Rarefaction on natural compound extracts diversity among genus

Lorentz Jántschi, a,b, Sorona D. Bolboaca c,*

a Technical University of Cluj-Napoca, 28 Memorandumului, 400144 Cluj-Napoca, Cluj, Romania
b University of Agricultural Science and Veterinary Medicine Cluj-Napoca, 3-5 Calea Mănăstur, 400372 Cluj-Napoca, Romania
c “Iuliu Hatieganu” University of Medicine and Pharmacy Cluj-Napoca, Department of Medical Informatics and Biostatistics, 6 Louis Pasteur, 400349 Cluj-Napoca, Cluj, Romania

A R T I C L E   I N F O

Article history:
Received 30 November 2012
Received in revised form 19 May 2013
Accepted 13 August 2013
Available online 24 August 2013

Keywords:
Biodiversity
Bootrap
Rarefaction
Compounds richness

A B S T R A C T

The aim of our study was to investigate the intrinsic diversity of genus based on chemical composition of natural extracts of Cacalia, Dracocephalum, Jatropha, Sauvurea and Senecio. Four rarefaction approaches (richness, Shannon’s entropy, Simpson’s diversity, and max-entropy) were implemented as an algorithm and run 10,000 times. Our results showed similar patterns of richness in natural compounds according to genus and of Simpson’s diversity index while max-entropy has a similar pattern as Shannon’s entropy. The analysis of both richness in active compounds and weight of most frequent extracted compound revealed that Dracocephalum genus is most suitable for propagation.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Plant phylogeny refers to the description of the evolutionary history of related species plant groups. Different information that is not limited to morphological characters [1,2], molecular [3] or genome-scale data [4] had their contribution to a phylogeny [5].

Plant extracts, and thus plant composition, are used as therapeutics from ancient times [6] and are sometimes seen as source of active compounds for new drugs [7,8]. Antibacterial effect of a series of medicinal plant extracts and/or oils have already been identified (Terminalia chebula [9], Garcinia kola [10], Pterygota macrocarpa and Cola gigantea [11], Cryptopleine sanguinolenta [12], Quercus dilatata L. [13], Rhus coriaria L. [14], Launaea resedifolia L. [15], etc.). Furthermore, the plants extracts are also used for their anti-inflammatory activities (mangrove plants [16], Caesalpinia Ferrea [17]). Beside those main effects, other effects such as anti-oxidant (Myrtus communis, Smilax aspera, Lavandula stoechas, and Calamintha nepeta [18]), anticancer (Zingiber officinale Rosc., Atractylodes lancea (Thun.) DC., Piper chaba Hunt. [19]), immunomodulation (triterpenoids ursolic acid and lupeol extracted from Eucalyptus tereticornis and Gentiana kurroo [20]), anti-viral (anti-herpes simplex virus type 1 activity of oxysresaristol extracted from Artocarpus lakoocha [21]) had been identified.

The biological activity of plant extracts are of interest due to their specific activities such as anti-viral [22] and antibacterial [23,24], anticancer therapeutics [25], neuroprotection [26], antiliver fibrosis [27], and the list could continue. The identification of those extracts with specific biological activity is of main interest in order to integrate the knowledge and to develop new plant-derived drugs [28,29]. The analysis of biological activity of plant extracts is an interdisciplinary approach and moved from a simple observation to integrated approaches (involving analytical chemistry, synthesis, molecular biology, proteomics, chemoinformatics, bioinformatics, etc.). Huge amounts of various data regarding activity and/or action mechanisms of plant extracts are available [30–34]. Integrating these amounts of disperses data could bring knowledge of important relevance and may provide a measureless resource for future discovery of biomedical relevant compounds.

The aim of our study was to conduct a differentiation analysis using rarefaction approach among genera regarding natural richness in active compounds and to provide the expression of the diversity as function of number of plants that belong to a genus.

2. Materials and methods

The chemical composition of five genera, represented by Cacalia [35] (NE of Asia & America & SW and NW of China), Dracocephalum [36] (North and South America & Europe & Asia), Jatropha [37] (America & Africa (tropical and subtropical)), Sauvurea [38] (Asia & Europe), and Senecio [39] (worldwide), represented the material of our research. The main characteristics of the genera included in the study are summarized in Table 1.

The names of compounds as well as the source of compounds for each investigated genus are available as supplementary materials.
Table 1
Main characteristics of investigated genera.

<table>
<thead>
<tr>
<th>Genus (no. of plants*)</th>
<th>Family*</th>
<th>Main activity</th>
<th>No. of Comp* (no. of Species*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cacalia (1008)</td>
<td>Compositae</td>
<td>Antioxidative; Antimicrobial; Antitumor; Citotoxic</td>
<td>109 (13)</td>
</tr>
<tr>
<td>Dracocephalum (231)</td>
<td>Lamiaceae</td>
<td>Antioxidant; Antithyroid; Antimicrobial; Antitumor; Cardiovascular</td>
<td>246 (12)</td>
</tr>
<tr>
<td>Jatropha (567)</td>
<td>Euphorbiaceae</td>
<td>Anticoagulant and coagulant; Antinflammatory; Antitumor; Antibacterial; Antioxidant; Antioxidant</td>
<td>143 (26)</td>
</tr>
<tr>
<td>Saussurea (1054)</td>
<td>Compositae</td>
<td>Antioxidant; Antinflammatory; Antitumor; Antimicrobial; Apocrine; Antifeedant; Antifungal; Antimicrobial</td>
<td>216 (22)</td>
</tr>
<tr>
<td>Senecio (6183)</td>
<td>Compositae</td>
<td>Antibacterial; Antifungal; Antinflammatory; Antitumor; Antioxidant; Antioxidant</td>
<td>631 (186)</td>
</tr>
</tbody>
</table>

*Source: [http://www.theplantlist.org](http://www.theplantlist.org) [online] (accessed 29.05.12).

Information extracted from the reference associated to each genus.

(r_cacalia.txt, r_dracocephalum.txt, r_jatropha.txt, r_saussurea.txt, and r_senecio.txt).

A simulation study was conducted based on chemical compounds extracted from investigated species to analyze intrinsic diversity by applying a rarefaction approach function.

The Bootstrap method, introduced by Efron [40], was used in this analyzes to estimate (based on independent observation represented by the compounds obtained from plant extracts) the distributions of a series of statistics in order to assess what happen in population. The following statistics were obtained and used in this analysis: mean value of the number of chemical compounds (revealing the richness within species), biodiversity expressed through two entropies (Shannon entropy [41,42] and max-entropy [43,44]), and one diversity measure (Simpson diversity [45]).

The following algorithm was developed and applied in our analysis:

For n=1 to n=50 (number of plants to be extracted)
  - For q=1 to q=10,000 (repeat the experiment by 10,000 times)
    - For i=1 to i=n (execute the experiment of repeated sampling)
      - plant_i = random(number of species)
      - obs_i = number of compounds for plant_i
    - End for i
  - Exp_q = the array containing distinct compounds and the number of plants from which were extracted (from n)
  - dc_q = Count(Exp_q) - the number of distinct compounds
  - For α = 0 (Hartley=log(Count)), 1 (Shannon), 2 (log(Simpson)), =
    - (max-entropy)
      - hα_q = Renyi(Exp_q, α) - Renyi entropy of Exp_q
    - End for q
  - Calculate the mean and standard deviation (from 10,000 repetitions) for Count and Entropies
End for n

The algorithm was run 10,000 times in order to emulate changes in population. The rarefaction method [46] was furthermore applied to compare genetic diversity regarding the extracts on the investigated genus. The rarefaction analysis was applied to answer the following question “How many extracts could have been found in a small sample?” [47]. Rarefaction curves were used for graphical visualization of diversity based on the obtained results.

3. Results

The distribution of richness in natural active compounds according to genus was obtained based on applied methodology and is presented in Fig. 1. The minimum number of active compounds (average of the expectance, 4.10) was observed for Senecio when just one plant was included in analysis. As expected, the maximum number of compounds (average of the expectance, 243.35 – Dracocephalum) was observed when 50 plants were included in the sample.

The increase in number of active compounds varied from 0 (Cacalia and Dracocephalum) to 23 (Dracocephalum) with a minimum variation observed for Senecio (from 2 to 4).

Shannon’s entropy represented the second method used to characterize diversity in our investigation. It had values from 1.1259 (Senecio – n = 1) to 5.3397 (Dracocephalum – n = 50) and its evolution as function of sample size is presented in Fig. 2.

The Shannon’s entropy increase almost 2 times for n = 2 compared with n = 1 for Jatropha, Saussurea and Senecio, being the highest increasing observed in our study. Without any exception, the Shannon’s entropy slightly increased with sample size but the increase proved not being uniform when the differences were analyzed.

The Simpson’s diversity index varied from 3.0830 (Senecio – n = 1) to 179.8364 (Dracocephalum – n = 50). The evolution of Simpson’s diversity index with sample size is presented in Fig. 3.

The entropy of most frequent compound (max-entropy) varied from 1.1259 (Senecio – n = 1) to 4.0842 (Dracocephalum – n = 50). The evolution of max-entropy with sample size is presented in Fig. 4.

The max-entropy results revealed that the value increased ~2 times when n = 2 is compared to n = 1 for Jatropha, Saussurea and Senecio genus. The max-entropy systematically increased with sample size just for Senecio. The rule of increase of max-entropy with sample size was broken 1 time by Saussurea, 2 times by Jatropha 3 times by Cacalia) and 9 times by Dracocephalum. The rarefaction curve for Senecio overlaps the Jatropha line for n < 5 when max-entropy was investigated.
Bioactive compounds in plants (compounds with pharmacological and toxicological effects in human and animals that are produced by plants) become of interest due to the necessity of identification of new drug with highest activity and lowest adverse effects as possible. The main groups of bioactive compounds according to biochemical pathways [48] are: glycosides (such as cardiac glycosides which are presented in plants of Scrophulariaceae, and Convallariaceae or cyanogenic glycosides that are present in species such as Rosaceae), flavonoids (produced by species of Fabaceae), tannins (species of Fagaceae and Polygonaceae), terpenoids (the most representative family is Lamiaceae), diterpenoids (present in Coffea arabica), alkaloids (such as alkaloids with acticholinergic activity present in species of Solanaceae), furcocoumarines (in Apiaceae – carrot family) and naphthodianthrone, proteins and peptides (Euphorbiaceae or Fabaceae). For example, morphine (painkiller; its use lead to addiction), codeine (use to treat cough or as painkiller), noscapine (use to treat cough), and papaverine (use to treat vasospasms – heart and brain) are extracted from Papaver somniferum [49]. Acetyl salicylic acid used as rheumatism, painkiller, fever reducing agent, and anti-coagulant is extracted from Salix-species [49]. Atropine used as eye drops to study the inner parts of the eyes, and scopolamine used as travel sickness drug are extracted from Atropa belladonna, Hyoscyamus niger, and Datura stramonium [49].

It is estimated that only 10% of the plants on earth have been investigated for their content of bioactive compounds [49]. The investigation of at least those plants used by traditional healers could provide the faster route to new active compounds.

In this study we used four rarefaction approaches (richness, Shannon’s entropy, Simpson’s diversity, and max-entropy) to analyze the bioactive compounds diversity of investigated species.

The results of the first approach revealed that the richest in bioactive compounds is Dracocephalum (see Fig. 1). Three pathways could be observed according to genus when the differences in number of active compounds are investigated relative to the number of investigated plants (Fig. 1): first pattern that involves Cacalia – Jatropha – Saussurea, second pattern that comprises Senecio and third pattern that comprise Dracocephalum. As far as the first pattern is concerned, it could be observed a slightly uniform decrease of the number of active compounds regarding the difference related to previous n, with increase of number of plants included in analysis for Cacalia (the difference varied inverses to sample size with values from 0 to 9), Jatropha (the difference varied inverses to sample size with values from 1 to 5) and Saussurea (the difference varied inverses to sample size with values from 1 to 10). Furthermore, the difference decreased till a certain n (Cacalia – n = 35; Jatropha – n = 37, and Saussurea – n = 42) and after that became flattened (0 for Cacalia – n > 35; 1 for both Jatropha – n > 37 and...
Dracocephalum – n > 42). As far as the second pattern is concerned, with one exception (n = 41), the difference in active compounds relative to the previous sample size was of 3 (systematically for n > 29) or 4 (systematically for n < 15). This observation could be related with the number of active compounds (this genus had the higher number of active compounds – see Table 1 – ~2.5 times higher compared to Dracocephalum) as well as with the richness in specie (the number of Senecio was more than 7 times higher compared to Jatropha – see Table 1) for Senecio genus. Regarding the third pattern, significant differences were observed till n = 7 (a difference of 23 compounds between n = 2 and n = 1; a difference of 12 compounds between n = 7 and n = 6) but for n > 7 the differences were not higher than 1 and become systematically 0 for n > 46. Moreover, the number of active compounds doubled when two plants were included in analysis and after this threshold the number of compounds increased slightly.

In terms of Shannon’s entropy, the second approach of our study, rarefaction was identical for Jatropha and Senecio for n > 5 and very close till the intersection of their rarefaction curves (n = 18) (see Fig. 2). Starting with this point, different patterns were identified (slightly modification for Jatropha and increases for Senecio). The Shannon’s entropy of Calacila and Saussurea follow a similar pattern as of Dracocephalum; both showing an increase with sample size but values of Shannon entropy associated to Dracocephalum was higher (Fig. 2).

Simpson’s diversity index, the third approach used in our investigation, follows somehow the patterns of number of compounds (Fig. 1). Thus, two pathways are observed in terms of index range: relative small ranges (Cacila with a range of 73.5009 and Jatropha with a range of 89.1564) and relative large ranges (Senecio – 133.9251, Saussurea – 138.4330 and Dracocephalum – 168.3506). As it could be observed from Fig. 3, without any exception, the diversity evaluated using the Simpson’s index increase with sample size. The highest increase in diversity was observed when n = 2 was compared to n = 1 (an increase of 2 times was observed for Cacila, Jatropha, and Senecio and of 3 times for Dracocephalum and Saussurea). Moreover, the diversity increased twice when n = 3 was compared to n = 2 for Dracocephalum, Jatropha, Saussurea, and Senecio. For n > 3 the increased in diversity was uniformly of ~1. Furthermore, the smallest variation regarding the differences is observed for Senecio genus (Fig. 3). Cacila is the genus with the diversity around 0 for n > 39. Small variation in differences is also observed on Jatropha while the higher variation is observed on Dracocephalum.

Similar patterns with the one described for the Shannon’s entropy was observed when the results of the fourth approach used in this analysis were investigated (max-entropy – Fig. 4).

The results of our study reflect the rarefaction obtained based on the information available when the research was initiated. A tendency of flattening is observed for Cacila and Dracocephalum for both number of compounds and Simpson diversity index. Similar patterns were observed on Shannon’s entropy and max-entropy as well as on Simpson diversity index and rarefaction of number of compounds. The obtained results must be interpreted taking into consideration that different species have not the same contribution to diversity [50–53]. The analysis could be extended by implementing other diversity indices such as phylogenetic diversity index [54], quadratic diversity index [55], taxonomic entropy [56], unnamed Hq index [57] or phylogenetic entropy Hq [58] to obtain a more exhaustive investigation. Such researches are currently conducted in our laboratory to obtain more relevant results on the topic.

5. Conclusion

Our results revealed that both in terms of richness in active compounds and weight of most frequent extracted compound Dracocephalum genus is most suitable for propagation being followed by Saussurea. Calacila and Jatropha proved to have the lowest diversity in terms of bioactive compounds related to the investigated genus.

Acknowledgements

The study was supported by the European Social Fund, Human Resources Development Operational Program, project number 89/15/62371 through a fellowship for L. Jäntschi. The funding source had no role in the study design, data collection and analysis, decision to publish, or in preparation of the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version, at http://dx.doi.org/10.1016/j.jjocs.2013.08.002.

References
