Journal of Ethnopharmacology 119 (2008) 356–364

Contents lists available at ScienceDirect

Journal of Ethnopharmacology

journal homepage: www.elsevier.com/locate/jethpharm

Regression analyses of southern African ethnomedicinal plants: informing the targeted selection of bioprospecting and pharmacological screening subjects

E. Douwes[a,b], N.R. Crouch[b,c,*], T.J. Edwards[a], D.A. Mulholland[c,d]

[a] School of Biological and Conservation Sciences, University of KwaZulu-Natal, Private Bag X01, Scottsville 3209, South Africa
[b] Ethnobotany Unit, South African National Biodiversity Institute, P.O. Box 52099, Berea Road 4007, South Africa
[c] School of Chemistry, University of KwaZulu-Natal, Durban 4041, 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

ARTICLE INFO

Article history:
Received 30 April 2008
Received in revised form 16 July 2008
Accepted 18 July 2008
Available online 7 August 2008

Keywords:
Bioprospecting
Ethnomedicinal plant selection
Regression analyses
Southern Africa

ABSTRACT

Ethnopharmacological relevance: Regression analyses of local medicinal floras are considered potentially useful when prioritising candidate plant taxa for pharmacological/bioprospecting investigations.
Aim of the study: To identify plant orders and subsequently families within the highly diverse ethnomedicinal flora of southern Africa, towards which biases by traditional healers are demonstrable. Taxa so identified can subsequently be weighted appropriately in semi-quantitative selection systems.
Methodology: Plant data sourced from the SANBI MedList database, the most comprehensive inventory of ethnomedicinal plants for the *Flora of southern Africa* region were grouped by order. A least squares regression analysis was applied 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.

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 traditionally to target specific diseases) (Farnsworth and Bingel, 1977). The low probability of finding useful compounds in random plant screening programmes (approximately one plant 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* (Soejarto, 1993; Macilwain, 1998). Taxol is one notable exception discovered through random screening (Cragg et al., 1993; Cox and Balick, 1994).

Focused selection and ethnobotanical screens in particular, have shown relatively high success rates (Cox, 1990, 1994; Farnsworth et al., 1985; Farnsworth, 1990). 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

* Corresponding author at: Ethnobotany Unit, South African National Biodiversity Institute, P.O. Box 52099, Berea Road 4007, South Africa. Tel.: +27 31 2024095; fax: +27 31 2023430.
E-mail address: crouch@sanbi.org (N.R. Crouch).

0378-8741/$ – see front matter © 2008 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.jep.2008.07.040

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 to the 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 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 (Trotter, 1986). 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 is 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 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. A least squares regression analysis was undertaken rather than employing simple counts that over-emphasise large families with high numbers of medicinally used taxa, or indices (e.g. percentages) that over-emphasise small ones. Methods based on percentages do not allow for refined analysis of the relationship between diversity, medicinal activity and selection by traditional healers. The identification of generic 'hot' taxonomic groups using regression analyses is expected to facilitate the prioritising of plant selections for broad-based bioprospecting, and pharmacological screening. We do not advocate the use of regression analyses data alone to rank potential bioprospecting subjects: reasonably weighted "hot" or "cold"' taxonomic groups should be included with other appropriate elements (e.g. popularity in trade, toxicity, uses relevant to bioactivity to be screened for, etc.) in semi-quantitative selection systems. This would ensure that taxa from all relevant families (large and small) still be assessed for their likelihood to yield further leads.

Further, assessments of growth forms of regional representatives of 'hot' orders, and their phytochemical diversity were undertaken to explore reasons for their selection by practitioners.

2. Methodology

The overall approach to data organisation and analysis is presented in Fig. 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, whilst 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), Chaw et al. (2000), and Germishuizen and Meyer (2003) respectively. The subsequent regrouping enumerated a total of 193