PRESENCE OF FLAVONOIDS AND INTRAFAMILIAL CLASSIFICATION OF THE *COMPOSITAE*

BOGUSŁAW WILKOMIRSKI, ANDRZEJ BATKO

Department of Plant Systematics and Geography, University of Warsaw
Warszawa, Al. Ujazdowskie 4, 00-478 Poland

(Received: May 19, 1992. Accepted: August 9, 1993)

ABSTRACT

The structure of flavonoids isolated from plants of the *Compositae* family was applied for comparison of two classifications (Wagenitz 1976 and Takhtadjan 1987). The methods of numerical taxonomy, both phenetic and phylogenetic, confirm that Takhtadjan's classification based on morphological and anatomical data is more adequate to the flavonoid diversity.

KEY WORDS: *Compositae*, numerical taxonomy, flavonoids.

INTRODUCTION

The presence of sesquiterpene lactones, poliacetylene compounds, triterpenes, flavonoids and inuline type fructans is characteristic for the whole *Compositae* family (Hegnauer 1977). Classification of the family *Compositae* is the subject of many taxonomical controversies, particularly in case of the suprageneric level.

The division of the family into 13 tribes which was done by Bentham and Hooker (1862-1883), and Bentham (1873) in the 19th century, remains the base of most proposed classifications, however, the ranges of individual tribes and, first of all, their grouping into subfamilies are variously formulated. In the last decade many reports appeared, devoted to the phytochemical diversity of the *Compositae*, among others to sesquiterpenes and flavonoids (Emerenciano et al. 1985, 1987). The data found in them are rich enough to enable confrontation with commonly used classifications of the *Compositae* (at the level of tribes and subfamilies) based mostly on morphoanatomical information. The aim of our work was the comparative study of two selected classifications of the *Compositae* using phytochemical data and numerical methods.

CLASSIFICATIONS OF THE *COMPOSITAE*

The task resolves itself into the choice of two alternative classifications and the comparison of their ability "to incorporate" the new knowledge about phytochemical diversity of the *Compositae* with application of the methods of modern numerical taxonomy.

The results obtained should enable the choice of a classification which is more stable due to new information.

The comparative examination of phytochemical data included in the literature (Emerenciano et al., 1985, 1987) belongs to the field of exploratory data analysis (Tuckey 1977). It reveals the logical structure of multidimensional knowledge about the diversity of the considered phenomena and presents its image in the form of graphs.

In the *Compositae* family we traditionally distinguish two subfamilies: *Cichorioideae* (vel *Lactucoideae* or *Liguliflorae*) and *Asteroideae* (vel *Tubuliflorae*) and several tribes. The ranges of subfamilies, however, are differently formulated; in many older classifications *Cichorioideae* in the monotypical subfamily i.e. includes only the tribe *Lactuceae*. Wagenitz's classification, presented graphically in Figure 1 supports this tradition.

Fig. 1. Classification I (Wagenitz 1976)

Thorne (1976, 1983) and Jeffrey (1977), however, distribute the tribes between two subfamilies more proportionally. Such a solution is also accepted by Takhtadjan whose classification is presented in Figure 2.

Fig. 2. Classification II (Takhtadjan 1987)

* Included into *Senecioneae*

The above diagram shows the simplified variant of Takhtadjan's classification, because the tribes *Arctioideae*, *Gundelieae*, *Echinopeae*, *Carlineae* and *Liabeae* are omitted. These tribes are not in classification I. For the sake of comparison, the tribe *Arniceae*, included by Takhtadjan among the *Senecioneae*, is considered as a separate taxon. The salient feature of Takhtadjan's classification is the recognition of the subfamily *Cichorioideae* comprising four tribes: *Lactuceae*, *Vernoniae*, *Cardueae* and *Mutisieae*. This belief was supported very recently by Bremer et al. (1992).

We are often confronted with the question "which of the two alternative classifications is better, and what criterion decides about it?" "One of the simplest answers to this difficult question is: that, which can longer accept to face the continuous increase of knowledge, influx of new data and growing scientific demands" (Abbott et al. 1985). In classification II *Cichorioideae* are no longer equal to traditional "ligulate flowers", because 3 out of 4 tribes included into this family do not provide diagnostic features of the *Liguliflorae* (the lack of tubular flowers and the presence of latex). The tribe *Mutisieae* (about 90 genera and 1000 species) is represented mostly in the flora of the southern hemisphere; gerbera belong to its well known members. This tribe is often recognized as the less advanced due to a primitive type of pollen exine, anatomical features of the stem and a two-lobed corolla (Thorne 1976, 1983; Wagenitz 1976, Jeffrey 1977, Takhtadjan 1987). The relatively small tribe *Vernoniae* (70 genera and about 1460 species) is also localized mostly in tropical America and to some degree in Africa and south-eastern Asia. Hence, the essential difference between the two classifications analysed in this study is the answer to the question: is it substantial to connect in one subfamily the tribe *Lactuceae* (among others genera *Cichorium*, *Crepis*, *Scorzonera*, *Tragopogon*, *Lactuca*, *Sonchus*, *Hieracium*, *Taraxacum*) with the big tribe *Cardueae* (represented among others by the genera *Arctium*, *Cousinia* (about 600 species), *Centaurea* (550), *Cirsium* (up to 300 species), *Carduus* (about 120)? Which solution reflects the phytochemical diversity of the *Compositae*? more adequately?

FLAVONOIDS AND TAXONOMY OF *COMPOSITAE*

The knowledge of plant chemistry is excellent for comparative verification of competing classifications, because it is independent of information establishing such classifications. Moreover-the information about the chemical composition of the *Compositae* plants is abundant, heterogeneous and obtained by modern methods (spectroscopy and chromatography), what allows to assume that it is highly correct.

Among the great diversity of secondary metabolites, the flavonoids are compounds providing good taxonomical characters. They are present in all vascular plants, provide many structural forms and genetically controlled variability. Their content is rather stable and they can be easily extracted and identified.

The great number of flavonoid forms (compounds with the base skeleton of 15 carbon atoms in the grouping C₆-C₃-C₆) derives from:

-differences in basic structure of the C₁₅ skeleton - number and position of hydroxyl groups in both rings - different level of oxidation of C₃ unit - glycoside bonds with simple sugars and sugar acids - methylation of hydroxyl groups

The tributary data analysis of micromolecular features is most powerful when based on statistical methodology. In the case of flavonoids it is possible to carry out the estimation of

oxidation of ring A and ring B as well as the difference between the extreme oxidation values of both rings (Δ) in flavonoids isolated from plants of a selected taxon (Gottlieb 1982, Emerenciano et. al. 1987.).

The degree of glycosylation, methylation as well as the total factor of hydroxyl group protection is also estimated. Statistical quantitative features are estimated on the base of frequency of occurrence of certain features and anticipated sequence of biosynthetic pathways eg. in most natural flavonoids ring A exhibits floriglucine with hydroxyl groups occurring in positions 5 and 7 nearly in all cases. Hence, such localization of hydroxyl groups is considered as primary, and some conventional figures are attributed to a lesser or greater number of hydroxyl groups. Other parameters which are used as the features in our numerical investigations are calculated similarly.

Although synthetical indexes, such as: OH - groups, protection index and ring A oxidation index could be a source of some errors, these are commonly used as chemotaxonomical parameters. However, it is not an oversimplification, at least as far as the phenetic approach is considered.

NUMERICAL ANALYSIS

The numerical analysis of data collected in the paper by Emerenciano et. al. (1987) was performed twice - in the convention of phenetic and phylogenetic taxonomy (Abbott et al. 1985, Batko and Moraczewski 1992). The aim of the phenetic analysis was to find in the set of 13 tribes the phenons - subsets linked by the overall similarity and to built up a hierarchical system of such phenons as a graph called dendrogram (phenogram). The phylogenetic analysis is aimed towards construction of another graph called Wagner tree. The Wagner tree is a hierarchical system of monophyletic "sister groups" tied by synapomorphies - common features inherited from the nearest common ancestor and shared with them only.

Both analyses tended towards the finding among many variants of classifications, the solutions possibly close to one of two alternative classification, under consideration and deciding which of them supports better the knowledge on the diversity of the *Compositae* extended with flavonoid data.

The starting point for the analysis was information concerning the set of 13 *Compositae* tribes described by 7 features:

List of objects

1. *Lactuceae*
2. *Vernoniae*
3. *Cardueae*
4. *Mutisieae*
5. *Anthemideae*
6. *Arniceae*
7. *Astereae*
8. *Calenduleae*
9. *Eupatorieae*
10. *Heliantheae*
11. *Inuleae*
12. *Senecioneae*
13. *Tageteae*

List of features

1. ring A oxidation index
2. ring B oxidation index
3. ring A oxidation D
4. ring B oxidation D
5. glycosilation index
6. methylation index
7. OH groups protection index

The values of all features were multiplied by a constant, to be represented by whole numbers, desired for technical reasons, and compiled in the data matrix:

THE PHENETIC APPROACH

The cluster analysis - the most commonly used tool of exploratory data analysis - was applied to the above data. In the first step of the procedure the measure of dissimilarity between objects is calculated. A lot of such measures are applied. In the second step the objects are grouped hierarchically on the basis of this base. Many algorithms and methods of clustering were proposed (see Kucharczyk 1982, Anderberg 1973, Abbott et al. 1985). Different combinations of preliminary data transformation and measures of dissimilarity with methods of clustering provide results slightly differing in details, but invariant in the "core" structure.

In the present study the numerical taxonomy programs package TAXAL 2 (Batko and Moraczewski 1992) was used. Among 56 obtained phenograms, those based on non-metric Smirnov taxonomic coefficient (Smirnov 1969) and two metric distances: euclidean distance and Canberra distance approximate one of two compared classifications. It is interesting, because Smirnov's measure depends on the frequency of particular character states in whole set estimating the similarity the higher, the more rare feature it concerns. Metric measures estimate some "absolute" similarity, i.e. - depending on features of two compared objects only.

A range of clustering methods (Ward method, UPGMA, centroidal sorting, flexible sorting and complete linkage method) working under well known algorithm SAHN (Abbott et al. 1985, Anderberg 1973) produced, among others, phenograms presented in Figures 3-7.

Fig. 3. UPGMA phenogram based on max-complement of Smirnov coefficient (data converted to nominal type)

Fig. 4. UPGMA - dendrogram based on euclidean distance calculated from normalized data

Fig. 5. Centroidal Sorting phenogram based on euclidean distance calculated from normalized data

Fig. 6. Ward phenogram based on euclidean distance calculated from logarithmized data

Fig. 7. Complete linkage phenogram based on Canberra distance

In Fig. 3 there is the cluster *Lactuceae* + *Vernoniae* + *Mutisiae* + *Carduae* the composition of which is the same as in the subfamily *Cichorioideae* by Tachtadjan.

In phenograms in Figs. 4-7 the cluster marked "!" contains three of four *Cichorioideae* tribes (sometimes complemented by *Senecioneae*). The analysis reveals, therefore, in flavonoids data the stable cloud of 3 tribes, compatible with Tachtadjan's *Cichorioideae* or (in the case of Smirnov coefficient) perfectly matching this subfamily. Only the fate of *Mutisiae* is to some degree unpredictable. At the same time we detected no dendrograms supporting the Wagenitz solutions, i.e. - separating the tribe *Lactuceae* from the rest of the family *Compositae*.

THE PHYLOGENETIC APPROACH

The numerical phylogenetic analysis (Wiley 1981) is a much more difficult task, both from the theoretical and mathematical point of view. The aim of this study is to find the most probable estimation of true (unknown) phylogeny for the set of taxa under consideration; hence, there is here some objective criterion of quality of the obtained result. This criterion is, however, entirely non-operational i.e. the similarity to an unknown phylogenetic tree. Therefore, the other - operational - criterion derived from the principle of evolutionary parsimony (Wiley 1981) is accepted as substitution.

The phylogenetic analysis requires prepared data, mostly of logical type (e.i. assuming the "false" or "true" value). Moreover, the "primitive" vs. "advanced" state should be specified. This makes it possible to define the "elementary evolutionary change" as the pass from the state "0" (false) to "1" (true) and the opposite pass as the "reversion" (Wiley 1981, Abbott et al. 1985). The measurement of tree length becomes also possible. All the mentioned conditions are necessary requirements to treat the resulting graphs as an estimation of the genuine genealogy.

The data analysed in our paper did not fulfill any of these conditions. This obstacle was partly overcome due to automatic conversion of our variables to the set of binary "factors". Such a conversion was performed by programs TPH and FACTOR. The former interfaces TAXAL with PHYLIP (Felsenstein 1988), the latter recalculates data. As a consequence of such status of data the resulting most parsimonious tree is rather the estimation of data structure than reconstruction of genealogy. Hence the presented "phylogenetic" analysis is obviously the preliminary one. However the comparison of its result with the phenetic approach is worthy of consideration.

The PHYLIP package includes the MOVE program working in the interactive mode. Using this program, "user defined" trees isomorphic with both discussed classifications were created (Figs. 8 and 9) and their lengths were measured on the base of a new data file.

Fig. 8. Tree isomorphic with classification I

Fig. 9. Tree isomorphic with classification II

The tree isomorphic with classification II appeared much shorter (885 against 987 steps) therefore the "better" estimation of phylogenesis. The other program - MIX - enables an automatic search for the shortest trees using two algorithms (Wagner or Camin-Sokal) or their combination. We used the Camin-Sokal algorithm and obtained the tree presented in figure 10.

This tree is shorter than both trees illustrating the discussed classifications. It is closer to classification II than to classification I. It is particularly concerned with the position of the tribe *Lactuceae*.

Fig. 10. Wagner tree calculated by programme MIX

SUMMARY

None of the used techniques of numerical taxonomy separated the tribe *Lactuceae* into an isolated subtree, hence the traditional division of *Compositae* to *Liguliflorae* and *Tubiflorae* was not supported by their flavonoid diversity.

The above statement is contradictory to conclusions of Emerenciano et al. (1987) based on the same data, but in accordance with the new opinion of systematics (Takhtadjan

1987, Thorne 1976). In our opinion, it proves that "confirmatory" statistical data analysis used by Emerenziano and co-workers is not the most adequate tool of exploration of the structure of diversity presented as a multidimensional data set. The cluster analysis and numerical cladistics seem to be more adequate in this case.

The analysis shows that the relatively small set of chemical data contains the information satisfactorily sound both from biological and mathematical points of view. This information is sufficient to draw some image of diversity of the taxon studied, which is close to its classification based on independent, mostly morphological and anatomical data. It confirms the high cognitive value of phytochemical data and heuristic power of contemporary taxonomical analysis.

ACKNOWLEDGMENT

Our sincere thanks are due to Professor Krzysztof Rostanski for his continuous support of this study and many useful suggestions concerning the bibliography of the problem under consideration.

LITERATURE CITED

- ABBOTT L.A., BISBY F.A., ROGERS D.J., 1985. Taxonomic Analysis in Biology, Computers, Models, and Databases. Columbia Univ. Press, New York, 336 pp.
- ANDERBERG M.R., 1973. Cluster Analysis for Applications. Acad. Press, New York.
- BATKO A., MORACZEWSKI I., 1992. TAXAL 2 - Software in Numerical Taxonomy. Documentation. Warszawa.
- BATKO A., 1988. Metody współczesnej systematyki: fenetyka i kladystyka. Informatyka w Szkole, 9.
- BENTHAM G., 1873. Notes on the classification, history and the geographical distribution of *Compositae*, Linn. Soc. J. Bot. 13: 335-577.
- BENTHAM G., HOOKER J.D., 1862-1883. Genera Plantarum. London.
- BREMER K., JANSEN R.K., KARIS P.O., KI-JOONG KIM, 1992. Current research on *Asteraceae* phylogeny and classification. Comp. Newsl. 20/21.
- EMERENCIANO V.P., KAPLAN M.A.C., GOTTLIEB O.R., 1985. Evolution of sesquiterpene lactones in angiosperms. Biochem. Syst. Ecol. 13: 145-166.
- EMERENCIANO V.P., FERREIRA Z.S., KAPLAN M.A.C., GOTTLIEB O.R., 1987. A chemosystematic analysis of tribes of *Asteraceae* involving sesquiterpene lactones and flavonoids, Phytochemistry 26: 3103-3115.
- FELSENSTEIN J., 1988. PHYLIP - Phylogeny Inference Package, version 3.1, New York [microcomputer programs].
- GOTTLIEB O.R., 1982. Micromolecular evolution, Systematics and Ecology, Springer, Berlin, 170 pp.
- HEGNAUER R., 1977. The chemistry of the Compositae, in: The Biology and Chemistry of the *Compositae*, Heywood V.C., Harborne J.B., Turner B.L., eds. Academic Press, pp., 283-335.
- JEFFREY C., 1977. Corolla forms in *Compositae* - some evolutionary and taxonomic speculations, in: The Biology and Chemistry of the *Compositae*, eds: Heywood V.C., Harborne J.B., Turner B.L., Academic Press, pp. 111-118.
- KUCHARCZYK J., 1982. Algorytmy analizy skupień w języku ALGOL 60. PWN Warszawa 1982.
- SMIRNOV E.S., 1969. Taksonomitsheskij Analiz. Moskva.
- TAKHTADJAN A., 1987. Sistema Magnoliifitov. Nauka, Leningrad, 438 pp.
- THORNE R.F., 1976. A phylogenetic classification of the *Angiospermae*, Evol. Biol., 9: 35-106.
- THORNE R.F., 1983. Proposed new realignments in the angiosperms, Nord. J. Bot. 3: 85-117.
- TUCKER J.W., 1977. Exploratory Data Analysis. Addison-Wesley Publ. Co, Inc.
- WAGENITZ G., 1976. Systematics and phylogeny of the *Compositae*, Plant Syst. Evol. 125: 29-46.
- WILEY, E.O., 1981. Phylogenetics: The theory and practice of phylogenetic systematics, New York and Chichester, Wiley.

WYSTĘPOWANIE FLAWONOIDÓW A KLASYFIKACJA RODZINY *COMPOSITAE*

STRESZCZENIE

Porównano dwa systemy klasyfikacji *Compositae* na poziomie plemion i podrodzin (Wagenitz 1976, Takhtadjan 1987) z opartymi na danych fitochemicznych klasyfikacjami numerycznymi tego samego zbioru plemion. Posłużono się danymi Emerenziano i in. (1987) odnoszącymi się do składu flawonoidowego omawianych roślin. Szczególną uwagę zwrócono na pozycję plemienia *Lactuceae*, które w klasyfikacji Wagenitza jest wyodrębniane w osobną podrodzinę, zaś wg Takhtadjana - łączone na tym poziomie z trzema innymi plemionami. Analiza wykazała znacznie większą zgodność danych fitochemicznych z poglądem Takhtadjana. Rzadko stosowana miara podobieństwa, jaką jest współczynnik t Smirnova w połączeniu z metodą skupiania UPGMA pozwoliły na wyodrębnienie fenomenu składającego się z: *Lactuceae*, *Vernonieae*, *Mutisieae* i *Cardueae*, identycznego z podrodziną *Cichorioideae* w ujęciu Takhtadjana.

SŁOWA KLUCZOWE: *Compositae*, klasyfikacja numeryczna, flawonoidy