User Manual

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

6.11.2024

Sarah Gronefeld, Axel Hochkirch

Contents

Before using the App

Determining Key Biodiversity Areas (KBAs) is a process that involves many steps (Fig. 1). After deciding to study a threatened or geographically restricted organism, you need sufficient data covering the entire range of the species. For using genetic distinctiveness as a parameter, this means that you need to sample the majority of populations of a species. Once a genetic dataset has been created and the information about the sampling locations is structured in a meaningful way, the basis to calculate distinct genetic diversity and identify sites that qualify for KBAs is laid. This app has been programmed to calculate distinct genetic diversity and identify suitable sites for KBAs.

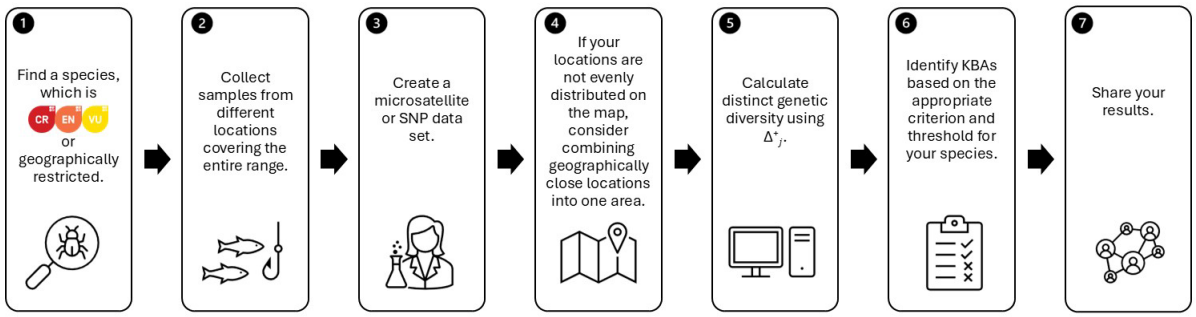

Fig. 1: Seven steps to identify KBAs using distinct genetic diversity.

Alternatives to Distinct Genetic Diversity App

This app calculates distinct genetic diversity using Δ^+_j, but other methods calculating distinct genetic diversity may also provide useful results, which are also accepted for KBA identification (IUCN 2020). If you have multiple files in the same format and you want to automatically process them, you may need to write your own script. You can use the following instructions which might make it easier and faster for you to create your own script: https://rpubs.com/Who_could_this_be_2/1217013

What the App does for you

The app was created to calculate distinct genetic diversity and find suitable sites for KBAs with data sets provided by the user (Fig. 2). You can upload SNP and microsatellite data sets to the app. If you are unsure whether your file is formatted correctly, you can download the example data sets and find more information in this manual. The app will calculate distinct genetic diversity for all sites of your data set and will give you the option to view the results of these calculations and download them as .csv file.

Furthermore, the app will help you to identify sites that qualify as KBA according to IUCN standard (IUCN, 2020).

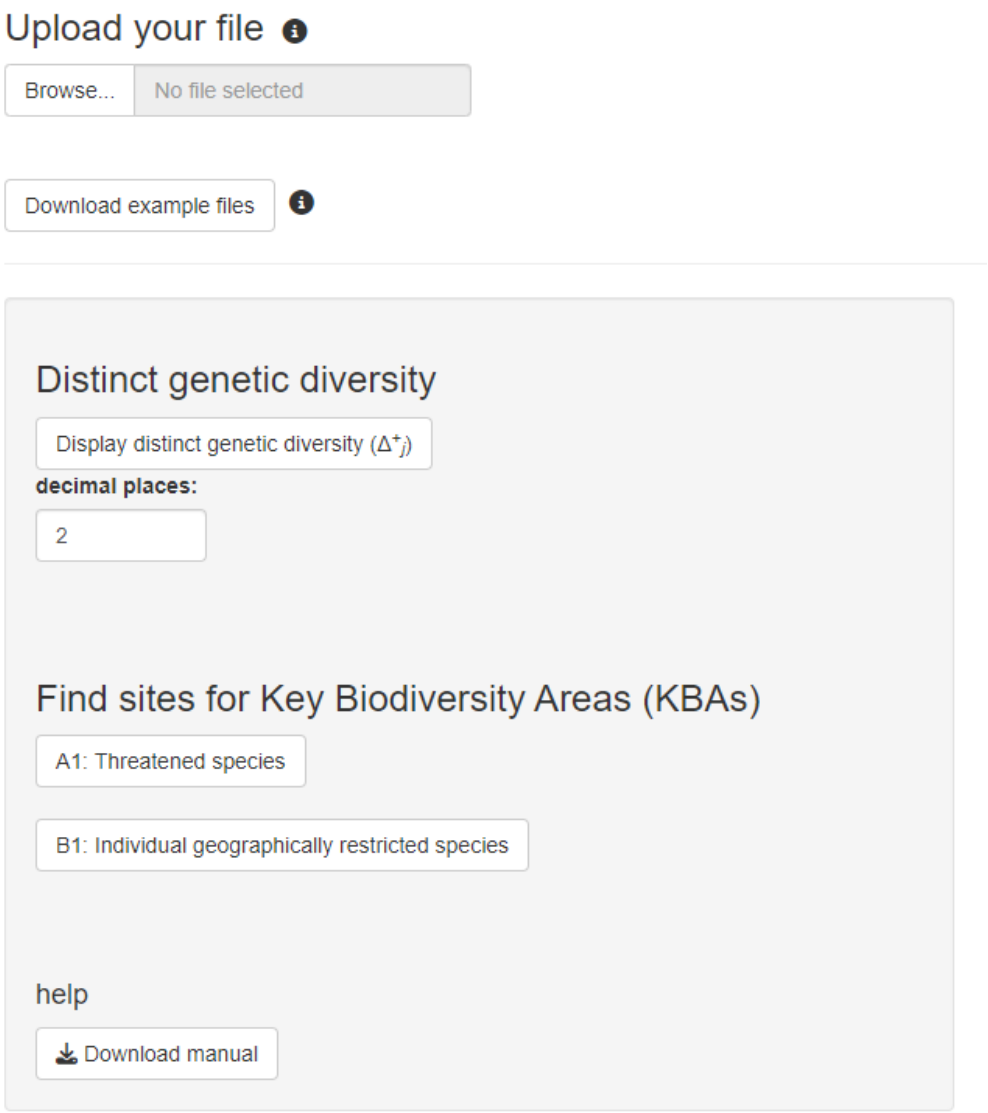

Fig. 2: All features off the app.

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). Hovering over the information icon (i) will provide you with more information.

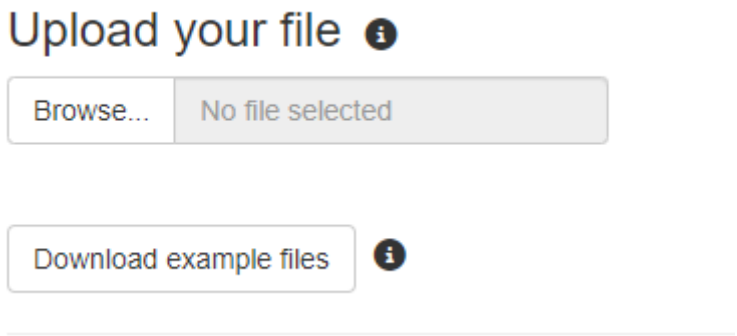

Fig. 3: Section to upload files.

By clicking the "Download example files" button, a menu will appear, allowing you to download various files in various formats (Fig. 4). Please note that the files offered for downloading 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. Downloading the example file in .csv format can be used in case your data is not formatted in any style being compatible with the programs GENETIX, GENPOP, FSTAT, or STRUCTURE.

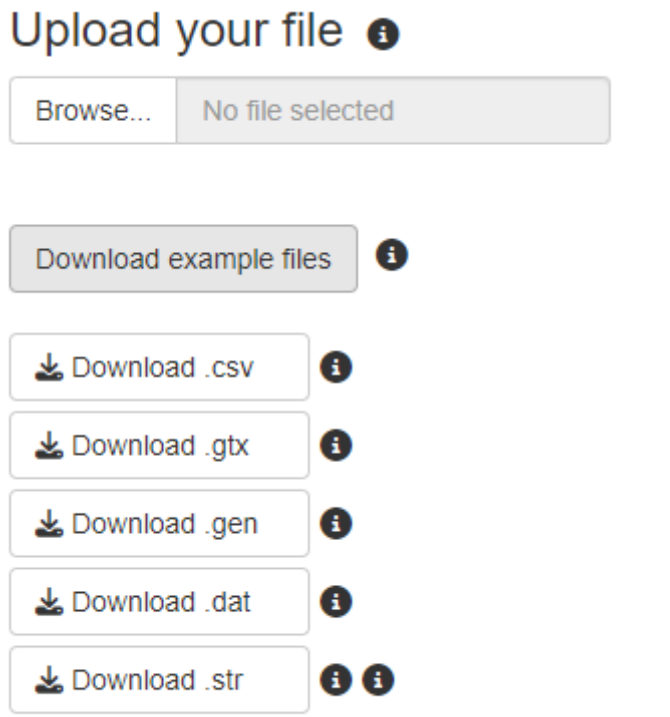

Fig. 4: Download example files.

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 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 latest uploaded file.

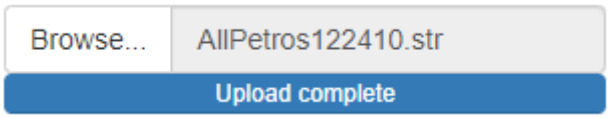

Fig. 5: Successfull file upload.

.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 the individuals (names can be numbers only), the second column the names of the sites and all following columns should contain genetic information (i.e. alleles). Missing values should be encoded as NA. If your organism is diploid or polyploid, more columns depending on your ploidy can be added per marker (Tab. 1).

Tab. 1: An example of a .csv input file, spaces added for readability.

,	"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"

.gtx

Information about the file structure of a GENETIX file can be obtained from the webpage https://kimura.univ-montp2.fr/genetix/.

.gen

Information about the structure of GENEPOP 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

.dat is an old data format of FSTAT, which is no longer required (Meeûs & Goudet, 2007). They can be also used by the App and an example file can be downloaded.

.str or .stru

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 (i.e. individuals) are there? How many markers are there? Which column contains labels for genotypes (‘0’ if absent)? Which column contains the site names? 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). Apart from

the keyboard, 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 unsure of 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 contains site names. The following columns contain genetic information. In the first row you will find column headers that include the marker names. There are two rows in the example file for one individual, reflecting the two alleles of the diploid organism. The required answers for the example data set can be found when you hover over the second (i) icon (Fig. 4).

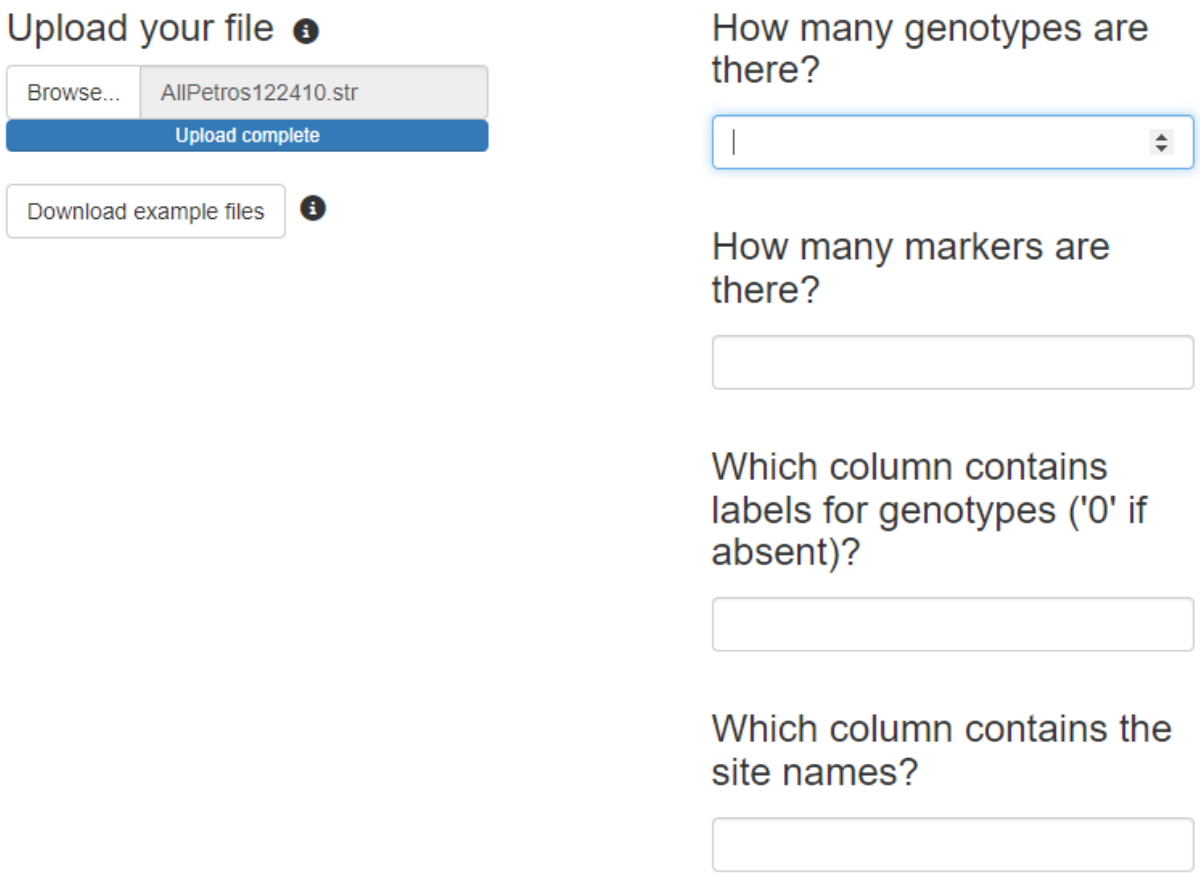

Fig. 6: Enter additional information to process .str or .stru files

References for example files

All data sets except the .csv file can also be directly downloaded from DRYAD (Tab. 2). The .csv file is a modified version of the nancycats data set included in the R package adegenet.

Tab. 2: References for example files.

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

Distinct genetic diversity

If you successfully uploaded your file distinct genetic diversity can be displayed and downloaded. Click the button "Display distinct genetic diversity (Δ^+_j)". As a result, a table will be displayed on the right-hand side of the side panel (Fig. 7).

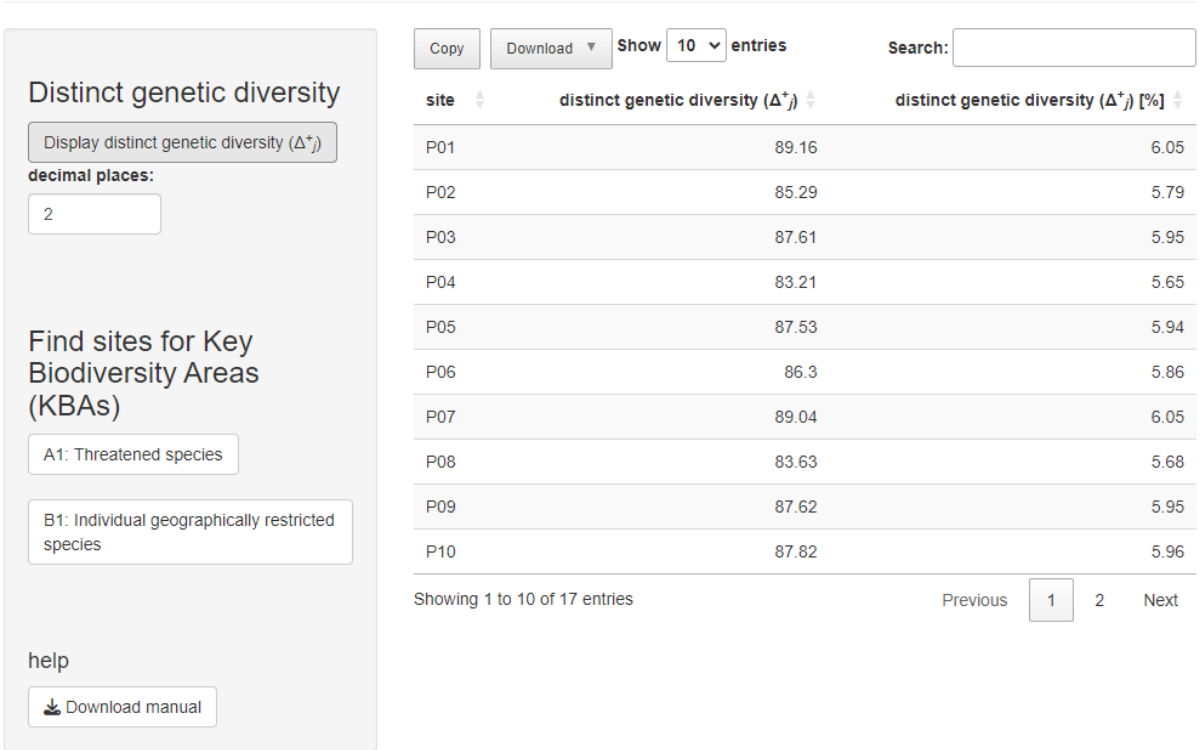

Fig. 7: Display and download distinct genetic diversity (Δ^+_j).

What was calculated

The table has three columns: "site", "distinct genetic diversity (Δ^+_j)", and "distinct genetic diversity (Δ^+_j) [%]" (Fig. 7). The column "site" contains the labels that you picked for your

locations. The column “distinct genetic diversity (Δ^+_j)” contains the raw results of Δ^+_j. Δ^+_j was calculated using a varying step length because it is closer to reality and comparisons to external data sets are not necessary, as Δ^+_j is only compared in the form of ratios within a species. Since both step length options produce similar results, the results will reflect Δ^+_j calculated with equal step length (Clarke & Warwick, 1998; Mistri et al., 2000). Δ^+_j calculates the mean genetic distance of one area to all other areas by assigning greater distances to larger differences in allele abundances between areas, summing all distances of one area to other areas, and dividing by the number of possible area combinations with the area (equation 1) (Clarke & Warwick, 2001; Oksanen et al., 2024).

$$\Delta_j^+ = \frac{2}{N-1}\sum_{i=1}^{N} d_{ij} \qquad 1$$

Δ^+_j = mean genetic distance of site *j* to all other sites

N = number of sites

d_{ij} = genetic distance between site *j* and site *i*

The last column of the table “distinct genetic diversity (Δ^+_j) [%]” contains the proportion of distinct genetic diversity for each site. This proportion is relevant for testing whether the site meets the IUCN thresholds, which can be be computed automatically by this program (see Identify Key Biodiversity Areas (KBAs)).

Display options

You can change the rows sorting, the number of decimal places, and the data entries being displayed. It is possible to change the number of data entries being displayed, switch between table pages, and you have the option of filtering your data according to specific sites you are interested in.

New arrangement of the table rows

There are small arrows next to each column name in the table. You can use the arrows to change the order of the rows (Fig. 8).

site	distinct genetic diversity (Δ^+_j)	distinct genetic diversity (Δ^+_j) [%]
P17	89.19	6.06
P01	89.16	6.05
P07	89.04	6.05
P16	88.45	6.01
P10	87.82	5.96
P03	87.61	5.95
P09	87.62	5.95
P05	87.53	5.94
P12	86.58	5.88
P15	86.49	5.87

Showing 1 to 10 of 17 entries Previous 1 2 Next

Fig. 8: Change order of rows of the table.

Changing the number of decimal places

In the side panel below the button for displaying the table of distinct genetic diversity, there is a field that allows you to change the number of decimal places (Fig. 9).

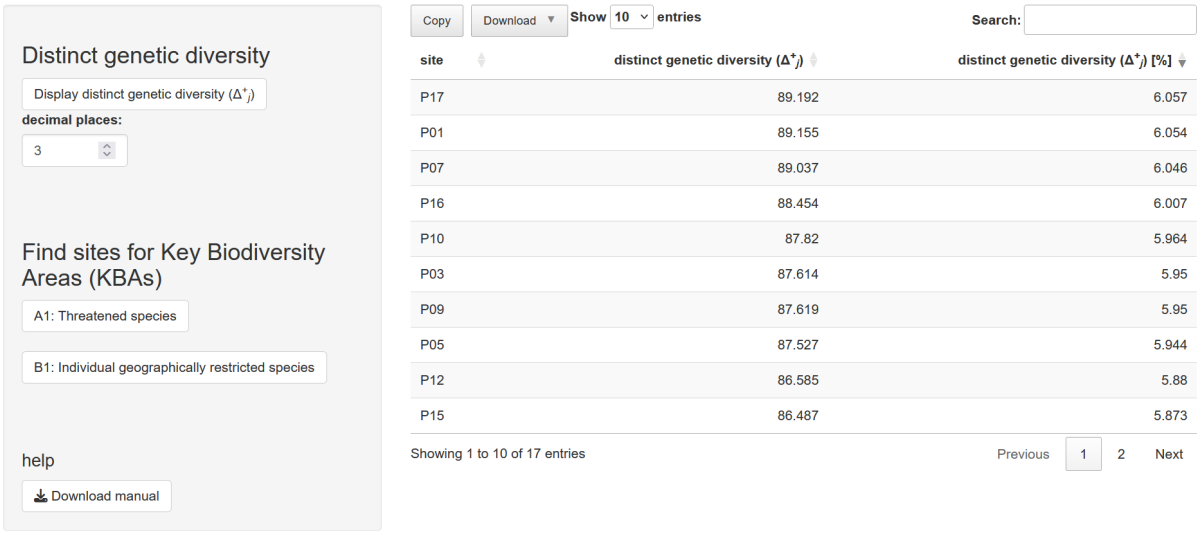

Fig. 9: Change number of decimal places of the table.

Change which and how many results are displayed

So far only 10 of the 17 sites have been displayed. You can see more results by changing the table page by pressing “Previous”, “Next” or the page number on the right below the table or by increasing the number of entries displayed on a table page (Fig. 10). You can decide whether you want to display 10, 20 or all entries on a table page.

Copy Download Show 20 entries Search:

site	distinct genetic diversity (Δ^+_j)	distinct genetic diversity (Δ^+_j) [%]
P01	89.16	6.05
P02	85.29	5.79
P03	87.61	5.95
P04	83.21	5.65
P05	87.53	5.94
P06	86.3	5.86
P07	89.04	6.05
P08	83.63	5.68
P09	87.62	5.95
P10	87.82	5.96
P11	84.26	5.72
P12	86.58	5.88
P13	85.46	5.8
P14	84.9	5.77
P15	86.49	5.87
P16	88.45	6.01
P17	89.19	6.06

Fig. 10: Display more than 10 entries of the table.

Filtering table

You can use the search field to filter your data. If you are specifically interested in the ratio of distinct genetic diversity of a particular site, you can enter the site name in the search field (Fig. 11). It is also possible to search for all sites that contain a certain combination of letters and thus have a certain selection of sites displayed.

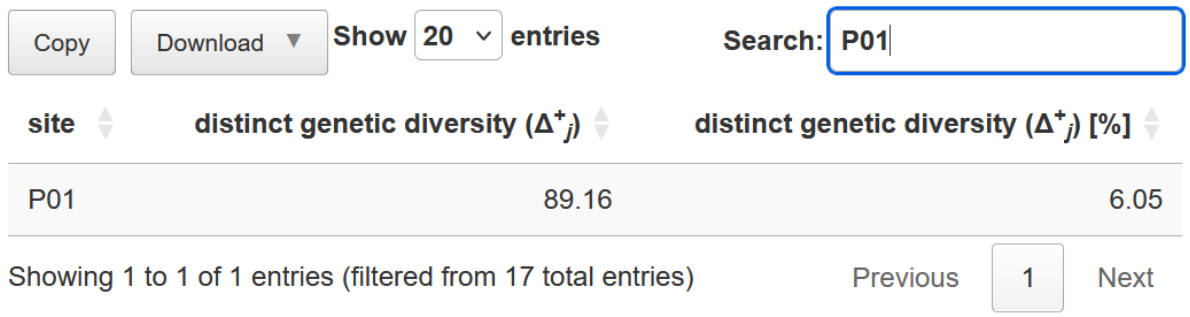

Fig. 11: Searching for a specific site.

Save results

You can copy the distinct genetic diversity table or download it as CSV, EXCEL (.xlsx), and PDF to the directory of your choice.

Copy to clipboard

The table will always be copied in the same way it is displayed (Tab. 3).

Tab. 3: First ten entries copied.

Exported data

site	distinct genetic diversity (Δ+j)	distinct genetic diversity (Δ+j) [%]
P01	89.16	6.05
P02	85.29	5.79
P03	87.61	5.95
P04	83.21	5.65
P05	87.53	5.94
P06	86.3	5.86
P07	89.04	6.05
P08	83.63	5.68
P09	87.62	5.95
P10	87.82	5.96

Download

If you click the download button you will be asked to choose between CSV, EXCEL (.xlsx), and PDF (Fig. 12). If you have selected the preferred format, you will be asked to select the directory in which you would like to save your table. The table will be downloaded as it is displayed.

Fig. 12: Choose file format for downloading the table.

If you select the CSV format, the last two columns are placed in quotation marks. Instead of Δ^{+}_{j} Î”+j is written. If you download your table as EXCEL (.xlsx) or PDF file, a line "Exported data" is added as a header. Δ^{+}_{j} is written as Δ+j.

Identify Key Biodiversity Areas (KBAs)

To identify sites that qualify as 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 restricted) (Fig. 13). Depending on the criterion, different thresholds will be applied that can influence the number of sites that will be selected (Tab. 4). If you click the button for criterion B1 your results will immediately be displayed.

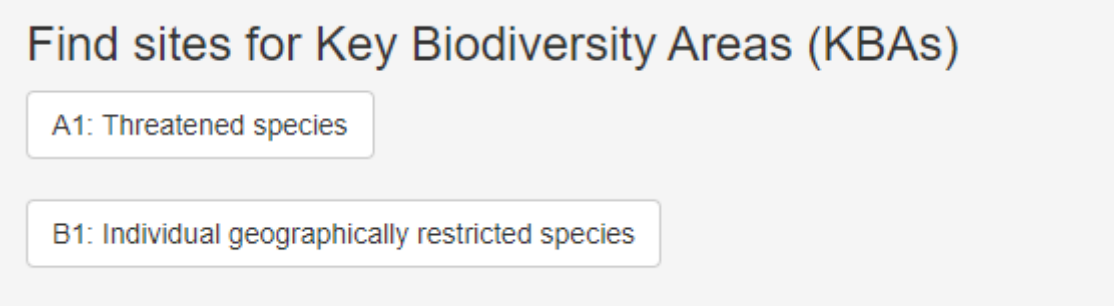

Fig. 13: Choose between criterion A1 and criterion B1.

If you click the button for criterion A1, you will have to pick the Red List status of the studied species (Fig. 14). The identified KBAs will be displayed after choosing the appropriate criterion.

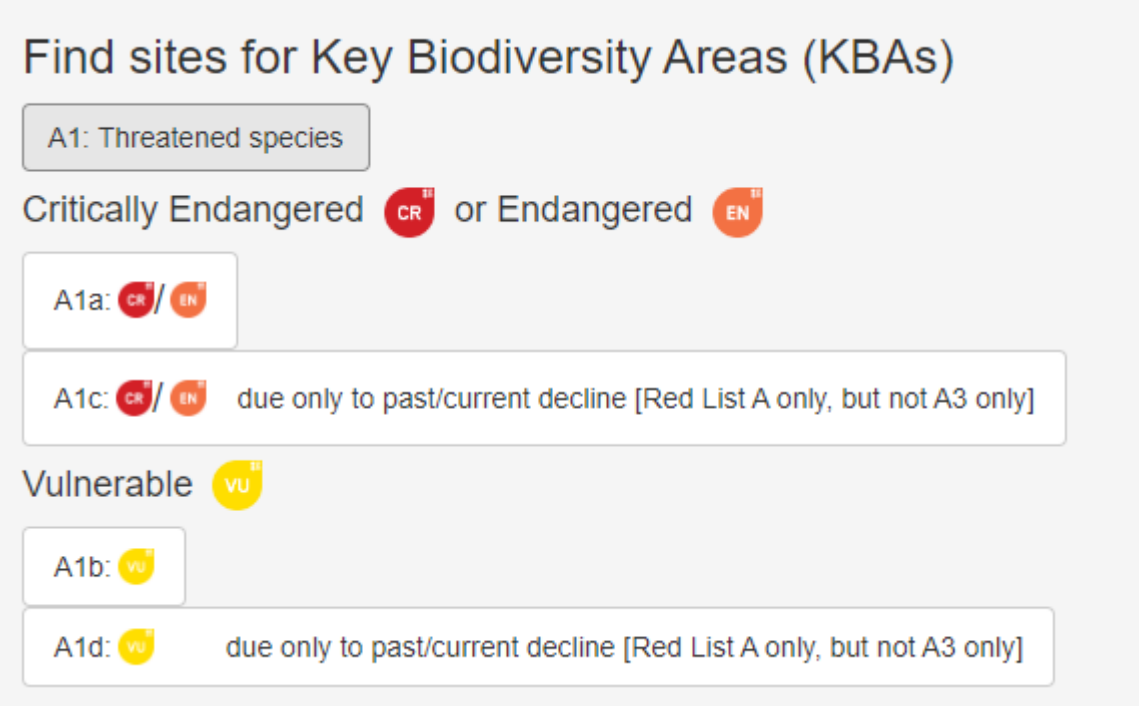

Fig. 14: Pick your species Red List status.

The KBA criteria do not differentiate between Critically Endangered (CR) or Endangered (EN) species, but if the species is Vulnerable (VU) a different threshold will be applied that will cause less sites to be selected. If the Red List category is based upon past or

current decline [Red List criteria A1, A2 or A4] a lower threshold applies so that more sites will be selected (Tab. 4) (IUCN, 2020).

Tab. 4: Summary of IUCN criteria and thresholds relevant to distinct genetic diversity

KBA criterion	threshold
A1a	≥0.5%
A1b	≥1%
A1c	≥0.1%
A1d	≥0.2%
B1	≥10%

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 sites qualify becoming a KBA." or start with "Following sites qualify becoming a KBA " and ending with a list of the sites identified. Note that distinct genetic diversity is a means to estimate the proportion of the global population size and does not provide any information on the number of reproductive units. The thresholds for reproductive units, therefore, must be assessed independently of the calculation of genetic distinctiveness.

help

Below the "help" header in the program you have the option to download this manual (Fig. 15). More information about the underlying calculations can be found in the paper and in the instructions on how to calculate distinct genetic diversity (https://rpubs.com/Who_could_this_be_2/1217013).

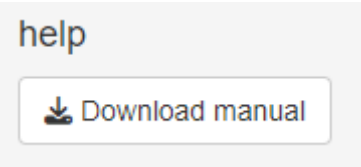

Fig. 15: Download manual.

How to continue with your results

Read the guidelines carefully. If you have information about sites selected by 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 you can compare these sites with the sites selected with distinct genetic diversity and decide which assessment parameter is best. You can combine the knowledge about your sites identified with other information you are able to gather about different species at the same site. 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! If you find any bugs, please inform the author...

Bibliography

Clarke, K.; Warwick, R. (Eds.) (2001): Change in Marine Communities. An Approach to Statistical Analysis and Interpretation. Plymouth, UK: Primer-E Ldt.

Clarke, K.; Warwick, Richard (1998): A taxonomic distinctness index and its statistical properties. In *Journal of Applied Ecology* 35.

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.

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.

Mistri, Michele; Ceccherelli, Victor Ugo; Rossi, Remigio (2000): Taxonomic distinctness and diversity measures: Responses in lagoonal macrobenthic communities. In *Italian Journal of Zoology* 67 (3), pp. 297–301. DOI: 10.1080/11250000009356327.

Acknowledgements

Our sincere thanks go to Pascal Barzen and Lukas Knob for programming support, Michael Mietzner for logo design, Andrew Plumptre for answering questions about how to continue after the app was used.