1. Dataset Title: Hazara University’s monocot specimen data in OpenHerbarium.org on 20 December 2023

2. Name and contact information of PI: Mary Barkworth ([mary.barkworth@usu.edu](mailto:mary.barkworth@usu.edu); +1-435-557-2121)

a. Name: Mary Barkworth

b. Institution: Utah State University

c. Address: Biology Department, 5305 Old Main Hill, Logan, Utah 84322-5305, U.S.A.

d. Email: mary.barkworth@usu.edu

e. ORCiD ID: http://orcid.org/0000-0001-9785-1538

3. Name and contact information of Co-PI: Liaqat Ullah (liaqatullah08@gmail.com)

a. Name: Liaqat Ullah

b. Institution: Hazara University

c. Address: Botany Department, Hazara University, Mansehra, Khyber Pakhtunkhwa, Pakistan

d. Email: liaqatullah08@gmail.com

e. ORCiD ID:

4. Name and contact information of Co-PI: Abdul Majid (abdulmajid@hu.edu.pk)

a. Name: Abdul Majid

b. Institution: Hazara University

c. Address: Botany Department, Hazara University, Mansehra, Khyber Pakhtunkhwa, Pakistan

d. Email: abdulmajid@hu.edu.pk

e. ORCiD ID

5. Name and contact information of Co-PI: Ghulam Mujtaba Shah (gmujtabashah72@yahoo.com)

a. Name: Ghulam Mujtaba Shah

b. Institution: Hazara University

c. Address: Botany Department, Hazara University, Mansehra, Khyber Pakhtunkhwa, Pakistan

d. Email: gmujtabashah72@yahoo.com

e. ORCiD ID

6. Name and contact information of Co-PI: Jan Alam (janalamkuh@yahoo.com)

a. Name: Jan Alam

b. Institution: Hazara University

c. Address: Botany Department, Hazara University, Mansehra, Khyber Pakhtunkhwa, Pakistan

d. Email: janalamkuh@yahoo.com

e. ORCiD ID

5. (Repeat if needed)

6. Funding source: None

7. Project abstract:

Northern Pakistan is interpreted here as including three administrative regions: Khyber Pakhtunkhwa, Gilgit-Baltistan, and Azad Jammu and Kashmir. It is dominated by multiple mountain ranges whose valleys drain, directly or indirectly, into the Indus River. Hazara University, which is in Khyber Pakhtunkhwa, was started in 2002 and its herbarium (HUP) in 2005.  Digitization of the herbarium's vascular plants began in 2021 with its records being shared in OpenHerbarium and GBIF. Its online presence is now growing more rapidly than any other Pakistani herbarium. This paper summarizes the taxonomic diversity and geographic origin of its monocot holdings in 2023. In 2023, the Hazara University Herbarium held 744 monocot specimens from Northern Pakistan. They belonged to 9 orders, 23 families, 126 genera, amd 299 species. Most specimens belonged to the Poales, with many more belonging to the Poaceae than to the Cyperaceae, the family with the next best representation in the collection. Two orders, Acorales and Dioscoreales were represented by only one specimen each. Slightly more than half the species were represented by only one specimen; eleven species were represented by 9 or more specimens. Comparison with a checklist developed from multiple sources, including a GBIF download, revealed that HUP did not have any reprsentatives of five families that other sources report as occurring in Northern Pakistan: Juncaginaceae, Potamogetonaceae, Nartheciaceae, Pontederiaceae, and Musaceae.Most of the herbarium's specimens were from Khyber Pakhtunkhwa, the university's home province, the fewest from Gilgit-Baltistan. The best represented districts (level 3 regions) were Districts Swat, Chitral, and Lower Dir in Khyber Pakhtunkhwa and Neelum in Azad Jammu and Kashmir, all of which are known for their scenic beauty and floristic wealth. The best represented district of Gilgit-Baltistan was Hunza which is home to three passes through the Karakoram Mountains.

8. Brief description of collection and processing of data: Data were transcribed from the labels on herbarium sheets into a spreadsheet. They were then uploaded to OpenHerbarium.org, a site that shares, displays, and visualizes occurrence data from collections of plants, fungi, and algae. The spreadsheet was designed to enable sharing the data using the Darwin Core Standard for sharing occurrence data. Geographic coordinates were added to records that did not have them, using other information on the label. The data in this file were downloaded on December 20, 2024.

9. Description of files (names, or if too numerous, number of files, file type(s): Only 1 file, csv format.

10. Definition of acronyms, codes, and abbreviations: WGS84: World Geodetic Standard set in 1984.

11. Description or definition any other unique information that would help others use your data: See Darwin Core. 2023. A standard to facilitate the sharing of information about biological diversity. <https://dwc.tdwg.org/> and Wieczorek et al. (2023) Darwin Core List of Terms. <https://dwc.tdwg.org/list/>.

12. Descriptions of parameters/variables:

a. Temporal (beginning and end dates of data collection: Data capture began in 2021; data processing was completed in February 2024.

b. Instruments used and units of measurements: no special instruments used; units of measurements in column headings or record..

c. Column headings of data files (for tabular data: NOTE: Many of the columns are empty. They result from downloading data from a site designed for sharing occurrence data from many different kinds of natural history collection. Leaving them in will facilitate integration with occurrence records from other kinds of collectiois.

id, institutionCode, collectionCode, ownerInstitutionCode, basisOfRecord, occurrenceID, catalogNumber, otherCatalogNumbers, higherClassification, kingdom, phylum, class, order, family, scientificName, taxonID, scientificNameAuthorship, genus, subgenus, specificEpithet, verbatimTaxonRank, infraspecificEpithet, taxonRank, identifiedBy, dateIdentified, identificationReferences, identificationRemarks, taxonRemarks, identificationQualifier, typeStatus, recordedBy, associatedCollectors, recordNumber, eventDate, eventDate2, year, month, day, startDayOfYear, endDayOfYear, verbatimEventDate, occurrenceRemarks, habitat, substrate, verbatimAttributes, fieldNumber, eventID, informationWithheld, dataGeneralizations, dynamicProperties, associatedOccurrences, associatedSequences, associatedTaxa, reproductiveCondition, establishmentMeans, cultivationStatus, lifeStage, sex, individualCount, preparations, locationID, continent, waterBody, islandGroup, island, country, stateProvince, county, municipality, locality, locationRemarks, localitySecurity, localitySecurityReason, decimalLatitude, decimalLongitude, geodeticDatum, coordinateUncertaintyInMeters, verbatimCoordinates, georeferencedBy, georeferenceProtocol, georeferenceSources, georeferenceVerificationStatus, georeferenceRemarks, minimumElevationInMeters, maximumElevationInMeters, minimumDepthInMeters, maximumDepthInMeters, verbatimDepth, verbatimElevation, disposition, language, recordEnteredBy, modified, sourcePrimaryKey-dbpk, collID, recordID, references

d. Location/GIS Coverage (if applicable to data): Khyber Pakhtunkhwa, Gilgit-Baltistan, and Azad Jammu and Kashmir, the three northern provinces of Pakistan.

e. Symbol used for missing data: Empty cell

13. Special software required to use data: None

14. Publications that cite or use this data: Liaqat et al. (to be submitted) Diversity and distribution of monocots in Northern Pakistan: a case study from the Hazara University Herbarium, Biodiversity Data Journal.

15. Was data derived from another data source? No. If so, what source?

OPTIONAL Fields

If these pertain to your data, consider including them to facilitate use your data in the future:

1. Uncertainty, precision, and accuracy of measurements, if known: The geographic coordinates vary widely in their accuracy, some being to centroid of the district (level 3 administrative unit), some to a smaller unit. They should be used with caution.

2. Quality assurance and quality control that have been applied: The names of the second and third level administrative units have been checked, but there is no gazetteer giving an “official” spelling for most localities in the country. The scientific names of the plants were all considered “accepted” at the time the data were recorded; some have now been placed in synonymy with other names.

3. Known problems that limit data use (quality control, sampling issues, etc.):

4. Related datasets outside of this dataset:

5. Example records for each data file or file type: For an updated listing of monocot records from Northern Pakistan in the Hazara University Herbarium, see https://openherbarium.org/collections/list.php?db=84&country=Pakistan&state=Khyber Pakhtunkhwa; Gilgit-Baltistan; Azad Jammu and Kashmir&taxa=Liliidae&usethes=1&taxontype=2

6. Description of relationship between files and/or any file dependencies:

7. Information about other files (names, locations) and documents (such as field notes, publications, etc.) that would be helpful to a person using your data.