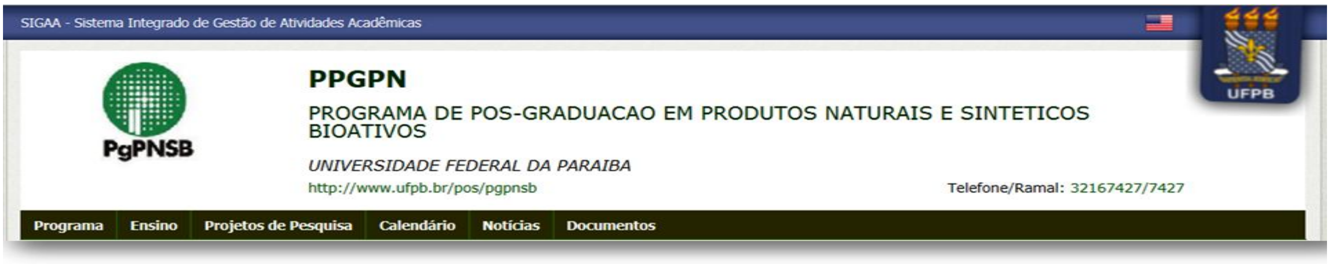


SIGAA - Sistema Integrado de Gestão de Atividades Acadêmicas

PgPNSB

PPGPN

PROGRAMA DE POS-GRADUACAO EM PRODUTOS NATURAIS E SINTETICOS BIOATIVOS

UNIVERSIDADE FEDERAL DA PARAIBA

http://www.ufpb.br/pos/pgpnsb

Telefone/Ramal: 32167427/7427

MOL2NET, International Conference Series on Multidisciplinary Sciences
http://sciforum.net/conference/mol2net-03

Application of Self-Organizing Maps generated from Molecular Descriptors of Flavonoid in the Chemotaxonomy of the Asteraceae Family

Élida Batista Vieira Sousa Cavalcanti (elidabvs@gmail.com)[a], Marcus Tullius Scotti (mtscotti@gmail.com)[a,*], Luciana Scotti (luciana.scotti@gmail.com)[a], Vicente de Paulo Emerenciano[b]

[a]Federal University of Paraíba, João Pessoa, Paraíba, Brazil; [b]University of São Paulo, São Paulo, SP, Brazil

.

*Correspondence: mtscotti@gmail.com

Abstract: The Asteraceae family belongs to the Asterales order, it consists of approximately 1,600 genera and 24,000 species, divided into 12 subfamilies and 44 tribes, is one of the largest families of angiosperms in the world. Asteraceae is remarkable the presence of flavonoids, these have the necessary requirements to be used successfully in chemotaxonomy because are found in abundance in the Asteraceae, presents structural diversity, are stable structures and relatively easy to identify, therefore can be used as taxonomic markers. The aim of this study is to classify Asteraceae tribes based on the number of occurrences of flavonoids from our in-house databank (available at www.sistematx.ufpb.br) using descriptors calculated by DRAGON 7.0 software. The 2371 botanical occurrences with respective 74 molecular fragment descriptors were used as input data in SOM Toolbox 2.0 (Matlab) to generate Self-Organizing Maps (SOMs), classifying four tribes: tribes Anthemideae (A), Gnaphalieae (G), Tageteae (T) and Senecioneae (S). Some descriptors show higher contribution to differentiate the flavonoids: RFD, nCIC and NNRS. Since these SOM are built based on physicochemical properties, so it is possible to use this tool in the search for flavonoids with potential biological activities with the respective taxonomic information.

Keywords: Asteraceae, flavonoids, chemotaxonomy, databank, descriptors, Self-Organizing Maps

1. Introduction

The Asteraceae family (Compositeae) is one of the largest families of angiosperms in the world [1]. Some 1,600 genera and 24,000 species of this family have been described botanically and several revisions regarding its chemistry and biology were published [2]. The latest classification recognizes 12 subfamilies and 44 tribes are usually represented by herbaceous plants and small shrubs, rarely by trees [2,3].

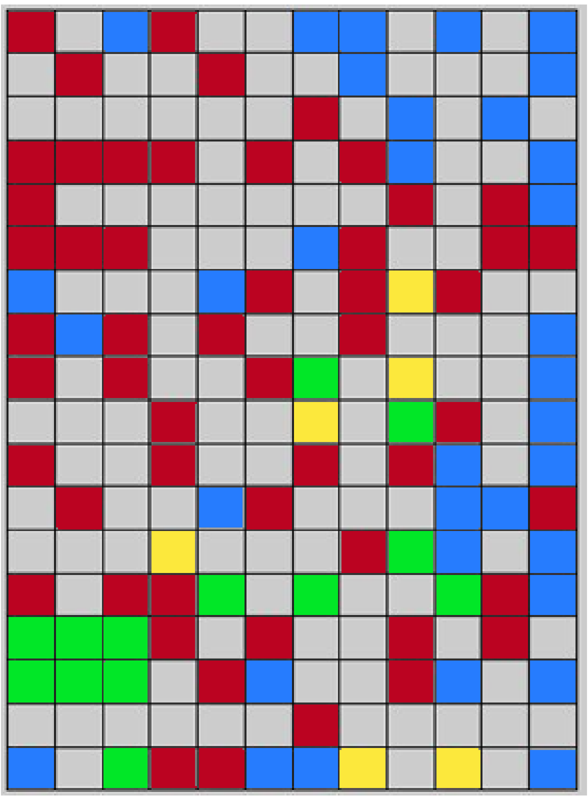
In Chemistry of Natural Products, secondary metabolites are important chemical markers, have a restricted distribution and specific botanical sources [4,5]. Among them stand out the flavonoids, contain a basic structure consists of 15 carbon atoms arranged into three rings (C6-C3-C6), have the necessary requirements to be used successfully in chemotaxonomy because this class are stable structures and relatively easy to identify, presents structural diversity, are found in abundance in the Asteraceae family [6], they can be used as taxonomic markers at lower hierarchical levels [7].

In this study, the Asteraceae tribes were classified based on the number of occurrences of flavonoids from our in-house databank using descriptors calculated by DRAGON 7.0 software [8]. With the Matlab software [9], chemical patterns were recognized and analyzed from unsupervised artificial neural networks, along with the SOM (Self Organizing Map) to create the maps.

2. Results and Discussion

From the botanical occurrence data collected, were generated 74 molecular descriptors for each molecule, through software DRAGON, then one can calculate the self-organizing matrix for each molecule, dividing the data into groups according to similarity. **Figure 1** shows the Self-Organizing Maps obtained with the flavonoids descriptors of the tribes Anthemideae, Gnaphalieae, Tageteae and Senecioneae. Flavonoid hit rates belonging to the A, G, S and T tribes were 91%, 80%, 79% and 73%, respectively.

Figure 2 shows The SOM cluster similar compounds regarding molecular fragments that were labeled according to the botanical occurrence in these four tribes.



Fiure 1. Self-Organizing maps obtained with the flavonoids of the tribes Anthemideae (red), Gnaphalieae (blue), Tageteae (green) and Senecioneae (yelow).

In the matrix we could observe the separation between the tribes Anthemideae, Gnaphalieae, Tageteae and Senecioneae. The tribe Anthemideae is represented by the lighter areas of the map, that is, show low values for the descriptors in general. The tribe Gnaphalieae is represented by darker areas of the map presenting high values generated descriptors. The tribe Tageteae is represented by darker area the left and the tribe Senecioneae, to present few molecules, it is more difficult to be seen in the matrix. The number of rings (or independent cycles) in a graph is commonly known as the cyclomatic number. The tribes Gnaphalieae present lower values for the descriptor **nCIC** (number of rings in a molecule), while the tribes Anthemideae, Tageteae and Senecioneae presents high values. The tribe Gnaphalieae present high values for the descriptor **NNRS** (normalized number of ring systems), while the tribes Anthemideae, Tageteae and Senecioneae present lower values. The tribe Gnaphalieae present lower values for the descriptor **RFD** (ring fusion density), while the tribes Tageteae and

Senecioneae, mainly Anthemideae, present high values.

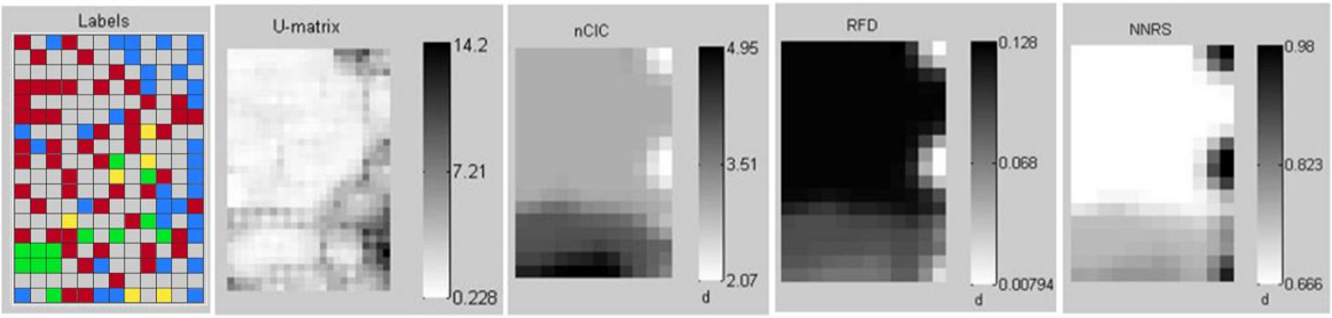


Figure 2. Map, matrix and descriptors generated: nCIC, NNRS and RFD.

3. Materials and Methods

2371 occurrences of flavonoids, which were extracted from 567 species, 47 genera and 5 tribes of the Asteraceae family, were registered in two dimensions using ChemAxon, then were used as input data in the Dragon software to generate 74 molecular descriptors. Then the descriptors were used as input data in Matlab, using the SOM Toolbox 2.0 for formation of Self Organizing Maps (SOMs), separating the tribes Anthemideae (A), Gnaphalieae (G), Tageteae (T), Senecioneae (S) and Carduoideae (CR).

4. Conclusions

The Self-Organising Map obtained separated the four tribes of the Asteraceae when using molecular descriptors. . Since these SOM are built based on physicochemical properties, so it is possible to use this tool in the search for flavonoids with potential biological activities with the respective taxonomic information.

References

1. Hattori, E.K.O. & Nakajima J.N. A família Asteraceae na estação de pesquisa e desenvolvimento ambiental galheiro, perdizes, Minas Gerais, Brasil. *Rodriguésia*. **2008** (59)4, 687-749.

2. Funk, V.A., Susanna, A., Stuessy, T.F. & Robinson, H. Classification of Compositae. 171-189. *In:* Funk, V.A.; Susanna A.; Stuessy, T.F. & Bayer, R.J. (Eds.). Systematics, Evolution, and Biogeography of Compositae. Vienna, Austria: *International Association for Plant Taxonomy*, **2009**.

3. Panero, J. L., Freire, S. E., Espinar, L. A., Crozier, B. S., Barboza, G. E. & Cantero, J. J. Resolution of deep nodes yelds an improved backbone phylogeny and a new basal lineage to study early evolution of Asteraceae. *Mol. Phylogenet. Evo*., **2014** (80), 43-53.

4. Geissman, T.A., Crout, D.H.G.,. Organic Chemistry of Secondary Plant Metabolism; Freeman Cooper & Company: San Francisco, CA, USA, **1969**.

5. Harborne, J.B. Ecological Biochemistry; Academic Press: London, UK., **1988**.

6. Stuessy, T.F., Bohm, B.A. Flavonoids of the Sunflower Family (Asteraceae): Springer Vienna, **2001**.

7. Emerenciano, V.P., Militão, J.S.L.T., Campos, C.C., Romoff, P., Kaplan, M.A.C. Flavonoids as chemotaxonomic markers for Asteraceae. *Biochem. Syst. Ecol*., **2001** (29), 947-957.

8. Talete, S.R.L. DRAGON for Windows (Software for Molecular Descriptor Calculations), **2011**. Version 6.0, Available online: http://www.talete.mi.it (accessed on 08 Nov 2017).

9. Vesanto, J., Himberg, J., Alhoniemi, E., Parhankangas, J. Self-organizing map in Matlab: the SOM Toolbox. Proceedings of the Matlab DSP Conference, Espoo, Finland, **1999**, 35-40.