# User Manual

Distinct Genetic Diversity App

An app to facilitate the calculation of distinct genetic diversity to identify Key Biodiversity Areas.

# Contents

# Before using the program/what the program offers

Identifying Key Biodiversity Areas (KBAs) is a process that involves many steps. After you decided for your study organism and collected enough samples, created a microsatellite or SNP data set that includes useful information about locations you can proceed using this program.

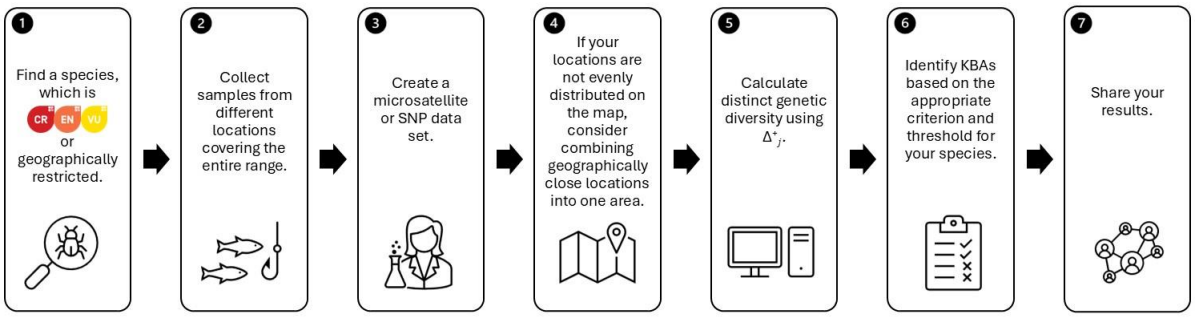



*Fig 1: 7 steps to identify KBAs using distinct genetic diversity.*

This app will help you to calculate distinct genetic diversity with Delta+j and identify KBAs based on the criterion you pick.

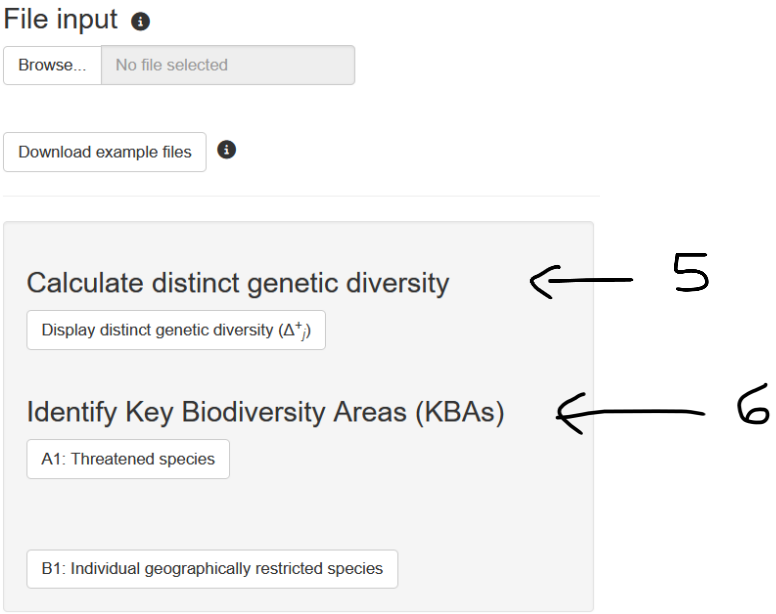



*Fig 2: What the program offers*

# Components of the app

The app is based on the programming language R (R Core Team, 2021) and several packages. adegenet was used to facilitate processing of genetic data sets (Jombart, 2008; Jombart & Ahmed, 2011). vegan was used to calculate distinct genetic diversity based on average taxonomic distinctiveness (Delta+ **($\Delta^+_j$)**) explain (Oksanen et al., 2024) …. shiny is a substantial to integrate the calculations into an app (Chang et al., 2024). shinyBS was used to display information in the app (Bailey, 2022).

The app is built up from two scripts: one main script that loads the required packages and function and combines server with ui into an app and an additional script (names!!) containing a function to calculate distinct genetic diversity. Furthermore, the app includes five example data sets, in different file formats, pictures and this manual. Pictures are saved in a folder called WWW. All components integrated in this app can be downloaded from zenodo and github.links

# Import data

To use this application, you must have SNP or microsatellite data in the appropriate file format. The app accepts the following formats: .csv, GENETIX (.gtx), GENPOP (.gen), FSTAT (.dat), or STRUCTURE (.str or .stru). The file upload section features a widget that allows you to browse for your files, along with a button for downloading files in different formats compatible with the app (Fig. 3). Please note that the files offered for download are not suitable for determining KBAs. They are intended to test various files and improve your understanding of how you should format your own data. Hovering over the information icon (i) will provide you with details on which files can be uploaded or information about the example data sets.

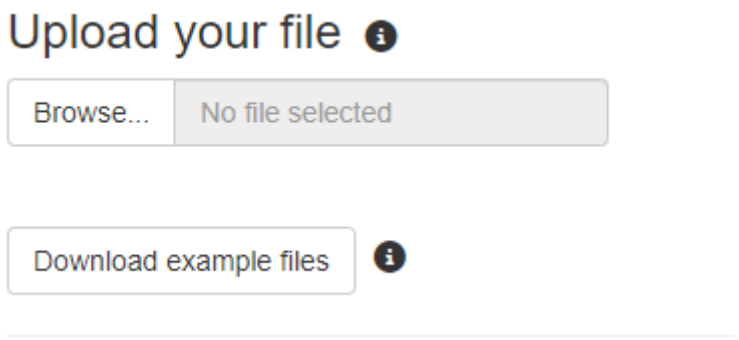



*Fig 3: Section to upload files*

By clicking the “Download example files” button, a menu will appear, allowing you to download various files in different formats (Fig. 4). Downloading the example file in .csv format is particularly beneficial, in case your data is not formatted in any style being compatible with the programs GENETIX, GENPOP, FSTAT, or STRUCTURE. If you move the mouse pointer over the information icon, the reference of the data set that can be

downloaded is displayed. In addition to the option to download the .str file, you will find more details about how the file is built up.

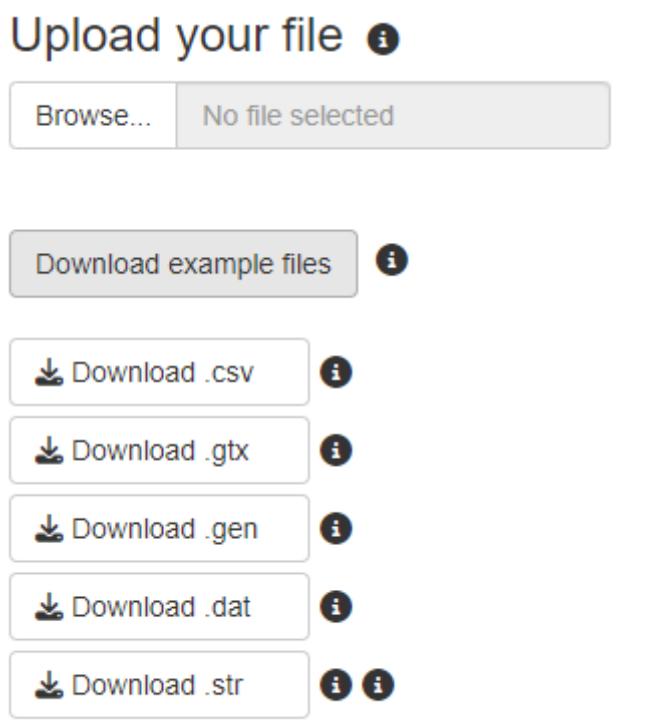



*Fig 4: Download example files.*

Upload your file by clicking on "Browse...". If your file upload was successful, the app will display the name of the file you selected with a message underneath that your upload has been completed (Fig. 5). After you have uploaded a file, you can upload a new file by clicking on "Browse..." again. The app will always process the currently uploaded file.

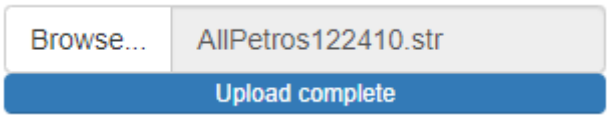



*Fig 5: Successfull file upload*

### .CSV

If your input file is a .csv file, the first row contains column names. The app will accept the file if your column names differ, but the first column should contain names of your individuals, the second column the name of your site and all following columns should contain information about your alleles. Missing values should be always encoded as NA. If your organism is diploid or polyploid, more columns depending on your ploidy will be added per marker.

```
,          "site_names",    "fca8","fca8_2",    "fca23","fca23_2",    "fca43","fca43_2"
N215,    "P01",             NA     ,NA,         "136",    "146",       "139",   "139"
N216,    "P01",             NA     ,NA,         "146",    "146",       "139",   "145"
```

Example of a .csv file, spaces added for readability

## .gtx

GENETIX is set of programs computes several basic parameters of population genetics such as Nei's D and H, Wright's F-statistics (the Weir-Cockerham's and Robertson-Hill's estimators), and linkage disequilibrium D according to Black & Krafsur. Information about the file structure of a .gtx file can be obtained from the webpage https://kimura.univ-montp2.fr/genetix/.

## .gen

GENEPOP is a population genetics software, distributed both as stand-alone software

and as an R package. It can be used to compute various tests and estimates and perform analyses of isolation by distance. Information about the structure of .gen files can be found here: https://genepop.curtin.edu.au/help_input.html or in the manual: https://kimura.univ-montp2.fr/~rousset/Genepop4.7.pdf.

## .dat

FSTAT was a computer package for PCs which estimates and tests gene diversities and differentiation statistics from codominant genetic markers that is now superseeded with the R package Hierfstat. To .dat format is no longer required (Meeûs & Goudet, 2007). If an old dataset is to be analyzed, it is still possible to upload .dat files. If distinct genetic diversity is not calculated, you can download the .dat file for comparison with your file or try a different file format.

## .str or .stru

STRUCTURE is a software to investigate population structure. Information on the .str or .stru file format can be found in its manual: https://www.ccg.unam.mx/~vinuesa/tlem09/docs/structure_doc.pdf. Files ending with .structure are not accepted and will cause an error message. If you upload a .str or a .stru file the app will ask you to enter more information on the file you uploaded (Fig. 6): How many genotypes are there? How many markers are there? Which column contains labels for genotypes (‘0’ if absent)? Which column contains the area names? After you answered the questions distinct genetic diversity can be calculated and KBAs can be identified. Genotypes could be also referred to as individuals. To enter your answer, click on the field below the question. As a result, a blue frame will appear around the field, and you can enter your number (Fig. 6). You can use the small arrows on the right-hand side of the field to increase or decrease your number (Fig. 6). If you are insecure on how to answer these questions it might help to have a closer look at the example file. The first

column of the example file contains labels for genotypes. The second column of the example file contains information that will be treated as area names suitable for KBA identification. The following columns contain information on alleles on different loci. In the first row you will find column headers that include the marker names. Each genetic marker can be always assigned to one column. If you examine a diploid or polyploid organism you will have several rows for each genotype reflecting the ploidy. The required answers for the example data set can be found when you hover over the second (i) icon (Fig. 4).

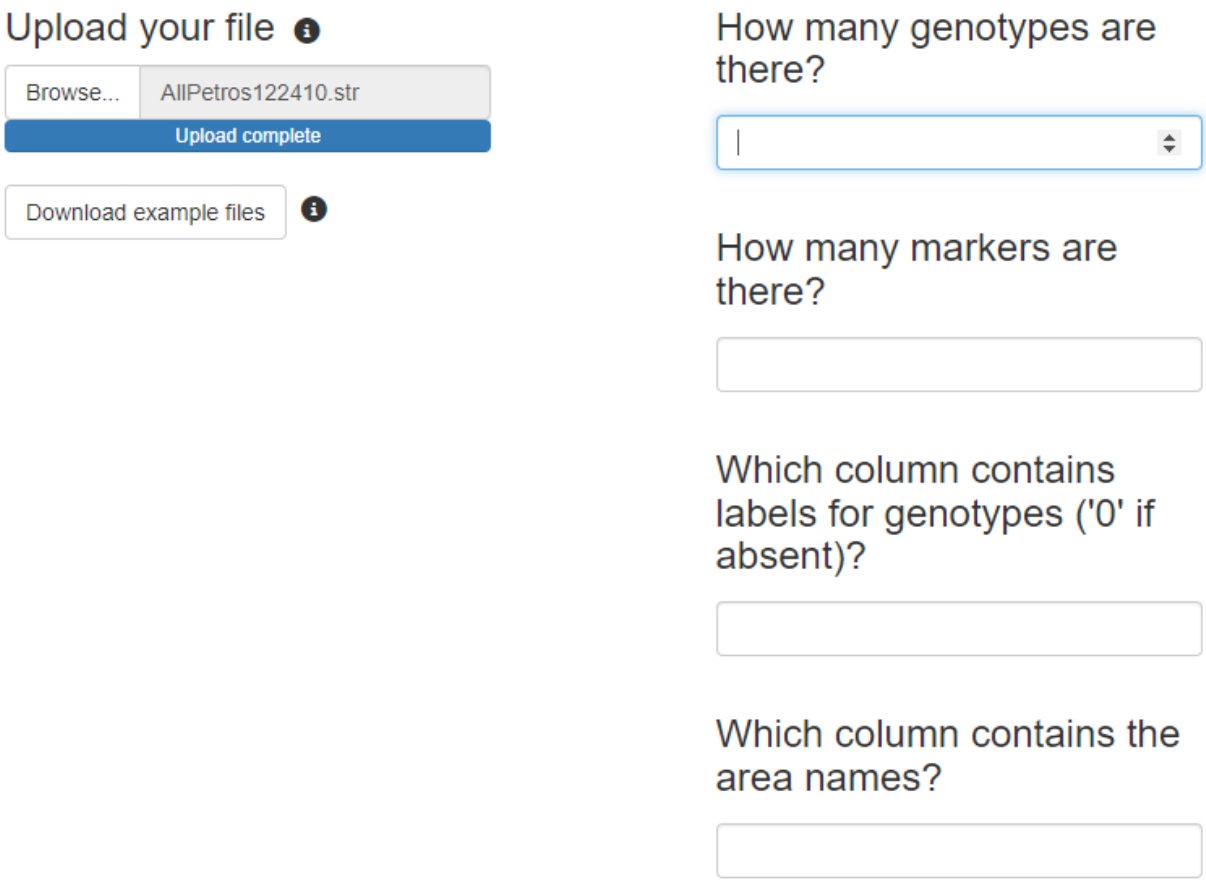



*Fig 6: Enter additional information to process .str or .stru files*

## References for example files

All data sets except the .csv file can be also directly downloaded from DRYAD (Tab. 1). The .csv file is a modified version of the nancycats data set included in the R package adegenet.

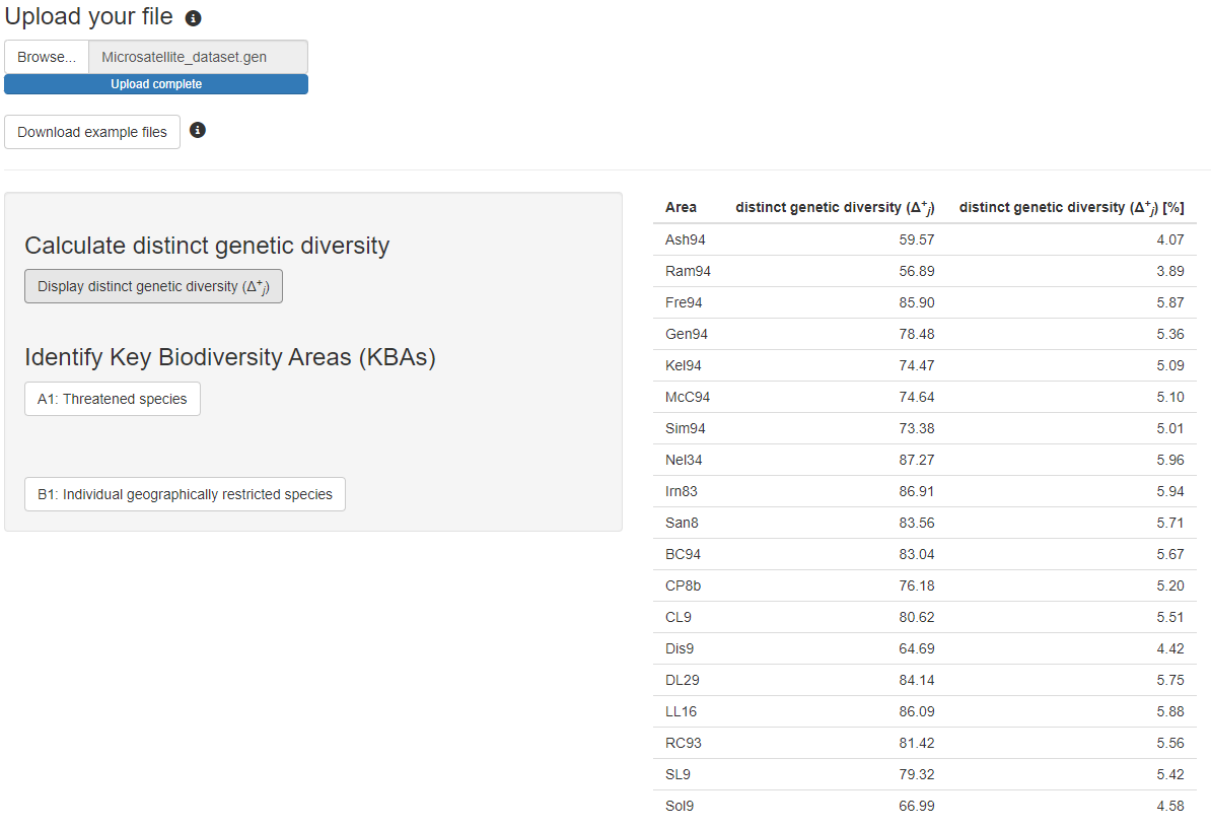
| File ending | Reference |
|---|---|
| .gtx | Salmona, Jordi; Heller, Rasmus; Quéméré, Erwan; Chikhi, Lounès (2017): Data from: Climate change and human colonization triggered |

| | |
|---|---|
| | habitat loss and fragmentation in Madagascar. https://doi.org/10.5061/dryad.8f45n |
| .gen | Millette, Katie L.; Gonzalez, Andrew; Cristescu, Melania E. (2019): Breaking ecological barriers: anthropogenic disturbance leads to habitat transitions, hybridization, and high genetic diversity. https://doi.org/10.5061/dryad.50557nm |
| .dat | DeFaveri, Jacquelin; Shikano, Takahito; Shimada, Yukinori; Merilä, Juha (2013): Data from: High degree of genetic differentiation in marine three-spined sticklebacks (Gasterosteus aculeatus). https://doi.org/10.5061/dryad.493jh |
| .str | Wagner, Catherine E.; McCune, Amy R.; Lovette, Irby J. (2012): Data from: Recent speciation between sympatric Tanganyikan cichlid color morphs. https://doi.org/10.5061/dryad.t6s441n0 |

*Tab. 1 References for example files*

# Distinct genetic diversity for each area

If you successfully uploaded your file distinct genetic diversity can be displayed and downloaded (Fig.…).




| Area | distinct genetic diversity ($\Delta^+_j$) | distinct genetic diversity ($\Delta^+_j$) [%] |
|---|---|---|
| Ash94 | 59.57 | 4.07 |
| Ram94 | 56.89 | 3.89 |
| Fre94 | 85.90 | 5.87 |
| Gen94 | 78.48 | 5.36 |
| Kel94 | 74.47 | 5.09 |
| McC94 | 74.64 | 5.10 |
| Sim94 | 73.38 | 5.01 |
| Nel34 | 87.27 | 5.96 |
| Irn83 | 86.91 | 5.94 |
| San8 | 83.56 | 5.71 |
| BC94 | 83.04 | 5.67 |
| CP8b | 76.18 | 5.20 |
| CL9 | 80.62 | 5.51 |
| Dis9 | 64.69 | 4.42 |
| DL29 | 84.14 | 5.75 |
| LL16 | 86.09 | 5.88 |
| RC93 | 81.42 | 5.56 |
| SL9 | 79.32 | 5.42 |
| Sol9 | 66.99 | 4.58 |

To display the distinct genetic diversity for each area, click the button “Display distinct genetic diversity ($\Delta^+_j$)”. As a result, a table will be displayed on the right-hand side of the

side panel. The table has three columns: Area, distinct genetic diversity ($\Delta^{+}_{j}$), and distinct genetic diversity ($\Delta^{+}_{j}$) [%]. Area contains the area labels that you

| Area | distinct genetic diversity ($\Delta^{+}_{j}$) | distinct genetic diversity ($\Delta^{+}_{j}$) [%] |
|---|---|---|
| Ash94 | 59.57 | 4.07 |
| Ram94 | 56.89 | 3.89 |
| Fre94 | 85.90 | 5.87 |
| Gen94 | 78.48 | 5.36 |
| Kel94 | 74.47 | 5.09 |
| McC94 | 74.64 | 5.10 |
| Sim94 | 73.38 | 5.01 |
| Nel34 | 87.27 | 5.96 |
| Irn83 | 86.91 | 5.94 |
| San8 | 83.56 | 5.71 |
| BC94 | 83.04 | 5.67 |
| CP8b | 76.18 | 5.20 |
| CL9 | 80.62 | 5.51 |
| Dis9 | 64.69 | 4.42 |
| DL29 | 84.14 | 5.75 |
| LL16 | 86.09 | 5.88 |
| RC93 | 81.42 | 5.56 |
| SL9 | 79.32 | 5.42 |
| Sol9 | 66.99 | 4.58 |

## Identify Key Biodiversity Areas (KBAs)

To identify areas suitable for KBAs, it is necessary to know which KBA criteria can be applied to your species. Choose between the two options criterion A1 (your species is threatened) and criterion B1 (your species is geographically restriced) (Fig. 7). Depending on the criterion you pick different thesholds will be applied, that can

influence the number of areas that will be selected as KBAs (Tab. 2). If you click the button for criterion B1 your results will immediately be displayed.

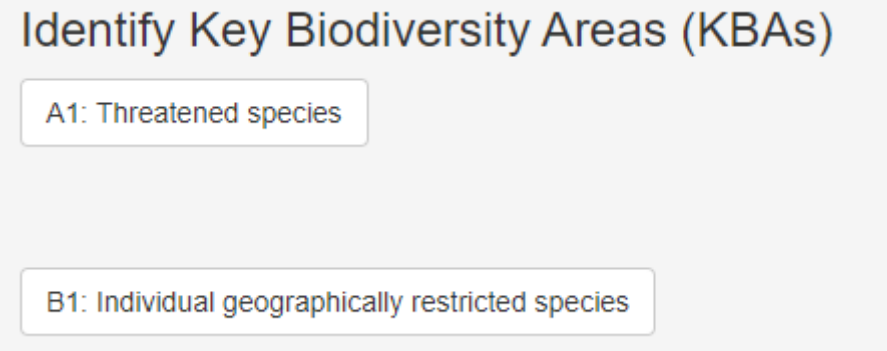



*Fig 7: Choose between criterion A1 and criterion B1.*

If you click the button for criterion A1, you will have to pick your species Red List status (Fig. 8). Your identified KBAs will be displayed after you choose between the criterion suitable for your species.

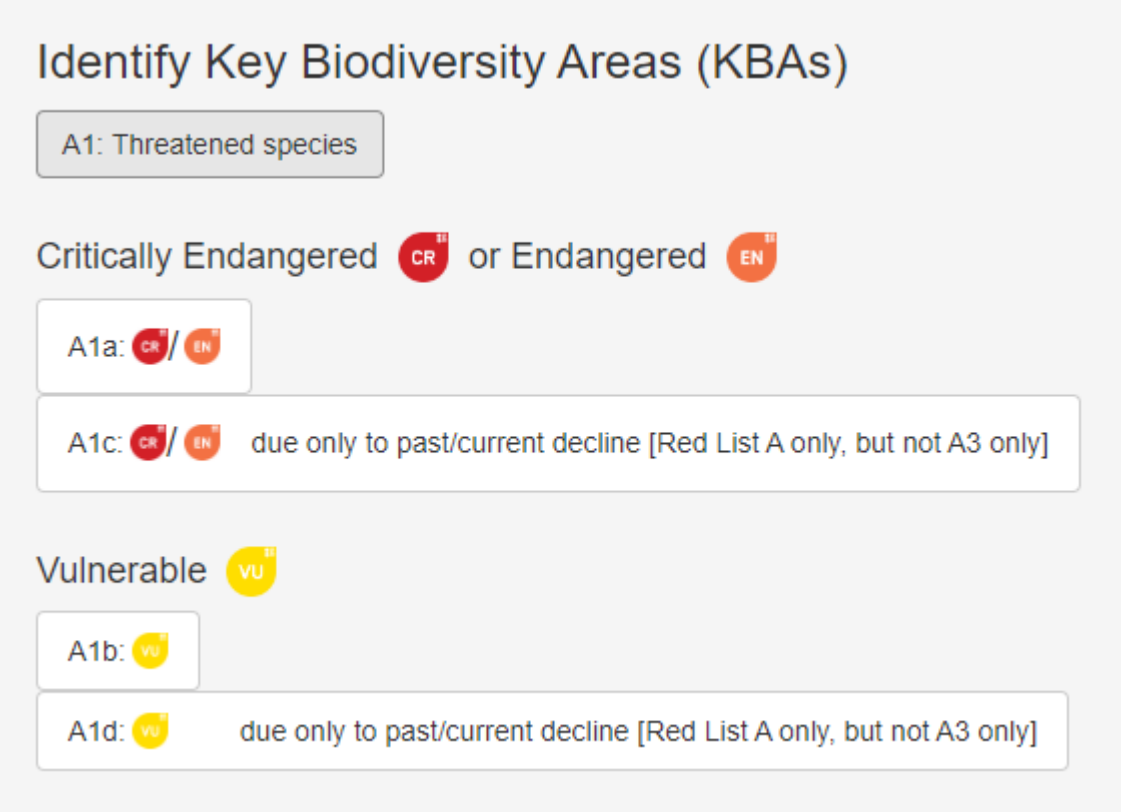



*Fig 8: Pick your species Red List status.*

The KBA criteria do not differentiate between species which are Critically Endangered (CR) or Endangered (EN). But if your species is Vulnerable (VU) a different threshold will be applied, that might cause less areas to be selected. If the Red List category was given due only to past/current decline [Red List A only, but not A3 only] another threshold allows more areas to be selected (Tab. 2) (IUCN, 2020).

| criterion | threshold |
|---|---|

| | |
|---|---|
| A1a | ≥0.5% |
| A1c | ≥0.1% |
| A1b | ≥1% |
| A1d | ≥0.2% |
| B1 | ≥10% |

*Tab. 2: Summary of IUCN criteria and thresholds relevant to distinct genetic diversity*

Your KBAs identified will be displayed as a sentence on the right-hand-site of the panel which allowed you to choose the criterion fitting your species. The sentences will either be "None of the areas qualifies as KBA." or start with "Following areas qualify as KBAs: " and ending with a list of the areas identified.

# How to continue with your results

Combine the knowledge about KBAs identified with distinct genetic diversity with other information you can gathered about the (i) number of mature individuals, (ii) area of occupancy, (iii) extent of suitable habitat, (iv) range, or (v) number of localities of your species. Involve stakeholders to define the geographic boundaries of your KBA. Inform yourself about documentation standards for proposing/ updating KBAs, read the guide, and complete the KBA proposal form (https://www.keybiodiversityareas.org/). Regional focal points can help you with questions. If you successfully proposed an area a monitoring program needs to be developed and then the process is initiated to present the KBA in the World Database of Key Biodiversity Areas. Good luck with your project!

# Bibliography


Bailey, E. (2022): shinyBS. Twitter Bootstrap Components for Shiny. Version 0.61.1. Available online at https://CRAN.R-project.org/package=shinyBS.

Chang, W.; Cheng, J.; Allaire, J.; Sievert, C.; Schloerke, B.; Xie, Y. et al. (2024): shiny. Web Application Framework for R. Version 1.9.1. Available online at https://CRAN.R-project.org/package=shiny.

IUCN (2020): Guidelines for using A global standard for the identification of Key Biodiversity Areas: version 1.1: IUCN, International Union for Conservation of Nature.

Jombart, T. (2008): adegenet. a R package for the multivariate analysis of genetic markers, pp. 1403–1405. Available online at 10.1093/bioinformatics/btn129.

Jombart, T.; Ahmed, I. (2011): adegenet. new tools for the analysis of genome-wide SNP data. Version 1.3-1: Bioinformatics. Available online at 10.1093/bioinformatics/btr521.

Meeûs, T. de; Goudet, J. (2007): A step-by-step tutorial to use HierFstat to analyse populations hierarchically structured at multiple levels. In *Infection, Genetics and Evolution* 7 (6), pp. 731–735. DOI: 10.1016/j.meegid.2007.07.005.

Oksanen, J.; Simpson, G.; Blanchet, F.; Kindt, R.; Legendre, P.; Minchin, P. et al. (2024): vegan. Community Ecology Package. Version 2.6-8. Available online at https://CRAN.R-project.org/package=vegan.

R Core Team (2021): R. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available online at https://www.R-project.org/.