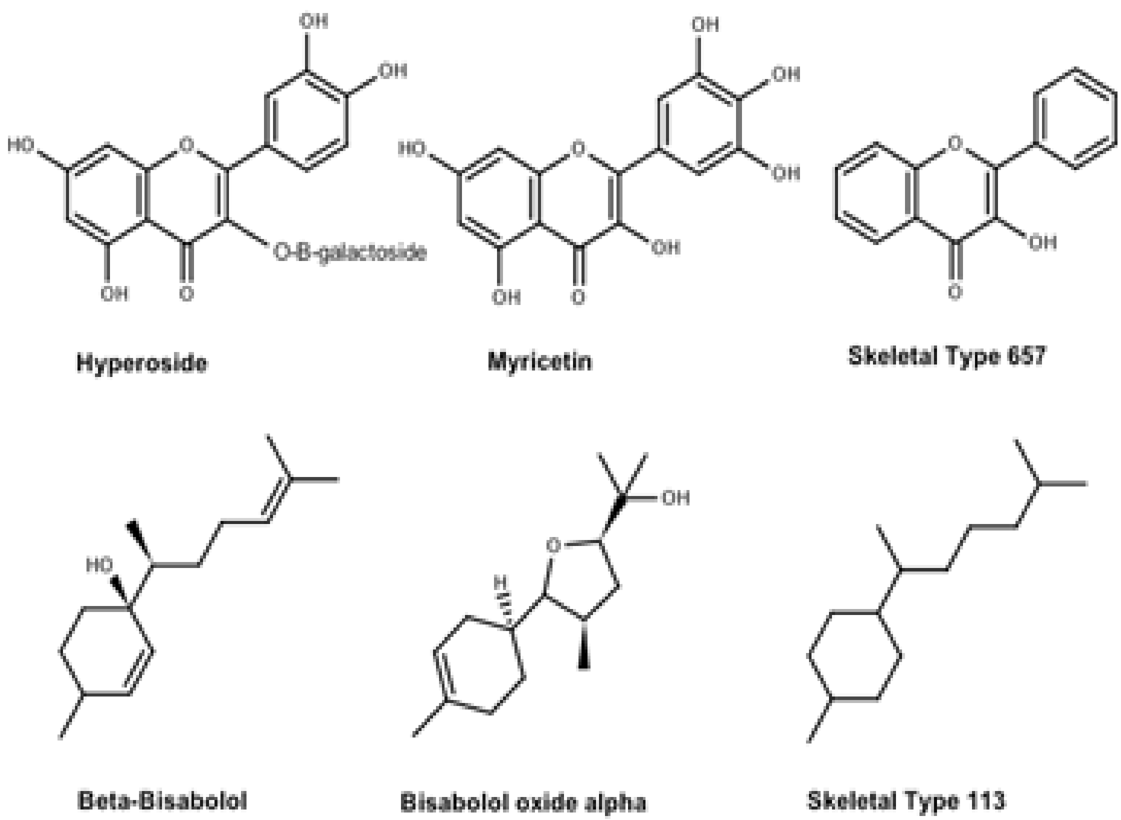

Figure 2. Examples of the skeletons and compounds that shared its.

neighborhood function value $h_{ic}(t)$ at its best match unit. Within this algorithm, the new weight vector is simple averages and there is no learning rate (Kohonen, 2001). This feature allows missing values to be ignored by the net. The number of epochs is automatically chosen by the Toolbox, that is, 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 sheet was used to indicate the global map shape,

Table 2. Abbreviation of the tribes studied.

Tribes	Abbreviation
Anthemideae	ANT
Astereae	AST
Bahieae	BAH
Calenduleae	CAL
Eupatorieae	EUP
Gnaphalieae	GNA
Heliantheae	HLT
Madieae	MAD
Neurolaeneae	NEU
Tageteae	TAG

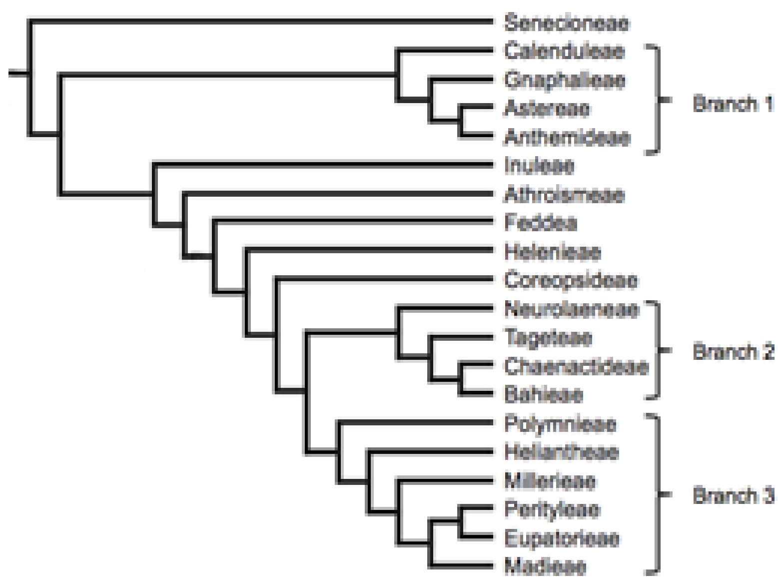

Figure 3. Summary diagram of the phylogenetic of the subfamily *Asteroideae* according to Funk et al. (2005).

using Gaussian neighborhood function. The literature shows that the determination of the size of the SOMs is an empiric process (Kohonen, 2001). 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 before.

SOM toolbox automatically labels the map based on the already labeled data. The label with most instances is added to the map unit. In the case of a draw, the first encountered label is used. A hit is a sample which has the same label as the map unit where it is located. Each map has been performed 10 cross-validation, splitting 10% of the data.

RESULTS

The initial test developed with the self-organizing maps tries to divide the three branches of the *Asteroideae* subfamily proposed by Funk et al. (2009). In the first experiment from a sample of the dataset (in each branches was taken three tribes and one gender of each tribe), we tried to divide branches 1 (820 occurrences), 2 (846 occurrences) and 3 (876 occurrences) according to Figure 3. The Kohonen map (Figure 4) shows the separation between the three branches of tribes. The performance of the set, including the cross validation

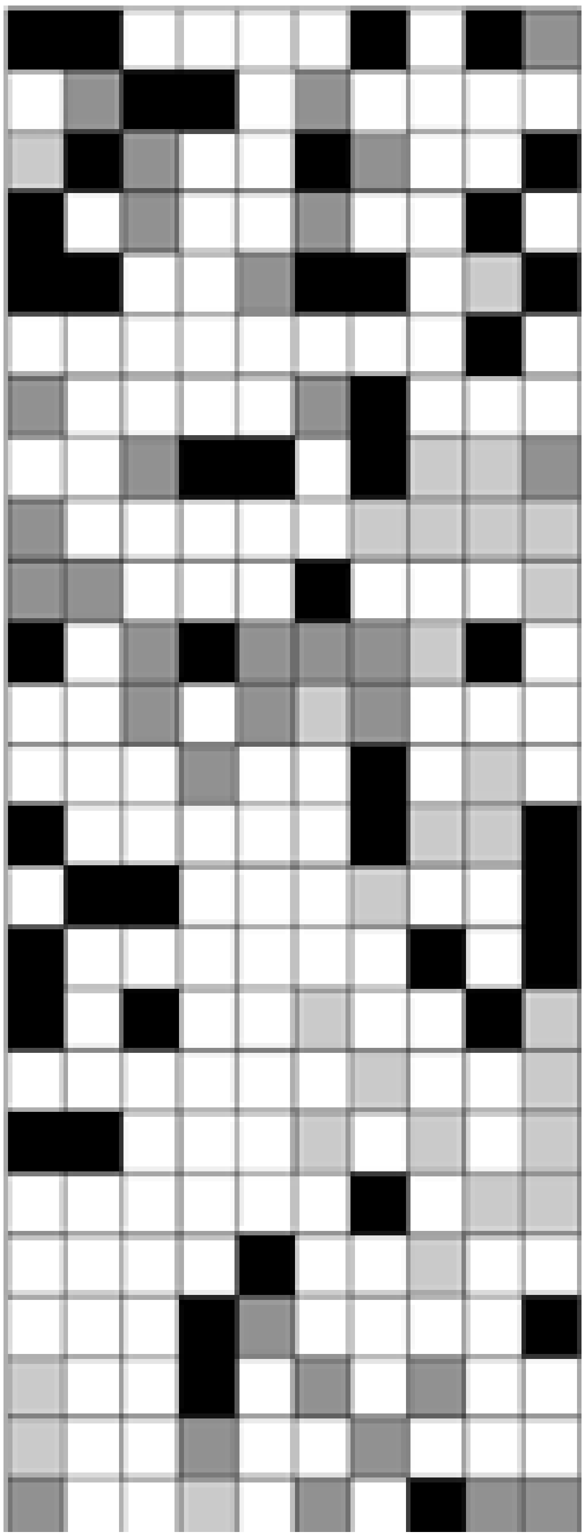

Figure 4. Self organizing map obtain for the classification of the branches of subfamily *Asteroideae* : Branch 1 (black), Branch 2 (dark gray), Branch 3 (light gray).

results, is presented in Table 3. In the second experiment, we tried to classify four tribes coming from branch 1 (in this branch was used 3 genera for each tribe). The tribes selected for this approach were the *Anthemideae* (427 occurrence), *Astereae* (375 occurrence), *Calenduleae* (158 occurrence) and *Gnaphalieae* (349 occurrence). The results are shown in Figure 5 and Table 4, from the obtained results, above 65% of percentage hit was observed for the separation of the tribes pertaining to the same branch.

The same procedure was applied to branches 2 and 3. In the branch 2 three tribes was selected for the experiment (in this case all the genera was used), tribes *Bahieae* (219 occurrence), *Neurolaeneae* (399 occurrence) and *Tageteae* (542 occurrence), the separation map is shown in Figure 6 and the performance in Table 5. In the branch 3, the same way than in the second experiment, three tribe was used and for each tribe three genera except for the tribe MAD that 2 genera was used, the selected tribes are *Eupatorieae* (905 occurrence), *Heliantheae* (705 occurrence) and *Madieae* (689 occurrence). The results are given in Table 6 and the Kohonen map is show in Figure 7. The separation of the three tribes was obtained with good hit percentages, as well as the regions for the tribes in the map being well characterized and distinguished.

In the fifth and sixth experiment, we tried to separate three genera classified in the same tribe at the lowest hierarchical level employed in this study. For the experiment fifth the genera *Ageratina* (258 occurrences), *Ageratum* (261 occurrences) and *Stevia* (324 occurrences), from the tribe *Eupatorieae*, were chosen for this approach as they were considered good for his experiment due to the large number of the occurrence in the database. The summary of results is displayed in Table 7 and the Kohonen map is given in Figure 8. From the results, the high hit degree can be observed for the genera separation pertaining to the same tribe. This fact could be proven with the three distinct regions observed in the Kohonen map. The genera *Erigeron* (219 occurrences), *Gutierrezia* (259 occurrences) and *Solidago* (339 occurrences) from the tribe *Anthemideae* were chosen for the sixth experiment, the results were similar to the fifth experiment. The summary of results is displayed in Table 8 and Figure 9.

DISCUSSION

Classification using chemical data most often encounters several problems such as the incomplete and/or inconsistent chemical reports; and also many species are not yet study; usually the studies are focused in finding

Table 3. Result obtain from the first experiment.

Branches	Tribes	Genera	Data Set 1	% Total	Hits	% Hits	% CV
	ANT	*CHRYSANTHEMUM*					
Branch 1	AST	*ASTER*	820	32.2	599	73.0	70
	CAL	*OSTEOSPERMUM*					
	BAH	*SCHKURIA*					
Branch 2	NEU	*CALEA*	846	33.3	651	79.4	81
	TAG	*TAGETES*					
	HLT	*HELIANTHUS*					
Branch 3	EUP	*STEVIA*	876	34.5	639	77.9	77.1
	MAD	*MADIA*					
Total			2542	100	1889	74.3	76.1

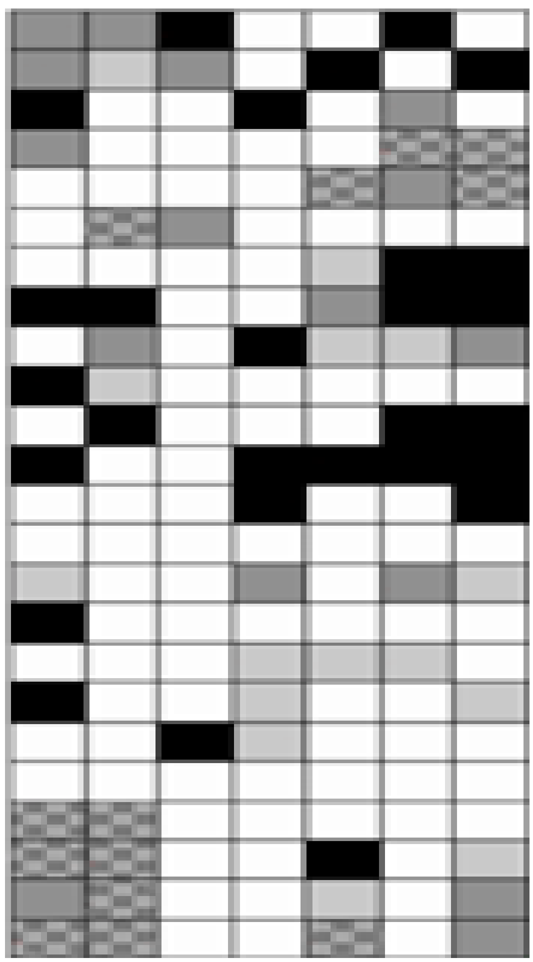

Figure 5. Self organizing map obtain for the classification of the tribes from branch 1: ANT (black), AST (dark gray), CAL (light gray), GNA (tile).

Table 4. Result obtain from the second experiment.

Tribe	Genera	Data Set	% Total	Hits	% Hits	% CV
ANT	*ANTHEMIS* *MATRICARIA* *URSINIA*	427	32.6	299	70.0	71.1
AST	*BELLIS* *CHRYSOTHAMNUS* *HETEROTHECA*	375	28.6	244	65.07	68.9
CAL	*CALENDULA* *GARULEUM* *OSTEOSPERMUM*	158	12.1	111	70.3	71.1
GNA	*ACHYROCLINE* *CASSINIA* *RAOULIA*	349	26.7	272	77.9	70.1
Total		1309	100	926	70.7	70.2

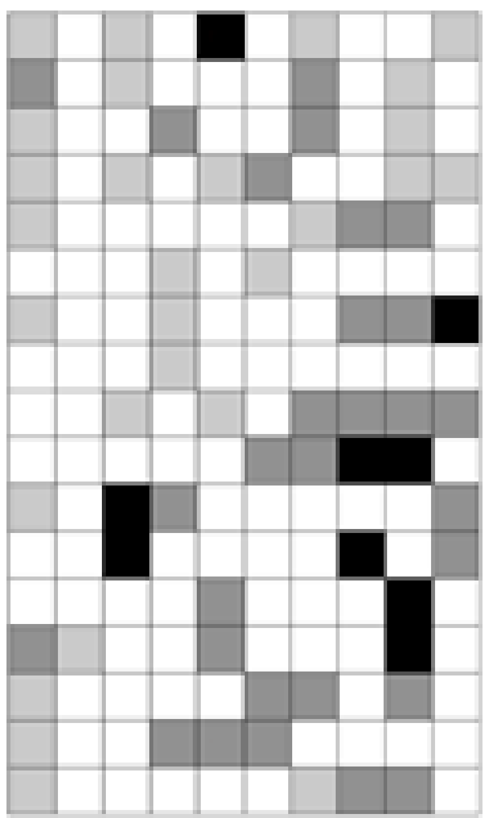

Figure 6. Self organizing map obtain for the classification of the tribes from branch2: BAH (black), NEU(dark gray), TAG (light gray).

Table 5. Results obtain from the third experiment.

Tribe	Data set	% Total	Hits	% Hits	% CV
BAH	219	18.9	151	68.9	65.8
NEU	399	34.4	347	87	81.1
TAG	542	46.7	472	87.1	91
Total	1160	100.0	970	83.6	82.8

Table 6. Results obtain from the fourth experiment.

Tribe	Genera	Data set	% Total	Hits	% Hits	% CV
EUP	*AGERATINA* (258) *BRICKELLIA* (191) *MIKANIA* (456)	905	39.4	688	76	75.1
HLT	*AMBROSIA* (348) *PARTHENIUM* (226) *XANTHIUM* (131)	705	30.7	522	74	76
MAD	*ARNICA* (443) *HEMIZONIA* (246)	689	30.0	558	80.9	81
Total		2299	100	1768	76.9	77.1

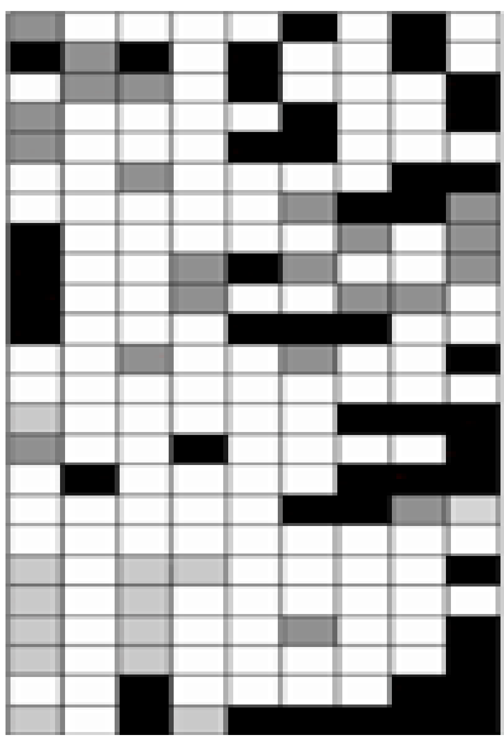

Figure 7. Self organizing map obtain for the classification of the tribes from branch 3: EUP (black), HLT (dark gray), MAD (light gray).

Table 7. Results obtained in the experiment developed with genera of tribe *Eupatorieae.*

Genera	Data set	% Total	Hits	% Hits	% CV
Ageratina	258	30.6	186	72.1	77.2
Ageratum	261	31.0	217	83.1	77
Stevia	324	38.4	298	91.9	93.2
Total	843	100	701	83.2	83.3

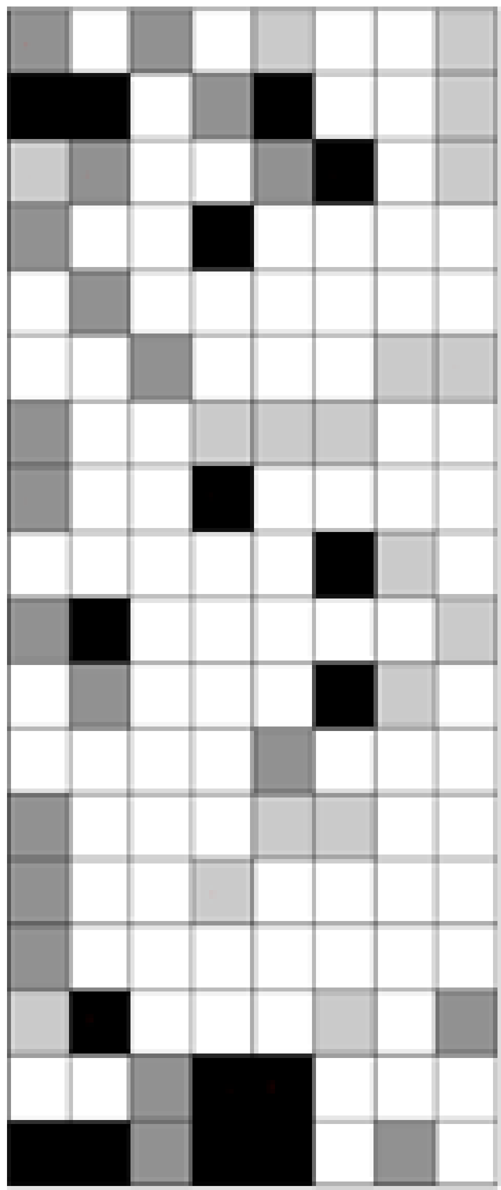

Figure 8. Self organizing map obtain for the classification of the genera from the *Eupatorieae* tribe: *Ageratina* (black), *Ageratum* (dark gray), *Stevia* (light gray).

Table 8. Results obtained in the experiment developed with genera of tribe *Anthemideae*.

Genera	Data set	% Total	Hits	% Hits	% CV
Erigeron	219	26.8	188	85.8	82.2
Gutierrezia	259	31.7	228	86.1	88.8
Solidago	339	41.5	298	77.8	82
Total	817	100	701	85.8	84.2

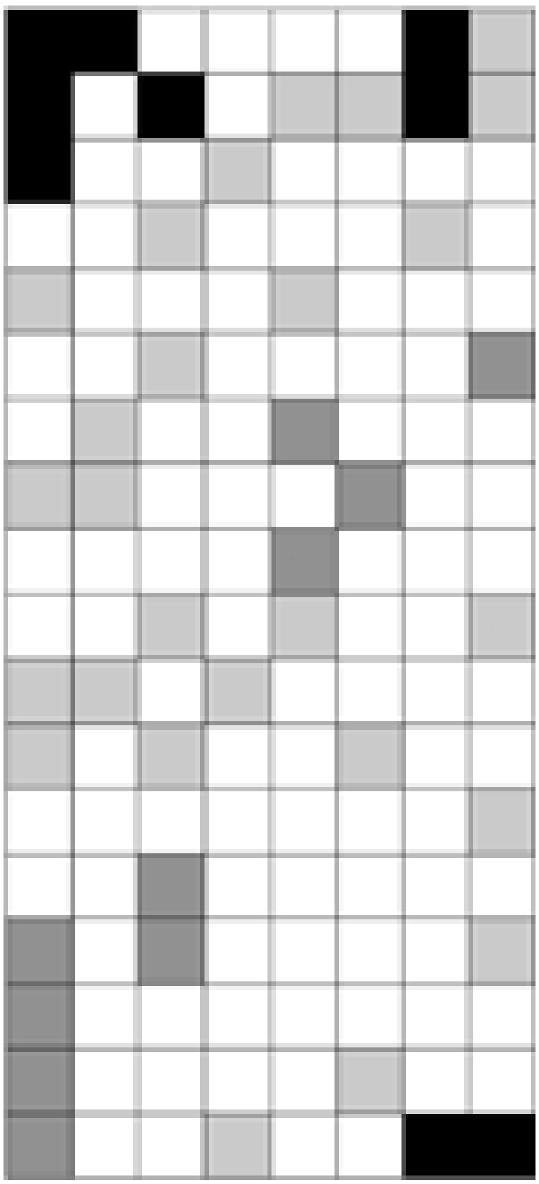

Figure 9. Self organizing map obtain for the classification of the genera from the *Anthemideae* tribe: *Erigeron* (black), *Gutierrezia* (dark gray), *Solidago* (light gray).

new compounds, not reporting known compounds. However, in this work it was possible to classify the *Asteroideae* subfamily using the chemical taxonomic marker. The three branches described in the Figure 3 were separated with an overall match of approximately 74%. It was also possible to separate satisfactorily four tribes in the branch 1 (*Anthemideae, Astereae, Calenduleae* and *Gnaphalieae*), three tribes in the branch 2 (*Bahieae, Neurolaeneae* and *Tageteae*), three tribes in the branch 3 (*Eupatorieae, Heliantheae* and *Madieae*); to separate three genera from tribe *Eupatorieae,* and other three genera from the tribe *Anthemideae*. Therefore, secondary metabolites as taxonomic markers corroborate the Funk classification (Funk et al., 2009), which is based on the DNA sequences.

From the hit percentages displayed in the experiments carried out and past study (Correia et al., 2008), one can highlight that ANN can be used as a powerful tool for plant classification. The chemical data employed in this approach were able to differentiate the *Asteroideae* tribes and genera with a similar performance as an expert.

To conclude, the results of this study show that the data of chemical classes, skeletal type, occurrence of secondary metabolites and evolutionary descriptors (Oxidation Step and Skeletal Specialization) are useful to aid taxonomic classification, as well as to associate classifications using other taxonomic markers, such as morphological and/or macromolecular (DNA or RNA).

REFERENCES

Baldwin BG, Wessa BL, Panero JL (2002). Nuclear rDNA evidence for major lineages of helenioid Heliantheae (*Compositae*). Syst. Bot., 27: 161-198.

Bayer RJ, Starr JR (1998). Tribal phylogeny of the *Asteraceae* based on two non-coding chloroplast sequences, the trnL intron and trnL/trnF intergenic spacer. Ann. Mol. Bot. Gard., 85: 242-256.

Bentham G (1873). Notes on the classification, history, and geographical distribution of the *Compositae.* Bot. J. Linn. Soc., 13: 335-577.

Bremer K, Anderberg AA (1994). *Asteraceae* - Cladistic and Classification. Timber Press (OR), P. 752.

Bremer K (1996). Major clades and grades in *Asteraceae.* In: Hin DJN and Beentje HJ (eds.). *Compositae:* Systematics Proceedings of the International *Compositae* Conference. Royal Bot. Gard.,Kew, pp. 1-7.

Calabria LM, Emerenciano VP, Ferreira MJP, Scotti MT (2007). A phylogenetic analysis of tribes of the *Asteraceae* based on phytochemical data. Nat. Prod. Comm., 2: 277-285.

Carlquist S (1976). Tribal interrelationships and phylogeny of the *Asteraceae*. Aliso, 8: 465-492.

Correia MV, Scotti MT, Ferreira MJP, Alvarenga SAV, Rodrigues GV, Emerenciano VP (2008). Self-Organizing Maps as a New Tool for Classification of Plants at Lower Hierarchical Levels. Nat. Prod. Comm., 3: 1723-1730.

Cronquist A (1977). *Compositae* Revisited. Brittonia, 29: 137-153.

Cronquist A (1988). The evolution and Classification of the Flowering Plants. New York Botanical Garden, New York, P. 555.

Dahlgren RMT (1980). A Revised System Of Classification Of The Angiosperms. Bot. J. Linn. Soc., 80: 91-124.

Emerenciano VP, Cabrol-Bass D, Ferreira MJP, Alvarenga SA, Brant AJC, Scotti MT, Barbosa KO (2006). Chemical evolution in the Asteraceae. The Oxidation-Reduction mechanism and production of secondary metabolites. Nat. Prod. Com., 1: 495-507.

Emerenciano VP, Militão J, Campos CC, Rornoff P, Kaplan MAC, Zambon M, Brant AJC (2001). Flavonoids as chemotaxonomic markers for *Asteraceae.* Biochem. Syst. Ecol., 29: 947-957.

Emerenciano VP, Rodrigues GV, Alvarenga SAV, Macari PAT (1998). A new method for grouping chemotaxonomic parameters. Quim. Nova, 2: 125-129.

Funk VA, Bayer RJ, Keeley S, Chan R, Watson L, Gemeinholzer B, Schilling E, Panero JL, Baldwin BG, Garcia-Jacas N, Susanna A, Jansen RK (2005). Everywhere but Antarctica: Using a supertree to understand the diversity and distribution of the *Compositae.* Biol. Skrif., 55: 343-374

Funk VA, Susanna A, Stuessy TF, Bayer RJ (eds.) (2009). Systematics, Evolution, and Biogeography of *Compositae.* International Association for Plant Taxonomy, Vienna, P. 777.

Gastmans JP, Furlan M, Lopes MN, Borges JHG, Emerenciano VP (1990). Artificial intelligence applied to natural products chemistry. The systematic program Part II - Organization of the Program and Applications. Nova Chem., 13: 75-80.

Gershenzon J, Mabry TJ (1983). Secondary metabolites and the higher classification of angiosperms. Nord. J. Bot., 3: 5-34.

Gottlieb OR (1989). The Role of Oxygen in Phytochemical Evolution towards Diversity. Phytochemistry, 28: 2545-2558.

Jansen RK, Holsinger KE, Michaelis HJ, Palmer JD (1990). Phylogenetic Analysis Of Chloroplast DNA Restriction Site Data At Higher Taxonomic Levels - An Example From The *Asteraceae*. Evolution, 44: 2089-2105.

Karis PO (1993). *Heliantheae Sensu-Lato* (*Asteraceae*), Clades And Classification. Plant Syst. Evol., 188: 139-195.

Kohonen T (1982). Self-Organized Formation of Topologically Correct Feature Maps. Biol. Cybernet., 43: 59-69.

Kohonen T (2001). Self-Organizing Maps Springer Verlag, Heidelberg, P. 501.

Massart DL, Vandeginste BGM, Deming SN, MichotteY, Kaufman L (1988). Chemometrics: a textbook. Elsevier, New York, Pp. 298-299.

Mathworks Inc (2004). Available http://www.mathworks.com accessed in 13/07/2011.

Pelser PB, Watson LE (2009). Introduction to *Asteroideae*. In: Funk VA, Susanna A, Stuessy TF and Bayer RJ (eds.). Systematics, Evolution, and Biogeography of *Compositae.* Int. Assoc. Plant Taxon., Vienna, Pp. 495-502.

Richardson PM, Young DA (1982). The Phylogenetic Content Of Flavonoid Point Scores. Biochem. Syst. Ecol., 10: 251-255.

Robinson H (1981). A revision of the tribal and subtribal limits of *Heliantheae* (*Asteraceae*). Smithsonian Cont. Bot., 51: 1-102.

Sneath PHA, Sokal RR (1973) Numerical Taxonomy. Freeman and Company, San Francisco, P. 573.

Statsoft Inc (2000). Statistic Neural Networks, Tulsa, OK, EUA.

Stuessy TF (1977). Heliantheae - Systematic Review. In: Heywood VH, Harbone JB and Turner BL (eds.). The Biology and Chemistry of the *Compositae.* Academic Press, London, Pp. 621-671.

Taketo O, Afendi M, Altaf-UI-Amin M, Hiroki T, Kensuke N, Shigehilo K (2010). Metabolomics of Medicinal Plants: The Importance of Multivariate Analysis of Analytical Chemistry Data. Comp. Aided Drug Design, 6: 179-196.

Takhtajan A (1973). The chemical approach to plant classification with special reference to higher taxa of *Magnolyiphyta.* In: Bendz G and Santesson J (eds.) Chemistry in Botanical Classification. Academic Press, London, Pp. 17-26.

SOM Toolbox for Matlab 5 (2011), Available http://www.cis.hut.fi/projects/somtoolbox. Accessed 07/13/2011.

Wagenitz G (1976). Systematics And Phylogeny Of *Compositae* (*ASTERACEAE*). Plant Syst. Evol., 125: 29-46.

Waterman PG, Gray AI (1987). Chemical Systematics. Nat. Prod. Rep. 4: 175-203.

Zupan E, Gasteiger J (1993). Neural Networks for Chemists - An Introduction. Weinheim: VCH, P. 305.