Regression analyses of southern African ethnomedicinal plants: prioritising plant selections for bioprospecting and pharmacological screening

Errol Douwes [a], Neil R. Crouch [b, c, *], Trevor J. Edwards [a], Dulcie A. Mulholland [c, d]

[a] School of Biological and Conservation Sciences, University of KwaZulu-Natal, Private Bag X01, 3209 Scottsville, South Africa

[b] Ethnobotany Unit, South African National Biodiversity Institute, P.O. Box 52099, 4007 Berea Road, South Africa

[c] School of Chemistry, University of KwaZulu-Natal, Howard College Campus, 4041 Durban, South Africa

[d] Natural Products Research Group, Division of Chemical Sciences, Faculty of Health and Medical Sciences, University of Surrey, Guildford, Surrey, GU2 7XH, United Kingdom

Received ; accepted

*Corresponding author. *E-mail address*: **crouch@sanbi.org** (N.R. Crouch)

Structured abstract

Ethnopharmacological relevance

Regression analyses of local medicinal floras are considered potentially useful when prioritising candidate plant taxa for pharmacological / bioprospecting investigations.

Aim of study

To identify plant orders and subsequently families most highly selected for by traditional healers in southern Africa.

Materials and methods

Using data sourced from the SANBI MedList database, the most comprehensive inventory of ethnomedicinal plants for the *Flora of southern Africa* region, taxa were grouped by order. A least squares regression analysis was used to test the null hypothesis

that the use of these plants by traditional healers is strictly random. Of 'hot' orders subsequently identified, characteristics of taxa therein were assessed to better determine the roles played by i) growth forms, and ii) inherent chemical diversity, in plant selections by ethnomedicinal practitioners.

Results

Analyses identified seven principally 'hot' plant orders (Malpigiales, Fabales, Gentianales, Asteraceae, Solanales, Malvales and Sapindales) and 'hot' families therein from a total of 55 regional ethnomedicinal orders. Five 'cold' ethnomedicinal orders (Rosales, Proteales, Poales, Asparagales and Caryophyllales) were shown to be significantly less represented in the medicinal flora than predicted. No clear growth form preferences were identified across orders. The presence of highly diverse bioactives was evident in the 'hottest' plant families from 'hot' plant orders.

Conclusions

These 12 outliers identified by the regression analyses allowed for the falsification of the null hypothesis. Indications are that 'hot' taxa are selected traditionally on the basis of bioactivity, which is reflected in chemical diversity.

Abstract

Using data sourced from the SANBI MedList database, the most comprehensive inventory of ethnomedicinal plants for the *Flora of southern Africa* region, taxa were grouped by order. A least squares regression analysis was used to test the null hypothesis that the use of these plants by traditional healers is strictly random. The above approach, following a North American study by Moerman (1979), sought to identify plant orders and subsequently families most highly selected for by traditional healers in southern Africa. The analyses identified seven principally 'hot' plant orders (Malpigiales, Fabales, Gentianales, Asteraceae, Solanales, Malvales and Sapindales) and 'hot' families therein from a total of 55 regional ethnomedicinal orders. Five 'cold' ethnomedicinal orders (Rosales, Proteales, Poales, Asparagales and Caryophyllales) were shown to be significantly less represented in the medicinal flora than predicted. These 12 outliers allowed for the falsification of the null hypothesis. Such regression analyses of local

medicinal floras are considered potentially useful when prioritising candidate plant taxa for pharmacological / bioprospecting investigations. Growth forms of taxa in 'hot' orders were assessed to better determine the role this factor plays in plant selection by ethnomedicinal practitioners. Results indicated no clear growth form preferences across orders, but rather selection in line with growth forms naturally dominating those orders. However, when the dominant chemical compound classes of the 'hottest' plant families from 'hot' plant orders were considered, the presence of highly diverse bioactives was revealed. Thus 'hot' taxa are probably selected traditionally on the basis of bioactivity which is reflected in chemical diversity.

1. Introduction

The historical development of pharmaceuticals and other novel drugs has proceeded primarily through the extraction of efficacious compounds from plants (Farnsworth and Bingel, 1977), identified through a variety of bioactivity screening programs (Hunter, 2001). Selection of plants for extract screening can be achieved in one of two ways (Cox, 1990): i) random selection, where no regard is taken of the taxonomic affinities, ethnobotanical context or other intrinsic qualities; or ii) targeted or focused selection, by means of phylogenetic surveys (close relatives of plants known to contain useful compounds are sampled), ecological surveys (plants in particular habitats with particular growth habits), or ethnopharmacological surveys (identifying plants used by indigenous peoples to target specific diseases)(Farnsworth and Bingel, 1977). The low probability of finding useful compounds in random plant screening programmes (approx. one plant sample in 10 000 will show promising activity of interest to researchers), particularly in areas of high biodiversity, is one reason why private drug companies are reluctant to engage in bioprospecting *de novo* (Macilwain, 1998; Soejarto, 1993). Taxol is a notable exception discovered through random screening (Cox and Balick, 1994; Cragg et al., 1993).

Focused selection and ethnobotanical screens in particular, have shown relatively high success rates (Cox, 1990, 1994; Farnsworth, 1990; Farnsworth et al., 1985). Ethno-directed research has reportedly contributed approximately 74% of all pharmaceutical

drugs derived from plants (Farnsworth et al., 1985). Prospecting of plants and most particularly medicinal floras is likely to continue into the foreseeable future, due to complementary advances in bioassay techniques (Tyler, 1986), and historical successes. However, given the high research and development costs associated with novel drug development, optimisation of the discovery phase of the bioprospecting process is continually sought after, especially in situations where the resource base is extensive as well as highly diverse. One such bioresource-rich environment is the *Flora of southern Africa* (*FSA*) region which includes more than 70 major vegetation units (Acocks, 1953) nested within the subcontinent's seven floristically distinct biomes (Rutherford, 1997). The *FSA* region includes the following countries: Namibia, Botswana, Swaziland, Lesotho and South Africa (Germishuizen and Meyer, 2003), occupying the area south of the Kunene, Okavango and Limpopo Rivers. An estimated 80% of the 24,300 plant taxa recorded for the *FSA* region are endemic (Goldblatt, 1978). The flora is estimated to constitute approximately 10% of global plant diversity, of which only a relatively small percentage has been investigated pharmacologically (Eloff, 1998).

The various focused methods that have been employed to identify efficacious ethnomedicinal plants worthy of research vary (Trotter, 1986) include: i) cross-cultural comparisons, where plant efficacies are inferred from the extent to which they are used across different ethnic groups or cultures; ii) the extent of selective borrowing and diffusion of herbal remedies by various ethnic groups or cultures; iii) market and household garden-based studies which identify popular plants or those with high trade volumes; iv) the collection and analysis of case histories and related plant use anecdotes which may prove to be instructive. In addition, Buenz et al. (2005) reported that correlations between ancient and current plant use practices suggest that the taxa in question are indeed effective treatments. The question of how ethnomedicinal practitioners select plants has also often been posed (Moerman, 1979, 1991). Adler and Hammett (1973) postulated that such plant selection is undertaken on a strictly symbolic basis, and that reported therapeutic benefits are of a placebo effect. If so, it could be assumed that symbolic selection of plant taxa is random, in so far as the proportion of taxa selected from any given family or order will be equal. Moerman (1991) proposed this null-hypothesis in an analysis of the patterns of collective ethnobotanical plant use by

Native Americans. However, by means of a least squares regression analysis, he identified a distinct bias towards the use of certain taxonomic groups ('hot' taxa) in the treatment of particular diseases, and so disproved the null hypothesis. Moerman demonstrated that the use of regression analyses is a simple yet effective means of reducing a large number of disparate ethnomedicinal taxa to a manageable group which are likely to display relevant bioactivities. Analyses comparing the actual number of medicinal taxa in a family with the probability distribution for numbers of medicinal taxa in that family (using a random test hypothesis) showed results comparable with the least squares regression analysis (Moerman and Estabrook, 2003). On the assumption that such 'hot' taxa are efficacious, their preferential selection for screening in pharmacological and bioprospecting programmes would be justified.

Clark et al. (1997) identified relevant criteria and formulated a semi-quantitative scoring system to help streamline plant selection for plant molluscicidal agents. Examples of desirable characteristics (criteria) included relative toxicity, availability of plants, plant growth characteristics, localisation of activity (plant part), physical and chemical stability, ethnobotanical use, ease of extraction and ease of application. The system of Clark and co-workers allowed for the identification of 63 short-listed taxa, of which six were prioritised for preliminary screening. Their system sought to identify species that could be utilised in a relatively crude way by communities and as such has limited application for more sophisticated bioprospecting approaches. However, the objectivity of such plant candidate selections and the ease with which the weighting system could be modified were highlighted by the authors as key advantages. A plant selection procedure applied to select antimalarial drug candidates (Clarkson et al., 2004) was subsequently modelled on that of Clark and her co-workers, and proved highly successful: extracts of 49% of species assayed exhibited promising antiplasmodial activity ($IC_{50} \leq 10\mu g/ml$). Whilst criteria related to ethno-directed selection were duly included in the system of Clarkson et al. (2004), *viz.* the traditional use of the species against the target disease, and popularity in the local ethnomedicinal plant trade, a further generic element could have been included to further maximise the likelihood of identifying positive screening leads. This criterion relates to the identification (and subsequent) weighting of 'hot' taxonomic groups identified through elucidation of biases towards their use by traditional medical

practitioners. As a proxy, Clarkson et al. (2004) had weighted the chemotherapeutic (antiplasmodial) potentials of the plant families in view, based on documented pharmacological activities of their constituent compound classes.

The present study has sought to identify plant orders and families within the highly diverse ethnomedicinal flora of southern Africa, towards which biases by user groups are demonstrable. The identification of such generic 'hot' taxonomic groups is expected to facilitate the prioritising of plant selections for broad-based bioprospecting, and pharmacological screening. Further, assessments of growth forms of regional representatives of 'hot' orders, and their phytochemical diversity are undertaken to explore reasons for their selection by practitioners.

2. Materials and methods

The overall approach to data organisation and analysis is presented in Figure 1.

Insert Figure 1

2.1 Data source and organisation

The SANBI MedList database (SANBI, 2004) held data on 3371 taxa (both indigenous and naturalised), from 1227 genera grouped into 211 families. This dataset was an electronic update of the annotated ethnomedicinals checklist of Arnold et al. (2002). In the current study, taxonomic groupings at genus and species levels conformed to the PRECIS database (SANBI, 2005), a curated list of all valid plant taxon names for the *FSA* region, while groupings at order and family levels followed APG II (2003) for angiosperms, with the exception of the Balanophoraceae, Bruniaceae and Vahliaceae. To accommodate currently these three families, which were not grouped into any order by APG II (2003), the Balanophoraceae were grouped with the Santalales, and the Bruniaceae and Vahliaceae with the Rosales, consistent with Cronquist (1988). Gymnosperms and pteridophytes were classified according to Bowe et al. (2000) and Chaw et al. (2000), and Germishuizen and Meyer (2003) respectively. The subsequent

regrouping enumerated a total of 193 families in 55 orders. Plants in the database exhibited a wide spectrum of growth forms, including trees, shrubs, climbers, herbs and geophytes.

2.3 Primary regression analyses

A least squares regression analysis measuring the association between the ethnomedicinal taxa and the total number of taxa present in the *FSA* region (both indigenous and naturalised plants included)(SANBI 2005) was performed. The entire dataset was incorporated into the primary analysis. A mathematical model for predicting the association between plant orders with ethnomedicinal taxa and the total number of taxa in those orders was obtained from the least-squares regression analysis. Two assumptions were made: firstly, that due to the extensive literature review conducted during the compilation of the SANBI MedList database (Arnold et al., 2002), the data constituted a census rather than a sample of the ethnomedicinal taxa in southern Africa. This assumption eliminated the need for statistical tests of significance designed to give confidence that the sample was representative of a larger data body. Census data implies that all individuals in the population were accounted for. The second assumption was that ethnomedicinal taxa used in the analysis are the only southern African plants with any ethnomedicinal potential. Data therefore included i) all recorded ethnomedicinal taxa in the *FSA* (grouped by order), and ii) the total number of taxa in the *FSA* (grouped by order). The population correlation coefficient (ρ) indicates the strength of the relationship between these two groups of variables. Total number of orders and families were considered independent variables, and ethnomedicinal taxa as dependant variables. Residual values were calculated by subtracting the predicted number of ethnomedicinal taxa used per order from the actual number of ethnomedicinal taxa used per order. The population variance calculated from these residuals was used to identify all outliers, i.e. orders which showed notably different values from those predicted. Ethnomedicinal taxa (grouped by order) were plotted against total taxa (grouped by order), and the regression line (equation obtained from the regression analysis) was overlaid to allow for visual

assessment of i) any obvious patterns/relationships and ii) the position of any outliers. Residual values correspond to the vertical distance from each datum point to the regression line ($y-\hat{y}$). Positive outliers (orders selected significantly more often than predicted) were further analysed at family level. This required a regression analysis for all ethnomedicinal taxa (grouped by family) against total taxa (grouped by family). Data for families within the selected orders were then filtered out for further scrutiny.

2.3 Secondary regression analyses

Outlying orders and families identified in the primary regression analyses were removed from the data set, and the regression analyses performed again to allow further partitioning of the data. The population variance of residual values was determined and used as a cut-off to identify outlying orders and families. Total taxa (grouped by orders or families) were considered independent variables and ethnomedicinal taxa (grouped by orders or families) as dependant variables.

2.4 Analysis of plant growth forms

In addition to the above analyses, an investigation of plant growth forms of plants present in the highly selected orders was undertaken. Plant growth form data were extracted from Germishuizen and Meyer (2003) and grouped according to four nominal categories, namely: geophyte, climber, tree/shrub and herb/dwarf shrub. All taxa in the respective orders were included, regardless of their variable status as annuals or perennials. For this analysis, the chemical defence strategies of annual and perennial taxa were assumed to not differ.

2.5 Evaluation of phytochemical trends in principal 'hot' families

A data mining exercise which summarised the important compound classes known to occur in selected 'hot' families was undertaken. 'Hot' families were those with the highest residual value in each of the highly selected orders. Compounds known to have been isolated from taxa in 'hot' families were compiled from the Dictionary of Natural Products (DNP)(DNP, 2005), and then grouped according to class, as defined in the DNP. Proportions of compound classes present in each selected family were determined to assess prevalence. The phytochemical data are limited and undoubtedly do not represent all compounds/compound classes present. However, the DNP was the most comprehensive data source available and data are assumed to be sufficiently representative to allow an overview of the important classes. Compound classes notably absent or infrequently listed for the relevant families were also identified. Notably, molecular trees based on *rbc*L DNA analyses provide a useful framework for assessing the comparative merits of secondary compound classes as chemotaxonomic characters (Grayer et al., 1999). For this reason plant families and orders (excluding the Pteridophyta) in the current study were grouped according to recently published phylogenetic trees, as described above (2.1).

3. Results and discussion

The results of the primary least squares regression analysis (Table 1) indicated the presence of a particularly strong linear relationship between ethnomedicinal plant taxa (grouped by order), and the total number of taxa in those orders, i.e. the value of ρ is very close to +1. Figure 2 provides further evidence of this positive relationship. Similar results were later obtained for ethnomedicinal taxa grouped by family (Table 3).

Insert Table 1

Residual values obtained from the regression analysis of ethnomedicinal taxa grouped by plant order ranged from –118.9 to +103.5 (residual values for each of the 55 orders are not presented, rather only those 12 orders significantly greater or less than predicted). The model was able to account for 86% ($\rho^2 = 0.86$) of the variation in the y-values (Table 1). As such, it was necessary to distinguish which orders could be considered outliers, i.e. farthest from the regression line. The population variance of all 55 order residuals (37.47) was employed as a cut-off, leaving 12 orders as outliers (seven positive and five negative)(Table 2). Plants in these orders were considered to have been selected either far more or far less than plants from other orders in the region. The magnitude of the outlying residuals falsified the null hypothesis that traditional user-groups in the region select plants for medicinal purposes in a wholly random manner. The high percentage of variation that the statistical model was able to account for, and the strong positive correlations observed in the regression plots, signified that the model performed well as a predictive tool. A direct comparison of the current results with those presented by Moerman (1991) has not been undertaken, due to the very different floras present in the two regions (North America and southern Africa).

Insert Table 2

The 55 orders containing ethnomedicinal taxa were plotted against the total number of taxa present within those orders in the *FSA* region (Figure 2). The strength of the positive relationship (ρ) is particularly evident. The prominent seven positive, and five negative, outlying orders which influence both the coefficient of determination (ρ^2) and the reliability of predictions made from the line of best fit, are indicated.

Insert Figure 2

Given the extensive botanical diversity of the *FSA* region, taxa identified through the analysis of plant orders are numerous – far more than could reasonably be included in most screening programmes. For this reason, residual values of families within these orders were also calculated (Table 3). Although analyses at family level provides for additional focus, taxa numbers at this level may still be too great e.g. the number of *FSA*

taxa in the Asteraceae alone is 2681 (SANBI, 2005). In such instances, it may be necessary to perform analyses at the generic level to obtain more definition for focused screening. To better understand how families constituting the outlying orders (Table 2) contributed towards their popularity in ethnomedicinal use, a regression analysis for all families was performed; only data for families which constitute the positive outlying orders are presented (Table 3). Families with high positive residual values contributed most to the outlier status assigned to their respective orders.

Outlying 'cold' orders which occur below the regression line (Figure 2), and therefore selected less often by ethnomedicinal practitioners, are of particular interest. Five were identified from a total of 55 orders in the *FSA* region. It is reasonable to argue that these, the Rosales, Proteales, Poales, Asparagales and Caryophyllales should be weighted negatively in candidate selection systems. Plants in these orders may have characteristics which result in their more modest usage. The Poaceae, for example, which are highly utilised by browsers rely primarily on physical attributes (e.g., sharp awns, high lectin and/or high silica content), growth form and compensatory growth rather than secondary metabolite production for defence (Lindroth, 1988; Peumans and Van Damme, 1995). The Poaceae also contain numerous food plants (e.g. maize, millet, etc.) and as a group may generally produce insufficient quantities of bioactive compounds to be of medicinal interest to humans. If this is the case, the same result is likely to be evident globally; it is therefore not surprising that taxa most rarely used medicinally by Native Americans include the Caryophyllaceae and Poaceae (Moerman, 1991). Both these families are in orders selected significantly less often than others by ethnomedicinal practitioners in southern Africa (Table 2; Figure 2).

Insert table 3

When outlying orders and families previously identified as outliers were removed from the dataset and secondary regression analyses performed, the results indicated a strong linear relationship ($\rho = 0.96$) between ethnomedicinal taxa (grouped by order), and total taxa in those orders (Table 4). The population variance of the 43 order residuals (12.03) was employed as a cut-off; again, seven positive and five negative outlying

orders were identified (Table 5), which could be secondarily evaluated for ethnopharmacological insights.

Insert Table 4

Insert Table 5

Growth forms of the positive outlying orders identified in the primary regression analysis were analysed, with results presented as stacked bar charts (Figure 3). In southern Africa, the Asterales and Solanales comprise predominantly herb-like plants and/or dwarf shrubs, while the Sapindales has a greater representation of trees and/or shrubs. Other orders such as the Fabales and Malpighiales have an approximately even mix of growth forms between trees/shrubs and herbs/dwarf shrubs. The Gentianales have the highest percentage of geophytes and climbers within the group. It is not possible at order level to assign significance to any particular growth forms as such a wide range are utilised (Figure 3). It is likely that the distribution of growth forms of ethnomedicinal taxa in 'hot' orders are representative of the growth forms for each order as a whole. Moerman (1979) noted that there 'appears to be some kind of order to the collective ethnobotanical wisdom [of Native Americans] in that the plants they use do show a high likelihood of producing biologically active secondary products'. However, while there may similarly be an ordered basis for plant selection by ethnomedicinal practitioners in the *FSA* region, this is not apparent from an analysis of growth forms (Figure 3). As such, growth forms of southern African ethnomedicinals would not prove useful as weighted criteria in semi-quantitative selection procedures. However, plants have many other characteristics such as the colour/shape/smell and size of fruits, seeds, leaves and flowers, a number of which may influence the selection behaviour of traditional users. Etkin (1986) noted that plant selection by ethnomedicinal practitioners may be patterned in accordance with the belief that certain attributes (e.g. leaf shape or colour) serve to indicate utility relative to a particular ailment or disease. This concept is referred to generally as the Doctrine of Signatures.

Insert Figure 3

Data mining results of compound class data for selected plant families (DNP, 2005) were grouped by class and percentage, relative to the total number of compounds known from each family globally (Table 6).

Insert table 6

Of the families of most significant regional interest (Table 6), the Fabaceae have the greatest percentage of flavonoids, followed by the Anacardiaceae, whilst the Rubiaceae have the greatest percentage alkaloids, followed by the Convolvulaceae. Terpenoid natural products and their derivatives were the dominant compound class for the Asteraceae and Euphorbiaceae. Accordingly, many 'hot' ethnomedicinal families are rich in chemical classes with known bioactive metabolites such as flavonoids, alkaloids and terpenoids (Balandrin et al., 1985; Bruneton, 1995; Yao et al., 2004). This finding concurs with previous reports that show ethnobotanical plant selections, (relative to random selections), yield better results by enhancing hit rates of pharmacologically active compounds (Hamburger and Hostettmann, 1991; Macilwain, 1998; Marles and Farnsworth, 1994; Soejarto, 1993). A detailed knowledge of compound class proportions in plants could potentially be applied in weighted selection systems to prioritise candidate taxa for bioactivity evaluation.

It was assumed, for the purpose of this study, that individual ethnomedicinal taxa are efficacious against the diseases for which they are reportedly used. Their phytochemical traits were thus regarded as being correlated through common descent as opposed to convergent evolution. The assumptions were based on reports that most kinds of secondary compounds, including tannins and alkaloids, are phylogenetically conservative in their distribution (Silvertown and Dodd, 1996). The isolation of popular ethnomedicinal orders/families by means of regression analyses is therefore considered appropriate for i) the identification of related taxa with similar bioactive constituents and

ii) the prioritisation of taxa for bioprospecting purposes. As detailed phylogenies were not incorporated into the analyses, all species were treated as independent data points. The phylogenetic independence of claimed relationships was therefore not confirmed and some degree of pseudo replication is expected (Silvertown and Dodd, 1996).

Understanding why different plants produce different secondary metabolites is an important consideration in the field of bioprospecting, as such insights allow for optimising of plant selections. The previously accepted consensus, that secondary metabolite production was related primarily to enhancing the fitness of the producer, has been undermined by data from pharmaceutical and agrochemical industries (Firn, 2003). Firn reported that the pharmaceutical and agrochemical industries, through experience with numerous screening programmes, have realised that a very low probability exists of finding useful compounds from either man-made or naturally made chemicals. This, according to Firn and Jones (2000) is due to the requirement of a very precise three-dimensional match between charge distribution on an efficacious biochemical compound and the surface of the target protein it is required to interact with. Jones and Firn (1991) have proposed that evolution favours organisms that can generate and retain the greatest sustainable chemical diversity at lowest cost. Such organisms would have an increased likelihood of enhanced fitness due to the greater chances of producing rare chemicals with potent biological activity. As such, the majority of natural products found in plants are unlikely to possess potent biological activity. If this argument is accepted, then bioprospectors need to identify those taxa with high chemical diversity as these would then provide the greatest potential for drug development. Table 6 reveals that the Euphorbiaceae is particularly diverse chemically, with no less than 1.7% representation of any major constituent class. This same family has contributed most to the outlier status of its respective order, the Malpigiales (Table 3), which has been used significantly greater than predicted (Table 2). In that the Euphorbiaceae is therefore a good bioprospecting taxon on which to focus, the value of combining data on ethnomedicinal applications and phylogenetics is exampled.

Native Americans have shown a predilection for selecting particular ethnomedicinal families, regardless of family size (Moerman et al., 2003). This is likely due to related plants showing similar efficacy against certain diseases of man due to

heritable similarities in secondary metabolites, in which case phylogenetic considerations are important in the selection of candidate bioprospecting subjects. Although disputed by some influential systematists it is generally accepted that related taxa share chemical characteristics, to the extent that phytochemicals can be used as taxonomic characters in classification (Cronquist, 1980; Grayer et al., 1999; Waterman, 1999). The early work by Robert Hegnauer is particularly relevant (Grayer et al., 1999) due to his attempt to understand the distribution of secondary (and some primary) metabolites in the plant kingdom and the phylogenetic relationships of plant families based on chemical profiles (Grayer et al., 1999). Hegnauer's early work was controversial but it was later endorsed by a number of systematists (Dahlgren, 1975; Thorne, 1981), who included chemical characters when constructing their classifications.

The structural diversity of plant compounds has likely increased along with other changes observed in the course of plant evolution (Hegnauer, 1967; Heinrich et al., 2004). Certain compound classes are sometimes restricted taxonomically, e.g. sesquiterpene lactones are limited to the Asteraceae, Apiaceae, Burseraceae, Lauraceae and Magnoliaceae (Dahlgren et al., 1981). Records documenting the occurrence of pharmacologically active secondary metabolites within monophyletic assemblages are therefore of particular interest. Homology in such groups, may lead to the evolution of compounds with similar pharmacological activity. Alternatively, production of the same or similar compounds in unrelated taxa through convergent evolution may be an indicator of endowed selective fitness due to compound efficacy (Dahlgren et al., 1981). Compound classes present across broad polyphyletic groups are generally unlikely to aid bioprospectors identify particularly efficacious taxa. However, convergent clades with known efficacious taxa may prove useful through the provision of independent sets of relatives to investigate. The divergent, convergent or parallelist nature of biosynthetic pathways producing such compounds may also prove insightful to chemists attempting laboratory syntheses.

Comparative methods (e.g. least squares regression analyses) are common tools for investigating trait correlations (Felsenstein, 1985; Harvey and Pagel, 1991; Westoby et al., 1995). However, the current lack of detailed phylogenies for the majority of southern African higher order taxa will likely generate some degree of pseudo replication

when applied locally. Further, whilst this study has assumed a census dataset it is acknowledged that historical regional human settlement patterns, and subsequent socio-cultural interactions, may have significantly shaped the documented body of traditional plant-use knowledge (e.g. Watt and Breyer-Brandwijk 1962; SANBI 2004). This last-mentioned factor could well skew the results of any regression analyses, and accordingly is worth investigating.

Conclusions

The use of regression analyses has allowed for the identification of plant orders and consequently families most highly selected for by traditional healers in southern Africa. The approach taken allows for flexibility in that assessments can be undertaken at a variety of taxonomic levels. The current analyses identified seven 'hot' plant orders (the Malpigiales, Fabales, Gentianales, Asteraceae, Solanales, Malvales and Sapindales in decreasing order of bias by user groups) and 'hot' families therein that hold significantly more ethnomedicinal taxa than predicted. The magnitude of the outlying residuals falsified the null hypothesis that traditional user-groups in the region select plants for medicinal purposes in a wholly random manner. An assessment of growth forms of taxa in 'hot' orders failed to elucidate the role this factor might play in plant selections by ethnomedicinal practitioners. An analysis of the dominant chemical compound classes of the 'hottest' plant families from 'hot' plant orders revealed that these taxa are documented to produce a wide variety of pharmacologically-active compounds. The results obtained overall are considered potentially useful in the rational prioritisation of plant subjects for regional ethnopharmacology studies, and in optimising bioprospecting ventures. This is particularly so when findings are applied in conjunction with the latest chemotaxonomic data available for the identified plant families.

Acknowledgements

We gratefully acknowledge the Department of Science and Technology of South Africa for the Innovation Fund Grant (TM1002FP). The staff of the Mary Gunn Library in

Pretoria is thanked for facilitating access to literature, and the Data Section at the National Herbarium of South Africa (SANBI) for provision of both PRECIS and MedList data. Dr Robin Mackey (UKZN) provided statistical assistance. E.D. acknowledges with gratitude the receipt of an NRF grant-holder bursary, a SANBI postgraduate bursary, the Ward Memorial bursary (WESSA), and financial support from the Ajax foundation, and Chris Davidson and Sharon Cristoph.

References

Acocks, J.P.H., 1953. Veld types of South Africa. Memoirs of the Botanical Survey of South Africa 28, 1–192.

Adler, H.M. and Hammett, V.B., 1973. The doctor-patient relationship revisited. An analysis of the placebo effect. Annals of Internal Medicine 78, 595–598.

APG II, 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. Botanical Journal of the Linnean Society 141, 399–436.

Arnold, T.H., Prentice, C.A., Hawker, L.C., Snyman, E.E., Tomalin, M., Crouch, N.R. and Pottas-Bircher, C. (Eds.), 2002. Medicinal and magical plants of southern Africa: an annotated checklist. National Botanical Institute, Pretoria, South Africa.

Balandrin, M.F., Klocke, J.A., Wurtele, E.S. and Bollinger, W.H., 1985. Natural plant chemicals: sources of industrial and medicinal materials. Science 228, 1154–1160.

Bowe, L.M., Coat, G. and DePamphilis, C.W., 2000. Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. Proceedings of the National Academy of Sciences 97, 4092–4097.

Bruneton, J., 1995. Pharmacognosy, phytochemistry, medicinal plants. Intercept, Hampshire.

Buenz, E.J., Johnson, H.E., Beekman, E.M., Motley, T.J. and Bauer, B.A., 2005. Bioprospecting Rumphius's Ambonese herbal: Volume I. Journal of Ethnopharmacology 96, 57–70.

Chaw, S.-M., Parkinson, C.L., Cheng, Y., Vincent, T.M. and Palmer, J.D., 2000. Seed plant phylogeny inferred from all three plant genomes: monophyly of extant

gymnosperms and origin of Gnetales from conifers. Proceedings of the National Academy of Sciences 97, 4086–4091.

Clark, T.E., Appleton, C.C. and Drewes, S.E., 1997. A semi-quantitative approach to the selection of appropriate candidate plant molluscicides--a South African application. Journal of Ethnopharmacology 56, 1–13.

Clarkson, C., Maharaj, V.J., Crouch, N.R., Grace, O.M., Pillay, P., Matsabisa, M.G., Bhagwandin, N., Smith, P.J. and Folb, P.I., 2004. In vitro antiplasmodial activity of medicinal plants native to or naturalised in South Africa. Journal of Ethnopharmacology 92, 177–191.

Cox, P.A., 1990. Ethnopharmacology and the search for new drugs. In: D.J. Chadwick and J. Marsh (Eds.), Ciba Foundation Symposium, John Wiley and Sons, Chichester, United Kingdom, p. 40–47; discussion 47–55.

Cox, P.A., 1994. The ethnobotanical approach to drug discovery: strengths and limitations. In: D.J. Chadwick and J. Marsh (Eds.), Ciba Foundation Symposium, John Wiley & Sons, Chichester, United Kingdom, pp. 25-36; discussion 36-41.

Cox, P.A. and Balick, M.J., 1994. The ethnobotanical approach to drug discovery. Scientific American 270, 82–87.

Cragg, G.M., Boyd, M.R., Cardellina, J.H., 2nd, Grever, M.R., Schepartz, S., Snader, K.M. and Suffness, M., 1993. The search for new pharmaceutical crops: drug discovery and development at the National Cancer Institute. In: J. Janick and J.E. Simon (Eds.), New Crops, Wiley, New York, USA, p. 161–167.

Cronquist, A., 1980. Chemistry in plant taxonomy: an assessment of where we stand. In: F.A. Bisby, J.G. Vaughan and C.A. Wright (Eds.), Chemosystematics: principles and practice., Academic Press, London, United Kingdom, p. 1–27.

Cronquist, A., 1988. The evolution and classification of flowering plants. New York Botanical Garden, New York, USA.

Dahlgren, R.M.T., 1975. A system of classification of the angiosperms to be used to demonstrate the distribution of characters. Botaniska Notiser 128, 119–147.

Dahlgren, R.M.T., Rosendal-Jensen, S. and Nielsen, B.J., 1981. A revised classification of the angiosperms with comments on correlation between chemical and other characters. In: D.A. Young and D.S. Seigler (Eds.), Phytochemistry and angiosperm phylogeny, Praeger, New York, USA, p. 149–204.

DNP, 2005. Dictionary of Natural Products. Version 13.2. Chapman and Hall/CRC. Hampden Data Services Ltd.

Eloff, J.N., 1998. Which extractant should be used for the screening and isolation of antimicrobial components from plants? Journal of Ethnopharmacology 60, 1–8.

Etkin, N.L., 1986. Multidisciplinary perspectives in the interpretation of plants used in indigenous medicine and diet. In: N.L. Etkin (Ed.), Plants in indigenous medicine and diet: biobehavioral approaches, Redgrave Publishing Company, New York, USA, p. 3–29.

Farnsworth, N.R., 1990. The role of ethnopharmacology in drug development. In: D.J. Chadwick and J. Marsh (Eds.), Ciba Foundation Symposium, John Wiley and Sons, Chichester, United Kingdom, pp. 2-21.

Farnsworth, N.R., Akerele, O., Bingel, A.S., Soejarto, D.D. and Guo, Z., 1985. Medicinal plants in therapy. Bulletin of the World Health Organisation 63, 965–981.

Farnsworth, N.R. and Bingel, A.S., 1977. Problems and prospects of discovering new drugs from higher plants by pharmacological screening. In: H. Wagner and P. Wolff (Eds), New natural products and plant drugs with pharmaceutical, biological or therapeutic activity, Springer-Verlag, Heidelberg, Germany.

Felsenstein, J., 1985. Phylogenies and the comparative method. The American Naturalist 125, 1–15.

Firn, R.D., 2003. Bioprospecting – why is it so unrewarding? Biodiversity and Conservation 12, 207–216.

Firn, R.D. and Jones, C.G., 2000. The evolution of secondary metabolism - a unifying model. Molecular Microbiology 37, 989–994.

Germishuizen, G. and Meyer, N.L. (Eds.), 2003. Plants of southern Africa: an annotated checklist. National Botanical Institute, Pretoria, South Africa.

Goldblatt, P., 1978. An analysis of the flora of southern Africa: its characteristics, relationships and origins. Annals of the Missouri Botanical Garden 65, 369–436.

Grayer, R.J., Chase, M.W. and Simmonds, M.S.J., 1999. A comparison between chemical and molecular characters for the determination of phylogenetic relationships among plant families: an appreciation of Hegnauer's "Chemotaxonomie der Pflanzen". Biochemical Systematics and Ecology 27, 369–393.

Halliwell, B., Rafter, J. and Jenner, A., 2005. Health promotion by flavonoids, tocopherols, tocotrienols, and other phenols: direct or indirect effects? Antioxidant or not? The American Journal of Clinical Nutrition 81, 268S–276S.

Hamburger, M. and Hostettmann, K., 1991. Bioactivity in plants: the link between phytochemistry and medicine. Phytochemistry 30, 3864–3874.

Harvey, P.H. and Pagel, M.D., 1991. The comparative method in evolutionary biology. Oxford University Press, New York, USA.

Hegnauer, R., 1967. Chemical characters in plant taxonomy: some possibilities and limitations. Pure Applied Chemistry 14, 173–187.

Heinrich, M., Barnes, J., Gibbons, S. and Williamson, E.M., 2004. Fundamentals of pharmacognosy and phytotherapy. Churchill Livingstone, Edinburgh, United Kingdom.

Hunter, D., 2001. Life in the fast lane: high-throughput chemistry for lead generation and optimisation. Journal of Cellular Biochemistry (Suppl.) 37: 22–27.

Hutchings, A., Haxton Scott, A., Lewis, G. and Cunningham, A.B., 1996. Zulu medicinal plants: an inventory. University of Natal Press, Pietermaritzburg, South Africa.

Jones, C.G. and Firn, R.D., 1991. On the evolution of secondary plant chemical diversity. Philosophical Transactions of the Royal Society of London (Series B) 333, 273–280.

Lindroth, R.L., 1988. Adaptations of mammalian herbivores to plant chemical defenses. In: K.C. Spencer (Ed.), Chemical mediation of coevolution, Academic Press, San Diego, USA, p. 415–446.

Macilwain, C., 1998. When rhetoric hits reality in debate on bioprospecting. Nature 392, 535–540.

Marles, R.J. and Farnsworth, N.R., 1994. Plants as sources of antidiabetic agents. In: H. Wagner and N.R. Farnsworth (Eds.), Economic and medicinal plant research, Academic Press, London, United Kingdom, p. 149–187.

Moerman, D.E., 1979. Symbols and selectivity: a statistical analysis of native American medical ethnobotany. Journal of Ethnopharmacology 1, 111–119.

Moerman, D.E., 1991. The medicinal flora of Native North America: an analysis. Journal of Ethnopharmacology 31, 1–42.

Moerman, D.E. and Estabrook, G.F., 2003. Native Americans' choice of species for medicinal use is dependent on plant family: confirmation with meta-significance analysis. Journal of Ethnopharmacology 87, 51–59.

Moerman, F., Lengeler, C., Chimumbwa, J., Talisuna, A., Erhart, A., Coosemans, M. and D'Alessandro, U., 2003. The contribution of health-care services to a sound and sustainable malaria-control policy. Lancet Infectious Diseases 3, 99–102.

Peumans, W.J. and Van Damme, E.J., 1995. The role of lectins in plant defence. Histochemistry Journal 27, 253–271.

Rutherford, M.C., 1997. Categorization of biomes. In: R.M. Cowling, D.M. Richardson and S.M. Pierce (Eds.), Vegetation of southern Africa, Cambridge University Press, Cambridge, United Kingdom, p. 91–98.

SANBI, 2004. Medicinal plants (MedList) database. South African National Biodiversity Institute, Pretoria, South Africa, Pretoria.

SANBI, 2005. PRECIS (PRE Computerized Information System) databank. South African National Biodiversity Institute, Pretoria, South Africa, Pretoria.

Silvertown, J. and Dodd, M., 1996. Comparing plants and connecting traits. Philosophical Transactions of the Royal Society of London 351, 1233–1239.

Soejarto, D.D., 1993. Logistics and politics in plant drug discovery. In: A.D. Kinghorn and M.F. Balandrin (Eds.), Human medical agents from plants, American Chemical Society, Washington D.C., USA, pp. 96-111.

Thorne, R.F., 1981. Phytochemistry and angiosperm phylogeny. A summary statement. In: D.A. Young and D.S. Seigler (Eds.), Phytochemistry and angiosperm phylogeny, Praeger Publishers, New York, USA, pp. 233-295.

Trotter, R.T., 1986. Informant consensus: a new approach for identifying potentially effective medicinal plants. In: N.L. Etkin (Ed.), Plants in indigenous medicine and diet: biobehavioral approaches, Redgrave Publishing Company, New York, USA, p. 91–112.

Tyler, V.E., 1986. Plant drugs in the twenty-first century. Economic Botany 40: 279–288.

Waterman, P.G., 1999. The chemical systematics of alkaloids: A review emphasising the contribution of Robert Hegnauer. Biochemical Systematics and Ecology 27, 395–406.

Watt, J.M. and Breyer-Brandwijk, M.G., 1962. The medicinal and poisonous plants of southern and eastern Africa. E & S Livingstone, London, United Kingdom.

Westoby, M., Leishman, M.R. and Lord, J.M., 1995. On misinterpreting the 'phylogenetic correction'. Journal of Ecology 83, 531–534.

Yao, L.H., Jiang, Y.M., Shi, J., Tomas-Barberan, F.A., Datta, N., Singanusong, R. and Chen, S.S., 2004. Flavonoids in food and their health benefits. Plant Foods For Human Nutrition 59, 113–122.