IDENTIFYING TROPICAL TREES USING COMPARISONS WITH SINGLE SPECIMENS OF SPECIES

Peter Hargreaves[1], Eduardo van den Berg[2], Douglas Antonio de Carvalho[2], Ary Teixeira de Oliveira Filho[3]

(received: november 13, 2007; accepted: november 28, 2008)

ABSTRACT: A taxonomic database system CARipé with integrated links to plant images was constructed to characterize the 800 forest tree species recorded for the catchments of the Rio Grande in the south of Minas Gerais state in Brazil using vegetative morphology. After a description was ready, its integrated system Empar, completed in a few seconds an identification attempt using automatic character correlation. Empar proved operationally successful in identification tests of 62 specimens, of as many species, against 554 reference specimens and species from the ESAL Herbarium. Empar employed a similarity coefficient with the option of a weighting based on character frequency and, in tests, the use of this rarity weighting further improved the performance. Three tree groups based on leaf type and phyllotaxy were analyzed separately. From the start, those with compound leaves received more descriptors and the result was little difference in identification performance between the tree groups. Indications were that experts and non-experts performed characterizations for identification purposes almost equally well. The most important result was that successful identifications were obtained using a reference base of just one specimen description per species.

Key words: Brazilian Atlantic forest, plant taxonomy, similarity index, species identification, vegetative morphology.

IDENTIFICAÇÃO DE ÁRVORES TROPICAIS USANDO COMPARAÇÕES COM APENAS UM ESPÉCIME POR ESPÉCIE

RESUMO: *Um banco de dados taxonômico com imagens, CARipé, foi construído para caracterizar, por morfologia vegetativa, as 800 espécies de árvores florestais registradas para a bacia do rio Grande, no sul de Minas Gerais. Depois que uma descrição foi completada, seu sistema integrado, Empar, realizou em alguns segundos uma identificação, usando a correlação automática de caracteres. Em operação, Empar provou ser efetivo em testes de identificação de 62 espécimes, do mesmo número de espécies, em comparações com 554 espécimes e espécies de referência a partir do Herbário ESAL. Empar usou um coeficiente de similaridade com a opção de um peso baseado na freqüência de caracteres. Em testes, o uso desse peso de raridade melhorou a performance. Três grupos caracterizados pelo tipo da folha e filotaxia foram analisados separadamente. No início, o grupo com folhas compostas recebeu mais descritores e o resultado foi pouca diferença entre os grupos na performance de identificações. Houve indicações de que os experts e os leigos realizariam boas caracterizações, para fins de identificação. O resultado mais importante foi que as identificações bem sucedidas foram obtidas usando-se uma base referencial de uma só descrição de espécime por espécie.*

Palavras-chave: Floresta atlântica, identificação de espécies, índice de similaridade, morfologia vegetativa, taxonomia vegetal.

1 INTRODUCTION

Species identification is the assigning of a specimen to a particular species. The plant is given a name by recognizing that it belongs to a previously described species, one probably held in a local herbarium (DAVIS & HEYWOOD, 1963). Species identification has an enormous number of actual and potential beneficiaries both inside and outside biology. Predominantly, they are people involved in conservation, ecology, agriculture, health and pharmaceuticals.

As all taxonomy relies on character correlations it was soon realized that the computer could process an enormous number of correlations in a relatively short time. The new interest in "numerical taxonomy" produced, among other things, identification systems based on automatic character correlation. Sneath & Sokal (1973) referred to this method as simultaneous keys, whereas in 1975, Pankhurst used the name "identification by matching" (PANKHURST, 1993). However, it was an idea that was advanced ahead of adequate desktop computer systems and this was one of the reasons why this method was relegated, at least by botanists, behind traditional identification using keys. For a discussion of both methods in this context see Pankhurst (1993, 1995). Only recently are all the necessary elements firmly in place for an

[1]Doutorando do Departamento de Ciências Florestais/DCF – Universidade Federal de Lavras/UFLA – Cx. P. 3037 – 37200-000 – Lavras, MG – ptharg@yahoo.com.br
[2]Professores do Departamento de Biologia/DBI – Universidade Federal de Lavras/UFLA – Cx. P. 3037 – 37200-000 – Lavras, MG – evandenb@ufla.br, douglasc@ufla.br
[3]Professor do Departamento de Botânica, Instituto de Ciências Biológicas/ICB – Universidade Federal de Minas Gerais/UFMG – Av. Antônio Carlos, 66627 – Pampulha – 31270-901 – Belo Horizonte, MG – ary@icb.ufmg.br

integrated identification system using automatic character correlation. The four main elements for efficient computer-aided identification based on plant morphology are: (1) rapid and accurate data input, (2) graphical online help to choose between character states, (3) touch-of-button software for automatic character correlation as part of a truly interactive system, and (4) image capture and display of botanical material.

The current paper briefly describes our taxonomic database system CARipé and, in more detail, its identification system Empar and the rigorous methods to evaluate it. The main hypotheses of the larger study were the following: (1) For a local flora it would be possible to characterize nearly all the 800 forest tree species of the region using only vegetative morphology; (2) with this system of characterization implemented on a computer, exsiccates of the herbarium and "live", recently collected specimens of trees could be quickly described; and (3) character correlation software linked to a flora including images would be an effective and efficient method for identifying tree specimens.

2 MATERIALS AND METHOD

2.1 The CARipé system

CARipé is a taxonomic database system that uses vegetative morphology to characterize tropical trees. It was constructed using the database management system of Microsoft Access but also includes stored image files which can be opened within Access by Windows mediated links to imaging software and individual image files (HARGREAVES, 2005). The data gathering started with an Excel spreadsheet, now maintained using Access, containing distribution and abundance data of 800 species of tree and shrub. This was the result of 20 years of phytosociological surveys of the Atlantic forest fragments of the upper Rio Grande in the state of Minas Gerais, Brazil, by staff and post-graduate students at the Departments of Forestry and Biology of the Federal University of Lavras (VAN DEN BERG & OLIVEIRA FILHO, 1999; OLIVEIRA-FILHO, 2008; OLIVEIRA-FILHO et al., 1994).

The principle user interface of CARipé consists of a screen form of several independently accessible pages for data entry of different groups of data. This CARipé form constitutes the main means to both update and query the descriptions of the database, which are presented in family and species order. It must be emphasized that the entry and retrieval of data is via familiar or easily learned botanical terms such as those in Ferri et al. (1981), Lawrence (1951) and Ribeiro et al. (1999); the coding of this data is entirely hidden from the user at all stages. Graphical online help is available to choose between character states.

Advances in information technology have also allowed the rapid development of molecular studies dealing directly with the genotype, which complement morphological and functional studies based on the phenotype. The molecular "barcode of life" project depends on biotechnology laboratories (RATNASINGHAM & HEBERT, 2007). CARipé does a similar thing with morphological data and requires little more than a desktop computer with scanner or digital camera.

2.2 Vegetative morphology

For a discussion of the 83 taxonomic characters employed see Hargreaves (2005), or for a discussion of some 22 morphological traits together with an analysis of their distribution among the 16 families represented by 10 or more species see Hargreaves (2006). A character state is said to be "constant" for a species when nearly all specimens of the species possess it. Leaf type and phyllotaxy have been found to be "constant" for many species so that they were confidently used as a first step in the identification system, and also to artificially classify the set of specimens into four tree groups before testing the identification system:

Group A: leaves simple or lobed, with alternate phyllotaxy (275)
Group B: leaves simple or lobed, with opposite or verticillate phyllotaxy (179)
Group C: leaves compound with alternate phyllotaxy (100)
Group D: leaves compound with opposite phyllotaxy (16)

The small Group D with little data was left out of the analysis. Also, Group C was less complete than the other groups as few descriptions of herbarium species of Rutaceae and Sapindaceae were included in the database.

2.3 The Empar identification system

Empar is integrated within the CARipé system and is a system for the identification of tree species using the automatic correlation of any taxonomic characters used by CARipé. Its mode of evaluation was the following. Single specimens of different test species were compared with a set of descriptions based on only one herbarium specimen per reference species. However, it must be pointed out that 78% of these herbarium specimens included one or more unfixed duplicates that were used in the description. On the other hand, the preference was to describe herbarium specimens that included no more than two duplicates, as more than this proved difficult to handle.

The best way to understand the identification process is to see the result of an identification event for *Miconia trianaei* Cogn., which is shown in Table 1 in the results section. A test specimen is compared to itself and thus comes out top of the list of specimens sorted by decreasing order of value of the similarity index, confirming what is necessary for a perfect correspondence in terms of numbers of shared characters.

Next in the list, the first position when giving rise to a unique value for the similarity index, comes the best ranked herbarium specimen which can be the correct identity, but not always. The possible occurrence of groups of identical scores is dealt with by calculating average rank positions, which however, are never far from actual listed positions ordered by similarity index then species name. Of course, the first-placed name is not always considered the correct identity; the next step is to activate a command to view a version of the main CARipé form in the sort order of the result of the Empar identification. Then, on the touch of a button, relevant image files can be opened via a Visual Basic mediated link to whatever image software is available on the computer. In this virtual-herbarium mode the decision is made to accept one of the high-ranking specimens as a species assignment, or not – as we may possibly have a very untypical specimen, or have a species that is not in the database as happened with *Miconia inconspicua* Miq. mentioned below. For a single identification, if images relating to more than 10 species must be appraised, the process is likely to seem tedious and uncertain, but in CARipé it is not too onerous a task to persevere through 30 species before either having to visit a herbarium or call on expert or specialist taxonomists. Naturally, all first-time identifications should be checked with material in a herbarium, but once there, working from a list of likely names from CARipé, most of the work will already have been done.

2.4 New forest survey

Perhaps the most important requirement for testing the identification system Empar was to obtain an independent set of specimens of as many species as possible while being careful to avoid bias. A new survey was considered the best way to achieve this and it also better reflected the usual conditions when identifications are necessary and undertaken. A new site was chosen in the region being studied and a survey took place between October and December 2004. Four transects were marked within the forest using a preset starting point and preset compass bearings. Every individual tree was identified within 2 m of the right-hand side of the transect lines (of both sides simultaneously of the fourth transect) when it met the phytosociological survey criterion of having a trunk of at least 5 cm diameter at standard breast height (DBH), taken as 1.3 m above soil level. All newly encountered species were sampled during a morning survey and described in the laboratory that same afternoon. To help field identification of species encountered more than once, all individuals were described in the field using bark surface characteristics, but these were limited to a maximum of 10 descriptions per species. As a rule the first individual of a species encountered was sampled if possible, described by further field characters such as smell of crushed leaves and stem, and presence of exudates, and then described in the laboratory using the main character sets. The four transects were deliberately dispersed throughout the site to sample the full extent and also the variety of topography with a view simply to sample different species. This method was a success and 65 species were found among the 269 individuals encountered. One of these species, *Miconia inconspicua* Miq., was found abundant and in flower in the understorey. It had been recorded for only one other site in the region but from outside of survey plots, and the specimen was not located in the herbarium. Group C trees, *Cupania vernalis* Cambess and *Matayba guianensis* Aublet sampled in the new forest survey, had to have their corresponding herbarium specimens especially but "blindly" described by an assistant, as they belong to the Sapindaceae which were not otherwise described. Also, *Lamanonia ternata* Vell. and *Vitex polygama* Cham. were excluded Group D trees. Voucher samples of the representative individual of each species have been deposited, with individual duplicates, in the ESAL Herbarium.

The 62 remaining specimens each representing a different species provided an adequate base for testing the identification system. Not only was it as random as possible in relation to the ease of identification of the material but it was separated in space and time from the other material and descriptions. Furthermore, it introduced an interesting difficulty of being of live material as opposed to herbarium material. For example, when deciduous stipules are a species characteristic they are nearly always observable in recently collected material but are quite often lost without trace in preserved material.

2.5 Evaluation of Empar as an identification system

As explained above, during the course of the new forest survey the first individual of a different species was always sampled and designated as representative of the species. These single live test specimens were paired with those single herbarium specimens previously designated representatives of the corresponding species. There were 30 species of Group A, 18 of Group B and 14 of Group C. These were compared against a herbarium species base of 275 Group A, 179 Group B and 100 Group C, that is, a total of 554 species.

For the automatic correlation of the taxonomic characters of the test specimens with those of the reference specimens, the coefficient known as Simple Matching was our *a priori* selection. We are using the term "similarity index" to represent the resulting value of the similarity coefficient, with or without additional rarity weighting. All the analyses described here used a combination of 65 non-numerical and six numerical characters. An additional four non-numerical and eight numerical traits were available for the species with compound leaves giving a total of 83. Clearly, any analysis using all the available data will reflect the real experience of an exercise in tree identification in our region as the "sample" of species was some 70% of our total list. However, the existence of different sample sizes for Groups A, B and C meant that they were only more generally comparable after some form of standardization. This was achieved in a second analysis by randomly sub sampling the data of Groups A and B to simulate 100 different analyses of both groups with sample sizes equal to the limiting sample sizes of Group C, that is, 14 specimen identifications using 100 previously described species. This approach was similar to a bootstrap technique but we only aimed to compare the tree groups and give some indication of confidence limits on the results.

3 RESULTS AND DISCUSSION

A single identification event for *Miconia trianaei* Cogn. is shown in Table 1.

Summary results of the 18 identification events in Group B are shown in Table 2. The rank positions of species representing the correct identification in each of the separate identification events have here been collectively ranked again.

The resulting distribution of position values is strongly skewed towards the better results with the median lower than the mean. This pattern was also seen for Groups A and C. Furthermore, there were many identification events where the computer's first indication was confirmed as being the correct species identification, and this was the mode (mode = 1) for each of the tree groups. These results exceeded our expectations.

Nevertheless, the results showed that some live specimens varied a good deal from their paired herbarium specimen. The information in Table 2 on Group B highlights one method of lessening the impact of general character variation on identification success, and that is differential character weighting. We have already used some of the more constant characters of phyllotaxy and leaf type to separate four tree groups. For the purposes of analysis we had to discard Group D (compound leaves and opposite phyllotaxy) for lack of data whereas in practice it will be a useful category. From the point of view of a working system we clearly could have gone further with this key-like separation of groups, for in Table 2 *Amaioua guianensis* Aublet and *Vochysia tucanorum* Mart., the lowest placed identifications, both have rare verticillate phyllotaxy. Fortunately, rarity weighting did ameliorate their positions substantially.

It appears reasonable to judge the success or efficiency of this type of identification by the number of images it would be necessary to view before arriving at correct identities for all the test specimens. Furthermore, the conspecific relations appear so strong that we cannot advise the designation of poor performances for individual pairwise comparisons as outliers not entering the statistical results. Therefore the mean position of all identification events for a tree group may be considered the best indicator of identification performance.

The most striking result was that a common set and an extended set of vegetative morphological characters, produced almost identical good performances, measured by mean position, across three tree groups (Table 3). This was despite substantial differences in number of reference species per group. It can also be seen that rarity weighting consistently improved the performance of the similarity index for all groups. From the results with rarity weighting an increase in performance can be seen to follow the actual decrease in number of reference species, as should be expected. Some of the better performance of Group C could also have come from the use of an extended character set, as adding

Table 1 – An identification event showing the result of the comparison of the test specimen of *Miconia trianaei* Cogn. (Melastomataceae) with 179 species of Group B. Each reference species was represented by only one specimen. Only the 22 species most similar to the test specimen are shown, ranked by a descending sort on the similarity index (SI). A perfect correspondence is shown by the test specimen compared to itself, when SI = 100.

Tabela 1 – *Um processo de identificação mostrando o resultado da comparação do espécime de teste de Miconia trianaei Cogn. (Melastomataceae) com 179 espécies do Grupo B. Cada espécie de referência foi representada por um espécime só. Aqui são mostradas apenas as 22 espécies mais semelhantes ao espécime, classificadas em ordem descendente no índice de similaridade (SI). Uma correspondência perfeita é mostrada pelo espécime de teste comparado com si mesmo, quando SI = 100.*

Specimen	Species	Position	SI
4120802	*Miconia trianaei* Cogn. ---------------------- **test specimen**		100.00
16549	*Miconia trianaei* Cogn.	1.5	81.69
13292	*Tibouchina adenostemon* (DC.) Cogn.	1.5	81.69
9476	*Heterocondylus vauthierianus* (DC.) R.M.King & H.Rob.	3.0	80.28
16943	*Aegiphila lhotskiana* Cham.	5.0	78.87
18941	*Miconia cinerascens* Miq.	5.0	78.87
17311	*Miconia pusilliflora* (DC.) Triana	5.0	78.87
17069	*Austroeupatorium inulifolium* (Kunth) R.M.King & H.Rob.	7.5	77.46
13305	*Leandra lacunosa* Cogn.	7.5	77.46
16257	*Miconia urophylla* DC.	7.5	77.46
18879	*Tibouchina granulosa* Cogn.	7.5	77.46
9913	*Leandra pectinata* Cogn.	13.0	76.06
17314	*Tibouchina arborea* (Gardner) Cogn.	13.0	76.06
16859	*Tibouchina candolleana* (DC.) Cogn.	13.0	76.06
6853	*Tibouchina multiflora* (Gardner) Cogn.	13.0	76.06
17315	*Tibouchina mutabilis* Cogn.	13.0	76.06
16857	*Leandra scabra* DC.	19.0	74.65
17309	*Miconia eichlerii* Cogn.	19.0	74.65
16255	*Miconia latecrenata* (DC.) Naudin	19.0	74.65
17146	*Miconia pepericarpa* DC.	19.0	74.65
9802	*Miconia theaezans* (Bonpl.) Cogn.	19.0	74.65
17319	*Mollinedia triflora* (Sprengel) Tul.	19.0	74.65
14590	*Siparuna guianensis* Aublet	19.0	74.65

characters to the data matrix generally helps to distinguish between species (GRAYBEAL, 1998).

When the analysis was standardized by using a common character set for all tree groups and by sub sampling to test 14 specimens against 100 reference species for each of the tree groups (Table 4), the reduced pool of reference species caused an absolute increase in performance as measured by mean position for Groups A and B. This did not happen for Group C because there was no change in its sample sizes from that shown in Table 3. What did occur for Group C was a reduction in absolute performance associated with a smaller character set (GRAYBEAL, 1998). As in the previous analysis, rarity weighting improved the performance of Groups A and B, but this time there was no improvement in the performance for Group C, which in fact deteriorated further.

Table 2 – Summary of identification results for 18 species of Group B when compared with 179 reference species. Each line represents the principal data of an identification event and includes the average rank position together with the similarity index. Compare the line showing *Miconia trianaei* Cogn. with that in Table 1. The identification events have been classified in order of average rank positions, then by species name. The lack of order of the similarity index (SI) highlights the fact that the identification events were separate, independent events.

Tabela 2 *– Sumário de resultados de identificações para 18 espécies do Grupo B quando comparadas com 179 espécies de referência. Cada linha representa os dados principais de um processo de identificação e inclui a posição média junto ao índice de similaridade. Compare a linha de Miconia trianaei Cogn. com a da Tabela 1. Os processos de identificação foram classificados primeiramente em ordem de posições médias, e depois, pelo nome das espécies. A falta de ordem dos índices de similaridade destaca que os processos de identificações foram processos separados e independentes.*

Position	SI	Species and author(s)	Test specimen	Reference specimen
1.0	83.10	*Chomelia obtusa* Cham. & Schltdl.	4122801	16462
1.0	88.73	*Miconia cinnamomifolia* (DC.) Naudin	4111203	12879
1.0	81.69	*Psidium rufum* DC.	4121705	16898
1.0	81.69	*Tibouchina stenocarpa* (DC.) Cogn.	4112606	16550
1.0	84.51	*Vismia brasiliensis* Choisy	4121003	12379
1.5	83.10	*Hyptidendron asperrimum* (Epling) Harley	4113003	17111
1.5	81.69	*Miconia trianaei* Cogn.	4120802	16549
1.5	83.10	*Mollinedia argyrogyna* Perkins	4102902	16558
1.5	83.10	*Myrcia tomentosa* (Aublet) DC.	24102901	16480
3.0	76.06	*Miconia eichlerii* Cogn.	4120205	17309
4.0	73.24	*Leandra scabra* DC.	4122803	16857
4.0	77.46	*Myrcia splendens* (Swartz) DC.	4120501	16891
5.0	78.87	*Gomidesia anacardiifolia* (Gardner) O.Berg	4113004	18955
7.0	81.69	*Ixora gardneriana* Benth.	4121303	16294
7.5	71.83	*Chomelia sericea* Müll.Arg.	4122702	18110
8.0	80.28	*Myrcia venulosa* DC.	4121002	16893
15.5	71.83	*Amaioua guianensis* Aublet	4112501	16913
33.0	71.83	*Vochysia tucanorum* Mart.	4120101	9855
Group mean:				
5.44	79.66			

This analysis, however, was designed to enable a comparison of identification success between tree groups. The identification performance for Group A was better than for Group B and both were much better than for Group C. The difference in performance between Groups A and B may not prove statistically significant, but the difference between them and Group C most certainly is.

Interestingly, all the test specimens together with the reference material for Group A were described by the same expert, while the reference material of Groups B and C was mainly described by an undergraduate assistant acting as a non-expert. Therefore, it is encouraging that the difference between identification success of Groups A and B was not very much. Overall it appears that briefly trained students can succeed with the CARipé system almost as well as an expert and the actual benefits to them of using the system must be commensurately more.

Table 3 – Identification performance comparisons, with and without rarity weighting, for three tree groups of different sample size. All positions and similarity indices (SI) are group means. Higher rank positions appear as lower values and represent more efficient identifications. A basic set of 71 characters was used for Groups A and B and extended to 83 characters for Group C. Lower values of "position" represent better performance.

Tabela 3 *– Comparações de performance de identificação, com e sem um peso de raridade, para três grupos arbóreos com números diferentes de amostras. Todas as posições e os índices de similaridade (SI) são médias de grupos. Posições mais favoráveis aparecem como valores baixos e representam identificações mais eficientes. Um conjunto básico de 71 caracteres foi usado para os Grupos A e B, e aumentado para 83 caracteres para o Grupo C. Valores menores de "posição" representam performance melhor.*

Group A Nr = 279		Group B Nr = 179		Group C Nr = 100		Means for 3 groups	
Position	SI	Position	SI	Position	SI	Position	SI
5.37	80.09	5.44	79.66	4.96	72.12	5.30	78.17
With rarity weighting:							
5.27	50.48	4.94	49.15	4.29	45.04	4.95	48.86

* Nr: number of reference species in a tree group.
** Nr: número de espécies de referência em um grupo arbóreo.*

Table 4 – Standardized identification performance. Results from 100 random subsamples of each of Groups A and B and from the entire sample of Group C. For each sample or sub-sample, 14 specimens were compared with 100 species and the same set of 71 characters was used throughout. For Groups A and B the minimum and maximum group mean position values from the sub-samples are shown. The mean positions are given for all tree groups. Lower values represent better performance.

Tabela 4 *– Performance padronizada de identificação. Resultados a partir das sub-amostras dos Grupos A e B, e da amostra inteira do Grupo C. Para cada amostra ou sub-amostra, 14 espécimes foram comparados com 100 espécies e o mesmo conjunto de 71 caracteres foi usado. Para grupos A e B são mostrados os valores mínimos e máximos das posições médias de grupos a partir das sub-amostras. As posições médias são apresentadas para todos os três grupos. Valores menores representam performance melhor.*

Group A			Group B			Group C
min	mean	max	min	mean	max	mean
1.71	2.68	3.86	1.86	3.48	4.54	7.79
With rarity weighting:						
1.50	2.55	3.79	2.29	3.23	4.25	8.14

4 CONCLUSIONS

The taxonomic database system CARipé greatly facilitated the description and imaging of over 700 specimens from 554 tree species previously acquired and maintained by the ESAL Herbarium. The evaluation of the identification system Empar involved painstaking fieldwork to acquire valid test material and further analysis. It has indeed been possible to alleviate the burden of identifying tropical trees. What would have been in the past a major task of initially populating the database with hundreds of descriptions of previously named herbarium specimens has been turned into a modest project feasible using briefly trained undergraduate assistants. Once this is in place, the description of unknown specimens is understandably a more agreeable process, especially when at the end of a description the computer responds at the touch of a button with a list of likely candidate species as the identity, together with ready access to increasingly high-quality images of herbarium exsiccates and, with time, various images of live material.

It appears that "barcodes" based on vegetative morphology are the obvious choice for identification of trees by region. It is also important to realize that the success of an identification system based on just one specimen per species is perfectly compatible with a population genetics view of the species, though data on more specimens are easily added to the database using CARipé and used to advantage by Empar. Contrast this with the essentially synthetic approach using several if not hundreds of specimens per species, on which keys and phylogenetic work absolutely depend. Therefore, we recommend to herbarium keepers, or curators, this approach of describing at least one specimen of each species in their guardianship. A methodology that is also being applied, together with the fully arrived digital photography, to living plants.

5 ACKNOWLEDGEMENTS

The first author is grateful for funding that enabled his research for this paper: the *Fundação de Amparo à Pesquisa do Estado de Minas Gerais*, FAPEMIG; the Kew Latin America Research Fellowship Programme, KLARF, for a Prance Fellowship in Neotropical Botany; and the Margaret Mee Foundation.

6 BIBLIOGRAPHIC REFERENCES

DAVIS, P. H.; HEYWOOD, V. H. **Principles of angiosperm taxonomy**. Edinburgh: Oliver and Boyd, 1963. 556 p.

FERRI, M. G.; MENEZES, N. L.; MONTEIRO, W. R. **Glossário ilustrado de botânica**. São Paulo: Livraria Nobel, 1981. 197 p.

GRAYBEAL, A. Is it better to add taxa or characters to a difficult phylogenetic problem? **Systematic Biology**, London, v. 47, n. 1, p. 9-17, Mar. 1998.

HARGREAVES, P. **Forest trees of the middle and upper Rio Grande in Southeast Brazil**: evaluation of vegetative morphological characters and of automated systems for the description and identification of species. 2005. 137 p. Thesis (Doctorate in Forestry) – Universidade Federal de Lavras, Lavras, 2005.

HARGREAVES, P. Vegetative morphology for species identification of tropical trees: family distribution. **Cerne**, Lavras, v. 12, n. 1, p. 1-7, Jan./Mar. 2006.

LAWRENCE, G. H. M. **Taxonomy of vascular plants**. New York: Macmillan, 1951. 823 p.

OLIVEIRA-FILHO, A. T. **TreeAtlan**. Disponível em: <(http://www.icb.ufmg.br/ treeatlan/) 2009.

OLIVEIRA-FILHO, A. T.; VILELA, E. A.; GAVILANES, M. L.; CARVALHO, D. A. Comparison of the woody flora and soils of six areas of montane semideciduous forest in southern Minas Gerais, Brazil. **Edinburgh Journal of Botany**, Edinburgh, v. 51, n. 3, p. 355-389, 1994.

PANKHURST, R. J. **PANKEY**: programs for the identification and descriptions of plants or animals. New York: Exeter Software, 1995. 154 p.

PANKHURST, R. J. Principles and problems of identification. In: FORTUNER, R. (Ed.) **Advances in computer methods for systematic biology**. Baltimore: Academic, 1993. p. 125-136.

RATNASINGHAM, S.; HEBERT, P. D. N. **BOLD**: the barcode of life data system. Disponível em: <http://www.barcodinglife.org>. Acesso em: 15 nov. 2007.

RIBEIRO, J. E. L. da S. et al. **Flora da Reserva Ducke**: guia de identificação das plantas vasculares de uma floresta de terra-firme na Amazônia Central. Manaus: INPA-DFID, 1999. 793 p.

SNEATH, P. H. A.; SOKAL, R. R. **Numerical taxonomy**. San Francisco: Freeman, 1973.

VAN DEN BERG, E. ; OLIVEIRA-FILHO, A. T. Spatial partitioning among tree species within an area of tropical montane gallery forest in south-eastern Brazil. **Flora**, Jena, v. 194, n. 2/3, p. 249-266, July 1999.