

# Agrobiodiversity Indicators and Measurement using R for Description, Monitoring, Comparison, Relatedness, Conservation and Utilization

**Bal Krishna Joshi**

Nepal Agricultural Research Council, Kathmandu, Nepal. Email: [joshibalak@yahoo.com](mailto:joshibalak@yahoo.com) | ORCID: <https://orcid.org/0000-0002-7848-5824>

**How to cite this paper:** Joshi, B.K. (2021). Agrobiodiversity Indicators and Measurement using R for Description, Monitoring, Comparison, Relatedness, Conservation and Utilization. *Agrobiodiversity & Agroecology*, 01(01): 47-64. Doi: <https://doi.org/10.33002/aa010103>

**Received:** 19 August 2021

**Reviewed:** 15 September 2021

**Accepted:** 30 September 2021

**Published:** 10 November 2021

Copyright © 2021 by author(s)

**Publisher's Note:** We stay neutral with regard to jurisdictional claims in published maps, permissions taken by authors and institutional affiliations.

**License:** This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). <http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Agrobiodiversity is the most important part of biodiversity. It can be described, quantified, compared, and related by using different statistical tools called agrobiodiversity statistics (agro-statistics). Six components and 25 groups of agrobiodiversity should be used for agrobiodiversity analysis. Six types and levels of agrobiodiversity can be quantified. Both quantitative and qualitative data are used for estimating scores and indices. The measurement objects for describing agrobiodiversity are community, household, site, crop group, species, landrace, etc. These objects are called operational agricultural units (OAU). Agromorphological, molecular, and perception data are used in agrobiodiversity studies. Among the many software, RStudio is very good. It is an integrated part of R and includes a console, syntax-highlighting editor, tools for plotting, history, debugging, and workspace management. Vegan and BiodiversityR packages are commonly used for estimating diversity indices and multivariate analysis. Richness, Shannon index and Simpson index are very common means of quantifying agrobiodiversity. Spatial and temporal analysis of agrobiodiversity helps monitor the status and plan the programs and activities.

## Keywords

Agrobiodiversity index; Agrobiodiversity statistics; Measurement; Indicators; R package

## 1. Introduction

Agrobiodiversity (also called agricultural genetic resources, AGRs) is a part of biodiversity and includes all genetic resources that are economically beneficial. In majority of the countries, native agrobiodiversity is neglected and underutilized due to their high priority to monomorphic and high yielding varieties. Many different factors are contributing to losing the AGRs. Among them the major factor is the rapid expansion of single improved homogenous varieties and breeds in the world. Such single improved variety is generally developed through studying a single species or variety or set of genotypes, and there are limited studies on the whole agrobiodiversity at a particular site. The general trend is that, rather than evaluating, describing and improving the native agrobiodiversity, improved variety or breed is easily adopted and expanded due to which many indicators are being affected. Indicators are any values, scores or status which explain about the agrobiodiversity of a particular location. Agrobiodiversity indicators have not been standardized across the world; and even the methodologies to estimate and measure the indicators are not available. Indicators are very important to manage

*Editor-in-Chief:*

Dr. Didier Bazile (France)

*Deputy Editors-in-Chief:*

Dr. Habil. Maria-Mihaela Antofie

(Romania); Dr. Gordana Đurić

(Bosnia i Hercegovina)

*Technical & Managing Editor:*

Dr. Hasrat Arjumend (Canada)

the agrobiodiversity better, to plan programs and activities, and to monitor the trends (Sthapit *et al.*, 2017; PAR, 2018).

For the conservation of forest biodiversity (non-agrobiodiversity), different indicators and approaches have been used, for example red listing of the species. Many types of species have been defined and given due attention. Different types of species include Alien, Charismatic, Dominant, Emblematic, Endangered, Endemic, Exotic, Flagship, Focal, Foundation, Indicator, Indigenous, Invasive, Keystone, Landscape, Priority, Rare, Specialty, Substitute, Surrogate, Target, Threatened, Tourism, Umbrella and Vulnerable species. Similar approaches can be applied at species and landrace level to support AGRs. Quantification of AGRs is another aspect that identifies such species or landraces.

Different types of scores and indices along with coefficients can be estimated and used as indicators (Joshi *et al.*, 2005; Jarvis *et al.*, 2000; Grum and Atieno, 2007)). Several statistical tools can be applied using computer software to quantify agrobiodiversity. Quantifications (measurements) of agrobiodiversity are generally done at different levels e.g., at the agroecosystem, species, varieties, and administrative units. Agrobiodiversity in any area should be estimated properly that leads to choosing the conservation approaches effectively. This paper, therefore, describes different operational agricultural units (OAU) for estimating diversity indices using R packages. Among the various components under agrobiodiversity statistics, this paper focuses on the measurement of agrobiodiversity. With the approaches described in this paper, one can rank any household, community, district, or the country and can locate a center of the diversity. A hotspot of agrobiodiversity and red zone for agrobiodiversity can be identified, in addition to identifying the indicator species and landraces.

## 2. Agrobiodiversity Components and Groups

Agrobiodiversity covers all genetic resources that have value for food, nutrition, health, and other economic uses to human beings. It has six components, and they are crops, forages, livestock, insects, microorganisms, and aquatic genetic resources (Joshi *et al.*, 2020c). Insects and microorganisms include only economic and beneficial species. Under aquatic genetic resources, only economically important species are included e.g., fish. Each of these components can further be divided into four sub-components. They are cultivated/ domesticated, semi-domesticated, wild relatives, and wild edible species (Joshi and Shrestha, 2017; Joshi and Shrestha, 2019).

Based on the economic uses, agricultural genetic resources can be grouped into 25 groups. They are 1. cereals, 2. pseudocereals, 3. millets, 4. sugar and starch crops, 5. grain legumes, 6. oilseed crops, 7. summer vegetables, 8. winter vegetables, 9. roots and tubers, 10. winter fruits, 11. summer fruits, 12. spices, 13. beverages and narcotics, 14. fibers, 15. forage trees, 16. forage grasses, 17. ornamental plants, 18. medicinal plants, 19. supportive plants, 20. economic and beneficial (EB) insects, 21. EB microorganisms, 22. fish/aquatic animals, 23. aquatic plants, 24. poultry, and 25. livestock (Joshi and Shrestha, 2019, Joshi and Shrestha 2017). Supportive plants include green manuring crops, cover crops, pesticide plants, and other economically important plants that are not included in the above groups.

These components, sub-components, and economic groups (Joshi *et al.*, 2020c; Joshi and Shrestha, 2019) are very useful to estimate different types of diversity indexes, indicators, and scores of a particular site, community, or household over a certain period. The AGRs may be of exotic and native types and both types can be considered for agrobiodiversity measurement, but measurement based on only native AGRs would be

more valuable and important. There are many other grouping systems of AGRs (Joshi and Shrestha, 2019), and these groups can also be considered to quantify agrobiodiversity.

### 3. Agrobiodiversity Levels and Types

Agrobiodiversity can be measured and studied at different levels or hierarchies by using different traits. Based on levels (coverage of objects), there are six types of agrobiodiversity (Figure 1) (Joshi *et al.*, 2020b; Bajracharya *et al.*, 2012). Genetic diversity includes three levels of diversity i.e., varietal diversity, genotypic diversity, and allelic diversity. Agrobiodiversity can also be described under six types of diversity based on traits and use-values. These include functional diversity, morphological diversity, molecular diversity, use-value diversity, nutritional diversity, and food diversity. All these 12 types of diversity should be measured and studied at a particular site in a given period. Based on the data types, objectives, and objects, different measures are used to estimate and compare these different types of agrobiodiversity. Diversity can also be assessed based on cropping patterns, growing season, land type and habitat. at species and varietal levels. Morpho type is very simple indicator to measure the diversity.

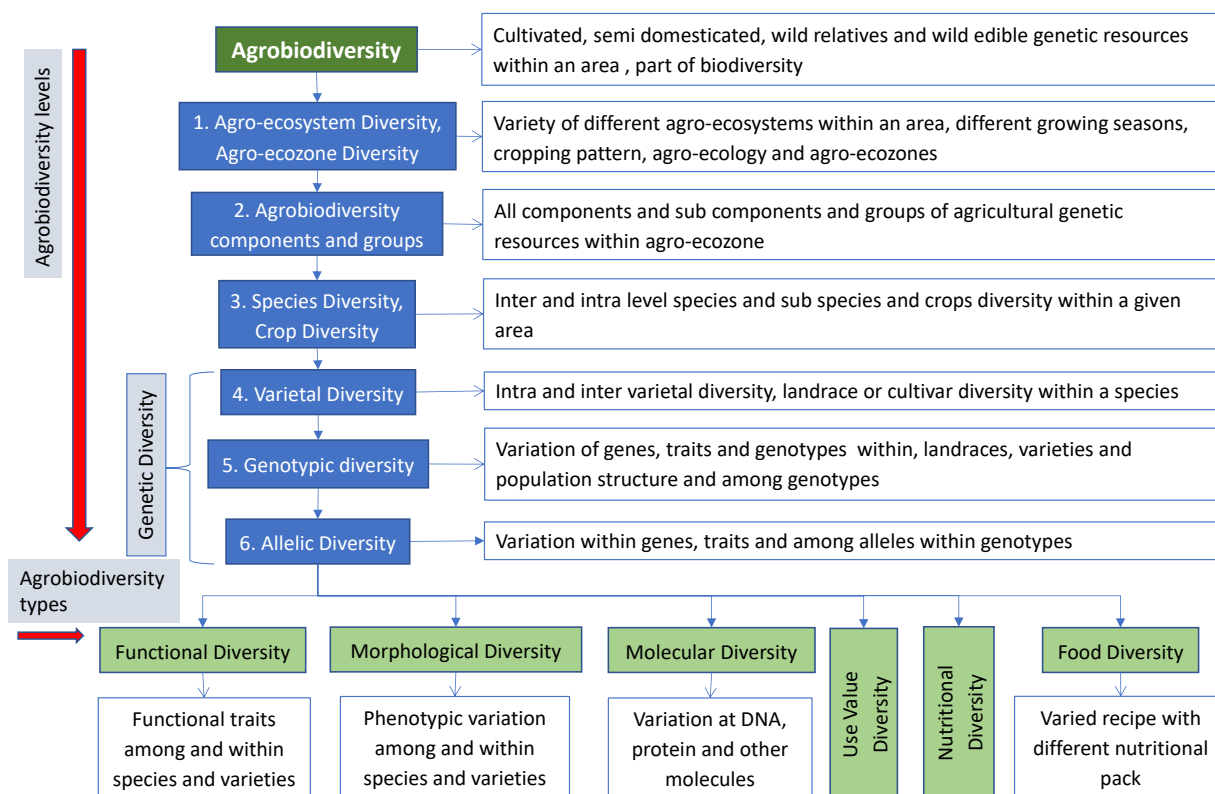


Figure 1: Types of agrobiodiversity based on levels, traits, and use-values. *Source: Joshi et al. (2020b)*

### 4. Agrobiodiversity Statistics (Agro-statistics)

Agro-statistics is a science of studying agrobiodiversity using different statistical tools, methods, and principles. Many common statistical tools are useful for measurement (quantification), characterization (description), classification (grouping),

evaluation (comparison) and association (relationship) of agrobiodiversity (Figure 2) (Bajracharya *et al.*, 2012; Grum and Atieno, 2007; Jarvis *et al.*, 2000; Joshi *et al.*, 2005). With the development of different molecular markers and computing software, genetic parameters are also commonly estimated. Description of these tools has been described by Joshi *et al.* (2005). Both parametric and non-parametric tests are also commonly used to compare agrobiodiversity. Appropriate test statistics are given in figure 3 based on data types and the number of objects (factors) used. Both temporal and spatial analysis (called trend analysis) can be carried out to see the status and changes in agrobiodiversity.

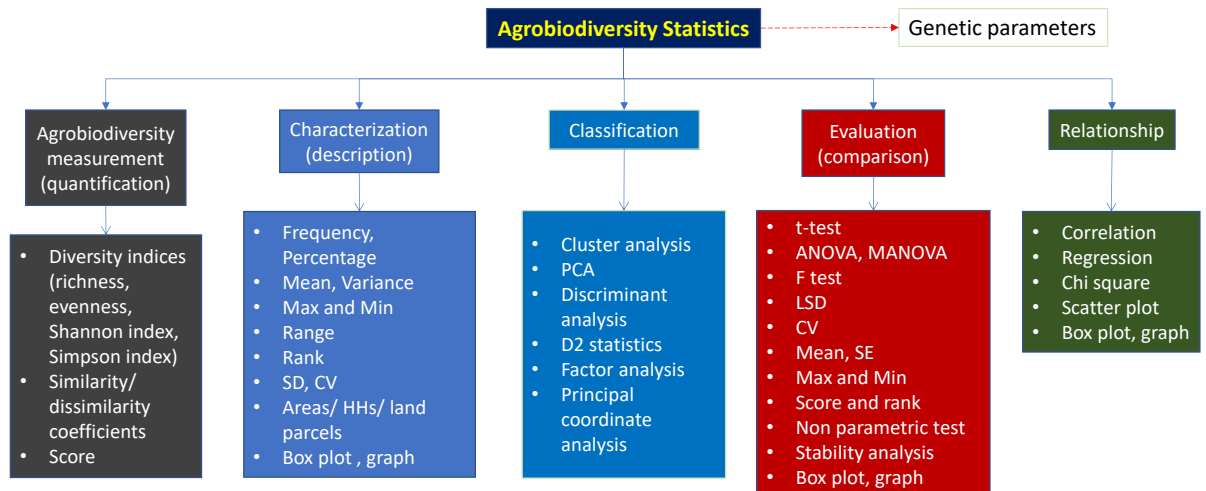


Figure 2: Different statistical tools for agrobiodiversity study.

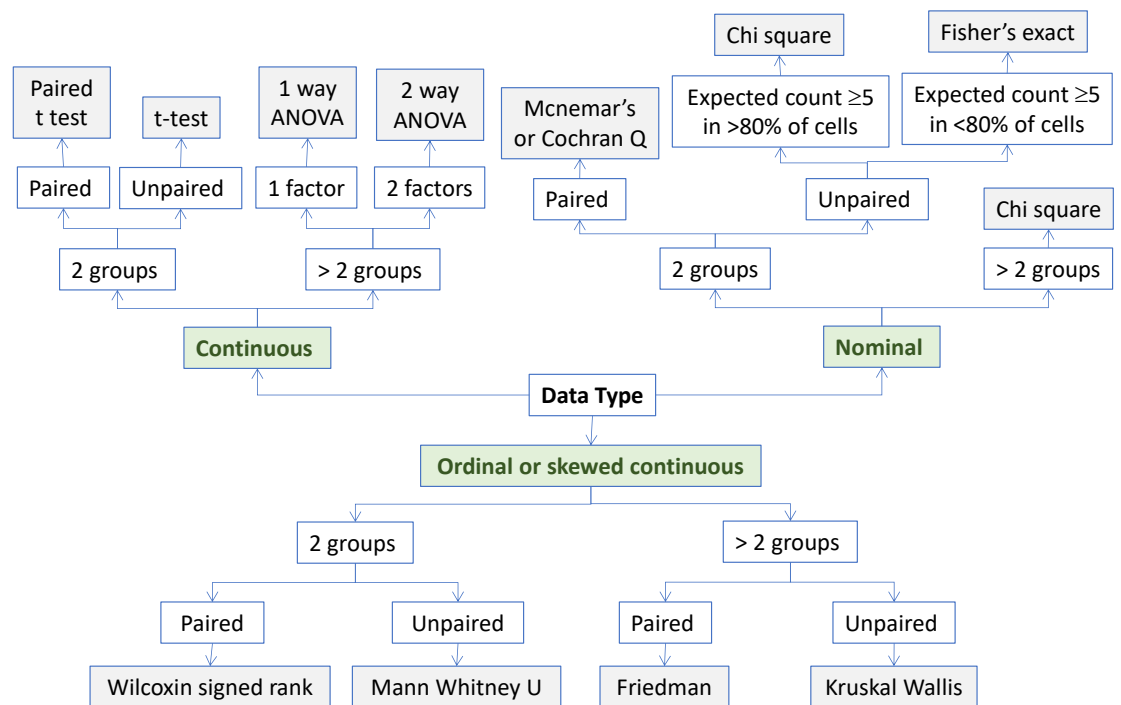


Figure 3: Statistical testing tools (parametric and non-parametric) for comparing agrobiodiversity based on data types

## 5. Agrobiodiversity Measurement (Quantification)

Agrobiodiversity measurement includes the quantification of AGRs at different levels. Based on the quantification, AGRs can be grouped at the level of different strata e.g., red list, endangered, rare, common, etc. (Joshi and Shrestha, 2019). The main measures of agrobiodiversity are richness, evenness, diversity indices (Shannon, Simpson indices), similarity coefficients, dissimilarity coefficients, scores (Joshi *et al.*, 2005; Kindt and Coe, 2005; Joshi *et al.*, 2018; Jarvis *et al.*, 2000; Grum and Atieno, 2007). Another measure is species density, which takes into account the number of species in an area. Similarly, landrace density can also be estimated. These measures should be measured at six different levels and types of agrobiodiversity (Figure 1) e.g., household, community, ward, municipality, district, province, and country. Such estimates are generally calculated based on native agrobiodiversity and are, therefore, useful for identifying the hotspot areas for agrobiodiversity. Quantification helps locate the center of diversity, identify the hotspot and red zone areas for agrobiodiversity. Hotspot areas are those areas that have the higher diversity score and indices, high diversity on wild relatives, endemic species, many rare and unique landraces, and species, and different types of land and cropping patterns.

Measurement (quantification) may be based on phenotypic, genotypic, perception, and survey data. Such data can be collected and measured through community biodiversity register and community seed bank, diversity block, diversity collection, diversity fair, field/transect walk, focus group discussions, food fair, household survey, key informant interviews, online survey, lab experiment, literature review, local market, on-farm, and on-station trials. Diversity changes over time and space are also estimated using different diversity measures, which are important for monitoring and applying appropriate methods for conservation and utilization.

For the index calculation at different levels, one can count the number of species within-group, or several landraces within species as well as group (PAR, 2018; Pudasaini *et al.*, 2016; Borcard, Gillet and Legendre, 2011; Grum and Atieno, 2007; Joshi and Baniya, 2006). Taking the natural logarithms of species richness or landrace richness, an index can be calculated. The proportion of each group, species, or landraces can be calculated by dividing the number of that groups, species, or landraces by the total number of all groups, species, or landraces in a given area. The formula for calculating the Shannon diversity index, Simpson index, evenness, and other indices can be applied on these data. Agrobiodiversity index (ABDI) can be of household (HH), village or community, district, province, agroecozone, and country. A weighted index using either agrobiodiversity components or groups can be estimated as described in the literature<sup>1</sup>. In some cases, microorganisms, insects, ornamental plants, and the medicinal plant may be excluded from the calculation due to data unavailability.

The percentage of species or landraces in each group or species can be calculated considering the total number of species or landraces in the country or studied areas (Pudasaini *et al.*, 2016; Joshi *et al.*, 2018; Joshi *et al.*, 2007). Based on the data obtained, each household or area or district can be ranked. For example, ABDI (based on landraces) for each household is equal to the number of landraces in each species or group divided by the total number of landraces in a community or district.

## 6. Agrobiodiversity Indicators (Score and Index)

<sup>1</sup> <https://news.mongabay.com/2016/05/top-10-biodiverse-countries/>

Agrobiodiversity indicators are any scores, indices, signs, symptoms, values, drivers, or marks that speak about the status of total diversity, trends on diversity, the status of intra- and inter-level diversity of species, and landraces in a particular area. It indicates that the agrobiodiversity is increasing, remaining constant, or decreasing. There is a wide range of methods of measuring various dimensions of agrobiodiversity, which is often referred to as the agrobiodiversity indicators, scores, and indices (Boversity International, 2017; Sthapit *et al.*, 2017; PAR, 2018; Kindt and Coe, 2005; Joshi *et al.*, 2020b). Diversity indicators, indices, and scores can be used to compare within and between different populations at species, landraces, and genetic levels over locations and years.

Agrobiodiversity indicators can be assessed at three different systems, namely, in consumption and market system, in production system, and in genetic resource management system (Sthapit *et al.*, 2017). Some indicators include the red zone, red list, landraces coverage (based on five cell analysis), cropping pattern, mixture, monocrop vs. multicrops, land type, food items, native products in the market, the richness of species and landraces, population size, etc. A red list is the list of names of genetic resources (at genotype, landrace, variety, strain, and breed levels) under different groups based on the analysis of distribution and population size (also called five cell analysis), and trait distribution. Among these indicators, scores and indices are more commonly estimated and used.

Diversity indices and scores are calculated using both qualitative and quantitative data. In case of quantitative data, it needs to be converted into qualitative groups. The proportion of entries in  $i^{\text{th}}$  class can be calculated using morphological data considering the different phenotypic classes of traits. Similarly, frequency data on genebank collection can be used to estimate different indices. Many ways can be used to estimate several types of household scores and indices. Household-level diversity can be of household diversity score and index as given below.

### 6.1 A1. Household Agrobiodiversity Score (HHABDS)

1. Number of species (species richness,  $n$ ) in each of 6 agrobiodiversity components (crops, forages, livestock, economical insects, economically important microorganisms, aquatic agricultural species) over a year
2. Number of landraces (landrace richness,  $n$ ) per species for each of 6 components in a year
3. Land type,  $n$  (marshy/ wetland, pond/aquatic, slopy upland, terrace upland, slopy low land, terrace low land, riverside, agroforestry land, grassland)
4. Functional diversity (number of special functions using special landraces) in a year
5. Unique diversity value (the number of specialty/ unique landraces divided by the total number of landraces)
6. Agrobiodiversity group score (or agrobiodiversity group richness) (based on 25 agrobiodiversity groups i.e., cereals, pseudocereals, millets, sugar and starch crops, grain legumes, oilseed crops, summer vegetables, winter vegetables, roots and tubers, winter fruits, summer fruits, spices, beverages and narcotics, fibers, forage trees, forage grasses, ornamental plants, medicinal plants, supportive plants, economical and beneficial (EB) insects, EB microorganisms, fish and aquatic animals, aquatic plants, poultry, and livestock), at 0 or 1 scale over a year with maximum 25 score
7. Dietary diversity score (based on 15 groups: cereals, pseudocereals, millets, roots and tubers, vegetables, fruits, nuts, meat and poultry, eggs, fish and aquatic animals,



- pulses and legumes, milk and milk products, oil/fat and ghee<sup>2</sup>, sugar and honey, and miscellaneous) at 0 or 1 scale on half-year basis with maximum 15 score
8. Social agrobiodiversity score (number of religious or culturally associated landraces, considering all 6 agrobiodiversity components)
  9. Food diversity score (number of food items/recipes eaten per meal, average of morning, day, and evening foods)
  10. Food component score (number of species in food per meal, average of morning, day, and evening foods)
  11. The average area per species (crops and forages) in square meter
  12. HH agrobiodiversity score: sum from above 1 to 10 scores.

## 6.2 A2. Household Agrobiodiversity Index (HHABDI)

### A. Based on species within agrobiodiversity group

- HH agrobiodiversity group richness, n
  1. HH Shannon diversity index (based on number of species within a group)
  2. HH Simpson index (based on number of species within a group)
  3. HH species evenness (specie within a group)

### B. Based on landraces within the agrobiodiversity group

- HH agrobiodiversity group richness, n
  4. HH Shannon diversity index (based on number of landraces within a group)
  5. HH Simpson index (based on number of landraces within a group)
  6. HH landraces evenness (specie within a group)

### C. Based on landraces within species

- HH agrobiodiversity species richness, n
  7. HH Shannon diversity index (based on number of landraces within a species)
  8. HH Simpson index (based on number of landraces within a species)
  9. HH species evenness (specie within a group)

HH agrobiodiversity index (HHABDI): sum of above 1 to 9 index values.

In the similar way of household scores and indices, one can estimate village or community agrobiodiversity scores and indices as follows.

## 6.3 B.1. Village Agrobiodiversity Score (VABDS)

1. Number of species (species richness, n) in each of 6 agrobiodiversity components (crops, forages, livestock, economical insects, economical microorganisms, aquatic agricultural species) over a year
2. Number of landraces (landrace richness, n) per species for each of 6 agrobiodiversity components over a year
3. Land type, n (marshy/ wetland, pond/aquatic, sloppy upland, terrace upland, sloppy low land, terrace low land, riverside, agroforestry land, grassland)
4. Functional diversity (number of special functions using special landraces) in a year
5. Unique diversity value (number of specialty/ unique landraces, functional trait-specific genotypes divided by total number of species)
6. Village agrobiodiversity score (based on 25 agrobiodiversity groups, i.e. cereals, pseudocereals, millets, sugar and starch crops, grain legumes, oilseed crops, summer vegetables, winter vegetables, roots and tubers, winter fruits, summer fruits, spices, beverages and narcotics, fibers, forage trees, forage grasses, ornamental plants, medicinal plants, supportive plants, economical and beneficial (EB) insects, EB microorganisms, fish and aquatic animals, aquatic plants, poultry, and livestock) at 0 or 1 scale over a year with maximum 25 score

<sup>2</sup> It is made by melting butter.

7. Village dietary diversity score (based on 15 groups: cereals, pseudocereals, millets, roots and tubers, vegetables, fruits, nuts., meat and poultry, eggs, fish and aquatic animals, pulses and legumes, milk and milk products, oil/ fat and ghee, sugar and honey, and miscellaneous) at 0 or 1 scale on half-year basis with maximum 15 score
8. Social agrobiodiversity score (number of religious or culturally associated landraces, considering all 6 agrobiodiversity components)
9. Food diversity score (number of food items/recipes eaten per meal, average of morning, day, and evening foods)
10. Food component score (number of species in food per meal, average of morning, day, and evening foods)
11. Village agrobiodiversity score: sum of above 1 to 10 values
12. The average area per species (crops and forages) in square meter
13. Average agrobiodiversity HH score
14. Average social agrobiodiversity HH score
15. The average number of species per HH
16. The average number of landraces per HH
17. Average areas per HH.

#### **6.4 B.2. Village Agrobiodiversity Index (VABDI)**

- A. Based on species within agrobiodiversity group
  - Agrobiodiversity group richness, n
  - 1. Village Shannon diversity index (based on number of species within a group)
  - 2. Village species evenness (specie within a group)
  - 3. Village Simpson's index
- B. Based on landraces within the agrobiodiversity group
  - Agrobiodiversity group richness, n
  - 4. Village Shannon diversity index (based on number of landraces within a group)
  - 5. Village landraces evenness (specie within a group)
  - 6. Village Simpson's index
- C. Based on landraces within species
  - Agrobiodiversity species richness, n
  - 7. Village Shannon diversity index (based on number of landraces within a species)
  - 8. Village species evenness (specie within a group)
  - 9. Village Simpson's index

Village agrobiodiversity index (VABDI): Sum of above 1 to 9 values

Similarly, we can estimate agrobiodiversity indices and scores at district, province/ state levels or any defined specific areas. OAU's can be further ranked based on these scores and indices. The followings are additional measures of agrobiodiversity.

- Agrobiodiversity index at HH, community, district, province, ward levels using the number of species or landraces divided by the total number of species or landraces in a country
- Analog site index of a particular landrace or species, calculated from climate analog tool based on reference site of a particular landrace or species
- Driver index can be estimated for each of different drivers (factors) in a particular area over the particular time frame, using the formula, lost landraces divided by the total number of landraces available before the effect of this driver.



## 7. Data Types and Collections

Different types of data are generated and collected for the measurement and other studies of agrobiodiversity. Different data types for agrobiodiversity study are given in figure 4. Data could be agro-morphological, molecular, and perception, which can be generally collected from on-station research, on-farm trial, surveys, and lab research. Several methods and techniques can be used to collect data and information (see Joshi *et al.*, 2005 for detail).

Apps and software are available for collecting data and information electronically both online as well offline. FieldLab is an application for Android tablets that are used for data collection in the field. It is developed by IRRI<sup>3</sup> and is available freely. Field Book is a simple app for taking phenotypic notes. It is an open-source application for field data collection on Android<sup>4</sup> and is available from Google Play<sup>5</sup>. The Fieldbook2020 software developed by CIMMYT<sup>6</sup> provides offline capabilities for managing pedigrees, phenotypic data, seed stocks, and field books for a breeding program. It provides integrated management of global information on genetic resources, crop improvement, and evaluation for individual crops. R Package<sup>7</sup> included in this software is useful for statistical analyses. Biologer<sup>8</sup> is a simple and free software designed for collecting data on biological diversity.

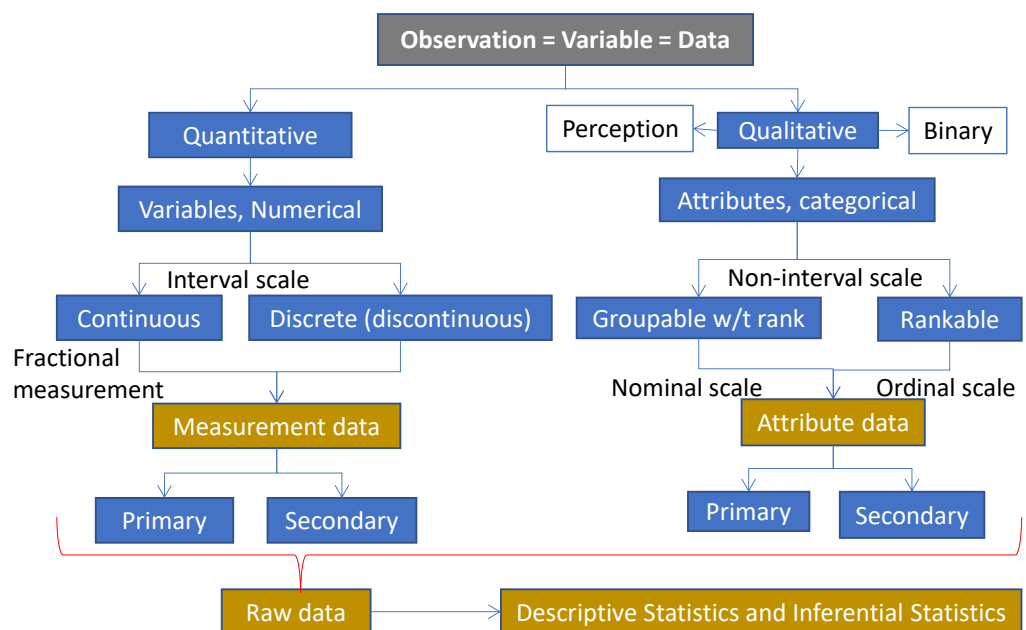


Figure 4: Data types for measuring on-farm agrobiodiversity at ecosystem, species, and cultivar levels

Perception data is generally collected from a survey. Along with the advancement of information technology, many data collections survey tools are available. These

<sup>3</sup> <http://bbi.irri.org/products/fieldlab>

<sup>4</sup> <http://dx.doi.org/10.2135/cropsci2013.08.0579>

<sup>5</sup> <https://play.google.com/store/apps/details?id=com.fieldbook.tracker&hl=en&gl=US>

<sup>6</sup> <https://www.cimmyt.org/>

<sup>7</sup> <https://data.cimmyt.org/dataset.xhtml?persistentId=hdl:11529/10548370>

<sup>8</sup> <https://biologer.org/>

online tools are very useful to minimize errors and speed up data processing. Some electronic media-based survey tools are given below.

- SurveyMonkey<sup>9</sup>: A cloud-based survey tool that helps users create, share, collect and analyze surveys.
- Google forms<sup>10</sup>: It is used to create online forms and surveys.
- SoGoSurvey<sup>11</sup>: A cloud-based platform that enables creation, distribution, and multilingual analysis of surveys, forms, polls, quizzes, and assessments.
- mWater Portal<sup>12</sup>: Free platform for data collection, data visualizations, and data-driven management of infrastructure in emerging economies.
- ODK<sup>13</sup>: It is an Open Data Kit, open-source software for collecting, managing, and using data in resource-constrained environments.

## 8. Measurement Objects

The information for measuring agrobiodiversity comes from different levels. These levels are alleles, genes, genotypes, cultivars (varieties and landraces), crops, species, components and groups, agroecosystems or agroecozones, parcels or plots, households (farmers), villages, communities, ethnicities, wards, municipalities, landscapes, regions, districts, provinces/ states, countries, and continents. These levels are measurement objects, called OAU (operational agricultural unit).

In addition, there are several crop groups that are OAU based on different criteria e.g., use-value base, economic importance base, national list base, habitat base, red list base, growing season base, national priority base, etc. Examples are cereals, vegetable fruits, released variety, registered variety, major, minor, primary, secondary, staple, commodity, high value, commercial, industrial, food crops, feed crops, manuring crops, pesticidal plants, cash crops, cover crops, trap crops, catch crop, cultivated, semi-domesticated, wild edible, field crops, garden crops, aquatic plants, common, rare, endangered, extinct, localized, vulnerable, winter crops, summer crops, and off-season (Joshi and Shrestha, 2019).

Object or OAU refers to the things being analyzed, interpreted, evaluated, or described. Variable or character refers to the properties used to describe the objects under study. Variables may be both qualitative and quantitative, and include agromorphological, genotypic, and perception data. These are measured or observed from an individual, representative samples, or population. In some cases, agromorphological markers, traits, and molecular markers can be treated as OAU.

## 9. Software for Agrobiodiversity Statistics

Many software are available for agrobiodiversity statistics. The general and molecular software are given below.

### I. General Statistical Software

- AGROBASE<sup>14</sup>: For data management, experiment management, and statistical analysis.

<sup>9</sup> <https://www.surveymonkey.com/>

<sup>10</sup> <https://www.google.com/forms/about/>

<sup>11</sup> <https://experience.sogosurvey.com/>

<sup>12</sup> <https://portal.mwater.co/#/>

<sup>13</sup> <https://opendatakit.org/>

<sup>14</sup> <https://www.agronomix.com/AGROBASE.aspx>

- CropStat<sup>15</sup>: For data management and basic statistical analysis of experimental data.
- DIPVEIW: For genebank data management and analysis.
- DIVA-GIS<sup>16</sup>: For mapping and geographic data analysis (a geographic information system (GIS)).
- Genstat<sup>17</sup>: For data analysis, particularly in the field of agriculture.
- GGEbiplot<sup>18</sup>: For biplot analysis, conventional statistical analysis, and decision making based on univariate and multivariate data.
- Instat<sup>19</sup>: A general statistical package.
- Minitab<sup>20</sup>: Simple and general statistical package.
- MS Excel<sup>21</sup>: Spreadsheet software program, a powerful data visualization, and analysis tool.
- MSTAT-C<sup>22</sup>: For the design, management, and analysis of agronomic research experiments.
- NTSYSpC<sup>23</sup>: Commonly used package for numerical taxonomy and multivariate analysis system.
- Past<sup>24</sup>: For scientific data analysis, with functions for data manipulation, plotting, univariate, multivariate statistics, ecological analysis, time series, and spatial analysis.
- R<sup>25</sup> and RStudio<sup>26</sup>: For statistical computing and graphics.
- SAS<sup>27</sup>: For data management, advanced analytics, and multivariate analysis.
- SPSS<sup>28</sup>: A software platform that offers advanced statistical analysis, a vast library of machine learning algorithms, and text analysis.
- STAR<sup>29</sup>: Statistical tool for agricultural research.
- Statistica<sup>30</sup>: A data analysis and visualization program.
- Statistix<sup>31</sup>: Statistical analysis program.
- PDA<sup>32</sup>: For biodiversity analysis and conservation prioritization problems.
- BioDiversity Pro<sup>33</sup>: A free statistical package program enabling many measures of diversity to be calculated for a dataset of taxa by samples.

## II. Molecular Data Analysis Software

- Arlequin<sup>34</sup>: Powerful genetic analysis packages performing a wide variety of tests, including hierarchical analysis of variance.

<sup>15</sup> <http://bbi.irri.org/products>

<sup>16</sup> <https://www.diva-gis.org/>

<sup>17</sup> <https://www.vsni.co.uk/software/genstat>

<sup>18</sup> <http://ggebiplot.com/>

<sup>19</sup> <https://instat.software.informer.com/3.3/>

<sup>20</sup> <https://www.minitab.com/en-us/>

<sup>21</sup> <https://www.microsoft.com/en-ww/microsoft-365/excel>

<sup>22</sup> [https://www.canr.msu.edu/afre/projects/microcomputer\\_statistical\\_package\\_mstat\\_1983\\_1985](https://www.canr.msu.edu/afre/projects/microcomputer_statistical_package_mstat_1983_1985)

<sup>23</sup> <http://www.appliedbiostat.com/ntsyspc/ntsyspc.html>

<sup>24</sup> <https://www.nhm.uio.no/english/research/infrastructure/past/index.html>

<sup>25</sup> <https://www.r-project.org/>

<sup>26</sup> <https://www.rstudio.com/>

<sup>27</sup> [https://www.sas.com/en\\_us/home.html](https://www.sas.com/en_us/home.html)

<sup>28</sup> <https://www.ibm.com/analytics/spss-statistics-software>

<sup>29</sup> <http://bbi.irri.org/products>

<sup>30</sup> <https://www.statistica.com/en/>

<sup>31</sup> <https://www.statistix.com/>

<sup>32</sup> <http://www.cibiv.at/software/pda/>

<sup>33</sup> <https://www.sams.ac.uk/science/outputs/>

<sup>34</sup> <http://cmpg.unibe.ch/software/arlequin35/>

- GDA<sup>35</sup>: For the analysis of discrete genetic data.
- GenAIEx<sup>36</sup>: Excel Add-In for the analysis of genetic data, particularly useful for dominant data such as RAPD and AFLP data.
- MEGA<sup>37</sup>: For reconstructing phylogenies using distance matrices and maximum parsimony methods, and includes neighbor-joining, branch-and-bound parsimony methods and bootstrapping.
- PHYLIP<sup>38</sup>: Extensive package of programs for inferring phylogenies.
- POPGENE<sup>39</sup>: For the analysis of genetic variation among and within populations using co-dominant and dominant markers, and quantitative data.
- PowerMarker<sup>40</sup>: A comprehensive set of statistical methods for genetic marker data analysis, designed especially for SSR/SNP data analysis.
- STRUCTURE<sup>41</sup>: Uses a clustering method to identify population structure and assigns individuals to those populations.

## 10. R Packages for Agrobiodiversity Measurement and Study

Most of the software and R packages used in biodiversity analysis can be used for agrobiodiversity analysis. Past is simple and free software that can be used for agrobiodiversity data. It is good for generating a graph, doing multivariate analysis, estimating different diversity indices, and analyzing time-series data. Some of the R packages useful for analysis of agrobiodiversity data are:

- adiv<sup>42</sup>: Analysis of Diversity, with functions, data sets, and examples for the calculation of various indices of biodiversity including species, functional and phylogenetic diversity.
- agricolae<sup>43</sup>: Statistical Procedures for Agricultural Research, offers extensive functionality on experimental design especially for agricultural and plant breeding experiments and other statistical analysis.
- analogues<sup>44</sup>: To calculate the climatic similarity between a reference site and a prescribed area, helps identifying locations with similar climates.
- BAT<sup>45</sup>: Biodiversity assessment tools, assess alpha and beta diversity in all their dimensions (taxonomic, phylogenetic and functional).
- BiodiversityR<sup>46</sup>: For statistical analysis of biodiversity and ecological communities.
- BioFTF<sup>47</sup>: To study biodiversity with the functional data analysis.
- BIO-R<sup>48</sup>: Biodiversity analysis using molecular data.
- GGEBiplotGUI<sup>49</sup>: A graphical user interface for the construction of, interaction with, and manipulation of GGE biplots.

<sup>35</sup> <https://phylogeny.uconn.edu/software/>

<sup>36</sup> <https://biology-assets.anu.edu.au/GenAIEx/Welcome.html>

<sup>37</sup> <https://www.megasoftware.net/>

<sup>38</sup> <https://evolution.genetics.washington.edu/phylip.html>

<sup>39</sup> <https://sites.ualberta.ca/~fyeh/popgene.html>

<sup>40</sup> <https://brwebportal.cos.ncsu.edu/powermarker/>

<sup>41</sup> <https://web.stanford.edu/group/pritchardlab/structure.html>

<sup>42</sup> <https://cran.r-project.org/web/packages/adir/index.html>

<sup>43</sup> <https://cran.r-project.org/web/packages/agricolae/index.html>

<sup>44</sup> <https://github.com/CIAT-DAPA/analogues>

<sup>45</sup> <https://biodiversityresearch.org/software/>

<sup>46</sup> <https://www.worldagroforestry.org/output/tree-diversity-analysis>

<sup>47</sup> <https://cran.r-project.org/web/packages/BioFTF/index.html>

<sup>48</sup> <https://data.cimmyt.org/dataset.xhtml?persistentId=hdl:11529/10820>

<sup>49</sup> <https://cran.r-project.org/web/packages/GGEBiplotGUI/index.html>

- `hclust`<sup>50</sup>: Hierarchical cluster analysis on a set of dissimilarities and methods for analyzing it.
- `prcomp`<sup>51</sup>: Performs a principal components analysis on the given data matrix and returns the results as an object of class `prcomp`.
- `psych`<sup>52</sup>: Procedures for psychological, psychometric, and personality research.
- `rich`<sup>53</sup>: For the analysis of species richness.
- `vegan`<sup>54</sup>: For community ecologists with multivariate and diversity analysis and other functions.

## 11. Data Preparation, Import and Analysis in R

A very common data frame in agrobiodiversity study is a data matrix that contains information about the properties, traits, characters, variables of several OAU (individuals, samples, specimens and population). For example, data is a household data matrix (household by several landraces within a species) and it is a count data set. The first column is household name or number, and it may be a community, site, household, species, agrobiodiversity component, agrobiodiversity group, or any other OAU. Other columns are the number of landraces under different crop species, and it may be species, cultivars, or any other variables. Data is generally prepared in MS Excel, and it is good to cross-check and verify the data before importing it into the R environment. The useful commands in Excel for data check are freezing or splitting panes, filter, sort, text to a column, data validation, exploratory data analysis, scatter plot, etc.

RStudio is more user-friendly, and the following analysis and process are based on RStudio. RStudio has four windows, script/editor window, data import/workspace window, console/ command window, and file/plot/package window. Among many R packages, `vegan` and `BiodiversityR` are more useful for estimating agrobiodiversity indices (Kindt and Coe, 2005), and, therefore, methods including R script are described below. To import data, the import dataset menu under environment is used. Here example data file is `hhdata`. The followings are the R scripts to import, view data, and converting imported data into a data frame.

```
library(readxl)#loading readxl package
hhdata <- read_excel("C:/Users/BK Joshi/Downloads/canada
training/ram/hhdata.xlsx")#importing data from given drive and saving this data
into hhdata
View(hhdata)#to see the data
hhdata<- as.data.frame (hhdata)#converting imported excel data into R data
frame
rownames(hhdata) <- hhdata[,1] #assigning row names from 1st column
hhdata[,1] <- NULL #removing the first column
hhdata #to display data contents
```

Followings are the R script for installation and estimating diversity indices using R package, `vegan`

```
#install vegan package from a menu, Package then install in RStudio
S=apply(hhdata>0,1,sum)# estimate species richness (S) without loading vegan
```

<sup>50</sup> <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/hclust>

<sup>51</sup> <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/prcomp>

<sup>52</sup> <https://cran.r-project.org/web/packages/psych/index.html>

<sup>53</sup> <https://cran.r-project.org/web/packages/rich/index.html>

<sup>54</sup> <https://cran.r-project.org/web/packages/vegan/index.html>

```

S # to display a richness
library(vegan) #loading vegan package
H=diversity(hhdata)#estimate Shannon diversity index
help(diversity)# look for description of function diversity
simp=diversity (hhdata, index="simpson") #estimate simpson index
J = diversity (hhdata, index ="simpson")/log(S) #estimate Pielou's evenness (J)
diversity(hhdata[-1], index="shannon")#exclude first column in case of data file
with first column as row name
barplot(simp) #plot simpson index
pairs(cbind(H, simp), pch="+", col="blue") #plot all
## Species richness (S) and Pielou's evenness (J):
S <- specnumber(hhdata) #estimate richness
cor(H,simp) #correlation coefficient between the Shannon and Simpson indices

```

A useful picture of diversity across several units is the function anosim() in the package, vegan. This analysis ranks all the dissimilarities among accessions and produces a boxplot of the ranks of dissimilarities within a given unit e.g., household. As an example, iris data set within this package is given below.

```

data(iris) #loading data in R memory
distiris<-dist(iris[,1:4]) #distance matrix computed by using the specified
distance measure to compute the distances between the rows of a data matrix
anoiris<-anosim(distiris,iris$Species) #analysis of similarities (anosim)
provides a way to test statistically whether there is a significant difference
between two or more groups of sampling units.
plot(anoiris) #produces a boxplot of the ranks of dissimilarities within a given
unit.

```

Another useful R package is BiodiversityR, which is a graphical user interface for statistical analysis of biodiversity and ecological communities, including species accumulation curves, diversity indices, Renyi profiles, GLMs for analysis of species abundance and presence-absence, distance matrices, Mantel tests, and cluster, constrained and unconstrained ordination analysis. It is menu-driven built within Rcmdr package. BiodiversityR analyzes two datasets simultaneously as does the vegan community ecology package. These data sets are the community datasets (rows correspond to sample units and columns correspond to species) and the environmental datasets.

It is suggested to install the package in R following the guidelines<sup>55</sup> as described in the installation guide. The manual<sup>56</sup> can also be accessed.

Followings are the commands and steps for analysis in BiodiversityR. An analysis can be carried out either through menu driven or using commands:

```

library (BiodiversityR) #load BiodiversityR package
library (Rcmdr) #load Rcmdr package
BiodiversityRGUI() #open graphical interface
help("BiodiversityRGUI", help_type="html") #to see details.

```

These are the steps for doing analyses with the menu options of BiodiversityR. To select the species and environmental matrices, follow these menu-driven steps:

```

BiodiversityR > Environmental Matrix > Select environmental matrix
Select the dune.env dataset as an example
Biodiversity > Community Matrix > Select community matrix
Select the dune dataset as an example.

```

<sup>55</sup> <https://www.worldagroforestry.org/sites/default/files/users/admin/Installation%20of%20BiodiversityR%202018.pdf>

<sup>56</sup> <http://apps.worldagroforestry.org/downloads/Publications/PDFS/b13695.pdf>



To calculate diversity indices for each site, follow these steps:

```
BiodiversityR > Analysis of diversity > Diversity indices ...
```

```
Diversity index: Shannon
```

```
Calculation method: separate per site.
```

To calculate diversity indices for each site using the command options of BiodiversityR, use the following scripts:

```
Diversity.1 <- diversityresult(dune, index="Shannon",method="each site")
```

```
Diversity.1
```

```
Diversity.2 <- diversityresult(dune, index="Simpson",method="each site")
```

### 11.1 Interpretation

Richness (S) is a number of species, landraces, particular traits in household, community, sites, or landrace. It quantifies types of the dataset. Shannon index (Shannon diversity index or Shannon Weaver index,  $H'$ ) includes both species number and evenness, where a greater number of species increase diversity, as does a more equitable distribution of individuals among species. High  $H'$  is representative of a diverse and equally distributed community.  $H'$  is strongly influenced by species richness and by rare species. Simpson index (D) is a measure of diversity, which takes into account both richness and evenness. The value of D ranges from 0 to 1, the greater the value the greater the diversity. The Simpson index gives more weight to evenness and common species. Evenness (Pielou's evenness, E) is a measure of the relative abundance of the different species making up the richness of an area. A community dominated by one or two species is considered to be less diverse than one in which several different species have a similar abundance. Its value ranges from 0 to 1 and 1 is complete equitability.

## 12. Conclusion

Native agrobiodiversity is generally neglected for conservation, quantification, evaluation, and monitoring. Different statistical tools can be used under agrobiodiversity statistics. Many software and R package are now available for agrobiodiversity study including measurement. Six types and levels of agrobiodiversity need to quantify and study for better management of agrobiodiversity. An operational agricultural unit is like a factor in which variables are generated and analyzed. Multivariate analysis and diversity indices are the major statistical components used in agrobiodiversity measurement. Estimates help generate the agrobiodiversity indicators that ultimately drive the program plans and activities. Many different types of scores and indices can be measured for household, community, any other administrative unit, and other OAU. Among the many software and R packages, vegan and BiodiversityR are very useful packages for estimating diversity indices and multivariate analysis along with many statistical features. Such estimates should be measured over a certain geo-region and period to monitor the status, plan the program, and rank the geo-regions.

## 13. Acknowledgments

The Grassroots Institute organized a Summer Field School on Mountain Ecosystem and Resource Management in September 2021. Based on the presentation in this Summer School, this review article was written. A special thank goes to Dr Hasrat Arjjumend for his initiation and Dr. Lila Khatiwada for valuable suggestions and grammar correction.

## 14. References

- Bajracharya, J., Rana, R., Joshi, B.K., Subedi, A. and Sthapit, B. (2012). Measuring on-farm crop diversity. In: B.R. Sthapit, P. Shrestha and M.P. Upadhyay (eds.), *On-farm management of agrobiodiversity in Nepal: Good practices*. Kathmandu: NARC, LIBIRD and Bioversity International, pp. 9-12.
- Bioversity International (2017). *Agrobiodiversity Index. Program Oversight*. A paper presented in 49<sup>th</sup> Board Meeting of Bioversity International, Rome, Italy, 10 – 12 May 2017.
- Borcard, D., Gillet, F. and Legendre, P. (2011). *Numerical ecology with R*. LLC: Springer Science+Business Media.
- Grum, M. and Atieno, F. (2007). *Statistical analysis for plant genetic resources: clustering and indices in R made simple*. Handbooks for Genebanks, No. 9. Rome: Bioversity International.
- Gurung, R., Sthapit, S.R., Gauchan, D., Joshi, B.K. and Sthapit, B.R. (2016). *Baseline Survey Report: II. Ghanpokhara, Lamjung*. Integrating Traditional Crop Genetic Diversity into Technology: Using a Biodiversity Portfolio Approach to Buffer against Unpredictable Environmental Change in the Nepal Himalayas. Pokhara: LI-BIRD, NARC and Bioversity International. Available online at <https://cgspace.cgiar.org/handle/10568/81039> [Accessed on 11 August 2021]
- Jarvis, D.I., Myer, L., Klemick, H., Guarino, L., Smale, M., Brown, A.H.D., Sadiki, M., Sthapit, B. and Hodgkin, T. (2000). *A Training Guide for In Situ Conservation On-farm*. Version 1. Rome: International Plant Genetic Resources Institute.
- Joshi, B.K. and Baniya, B.K. (2006). A diversity in qualitative traits of Nepalese cultivated buckwheat species. *Fagopyrum*, 23:23-27.
- Joshi, B.K. and Shrestha, B.B. (2017). Notes on plant and crop classification. In: B.K. Joshi, H.B. KC and A.K. Acharya (eds.), *Conservation and Utilization of Agricultural Plant Genetic Resources in Nepal*. Proceedings of 2<sup>nd</sup> National Workshop, 22-23 May 2017, Dhulikhel: NAGRC, FDD, DoA and MoAD, pp. 17-20.
- Joshi, B.K. and Shrestha, R. (eds.) (2019). Working Groups of Agricultural Plant Genetic Resources (APGRs) in Nepal. *Proceedings of National Workshop*, 21-22 June 2018, Kathmandu: NAGRC, NARC.
- Joshi, B.K., Ghimire, K.H., Gurung, R., Pudasaini, N., Pant, S., Paneru, P., Gauchan, D., Mishra, K.K. and Jarvis, D. (2020b). On-farm Agrobiodiversity Measurement and Conservation. In: B.K. Joshi, D. Gauchan, B. Bhandari and D. Jarvis (eds.), *Good Practices for Agrobiodiversity Management*. Kathmandu: NAGRC, LI-BIRD and Alliance of Bioversity International and CIAT, pp. 15-24.
- Joshi, B.K., Gorkhali, N.A., Pradhan, N., Ghimire, K.H., Gotame, T.P., KC, P., Mainali, R.P., Karkee, A. and Paneru, R.B. (2020c). Agrobiodiversity and its Conservation in Nepal. *Journal of Nepal Agricultural Research Council* 6: 14-33. DOI: <https://doi.org/10.3126/jnarc.v6i0.28111>
- Joshi, B.K., Gurung, S.B., Mahat, P.M., Bhandari, B. and Gauchan, D. (2018). Intra-Varietal Diversity in Landrace and Modern Variety of Rice and Buckwheat. *The Journal of Agriculture and Development*, 19: 1-8. Available online at <https://cgspace.cgiar.org/handle/10568/97576> [Accessed on 30 August 2021]
- Joshi, B.K., Shrestha, P., Upadhyay, M.P., Chaudhary, B., Mudwari, A., Baniya, B.K., and KC, H.B., (2007). On-farm variation and household diversity of pigeon pea landraces in Kachorwa, Nepal. *Nepal Agric. Res. J.*, 8:28-34. Available online at <https://www.nepjol.info/index.php/NARJ/article/view/11567> [Accessed on 30 August 2021]

- Joshi, B.K., Upadhyay, M.P., Bimb, H.P., Gauchan, D. and Baniya, B.K. (2005). Data analysis methods adopted under in-situ global project in Nepal. *Nepal Agric. Res. J.*, 6: 99-109. Available online at <https://www.nepjol.info/index.php/NARJ/article/view/3371> [Accessed on 30 August 2021]
- Kindt, R. and Coe, R. (2005). *Tree diversity analysis. A manual and software for common statistical methods for ecological and biodiversity studies*. Nairobi: World Agroforestry Centre (ICRAF).
- PAR (Platform for Agrobiodiversity Research) (2018). *Assessing Agrobiodiversity: A Compendium of Methods*. Platform for Agrobiodiversity Research, Rome.
- Pudasaini, N., Sthapit, S.R., Gauchan, D., Bhandari, D., Joshi, B.K. and Sthapit, B.R. (2016). *Baseline Survey Report: I. Jungu, Dolakha*. Integrating Traditional Crop Genetic Diversity into Technology: Using a Biodiversity Portfolio Approach to Buffer against Unpredictable Environmental Change in the Nepal Himalayas. Pokhara: LI-BIRD, NARC and Bioversity International.
- Sthapit, B.R., Gauchan, D., Joshi, B.K. and Chaudhary, P. (2017). Agrobiodiversity index to measure agricultural biodiversity for effectively managing it. In: B.K. Joshi, H.B. KC and A.K. Acharya (eds.), *Conservation and Utilization of Agriculture Plant Genetic Resources in Nepal*. Proceed. 2<sup>nd</sup> National Workshop, May 22-23, 2017, Dhulikhel: NAGRC/FDD/MoAD.

## Author' Declarations and Essential Ethical Compliances

### *Author' Contributions (in accordance with ICMJE criteria for authorship)*

This article is 100% contributed by the sole author. She conceived and designed the research or analysis, collected the data, contributed to data analysis & interpretation, wrote the article, performed critical revision of the article/paper, edited the article, and supervised and administered the field work.

### *Funding*

No funding was available for the research conducted for and writing of this paper.

### *Research involving human bodies (Helsinki Declaration)*

Has this research used human subjects for experimentation? No

### *Research involving animals (ARRIVE Checklist)*

Has this research involved animal subjects for experimentation? No

### *Research involving Plants*

During the research, the author followed the principles of the Convention on Biological Diversity and the Convention on the Trade in Endangered Species of Wild Fauna and Flora. Yes

### *Research on Indigenous Peoples and/or Traditional Knowledge*

Has this research involved Indigenous Peoples as participants or respondents? No

### *(Optional) PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)*

Has author complied with PRISMA standards? Yes

### *Competing Interests/Conflict of Interest*

Author has no competing financial, professional, or personal interests from other parties or in publishing this manuscript.

## Rights and Permissions

**Open Access.** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third-party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.