



Testing the non-random hypothesis of medicinal plant selection using the woody flora of the Mpumalanga Province, South Africa

Isidore Muleba¹ · Kowiyou Yessoufou¹ · Isaac T. Rampedi¹

Received: 7 August 2019 / Accepted: 29 April 2020 / Published online: 7 May 2020
© Springer Nature B.V. 2020

Abstract

Medicinal plants have been used by local communities to treat all sorts of diseases, and this unique indigenous knowledge has been documented in various studies. However, using this vast knowledge to formulate and test hypothesis in ethnobotany is not yet a common practice in the discipline despite recent calls for more hypothesis-driven ethnobotanical researches. Here, we collected ethnobotanical data on 811 woody plant species in the Mpumalanga Province of South Africa to test the non-random hypothesis of medicinal plant selection, which predicts a positive correlation between the size of plant families and the number of medicinal plants in the families. We tested this hypothesis by fitting the commonly used simple linear regression model and the negative binomial model. Our analysis confirmed the hypothesis and revealed that some plant families are over-utilised—i.e. contain more medicinal plants than expected. The identification of over-utilised families is the first step towards the prioritisation of research efforts for drug discovery. The proportion of over-utilised families ranges from 50% (linear regression with untransformed data) and 55% (linear regression after log–log transformation) to 34% (negative binomial model). With the simple linear model and untransformed data, the top over-utilised families are Fabaceae (residual = +34.44), Apocynaceae (+5.82) and Phyllanthaceae (+5.53). The log-transformed model confirms these three families as the top over-utilised families but in a slightly different sequence: Fabaceae (+1.55), Phyllanthaceae (+0.83) and Apocynaceae (+0.79). However, using the negative binomial model, Fabaceae is no longer even part of the top 10 over-utilised families, which are now Phyllanthaceae (+2.09), Apocynaceae (+1.51), Loganiaceae (+1.48), Rhamnaceae (+1.48), Sapotaceae (+1.48), Oleaceae (+1.39), Salicaceae (+1.39), Clusiaceae (+1.30), Boraginaceae (+1.28) and Lamiaceae (+1.18). This suggests that the relative medicinal value of some families may have been over-estimated in comparison with others. Our study is an illustration of the need to apply appropriate model while testing ethnobotanical hypotheses to inform priority setting for drug discovery.

Keywords Ethnobotanical hypothesis · Fabaceae · Generalised linear model with negative binomial · Phyllanthaceae

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10668-020-00763-5>) contains supplementary material, which is available to authorized users.

Extended author information available on the last page of the article

1 Introduction

A tremendous amount of data on medicinal plant uses has been documented worldwide over the years. Such data may include plants that are used to treat particular diseases, plant organs used, how the plant parts or organs are collected and how such medicines are prepared (York et al. 2011; Elansary et al. 2015; Leso et al. 2017). In the light of this wealthy dataset available in ethnobotany, some authors indicated that we now have enough information with which we should be formulating and testing theories and hypotheses that can advance the scope of ethnobotany as a scientific discipline (e.g. Albuquerque et al. 2006; Ford and Gaoue 2017; Gaoue et al. 2017; Hart et al. 2017). Such paradigm shifts towards a more hypothesis- or theory-driven ethnobotany is necessary to make ethnobotany a stronger scientific discipline with theories and hypotheses that can be used to predict new medicinal plant uses as well as explain plant–human interactions (Gaoue et al. 2017). Interesting questions that can be investigated for a better understanding of plant–human interactions are as follows: is traditional medicine a placebo? Why some plants in a particular family are predominantly used or over-utilised in some pharmacopoeias while other plants are less used (under-utilised)?

To answer this question, a hypothesis has been proposed, termed a “non-random hypothesis” (Moerman 1979, 1991, 1996), which predicts that large families are more likely to be richer in medicinal plants than small-sized families. This hypothesis implies that medicinal plants are not randomly selected by local communities such that a linear positive relationship is expected between the number of medicinal plants in families and the size of those families (Moerman 1979). Initially, Moerman (1979) formulated and tested this hypothesis to demonstrate that the traditional medicine of Native Americans was not a placebo. Because of this non-random selection, some plant families tend to be over- or under-represented in a given pharmacopoeia (Moerman 1979, 1991; Moerman and Estabrook 2003; Ford and Gaoue 2017). This implies that plant family can become a strong determinant of plant use value (Phillips and Gentry 1993), and in one of his early studies, Moerman (1991) already explained this by the fact that species in the same family, due to their evolutionary relatedness, share some characteristics of plant defence inherited from common ancestors, which influence their physiology and effectiveness as medicines. Using phylogenetic approach, recent studies confirmed that plant families that are closely related are more likely to have similar medicinal uses than those that are phylogenetically distant (Saslis-Lagoudakis et al. 2013; Yessoufou et al. 2015).

Several studies tested the non-random hypothesis in several geographic contexts, e.g. in Amazonian Ecuador (Bennett and Husby 2008), in Belize (Amiguet et al. 2006), in Kashmir (Kapur et al. 1992), and recently in Hawai’i, USA (Ford and Gaoue 2017) and Ecuadorian Amazon (Robles Arias et al. 2020). These studies reported strong support for the hypothesis. In particular, Robles Arias et al. (2020) demonstrated that the prediction of the hypothesis could be gender-specific. Nonetheless, such hypothesis-driven ethnobotanical studies are scant particularly in plant-rich countries with wealthy medicinal knowledge. South Africa is one of these species-rich families, with a remarkable plant diversity estimated at approximately 24,000 vascular plants but where ethnobotanical studies are less theory-driven. The different uses of medicinal plants are very well documented, e.g. ~3000 medicinal plants are recorded in the country, including 350 species known to be commonly used and traded (e.g. van Wyk and Gericke 2000; Fennell et al. 2004; van Wyk 2008; York et al. 2011; Elansary et al. 2015; Leso et al. 2017).

Furthermore, in studies that tested the non-random hypothesis, the methodological approaches used could be improved. For example, by fitting the simple linear model to the untransformed data he collected, Moerman (1979) did not account for normal residuals and homogeneity of variance. Recently, Ford and Gaoue (2017) have fitted the same model but on log-transformed data to account for that bias. Even so, the log-transformation performs poorly on “count data” (here, number of medicinal plants) in comparison with generalised linear model with negative binomial (see O’Hara and Kotze 2010). The application of these various statistics whilst ignoring their limitation is potential source of bias, not necessarily in the overall outcome of hypothesis testing, but more critically for the identification of over-utilised versus under-utilised families.

In the present study, the non-random hypothesis of medicinal plant selection in the Mpumalanga Province of South Africa was tested. Specifically, the different statistical approaches to explore the relationships between the number of known medicinal plants in families and the size of the family in the province were applied.

2 Materials and methods

2.1 Study area

Mpumalanga is one of the nine South African provinces within the Greater Maputaland-Pondoland Albany Biodiversity Hotspot, harbouring the southern half of the Kruger National Park and other centres of endemism. The Mpumalanga Province is divided into three districts, namely Gert Sibande, Nkangala and Ehlanzeni. Local communities are diverse in culture, and together with language discrepancies, there is a rich base of traditional knowledge. These communities include *Siswati* (30%), while 26% of the inhabitants speak *isiZulu* (26%), *isiNdebele* (10.3%), *Sepedi* (21.2%) and *Xitsonga* (11.6%) (Tshikalang et al. 2016).

Four major vegetation types are dominant in the study area, namely the highveld grasslands, escarpment grassland-forest mosaic, eastern Lowveld savannah and the north-western bushveld savannah (Schmidt et al. 2007). These vegetation types are represented in three distinct biomes: forest, savannah and grassland (Schmidt et al. 2007). The rainfall varies from a minimum of 440 mm in the north to a maximum of 740 mm in the south of the Kruger National Park (KNP) (Venter 1990). Mean annual temperature is around 21–23 °C, but in summer temperatures often exceed 38 °C, and frost can occur sporadically during winter.

2.2 Data collection

Data on the floristic composition of the Mpumalanga Province were collected through an intensive four-year fieldwork conducted from 2008 to 2012 by the last author of this paper (Yessoufou 2012). These data were supplemented by an existing database, i.e. the book entitled *Trees and shrubs of Mpumalanga and Kruger National Park* by Schmidt et al. (2007). This book contains both floristic and ethnobotanical knowledge of the region collected for more than 10 years of fieldwork. This book provided a unique botanical knowledge (including some medicinal uses) for a comprehensive checklist of 811 plant species representing 97 botanical families, of which 321 were reported to have some medicinal uses (Schmidt et al. 2007). In addition, data were also collected from *PRECIS* (SANBI,

2005), a comprehensive inventory of ethnomedicinal flora of Southern Africa containing 800,000 records of taxa grouped by order and regions (Magill et al. 1983; Germishuizen and Meyer 2003). More importantly, ethnobotanical data were further collected from *Prelude Medicinal Plants Database* (https://www.africamuseum.be/en/research/collection_s_libraries/biology/prelude), an electronic database of articles and various publications related to medicinal plants of Africa, hosted in the <https://www.africamuseum.be> website. From these data, two variables were derived, namely (1) the total number of plants species per family in the province and (2) the total number of medicinal plants recorded for each family in the province.

2.3 Data analysis

All analyses were done in R (R Development core Team 2017) using number of medicinal species recorded per families as response variable (count data) and the total number of species documented in the province for each family as predictive variable. Firstly, we fitted the simple linear model (model 1) to the untransformed data as commonly done in previous studies (Amiguet et al. 2006; Moerman 1996, 1979). Then, we tested for normality of the residuals. As this analysis indicated non-normality (Figure S1), we $\log(x+1)$ -transformed the response and predictor variables to address the normality issues (Figure S2). Then, we fitted the general linear model (model 2) to the transformed variables as done in a few recent studies (e.g. Ford and Gaoue 2017). Finally, because of the poor performance of simple linear regression with log-transformation of “count data” compared to the generalised linear model with negative binomial (see O’Hara and Kotze 2010), we also fitted a negative binomial model (model 3) to our dataset (Zeileis et al. 2008; O’Hara and Kotze 2010). For each of these models, we identified over-utilised families as those with positive residuals; this means these families contain a higher number of recorded medicinal species than would be expected from the model fitted.

3 Results

From the woody flora of the Mpumalanga Province, we recorded ~40% of medicinal plants, in ~76% of woody plant families in this study area (Table S1). Our analysis revealed that some plant families are over-utilised, while others are under-utilised (Fig. 1; Table 1). The proportion of over-utilised families ranges from 50% in line with Moerman’s linear regression approach through 55% (linear regression after log–log transformation) to 34% (negative binomial model). Following Moerman’s approach, the top over-utilised families are Fabaceae (residual = +34.44), Apocynaceae (+5.82) and Phyllanthaceae (+5.53). The log-transformed model confirms these three families as the top over-utilised families but in a slightly different sequence: Fabaceae (+1.55), Phyllanthaceae (+0.83) and Apocynaceae (+0.79). However, using the negative binomial model, Fabaceae is no longer even part of the top 10 over-utilised families, which are now: Phyllanthaceae (+2.09), Apocynaceae (+1.51), Loganiaceae (+1.48), Rhamnaceae (+1.48), Sapotaceae (+1.48), Oleaceae (+1.39), Salicaceae (+1.39), Clusiaceae (+1.30), Boraginaceae (+1.28) and Lamiaceae (+1.18) (Table 1). The top 10 under-utilised families comprise Celastraceae (–0.05), Monimiaceae (–0.06), Aquifoliaceae, Arecaceae, Canellaceae, Cornaceae, Gentianaceae, Hernandiaceae, Picrodendraceae and Piperaceae (–0.06, each).

Table 1 Residual values from various models fitting to medicinal data from Mpumalanga Province, South Africa

Families	Models fitted		
	GLM with negative binomial	LM with log transformed data	LM with untransformed data
Phyllanthaceae	2.09137599	0.835053937	5.53969331
Apocynaceae	1.512669883	0.797571948	5.822771068
Loganiaceae	1.483800427	0.54292917	2.654769982
Rhamnaceae	1.483800427	0.54292917	2.654769982
Sapotaceae	1.483800427	0.54292917	2.654769982
Oleaceae	1.395826781	0.764759196	2.884923327
Salicaceae	1.395826781	0.764759196	2.884923327
Clusiaceae	1.306267267	0.628796291	2.513231103
Boraginaceae	1.284008961	0.452975379	2.026462207
Lamiaceae	1.179202762	0.383068814	1.53969331
Ochnaceae	1.091813997	0.33070534	1.283077758
Vitaceae	1.091813997	0.33070534	1.283077758
Melianthaceae	1.054285932	0.716902316	2.256615552
Ebenaceae	1.028372859	0.653876636	3.707694395
Rutaceae	1.013622608	0.353145109	1.052924413
Capparaceae	0.987603461	0.447536572	1.566155516
Solanaceae	0.90086612	0.228918134	0.53969331
Annonaceae	0.890820041	0.40565274	1.513231103
Moraceae	0.780670125	0.219613716	0.052924413
Sapindaceae	0.67899315	0.095367151	- 0.088614466
Bignoniaceae	0.644334221	0.119276287	0.398154431
Proteaceae	0.585792491	0.15184317	- 0.690460035
Balanitaceae	0.541522014	0.429220243	1.256615552
Lauraceae	0.541522014	0.429220243	1.256615552
Polygalaceae	0.541522014	0.429220243	1.256615552
Urticaceae	0.541522014	0.429220243	1.256615552
Xanthorrhoeaceae	0.479787019	- 0.016686618	- 0.345230018
Cannabaceae	0.468902343	0.253933572	0.884923327
Malpighiaceae	0.468902343	0.253933572	0.884923327
Salvadoraceae	0.468902343	0.253933572	0.884923327
Combretaceae	0.464496089	0.317404399	- 0.292305605
Myrtaceae	0.233749663	- 0.176546974	- 1.46030669
Araliaceae	0.175847832	- 0.168405786	- 0.601845569
Celastraceae	- 0.050674529	0.336270859	- 1.265843398
Monimiaceae	- 0.060486287	0.270807822	0.628307776
Aquifoliaceae	- 0.060486287	0.270807822	0.628307776
Arecaceae	- 0.060486287	0.270807822	0.628307776
Canellaceae	- 0.060486287	0.270807822	0.628307776
Cornaceae	- 0.060486287	0.270807822	0.628307776
Gentianaceae	- 0.060486287	0.270807822	0.628307776
Hernandiaceae	- 0.060486287	0.270807822	0.628307776

Table 1 (continued)

Families	Models fitted		
	GLM with negative binomial	LM with log transformed data	LM with untransformed data
Picrodendraceae	- 0.060486287	0.270807822	0.628307776
Piperaceae	- 0.060486287	0.270807822	0.628307776
Pittosporaceae	- 0.060486287	0.270807822	0.628307776
Plumbaginaceae	- 0.060486287	0.270807822	0.628307776
Ranunculaceae	- 0.060486287	0.270807822	0.628307776
Velloziaceae	- 0.060486287	0.270807822	0.628307776
Asparagaceae	- 0.121431586	0.023755135	0.256615552
Chrysobalanaceae	- 0.121431586	0.023755135	0.256615552
Cyatheaceae	- 0.121431586	0.023755135	0.256615552
Podocarpaceae	- 0.121431586	0.023755135	0.256615552
Gerrardinaceae	- 0.121431586	0.023755135	0.256615552
Rhizophoraceae	- 0.121431586	0.023755135	0.256615552
Achariaceae	- 0.182802184	- 0.151531536	- 0.115076673
Stilbaceae	- 0.182802184	- 0.151531536	- 0.115076673
Ulmaceae	- 0.182802184	- 0.151531536	- 0.115076673
Scrophulariaceae	- 0.192652143	- 0.464229047	- 2.46030669
Icacinaceae	- 0.244562436	- 0.287494441	- 0.486768897
Myricaceae	- 0.244562436	- 0.287494441	- 0.486768897
Passifloraceae	- 0.244562436	- 0.287494441	- 0.486768897
Burseraceae	- 0.266145623	- 0.509383545	- 2.831998914
Verbenaceae	- 0.339455167	- 0.551421378	- 3.203691139
Fabaceae	- 0.34385417	1.558579578	34.441851
Apiaceae	- 0.369100661	- 0.492509295	- 1.230153345
Primulaceae	- 0.431799265	- 0.573870894	- 1.601845569
Thymelaeaceae	- 0.431799265	- 0.573870894	- 1.601845569
Malvaceae	- 0.48727429	0.376352255	- 1.867688967
Asteraceae	- 0.527738476	- 0.259957106	- 5.407382277
Rosaceae	- 0.684479905	- 0.820923581	- 3.088614466
Acanthaceae	- 0.874760528	- 0.956886486	- 4.203691139
Euphorbiaceae	- 0.882966212	0.683522402	2.043696567
Buxaceae	- 1.311885364	- 0.422339358	- 0.371692224
Connaraceae	- 1.311885364	- 0.422339358	- 0.371692224
Cupressaceae	- 1.311885364	- 0.422339358	- 0.371692224
Escalloniaceae	- 1.311885364	- 0.422339358	- 0.371692224
Hamamelidaceae	- 1.311885364	- 0.422339358	- 0.371692224
Linaceae	- 1.311885364	- 0.422339358	- 0.371692224
Lythraceae	- 1.311885364	- 0.422339358	- 0.371692224
Marattiaceae	- 1.311885364	- 0.422339358	- 0.371692224
Melastomataceae	- 1.311885364	- 0.422339358	- 0.371692224
Nyctaginaceae	- 1.311885364	- 0.422339358	- 0.371692224
Onagraceae	- 1.311885364	- 0.422339358	- 0.371692224
Osmundaceae	- 1.311885364	- 0.422339358	- 0.371692224

Table 1 (continued)

Families	Models fitted		
	GLM with negative binomial	LM with log transformed data	LM with untransformed data
Santalaceae	- 1.311885364	- 0.422339358	- 0.371692224
Pedaliaceae	- 1.311885364	- 0.422339358	- 0.371692224
Portulacaceae	- 1.311885364	- 0.422339358	- 0.371692224
Menispermaceae	- 1.311885364	- 0.422339358	- 0.371692224
Erythroxylaceae	- 1.351560192	- 0.669392045	- 0.743384448
Kirkiaceae	- 1.351560192	- 0.669392045	- 0.743384448
Putranjavaceae	- 1.432931235	- 0.980641621	- 1.486768897
Strelitziaceae	- 1.474574712	- 1.091731403	- 1.858461121
Penaecaceae	- 1.516818324	- 1.185656476	- 2.230153345
Ericaceae	- 1.559630896	- 1.267018074	- 2.601845569
Flacourtiaceae	- 2.057984046	- 1.794066979	- 6.690460035
Anacardiaceae	- 2.060053913	- 0.473216099	- 13.32799566
Rubiaceae	- 2.895742049	- 0.924598236	- 19.41661012

Families are ranked based on the residuals of the GLM with negative binomial

Family names in bold are those identified by GLM with negative binomial as medicinally over-utilized

GLM generalized linear model, LM linear model

is therefore a need for future studies to incorporate herbaceous plants into their analysis to further test the non-random hypothesis (or other theories).

Are medicinal plants a random selection of total flora in our study area? We investigated this question and found evidence that some medicinal families are over-utilised, i.e. they contain more medicinal plants than expected, whereas others are under-utilised, i.e. they have significantly lower number of medicinal plants. This finding supports the theory of non-random plant selection, which predicts a positive relationship between the total number of medicinal plants and the total number of species in a given family (Moerman 1979, 1991, 1996; Gaoue et al. 2017). Such a relationship has increasingly been reported in several studies in the Amazonian Ecuador (Bennett and Husby 2008), in Belize (Amiguet et al. 2006), in Kashmir (Kapur et al. 1992) and in Hawai'i, USA (Ford and Gaoue 2017), pointing potentially to the generalisation of the non-random hypothesis.

Despite this apparently general trend of non-random selection of medicinal plants in different contexts and regions of the globe, various methodological approaches have been used to test the theory. The most widely used approach is the general linear model proposed by Moerman et al. (1979). The Moerman approach has recently been modified using the log-log transformation of variables because it did not account for normal residuals and homogeneity of variance (Ford and Gaoue 2017), and some earlier studies employed the least squares regression analysis (e.g. Douwes et al. 2008). However, in the presence of just one zero observation, that is, when a plant species is not used for any medicinal treatment, log-transformation of the data becomes problematic, and we have to artificially create bias in the data by adding, for example, the number 1 to the observations to allow log-transformation. In any case, O'Hara and Kotze (2010), while comparing the different models (including those with variously transformed data),

demonstrated that models using Poisson or negative binomial models outperform any other models fitted to count data. In our case, as our response variable is count data (i.e. number of medicinal species), we fitted a negative binomial model to the medicinal plant data collected while also fitting the simple linear model with both untransformed and log-transformed data for comparison purpose.

All models fitted support not only the non-random plant selection hypothesis but also indicate that some families are over-utilised, whereas others are under-utilised. However, while Moerman and the log–log models yield similar proportions of over-utilised plant families (50 and 55%), the negative binomial model is very stringent as only 34% of plant families are categorised as over-utilised by this model. This is an indication that previous studies that employed Moerman approach may have included some (statistically) under-utilised families in their list of over-utilised families. For example, we identified the Fabaceae family as the top most over-utilised family when we applied Moerman and log-transformed models. Indeed, Fabaceae has been identified as one of the most over-utilised families in several studies that employed Moerman or similar approaches (Moerman 1999; Douwes et al. 2008; Gaoue et al. 2017; Kew 2017). In their recent report on the state of the world plants, RBKew (2017) indicated that Fabaceae, with its 11.2% of medicinal plants, is the 12th richest family in medicinal plants. They further indicated that the family contains important secondary compounds such as alkaloids.

However, while employing the negative binomial model, Fabaceae becomes under-utilised, thus indicating potential over-estimation of medicinal values of some taxa in previous studies. This does not imply that Fabaceae is not an important medicinal plant; rather, this implies that other families may outcompete Fabaceae in terms of people's preferences for medicinal uses. Indeed, in the negative binomial model, Phyllanthaceae is identified as the number one of all most over-utilised families followed by Apocynaceae, Loganiaceae, Rhamnaceae, Sapotaceae, Oleaceae, Salicaceae, Clusiaceae, Boraginaceae and Lamiaceae. Working in a similar floristic environment, Douwes et al. (2008) have already identified five of these families (Phyllanthaceae, Salicaceae, Apocynaceae, Loganiaceae and Boraginaceae) as over-utilised, although both studies employed different methodological approaches. It is possible that plants in these families are over-utilized medicinally in the study area for cultural reasons; that is, people may have developed cultural preferences for some plants. It could also be that the over-utilisation of these plant families is simply dictated by the environment (geography), i.e. people may be over-utilising what the environment made available to them in abundance (Saslis-Lagoudakis et al. 2014). Finally, the over-utilisation of these families may simply be indicative of the effectiveness of these plant families for medicinal purpose. Douwes et al. (2008) reported that most of these families are rich in terpenoids and their derivatives, flavonoids and alkaloids. It is therefore not surprising that these families are also identified in the present study among the most over-utilized.

Specifically, the family Phyllanthaceae belongs to the order Malpighiales, which contains a high level of secondary compounds such as aliphatic, alkaloids, amino acids and peptides, benzo-pyranoides, flavonoids, oxygen heterocycles, polycyclic aromatics, simple aromatics, terpenoids and derivatives (Douwes et al. 2008). In our dataset in the Mpumalanga Province (South Africa), species in the Phyllanthaceae family are reported to treat a variety of ailments. These ailments range from high blood pressure, oedema, bronchitis, intestinal disorders, diabetes, poison, skin infection, infertility, impotency, toothache, gingivitis, insecticide, heartburn, laxative, rheumatism, viral infections, HIV/AIDS epidemic, paralysis, bones diseases, kidney and bladder complaints (Bessong et al. 2005; Schmidt et al. 2007).

The recent work of Robles-Arias et al. (2020) employed the negative binomial model that we used in our study. Our study is similar to theirs in that both studies show that the relationships between medicinal plants and the total flora are not linear as suggested in Moerman (1979) (Fig. 1 in our study and Figure 2 in their study). Our study is different from theirs in two ways. First, we did not identify the same families as top over-utilized. This is because, as demonstrated in Saslis-Lagoudakis et al. (2014), the environment shapes the composition of medicinal floras. Second, Robles-Arias et al. (2020) showed that the outcome of the model prediction is gender-specific. This is a new knowledge that they brought into the non-random hypothesis. Although we did not test the influence of gender in this study, we suggest that future studies assess this influence in different geographic contexts for its generalisation.

In addition, Robles-Arias et al. (2020) also suggested that the presence of protected areas in an environment might hamper the development of medicinal knowledge in that environment. We tested and confirmed this negative effect of protected areas on medicinal plants (unpublished work). In the context of the present study, this means that more medicinal plants may have been reported in our study area if not for the presence of protected areas (e.g. Kruger National Park). Given this potential effect of protected areas on medicinal plants, we suggest that specimens of plant species that are found exclusively in protected areas could be grown *ex situ* in contact with local communities to increase the probability of the development of medicinal knowledge (availability hypothesis). From a similar perspective, a recent study linked land-use change with the change in human selection of medicinal plants (Kunwar et al. 2016); this is in support of our above claim that establishing protected areas may impact on the use and development of medicinal knowledge. The non-random hypothesis that we tested and supported in the present study may actually be driven by multiple other factors such as plant apparency (see da Silva et al. 2018).

Overall, the present study tested the non-random hypothesis of medicinal plant selection using the woody flora of the Mpumalanga Province of South Africa. This test was done using the most commonly statistical approaches (general linear model and log-linear model). Because of the limitations of these two models, a better model, the negative binomial model was also tested. The latter model seems to perform better than the former two models. In any case, our analysis showed that large families tend to have more plants being considered for local medicinal applications, a salient confirmation of the non-random plant selection for medicinal purpose. However, the negative binomial model identified the family Phyllanthaceae as the most over-utilized family in the province, while the other two families identified Fabaceae as the most over-utilized family. This is an illustration of the need to apply the most appropriate model while testing ethnobotanical hypotheses. This is paramount because the identification of over-utilized families is the first step towards the prioritisation of research efforts for drug discovery and wild plant conservation (Saslis-Lagoudakis et al. 2013).

Acknowledgements The first author gratefully acknowledges funding from the Global Excellence and Stature (GES) and a PhD merit bursary both offered by the University of Johannesburg, in South Africa. We acknowledge the South Africa's National Research Foundation for funding (Grants Nos: 103944; 112113; 111195) to Dr K. Yessoufou.

References

- Albuquerque, U. P. (2006). Re-examining hypotheses concerning the use and knowledge of medicinal plants: A study in the Caatinga vegetation of NE Brazil. *Journal of Ethnobiology and Ethnomedicine*, 2(1), 30.
- Albuquerque, U. P., & Lucena, R. F. P. (2005). Can apparency affect the use of plants by local people in tropical forests? *Interciencia*, 30, 506–511.
- Amiguet, V. T., Arnason, J. T., Maquin, P., Cal, V., Sánchez-Vindas, P., & Alvarez, L. P. (2006). A regression analysis of Q'eqchi' Maya medicinal plants from southern Belize. *Economic Botany*, 60(1), 24–38.
- Arnold, T. H., Prentice, C. A., Hawker, L. C., Snyman, E. E., Tomalin, M., Crouch, N. R., et al. (2002). *Medicinal and magical plants of southern Africa: An annotated checklist*. Lucknow: National Botanical Institute.
- Bennett, B. C., & Husby, C. E. (2008). Patterns of medicinal plant use: An examination of the Ecuadorian Shuar medicinal flora using contingency table and binomial analyses. *Journal of Ethno Pharmacology*, 116(3), 422–430.
- Bessong, P. O., Obi, C. L., Andréola, M. L., Rojas, L. B., Pouységu, L., Igumbor, E., et al. (2005). Evaluation of selected South African medicinal plants for inhibitory properties against human immunodeficiency virus type 1 reverse transcriptase and integrase. *Journal of ethnopharmacology*, 99(1), 83–91.
- da Silva, T. C., da Silva, J. M., & Ramos, M. A. (2018). What factors guide the selection of medicinal plants in a local pharmacopoeia? A case study in a rural community from a historically transformed atlantic forest landscape. *Evidence-Based Complementary and Alternative Medicine*, 2018, 2519212.
- Douwes, E., Crouch, N. R., Edwards, T. J., & Mulholland, D. A. (2008). Regression analyses of southern African ethnomedicinal plants: Informing the targeted selection of bioprospecting and pharmacological screening subjects. *Journal of Ethnopharmacology*, 119(3), 356–364.
- Elansary, H. O., Mahmoud, E. A., Shokralla, S., & Yessoufou, K. (2015). Diversity of plants, traditional knowledge and practices in local cosmetic industry in Egypt: A case study of Alexandria governorate in the Northern Egypt. *Economic Botany*, 69, 114–126.
- Fennell, C. W., Lindsey, K. L., McGaw, L. J., Sparg, S. G., Stafford, G. I., Elgorashi, E. E., et al. (2004). Assessing African medicinal plants for efficacy and safety: Pharmacological screening and toxicology. *Journal of ethnopharmacology*, 94(2–3), 205–217.
- Feeny, P. (1976). Plant apparency and chemical defense. In J. Wallace & R. Mansell (Eds.), *Recent advances in phytochemistry*. New York: Plenum Press.
- Ford, J., & Gaoue, O. G. (2017). Alkaloid-poor plant families, Poaceae and Cyperaceae, are over-utilized for medicine in Hawaiian pharmacopoeia. *Economic Botany*, 71(2), 123–132.
- Gaoue, O. G., Coe, M. A., Bond, M., Hart, G., Seyler, B. C., & McMillen, H. (2017). Theories and major hypotheses in ethnobotany. *Economic Botany*, 71(3), 269–287.
- Germishuizen, G., & Meyer, N. L. (2003). *Plants of southern Africa: An annotated checklist*. Pretoria: National Botanical Institute.
- Hart, G., Gaoue, O. G., de la Torre, L., Navarrete, H., Muriel, P., Macia, M. J., et al. (2017). Availability, diversification and versatility explain human selection of introduced plants in Ecuadorian traditional medicine. *PLoS ONE*, 12(9), e0184369.
- Kapur, S. K., Shahi, A. K., Sarin, Y. K., & Moerman, D. E. (1992). The medicinal flora of Majouri-kirchi forests (Jammu and Kashmir State), India. *Journal of Ethnopharmacology*, 36, 87–90.
- Kew, R. B. G. (2017). Seed Information Database (SID). Version 7.1.
- Kunwar, R. M., Baral, K., Paudel, P., Acharya, R. P., Thapa-Magar, K. B., Cameron, M., et al. (2016). Land-use and socioeconomic change, medicinal plant selection and biodiversity resilience in far Western Nepal. *PLoS ONE*, 11(12), e0167812.
- Leso, L. K., Stoffberg, G. H., Mearns, K., Elansary, H. O., & Yessoufou, K. (2017). Ethnobotany at local scale: Diversity of knowledge of medicinal plants and assessment of plant cultural importance in the Polokwane Local Municipality, South Africa. *Botany Letters*, 164, 93–102.
- Magill, R. E., Russell, G. G., Morris, J. W., & Gonsalves, P. (1983). PRECIS, the Botanical Research Institute herbarium data bank. *Bothalia*, 14(3/4), 481–495.
- Moerman, D. E. (1996). An analysis of the food plants and drug plants of native North America. *Journal of Ethnopharmacology*, 52(1), 1–22.
- Moerman, D. E. (1991). The medicinal flora of native North America: An analysis. *Journal of Ethnopharmacology*, 31(1), 1–42.
- Moerman, D. E. (1979). Symbols and selectivity: A statistical analysis of native American medical ethnobotany. *Journal of Ethnopharmacology*, 1(2), 111–119.
- Mpumalanga Provincial Government. About Mpumalanga Province. Available from: <http://www.mpumalanga.gov.za/about/province.htm>. Accessed January 23, 2018.

- O'Hara, R. B., & Kotze, D. J. (2010). Do not log-transform count data. *Methods in Ecology and Evolution*, *1*(2), 118–122.
- Phillips, O., & Gentry, A. H. (1993). The useful plants of Tambopata, Peru: I. Statistical hypotheses tests with a new quantitative technique. *Economic Botany*, *47*(1), 15–32.
- RBKew 2017. State of the world plants. <https://stateoftheworldsplants.com/2017/useful-plants.html>. Accessed June 2018.
- R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- SANBI. (2005). *PRECIS (PRE-Computerized Information System) databank*. Pretoria: South African National Biodiversity Institute.
- Saslis-Lagoudakis, C. H., & Clarke, A. C. (2013). Ethnobiology: the missing link in ecology and evolution. *Trends in Ecology & Evolution*, *28*(2), 67–68.
- Saslis-Lagoudakis, C. H., Hawkins, J. A., Greenhill, S. J., Pendry, C. A., Watson, M. F., Tuladhar-Douglas, W., et al. (2014). The evolution of traditional knowledge: Environment shapes medicinal plant use in Nepal. *Proceeding of the Royal Society, Biological Sciences*, *281*, 20132768.
- Schmidt, E., Lotter, M., & McClelland, W. (2007). *Trees and shrubs of Mpumalanga and Kruger National Park* (p. 702). Johannesburg: Jacana Press.
- Tshikalange, T. E., Mophuting, B. C., Mahore, J., Winterboer, S., & Lall, N. (2016). An ethnobotanical study of medicinal plants used in villages under Jongilanga Tribal Council, Mpumalanga, South Africa. *African Journal of Traditional, Complementary and Alternative Medicine*, *13*(6), 83–89.
- van Wyk, B. E. (2008). A broad review of commercially important southern African medicinal plants. *Journal of ethnopharmacology*, *119*(3), 342–355.
- van Wyk, B. E., & Gericke, N. (2000). *People's plants: A guide to useful plants of Southern Africa*. Pretoria: Briza Publications.
- Venter, F. J. (1990). *A classification of land for management planning in the Kruger National Park*. PhD Thesis, University of South Africa, Pretoria, South Africa.
- Williams, V. L., Victor, J. E., & Crouch, N. R. (2013). Red listed medicinal plants of South Africa: Status, trends, and assessment challenges. *South African Journal of Botany*, *86*, 23–35.
- Yessoufou, K., Daru, B. H., & Muasva, A. M. (2015). Phylogenetic exploration of commonly used medicinal plants in South Africa. *Molecular Ecology Resources*, *15*(2), 405–413.
- Yessoufou, K. (2012). Phylogenetic analysis of plant community assemblages in the Kruger National Park, South Africa (Doctoral dissertation, University of Johannesburg).
- York, T., De Wet, H., & Van Vuuren, S. F. (2011). Plants used for treating respiratory infections in rural Maputaland, KwaZulu-Natal, South Africa. *Journal of Ethnopharmacology*, *135*(3), 696–710.
- Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software*, *27*(8), 1–25.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Isidore Muleba¹ · Kowiyou Yessoufou¹ · Isaac T. Rampedi¹ 

✉ Kowiyou Yessoufou
kowiyou@uj.ac.za

Isidore Muleba
isidore_muleba@yahoo.com.au

Isaac T. Rampedi
isaacr@uj.ac.za

¹ Department of Geography, Environmental Management and Energy Studies, University of Johannesburg, APK Campus, Johannesburg, South Africa