

Regional floras: increasing their value while reducing their cost

Mary E. Barkworth^{1*}, Marina V. Olonova², Polina D. Gudkova³, Zahid Ullah⁴, and Curtis Dyreson⁵

¹Intermountain Herbarium, Department of Biology, Utah State University, Logan, Utah 84322-5305, USA

²Department of Botany, Altai State University, Russia

³Department of Environmental Management, Biological Institute, Tomsk State University, Russia

⁴Center for Plant Science and Biodiversity, University of Swat, Khyber Pakhtunkhwa, Pakistan,

⁵Department of Computer Science, Utah State University, 4205 Old Main Hill, Logan, Utah 84322-4205, USA

Abstract. Regional floras are primary resources for plant identification, an essential step in developing conservation strategies. They also provide students with a scientific window on the plants around them and help them learn botanical terminology, but they are expensive to maintain and publish. We are developing web-accessible updates for different floras, as part of which we are using online resources to help us work more effectively while rapidly providing richer resources. We use KeyBase for sharing dichotomous keys, linking the terminal taxa to subsidiary keys or descriptive taxon pages. Taxon pages are generated in OpenHerbarium which enables integrating specimen and observation data with descriptions, line drawings, and images and displaying maps based on georeferenced specimen data. Its nomenclatural backbone is easily modified to reflect new treatment and can also handle multiple taxonomies. We are examining the possibility of using a Wikipedia approach to provide a glossary.

1 Introduction

Regional floras are incredibly valuable. They enable mapping the distribution of plant diversity, a key step to understanding ecosystems at multiple scales and development of conservation plans. They are also important educational tools, familiarizing students with the plants around them, the characteristics of different genera and families, and botanical terminology. Learning to use a regional flora moves an individual from reliance on knowing someone to ask to increasing independence and becoming someone with the ability to build on what is known, contribute to the knowledge of plant species and their distributions, and becoming a person to whom others turn. Regional floras are also invaluable for practicing botanists, reminding them of the distinguishing morphological, ecological, and geographic features of individual taxa. The problem is that floras become outdated as new species are discovered or arrive in an area and systematic research leads to

* Corresponding author: mary.barkworth@usu.edu

taxonomic changes that impact the region’s flora. Four of us are working on providing web-accessible updated floristic resources for different floras: Flora of the Altai (MVO, PDG; [1, 2]), Manual of grasses for North America north of Mexico (MEB; [3]), Flora of Somalia (MEB; [4]), and Flora of Pakistan (ZU; [5]). We are using KeyBase [6] and OpenHerbarium [7], a Symbiota [8] based network, to share our updated resources. Dyreson is working, with others, on restructuring Symbiota [9] to better meet user needs. In this paper, we look at ways KeyBase and OpenHerbarium can help in developing rich floristic resources for a region by mobilizing and adding to what is known, enhancing or replacing posted information as more is learned.

2 Components of a flora

Floras vary in the categories of information they provide, and in the detail offered within each category (Table 1). Several computer programs exist for assisting in one or more category. This paper emphasizes those we use, primarily OpenHerbarium [7], Geolocate [10], KeyBase [6] and two nomenclatural resources, the International Plant Names Index [11] and Tropicos [12]. Of course, no resource is perfect. To improve them, users need to inform their owners of apparent errors or omissions. Our goal is to encourage others working, or thinking of working, on the creation or revision of a flora to integrate the tools becoming available into their workflows, and to suggest some developments that, if incorporated into Symbiota and similar software, would increase their ability to make writing and maintain floras easier, thereby freeing up time for focusing on critical issues while providing users with richer resources faster.

Table 1. Categories of information included in traditional regional floras.

	Category	Comment on occurrence
1	Scientific names	Name accepted by author always included, with its authorship.
7	Nomenclatural citations	Frequently included for accepted species names, less often for higher categories and synonyms
2	Identification mechanism	Usually dichotomous keys, rarely polychotomous
3	Descriptions	Usually written and edited as a block, vary in length and detail.
4	Ecological information	Varies considerably in detail provided
5	Geographic information	Varies considerably in detail provided
6	Illustrations	Varies from none to every taxon
7	Vernacular names	Frequently included
8	Cytology	Chromosome counts or base numbers often included but not the source or voucher.
9	Use by humans	Often included for species.
10	Glossary	Sometimes present; rarely illustrated.
11	Information sources	Rarely included unless explicitly cited.

2.1 OpenHerbarium

OpenHerbarium [7] integrates information from multiple herbaria with that from other sources. The software that runs it, Symbiota [8], also runs a network of over 300 US herbaria, SEINet [13]. OpenHerbarium was established for herbaria in the Old World, particularly for countries without an existing herbarium network, but it is sharing information developed for the grass volumes of the *Flora of North America north of Mexico* project [14, 15]

OpenHerbarium [7] can integrate specimen and image-based occurrence records with descriptions, images (photographic or line drawings), and maps developed from georeferenced records in its database (Figure 1). The taxon page integrates images from several photographers, many harvested from the Encyclopedia of Life [16]. The resources used are those for which the contributors have permission to share, usually under a CC-BY-SA license [17].

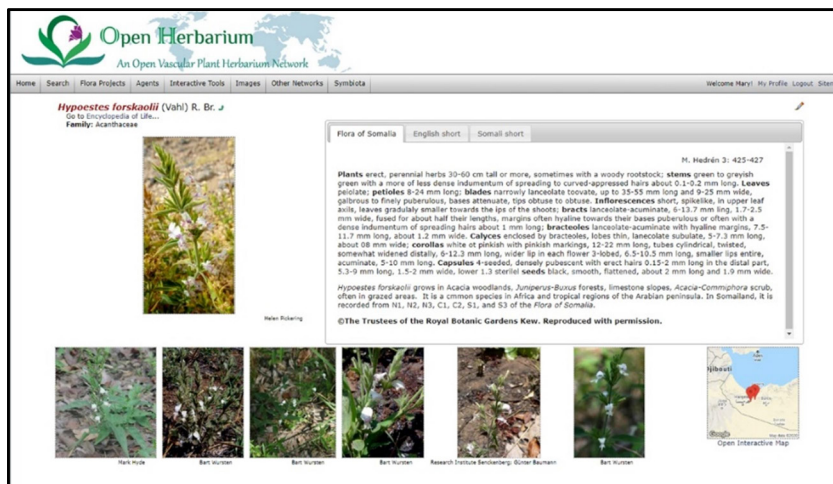


Fig. 1. Taxon page for *Hypoestes forskalii* generated by OpenHerbarium, generated 30 May 2020.

Taxon pages can include multiple descriptions, including descriptions in multiple languages and images. For example, Figure 1 shows a page with three descriptions, one from the *Flora of Somalia* [4], a short English description, and a Somali description. The first is designed primarily for professionals, the second and third for more general audiences, including students, a concern of the Somaliland Biodiversity Foundation [18]. The descriptions are written as paragraphs. OpenHerbarium [7] can also generate maps for multiple taxa, enabling a quick comparison of geographic distributions.

2.2 Taxonomy and nomenclature

When integrating floristic information from multiple sources, differences in taxonomic treatment, and the consequent differences in nomenclature, are a major problem. The World Flora Online [19] identifies which of several names is accepted, but they do not provide a reason for the decision made. Tropicos [12] takes a different approach, identifying what other names have been used for the taxon in various floras. Symbiota [8] takes a third approach, allowing citation of the primary source supporting a change. At present, this information is visible only to a network’s taxonomic editors but Symbiota2 will make it visible to all users.

Changing the nomenclatural backbone does not modify the name within a record but it does affect which records are found in a search. By default, searches locate records with the name specified and its synonyms as listed in the backbone. OpenHerbarium can accommodate multiple taxonomic treatments but, at present does not.

Building the nomenclatural backbone for a region is the top priority when expanding OpenHerbarium’s geographic range. It is time consuming because not only is it important to check the spelling of the name, one must also check the authorship, the abbreviations used for authors, and current thinking concerning its taxonomic position. At one time,

Barkworth imported names from the Integrated Taxonomic Information System [20], but its treatment caused multiple problems so that approach was abandoned.

IPNI [11] and Tropicos [12] are invaluable for checking the spelling of names and authors. IPNI was built from three pre-existing print resources. It has made enormous progress in eliminating redundancy, adding information, infraspecific taxa, and links to the original literature in the Biodiversity Heritage Library [21]. Many journals automatically send new names and combinations to it. Tropicos was designed and continues to be maintained by the Missouri Botanical Garden to meet that institution's needs. Consequently, its coverage is not as comprehensive as IPNI's, but it began adding links to protologues earlier than IPNI, routinely explains why a name is illegitimate, and shows how a name has been treated in different floras. Neither resource is perfect, but they are constantly improving.

2.3 Data sharing

OpenHerbarium makes data sharing simple. It can accept data either via browser based forms. These include tools for checking the accuracy/validity of the data entered in some fields and automatic completion of a few. Alternatively, data can be uploaded to a herbarium's portion of the database via a csv file. It can accept non-Latin scripts, including Cyrillic. Once in OpenHerbarium, data are immediately included in searches and displays.

OpenHerbarium also makes it easy to send data to GBIF [22]. All it requires is for the herbarium to become a GBIF data provider, forms for which are on the GBIF website. On approval, GBIF will provide a key that, when inserted into the herbarium's OpenHerbarium metadata, enables GBIF to “see” when it has published new records or revised old ones. One advantage of providing records to GBIF is that, when used in a publication, GBIF records the fact and adds it to information about the herbarium. This makes it easier to demonstrate the scientific value of a collection.

At present, OpenHerbarium is being used for four floristic projects, each with its own goals and aspirations. It would be better if each project had its own URL, leading to a home page that highlights the goals of the project. SEINet [13] for example, is accessed by more than 10 URLs (see, for example, <http://intermountainbiota.org/portal/>). Each URL leads to a different homepage, but all draw on and contribute to a common resource pool. Establishing a portal requires someone with a background in computer science. It will become somewhat easier once Symbiota2 [9] is available, in October 2021. Symbiota2 will also enable installation of local versions of OpenHerbarium that can work offline, synchronizing with the network when there is good internet access.

3 KeyBase

Identification keys are a critical element in any modern flora. Indeed, some floras contain little more than an identification key and a list of the species with the author of their name and a short summary of its ecology e.g., [23]. The keys are almost always dichotomous, each step requiring choosing between two alternatives. KeyBase [6] is an excellent resource for checking that the structure of a key is correct, in other words, that every lead goes somewhere and that there is a path to each terminal taxon. It will not display incorrectly structured keys.

The format for uploading keys to KeyBase is simple, a 3-column csv file (Table 2) that can, for example, be generated by saving an Excel spreadsheet to csv format. It accepts keys in many scripts, including Cyrillic, see, for example, the key to *Poa* and it segregates in the Flora Altaica project [24, 25].

Table 2. Format for uploading keys to KeyBase

1	Lemmas pubescent along keels and marginal veins, with a tuft of long hairs on the callus; lower sheaths and portions of culms compressed, almost 2-edged	Poa remota
1	Lemmas completely glabrous; sheaths and culms usually terete throughout, rarely slightly compressed neat the base (aggr. Poa sibirica)	2
2	Plants robust, growing predominantly in forest and subalpine belts; leaf blades 3-4 mm wide; lower internodes 2-4 mm in diameter	Poa insignis
2	Plants usually alpine; leaf blades 1.5-3 mm wide; lower internodes 1-2 mm in diameter	Poa sibirica

Once successfully uploaded, the keys can be viewed in three different formats, the default view in which 4 panes are used: one showing the choice that needs to be made, another the taxa not yet rejected by choices that have been made and, on the lower level, the choices that have been made and the taxa rejected. Such a format can only be used with a computer-based key, but KeyBase can also generate two other views, a bracketed view in which the leads are grouped by their number, and an indented view, in which one always moves to the choice below the lead being accepted. Indented keys make it easy to see which characters sort the taxa into different groups. These groups may be formal taxa, but they need not be.

KeyBase is also useful for checking (and reviewing) keys in a paper being submitted for publication. Of course, if being used to review a key in a manuscript, the key should be deleted once any problems with its format have been noted. Dr. Niels Klazenga, who developed and maintains KeyBase, is considering how to modify it so that keys can be hidden until a manuscript is accepted, just as GenBank[26] allows for sequence data.

4 Other resources

Table 1 cites many components of a flora that neither OpenHerbarium nor KeyBase that helps with. An illustrated glossary is a useful component of any flora. Dyreson will be exploring the feasibility of using Wikipedia [27, 28] for the glossary in a manner like that used in its own entries. Thus, a description would show the word “spikelet” as having a link which someone could click if they did not understand the term but ignore if they did. This approach is unobtrusive and would potentially benefit all users of Wikipedia, not just users of floras.

For the other components, such as medicinal uses and cytological information, the primary need is to be able to add links to the primary information source. We have not yet identified an approach for making this easier to do and freely accessible but there are undoubtedly some out there. While we look, we shall continue to provide information to OpenHerbarium and keys to KeyBase so that access to existing valuable, even if somewhat outdated, resources is increased.

5 Conclusions

Developing a regional flora an enormous amount of work. The programs and web sites suggested here can help with the task but, ultimately what is needed is intensive field work, including in the hard to reach places. All this paper tries to highlight is software that can help synthesize the information resulting from field and laboratory research and make it more accessible. Admittedly, OpenHerbarium although fully functional, still needs considerable development before its value will be clear, for example, many scientific names need to be added to its backbone. Nevertheless, as its resources increase, experience in the US indicates that it will become a critical component of a country’s research and

educational infrastructure. In the process, both contributors and users will gain a better understanding of the importance of good fields notes, herbarium specimens, and the features required for maximizing the value of collecting activities.

References

1. P. N. Krylov, Fl. Zapadnoi Sibiri, volumes 1-11 (Tomsk State University, Tomsk, Russia 1927–1949).
2. L. P. Sergievskaya, Fl. Zapadnoi Sibiri, volume 12, parts 1 and 2. (Tomsk State University, Tomsk, Russia 1951–1964).
3. Barkworth, M. E., L. K. Anderton, K. M. Capels, S. Long, and M. B. Piep. (Editors). Manual of grasses for North America. (Utah State University Press, Logan, Utah, USA, 2007).
4. M. Thulin (Editor), Flora of Somalia, volumes 1–4. (Royal Botanic Gardens, Kew, Richmond, England, UK, 1993-2006)
5. E. Nasir, S.I. Ali, M. Qaiser (Editors) Flora of Pakistan, volumes 1-225. (Centre for Plant Conservation, University of Karachi, Karachi, Pakistan, 1970-2020)
<http://legacy.tropicos.org/Project/Pakistan>
6. N. Klazenga, KeyBase: teaching old keys new tricks. <https://keybase.rbv.gov.au/>
7. OpenHerbarium: integrating records from multiple herbaria with images and research tools. <https://openherbarium.org/>.
8. Symbiota: promoting biocollaboration. Description: <http://symbiota.org/docs/>; code: <https://github.com/Symbiota/Symbiota>.
9. C. Dyreson, Collaborative Research: ABI Development: Symbiota2: Enabling greater collaboration and flexibility for mobilizing biodiversity data
<https://pdfs.semanticscholar.org/eda9/9c2709b1fb804f2fd243a10e76ea6ad32478.pdf> e;
code: <https://github.com/Symbiota2/Symbiota2>.
10. Geolocate: software for georeferencing natural history collections. <https://www.geolocate.org/>.
11. International Plant Names Index. <https://www.ipni.org/>
12. Tropicos: connecting the world to botanical data since 1982.
<https://www.tropicos.org/home>
13. SEINet Collaborative: A collection of separate projects that contribute to and share data. Access is provided by several different portals, e.g.,
<http://intermountainbiota.org/portal/>, <http://sernecportal.org/portal/>,
<http://www.soroherbaria.org/portal/#:~:text=Welcome%20to%20the%20Consortium%20of,%2C%20Idaho%2C%20Montana%20and%20Wyoming.>
14. M. E. Barkworth, K. M. Capels, S. Long, L. K. Anderton, and M. B. Piep, Flora of North America, volume 24. Poaceae, part 1. (Oxford University Press, New York, USA, 2007).
15. M.E. Barkworth, K.M. Capels, S. Long, M.B. Piep (editors). Flora of North America, volume 25. Poaceae, part 2. (Oxford University Press, New York, USA, 2003)
16. Encyclopedia of Life: global access to knowledge about life on earth. <https://eol.org/>
17. Creative Commons Licenses, <https://creativecommons.org/licenses/>.
18. Somaliland Biodiversity Foundation: building resources for studying biodiversity in Somaliland. <http://www.somalilandbiodiversity.org/home/>

19. World Flora Online; an online flora of all known plants.
<http://www.worldfloraonline.org/>
20. ITIS: Integrated Taxonomic Information System. <https://www.itis.gov/>
21. Biodiversity Heritage Library: inspiring discovery through free access to biodiversity knowledge. <https://www.biodiversitylibrary.org/>
22. Global Biodiversity Information Facility; free and open access to biodiversity data.
<https://www.biodiversitylibrary.org/>
23. R.J. Shaw, Vascular plants of northern Utah; an identification manual. Utah State University Press, Logan, Utah, U.S.A., (1909).
24. M.V. Olonova, *Key to Poa species aggregates and species in the Altai* [in Russian].
<https://keybase.rbg.vic.gov.au/keys/show/10483>.
25. Flora Altaica, <http://altaiflora.asu.ru/ru/>.
26. GenBank; NIH genetic sequence database. <https://www.ncbi.nlm.nih.gov/genbank/>.
27. Wikipedia, the free encyclopedia that anyone can edit.
https://en.wikipedia.org/wiki/Main_Page
28. Wikipedia, description of spikelet. <https://en.wikipedia.org/wiki/Spikelet>