Potential application of GBIF platform for flora inventory of Karaganda region

Margarita Ishmuratova[1*], *Maxim* Shashakov[1], and *Saltanat* Tleukenova[1]

[1]Faculty of Biology and Geography, Karaganda Buketov University, Karaganda, 100028, Kazakhstan

Abstract. The current biodiversity inventory has important scientific and practical implications for the country. In Kazakhstan, data on the flora of spore and vascular plants are outdated; there is a significant gap in assessing the current species composition of certain regions of Kazakhstan, including Central Kazakhstan (Karaganda and Ulytau regions). The problems of formation of a modern flora abstract are determined by the need to use the entire volume of accumulated information, both for individual findings of naturalists and multi-year collections in domestic and international herbarium collections. When carrying out research to analyze the current state of the flora of Karaganda region, the GBIF platform resources containing information on the biological diversity of flora and fauna were used. Data analysis showed that the platform has more than 130 thousand records of vascular plant finds for the territory of Kazakhstan, including 2241 records for Karaganda region with geo-referencing. The main funds containing the largest number of collections made on the territory of our region were determined. The stages of data processing and geo-linking of herbarium collections are defined. In general, a high potential for supplementing field collections with GBIF data to clarify the distribution of individual taxa and make range maps was determined.

1 Introduction

Present days, the only publication for plants is "Flora of Kazakhstan" (1956-1966). This is a monographic work, in which material for 5234 species of higher vascular plants was collected and systematized. During the past period, the flora has not been republished and has become very outdated. New species for science were described, floristic finds were revealed. The plant cover of Kazakhstan has undergone significant changes associated with climate change, expansion or narrowing of the range boundaries of individual taxa, and new botanical nomenclature was introduced (1986, 2005). Not all regions have information on distribution and phytoprotective status of plants, invasive species and economic properties of taxa are poorly studied. The above-mentioned changes make it necessary to update new data and re-publish the flora for individual regions. The studies on the assessment of the current floristic diversity of floras are widely presented worldwide [1-5].

* Correspondingauthor: margarita.ishmur@mail.ru

Currently, there are many information resources containing information on biodiversity at the level of individual findings of living organisms, including materials from biological collections. Many of them are available online; the largest resource that aggregates data from other sources is the Global Biodiversity Information Facility (GBIF [6-9].

2 Materials and Methods

The objects of research were: plants of natural flora of Karaganda region (higher spore and seed plants); herbarium collections in domestic and foreign herbarium funds, data on which are placed on international biodiversity platforms (GBIF и iNaturalist) [9, 10].

The statistical programming language R (4.3.1), the integrated development environment R-Studio (2023.03.0. Build 386) and the specialized package rgbif, designed for data exchange via the GBIF API, were used to automate the process of obtaining and processing the available information.

3 Results and Discussion

As part of the inventory of the flora of higher spore and vascular plants of Karaganda region, the team of the Department of Botany of Karaganda Buketov University started inventory of data on biodiversity platforms and in domestic and foreign herbarium funds.

It was found that in the global information system GBIF on October, 2023 presented 131,216 records of finds of vascular plants (Tracheophyta) for the territory of Kazakhstan. The main part is herbarium specimens - 86 168 records (Fig. 1), the main part of which is stored in the herbarium of the Lomonosov Moscow State University (53 151 specimens), acronym of herbarium – MW.

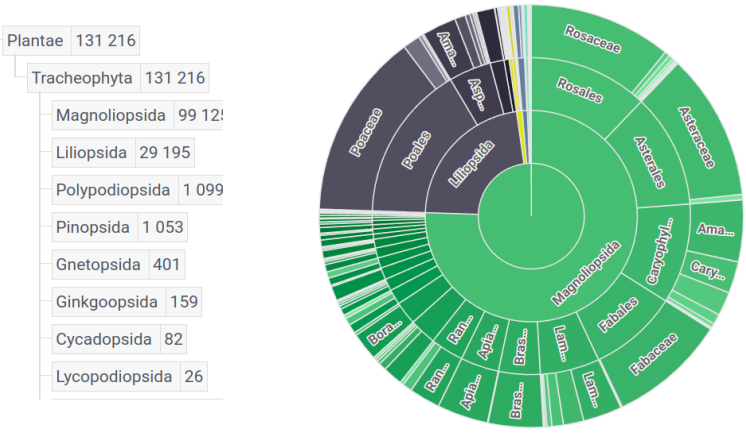

Fig. 1. Distribution of main taxonomic groups of plants for Kazakhstan.

According to the GBIF portal, herbarium sheet images are available for more than 65 thousand digitized specimens, of which geographic coordinates have about one fourth (about 19 thousand records).

Most of the information obtained through the GBIF portal is distributed under the open Creative Commons Attribution (CC-BY) license, which implies free use of the entire amount of information with attribution, the rest of the data under the CC-0 (attribution is not required) and CC-BY Non Commercial (CC-BY-NC) licenses, which does not allow commercial use of the data and any results. Thus, we can use in our work both tabular data

obtained through GBIF and images of herbarium sheets to decipher the information presented on the labels for the purpose of independent geo-linking.

A total of 60,216 herbarium leaf images were acquired. As the GBIF system is only an aggregator of biodiversity data, file storage and access to the images is provided by numerous organizations that publish data. For reasons of inaccessible links and other technical difficulties, images of about 5,000 herbarium sheets were inaccessible. Of the images obtained, 18,335 sheets are geo-linking records, while 41,881 sheets are not geo-linking.

Of all the collection specimens presented in GBIF, only about one third (28,836 records, not taking into account the presence of an image) have more or less accurate geo-linking, and it is obvious that the specified geographic coordinates are not always located in the territory of the country specified for the specimen (Fig. 2), there are at least several hundred such finds. On the territory of Central Kazakhstan (in the administrative boundaries of Karaganda region as of the beginning of 2022) there are 2241 geo-linking collection points.

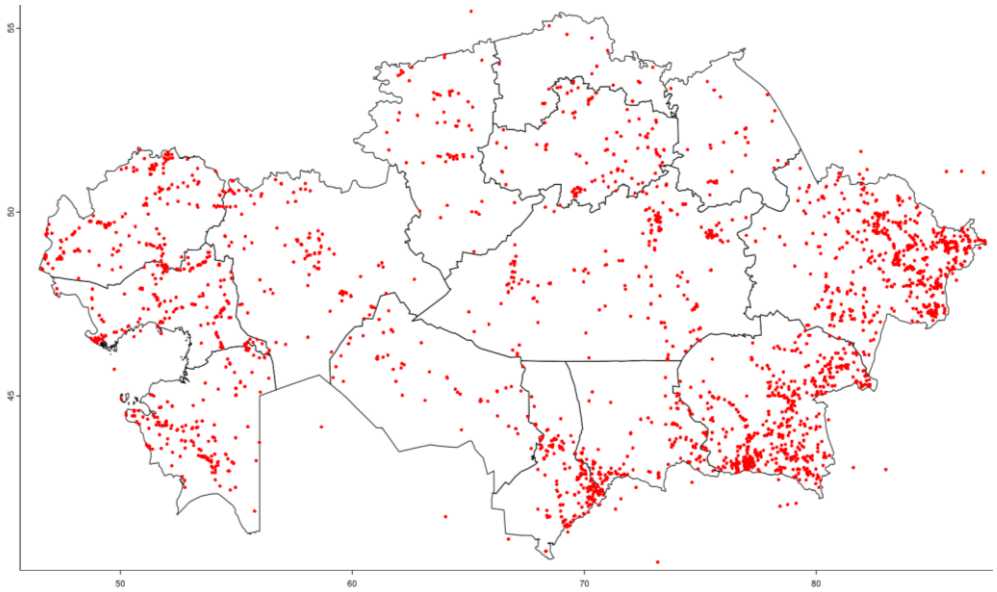

Fig. 2. Scheme of location of collections of geo-linking herbarium specimens, for which Kazakhstan is indicated as the place of collection (according to the data of the information system GBIF).

As of the end of 2023, 423392 records of species occurrences are available through the GBIF portal for the territory of Kazakhstan, of which 160315 come from scientific collections. Digitized data on more than 85 thousand herbarium sheets of vascular plants are available; the earliest collections belong to the second half of the XVIII century - 5 herbarium sheets, 5,417 specimens collected in the 19th century, 55,529 in the 20th century, and 13,232 in the 21st century. The distribution of the number of herbarium sheets for the territory of Kazakhstan according to the data available through the GBIF information system is presented in the figure below 3.

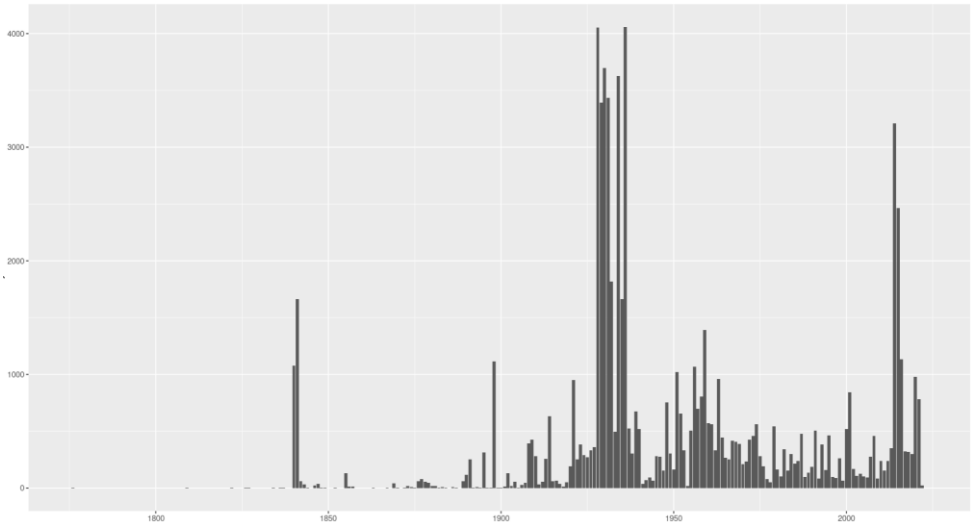

Fig. 3. Distribution of herbarized finds for the territory of Kazakhstan according to the data available through the platform GBIF.

The majority of records do not have geographical coordinates and in electronic form their geo-linking is indicated at the level of the country in which the corresponding specimens were collected. The main herbarium funds, which contain plant specimens collected on the territory of Karaganda oblast, have been identified (Tab. 1)

Table 1. Foreign herbaria that hold specimens collected from the territory of Kazakhstan (> 1000 sheets).

Herbarium, city, country	Acronym	Number of records
Lomonosov Moscow State University, Moscow, Russia	MW	53460
South Siberian Botanical Garden, Barnaul, Russia	ALTB	3257
Smithsonian Institution National Museum of Natural History, Washington, DC, USA	USNM	2107
Masaryk University, Brno, Czech Republic	BRNU	2103
Missouri Botanical Garden, USA	MO	1156
Meuse Botanical Garden, Brussels, Belgium	BR	1023

Since digitization of accumulated biodiversity data is a long and time-consuming process that requires expertise, the details of available data on individual finds are very heterogeneous. The most important information is an indication of the species, location, time and author of collection. Qualitative geo-linking is geographic coordinates with an assessment of the accuracy of the geo-linking. Specimens collected in the time before computer technology and satellite navigation, as a rule, are not provided with geographic coordinates, but contain only a text description of the collection site, which can be used to find the location using publicly available electronic cartography (Yandex Maps and Google Maps). However, such descriptions are not always available digitally, and for many finds and observations, often only the country in which they were collected is given, not always accurately, as for available specimens collected 100 or more years ago, the detail of the location description does not make it possible to find out with certainty where they are located within modern political boundaries. To establish the coordinates or confirm the

correctness of their determination, a digital image of the herbarium sheet (scan or photograph) is required, but this possibility is not always available. Therefore, even with the large amount of information available for the study area, additional work is needed to assess the quality of the data and to select finds suitable for the aims and objectives of the study.

Further work with this amount of data involves three steps:

1 Viewing images of scanned sheets and sorting them by region: Central Kazakhstan, Eastern, Western, Balkhash, etc. In the process, finds recorded erroneously for the territory of Kazakhstan are identified.

2 Geographical binding (determination of coordinates) of sheets from the priority region, according to the information provided on the label.

3 Checking the geographical linking of sheets for which it has already been specified.

The results of the inventory showed that among the collections for the territory of Karaganda region there are, including 2,231 type specimens stored in herbaria of Lomonosov Moscow State University, Paris Museum of Natural History, Vienna Museum of Natural History, Vienna University, New York Botanical Garden, Yerevan Institute of Botany named after A.L. Takhtajan, Natural Science Collections of Bavaria, etc.

It was also found out that only a little more than 28 thousand specimens have spatial linking in the form of geographical coordinates. Therefore, in order to form samples for analysis on separate regions of Kazakhstan, additional work on manual geo-linking of these materials to the regions of study is necessary. In this labor-intensive process, scanned samples should have priority in processing, as they allow for expert assessment of the correctness of determining the type.

4 Conclusion

The formation of a modern flora abstract should be based not only on the authors' own research and literature data, but it is necessary to apply international online resources, among which we can mention GBIF and iNaturalist. That is, in addition to scientific research, it is possible to use the potential of amateur observations, which significantly expands the information on the growth and morphology of the studied objects. Key resources containing the largest number of herbaria in the world have been identified, including Lomonosov Moscow State University, South Siberian Botanical Garden, Smithsonian Institution National Museum of Natural History, Masaryk University, Missouri Botanical Garden, and Meuse Botanical Garden.

Work with the data on the flora of Karaganda region allowed to reveal additional 2241 records on vascular plants with geo-linking. This allowed clarifying both species composition of separate groups of taxa and their areas. However, there is a high potential for expanding the information when organizing the geo-linking of the remaining finds.

The potential of the findings placed on the GBIF platform shows the possibility of refining the current list of flora of Karaganda region, taking into account the data of both private collections and the largest foreign herbarium funds. The obtained data after full geo-linking will allow making modern maps of species ranges for our region, clarify the localization of rare and endangered, as well as economically valuable species.

The data allow forming a modern monographic publication "Flora of Karaganda region" in the future, to make distribution maps for the most widespread taxa, as well as species with conservation status.

Acknowledgments

The research was carried out within the framework of the target funding program of Ministry of Science and Higher Education BR18574125 "Study of the current state of species diversity of vascular plants of Kazakhstan using modern methods of botany, molecular genetics and bioinformatics" (2023-2024).

Authors' contribution

Original Draft Preparation and Project Administration: M.I.; Conceptualization and Methodology: M.S. and M.I.; Visualization: M.S.; Investigation: S.T., H.G. and R.M.

References

1. S. Baasanmunkh, M. Urgamal, B. Oyuntsetseg, Flora of Mongolia: annotated checklist of native vascular plants. PhytoKeys, **192**, 63-169 (2022). https://doi.org/10.3897/phytokeys.192.79702

2. X. Chen, F.Q. Jiang, Y.J. Wang, Y.M. Li, R.J. Hu, Characteristics of the Eco-geographical pattern in arid land of central Asia. Arid Zone Res., **30**, 385–390 (2013)

3. M. Nobis, G. Domina, M. Meco et al., Contribution to the flora of Asian and Europe countries: a new regional vascular plant records. Botany Letters, **165**(1), 1-23 (2018). https://doi.org/10.1080/23818107.2017.1415817

4. V. Gundegmaa, T. Dashmaa, C. Bilegtmandakh et al., The vascular flora of Sutai Khairkhan Mountain Nature reserve, Mongolia. Botanica Pacifica. A Journal o Plant Science and Conservation, **1191**, 115-128 (2022). https://doi.org/10.17581/bp.2022.11102

5. R.F. Wagensommer, Floristic Study in the Light of Biodiversity Knowledge and Conservation. Plants, **12**(15), 1-3 (2023). https://doi.org/10.3390/plants12162973

6. S. Gaiji, V. Chavan, A.H. Arino et al., Content assessment of the primary biodiversity data published through GBIF network: Status, Challenges and Potentials. Biodiversity Information, **8**(2), 94-172 (2013). https://doi.org/10.17161/bi.v8i2.4124

7. N.V. Ivanova, M.P. Shashkov, The possibilities of GBIF data use in ecological research. Russian Journal of Ecology, **52**(1), 1-8 (2021). https://doi.org/10.1134/S1067413621010069

8. U. Burkhardt, D.J. Russell, P. Decker et al., The Edaphobase project of GBIF-Germany: A new online soil-zoological data warehouse. Appl. Soil Ecol., **83**, 3–12 (2014). https://doi.org/10.1016/j.apsoil.2014.03.021

9. H. Qian, J. Zhang, M. Jiang, Global patterns for fern species diversity: an evaluation of fern data in GBIF. Plant Diversity, **44**, 135-140 (2022). https://doi.org/10.1016/j.pld.2021.10.001

10. A. Echeverria, I. Ariz. J. Moreno, J.P. de Andres, E.M., Gonzalez. Leaning plant biodiversity in nature: the use of the citizen-science platform iNaturalist as a collaboration tool in secondary education. Sustainability, **13**(2), 1-12 (2021). https://doi.org/10.3390/su13020735