

Authors

Marie C Henniges^{1,2}, Andrew R Leitch², Ilia J Leitch¹

Affiliations

1. Royal Botanic Gardens, Kew, Richmond TW9 3AE, UK

2. School of Biological and Behavioural Sciences, Queen Mary University of London, London E1 4NS, UK

Corresponding author(s):

Marie C Henniges (marie.c.henniges@gmail.com)

Ilia J Leitch (i.leitch@kew.org)

Proposed running head: 'BIFloraExplorer': Vascular plants of Britain and Ireland

Chapter title

'BIFloraExplorer'--a taxonomic, genetic and ecological data resource for the vascular plants of Britain and Ireland

Summary/Abstract

The vascular flora of Britain and Ireland is a historically well documented and clearly delimited study system that offers itself to large-scale analyses of ecology and species assemblages. Such analyses require clean, curated, and taxonomically resolved data, which are often unavailable. In this chapter we describe how to access and use a key data resource

that combines a taxonomically stable species list with genetic data (genome size, chromosome counts and DNA barcode information), ecological information (such as life form, realized niche description and geographic origin) and distribution records. The data resource enables and encourages the study of natural ecological and evolutionary patterns and processes within the vascular flora of Britain and Ireland.

Key words

Vascular plants, taxonomy, native, alien, functional traits, Ellenberg indicator values, life strategy, life-form, biome, distribution, hybrid propensity, DNA barcodes, genome size, chromosome number

1 Description

Systematic botanical research on the islands of Britain and Ireland (comprising England, Scotland, Wales, Northern Ireland, Republic of Ireland, Isle of Man and the Channel Islands) dates back to John Ray's 1690 work *Synopsis methodica stirpium Britannicarum* **(1)** and has been driven forward to this day by the work of organizations such as the Botanical Society of Britain and Ireland (BSBI) **(2)** and individual recorders and researchers who have built an impressive knowledge base of the vascular flora, both native and introduced to the islands. While various printed works and online resources (notably **3, 4, 5, 6, 7, 8**) strive to catalogue and make available this wealth of information, integrated and large-scale analyses of trends

within the flora require a new approach to databasing this knowledge. As of January 2022, a new resource offers a taxonomically unified and harmonized overview of the genetic, ecological and distribution data currently available for the vascular flora of Britain and Ireland. The dataset, available as a static dataset on the Environmental Information Data Centre (EIDC) **(9)** and as an R data package on GitHub **(10)**, is fully open access, with detailed descriptions of how it was assembled, and the sources used for its compilation available in the original data descriptor **(11)**. Based on an internally consistent species list of all clearly delimited species, both native and introduced to Britain and Ireland following the *New Flora of the British Isles* (Fourth Edition) **(3)**, the taxonomy is further consolidated by linking species names with those accepted by the *World Checklist of Vascular Plants* **(12)** and assigning unique numeric taxon identifiers from the latter. The dataset enables access to a representative set of 27 traits, characters, and measurements describing the 3,227 species currently considered part of the flora (Fig. 1). The compiled information not only comprises data from published papers but also includes data uniquely available in this repository from previously unpublished and newly generated datasets. These include (i) novel genome size measurements and chromosome counts, (ii) imputed life strategy categories following Grime **(13)**, (iii) life-form descriptions following Raunkiaer **(14)** which were determined specifically for typical growth within Britain and Ireland, and (iv) a hybrid propensity metric **(15)**.

The resource aims to make large-scale studies of the British and Irish flora an easier undertaking, reducing difficulties due to taxonomic inconsistencies between available datasets and/or scattered, heterogenic data.

The available data are contained within one main dataset ('BI_main') and three tables containing additional chromosome numbers ('chrom_num_BI'), genome sizes ('GS_BI') and more detailed information about the new genome size measurements not previously published ('GS_Kew_BI').

Below we describe how to access the data via the static repository (2.1) and in the R package 'BIFloraExplorer' (2.2). We also give examples of how the resource might be used (3).

2 Accessing the data

The data are available either via download from the EIDC catalogue **(9)** or as an R data package on GitHub **(10)**. Both versions can be accessed online as described below.

2.1 Download of the static dataset (EIDC)

The static version contains all data and associated information as of the first publication of the dataset **(11)**. Future additions, changes and other developments will not be represented here. Because of this, the authors recommend using the R package 'BIFloraExplorer' (see Section 2.2 below) **(10)** since this version of the repository will be updated with new information as it becomes available. Nevertheless, for users less comfortable with GitHub repositories, the static version still offers the full range of information described in the initial publication **(11)** of the resource.

1. Navigate to <https://catalogue.ceh.ac.uk/documents/9f097d82-7560-4ed2-af13-604a9110cf6d>. The landing page contains some general information describing the dataset.
2. In the box “Get the data” users can choose to download supporting documentation only or the full dataset. Clicking on “Download the data” will download both supporting documents and data simultaneously. While the supporting documents can be downloaded without any requirement to register with the EIDC, downloading the dataset requires the user to provide an email address.
3. Click “Download the data”. This prompts the user to either log into a pre-existing account or to create a new account.
4. Should you be a pre-existing user of the EIDC catalogue, log in and click on download on the following screen. This will cause the full data package to be downloaded as a compressed folder that can be opened with any conventional zip tool program.
5. If you don’t have an EIDC account, click the option “Create an account” at the bottom of the login mask. After signing up on the following screen you will receive an email containing a hyperlink to activate your new account (usually within a few minutes of signing up). Follow this link and you will be prompted to log into your newly activated account. Use the email address used to set up the account and your new password to log in. Navigate back to <https://catalogue.ceh.ac.uk/documents/9f097d82-7560-4ed2-af13-604a9110cf6d> and click on “Download the data” (see Note 1). Accept the Access License (Open Government License) notice when prompted and allow downloads from “data-package.ceh.ac.uk” in the pop-up window (may not appear depending on browser settings). The data are then downloaded and can be decompressed with any conventional zip tool program (see Note 2).

2.2 Import of data into R from GitHub

The preferred way of accessing the repository is via the R package 'BIFloraExplorer' **(10)** (Fig. 2). The GitHub repository of the 'BIFloraExplorer' can be found under <https://github.com/RBGKew/BIFloraExplorer>. Any changes to the dataset as it evolves over time and new data that are added will be reflected here.

Accessing the data in this way requires a version of R to be installed on the user's machine. We recommend the use of a recent version of R for a smooth experience in working with the package.

1. In the R console, install the R package 'devtools' **(16)** and load it into the session's memory using the following commands.

```
> install.packages("devtools")  
> require(devtools)
```

2. Continuing in the R console, install the newest version of 'BIFloraExplorer':

```
> devtools::install_github("RBGKew/BIFloraExplorer")  
> require(BIFloraExplorer)
```

3. The package is now installed and ready for use. The contained data frames can now be interacted with in R. E.g.:

```
> View(BI_main)  
> head(BI_main)
```

4. To obtain information about the different categories of data contained in the resource, please refer to the data descriptor publication **(11)** or use the following command, inserting the table of interest ('BI_main', 'chrom_num_BI', 'GS_BI' or 'GS_Kew_BI'):

```
> help(BI_main)
```

3 Example uses (R package)

Most of the actions described below for the R package can also be performed on the static dataset after download from the EIDC repository by using spreadsheet or database software (such as Microsoft Excel, Microsoft Access, or Google Sheets).

For users of the 'BIFloraExplorer' R package, we recommend using functions from the 'dplyr' package **(17)** for easy and smooth use of the dataset. To avoid issues with the following process, users should ensure they are running a recent version of R with up-to-date dependencies of packages such as 'dplyr'.

1. Some species (e.g. those belonging to genus *Cotoneaster*, *Rubus* or *Taraxacum*) are part of aggregates, groups and/or can be difficult to tell from their close relatives in the field. This means that data associated with these species are potentially less reliable than for other species. Users who would like to use only data for taxa without such uncertainties can remove unclear species from the dataset, creating a new object without these uncertain taxa. The same method can be used to remove extinct species that are included in the repository without associated trait information.

```
> BI_main_clear <- filter(BI_main, unclear_species_marker != "Y")
```

```
> BI_main_extant <- filter(BI_main, extinct_species_marker != "E")
```

2. Depending on the user's needs, they may wish to use species names as they are used in the most recent version of the *New Flora of the British Isles* **(3)** ('BI_main\$taxon_name' and 'BI_main\$taxon_name_binom' for taxon names with and without taxonomic

authority respectively) or species names as they are used in the *World Checklist of Vascular Plants* **(12)** ('BI_main\$taxon_name_WCVP'). N.B. Hyperlinks within column 'BI_main\$WCVP_URL' allows users to view the species' page on the *World Checklist of Vascular Plants*.

3. Quick overviews of the information can be obtained by running e.g. `summary(BI_main$GS_1C_pg)` on numeric and `summary(as.factor(BI_main$StaceIV_nativity))` on categorical columns.
4. The dataset is ready for use in statistical modelling but can also be used to visualize the current state of the British and Irish flora (e.g. Fig. 3 a-d). These example visualizations were made using the R packages 'ggplot2' **(18)**, 'treemap' **(19)** and 'maps' **(20)**.
5. Users interested in more detailed information about chromosome counts and genome sizes can make use of the additional tables within the package. Objects 'chrom_num_BI' and 'GS_BI' offer further chromosome counts and genome sizes that allow users to gain an overview of the infraspecific variation that has been reported for some taxa.

4 Notes

1. When downloading the dataset from the EIDC without a pre-existing account, some users are redirected to a page with the message that the desired page cannot be found. Simply navigating to the dataset's landing page again allows users to download the data, regardless of this message.
2. The data in the static repository are available in csv format and can be opened and viewed in spreadsheet and database software such as Microsoft Excel, Microsoft Access, or Google Sheets.

Acknowledgement

Marie C Henniges is funded by the Natural Environment Research Council, grant no. NE/L002485/1.

References

(1) Ray J (1690) Synopsis methodica stirpium Britannicarum, in qua tum notae generum characteristicae traduntur, tum species singulae breviter describuntur. S. Smith (Londini).

(2) Botanical Society of Britain and Ireland. <https://bsbi.org/>. Accessed 14 Feb 2022

(3) Stace C (2019) New Flora of the British Isles – Fourth Edition. C&M Floristics

(4) Hill MO, Preston CD, Roy DB (2004) PLANTATT-attributes of British and Irish plants: status, size, life history, geography and habitats. Centre for Ecology & Hydrology

(5) Fitter AH, Peat HJ (1994) The ecological flora database. Journal of Ecology, 82(2), 415–425 (1994). <https://www.ecoflora.org.uk>. Accessed 14 Feb 2022

(6) Database for the Biological Flora of the British Isles. British Ecological Society.

<https://www.britishecologicalsociety.org/publications/journals/journal-of-ecology/biological-flora-database/>. Accessed 14 Feb 2022

(7) BSBI database search facility (taxon, literature and cytology database).

<https://websites.rbge.org.uk/BSBI/intro.php>. Accessed 14 Feb 2022

(8) BSBI Distribution Database. <https://database.bsbi.org/>. Accessed 14 Feb 2022

(9) Henniges MC, Powell RF, Mian S et al (2021) A taxonomic, genetic and ecological data resource for the vascular plants of Britain and Ireland. NERC Environmental Information Data Centre. <https://doi.org/10.5285/9f097d82-7560-4ed2-af13-604a9110cf6d>. Accessed 14 Feb 2022

(10) Henniges MC, Powell RF, Mian S et al (2022). BIFloraExplorer: A taxonomic, genetic and ecological data resource for the vascular plants of Britain and Ireland. R package version 0.1.0. <https://github.com/RBGKew/BIFloraExplorer>. Accessed 14 Feb 2022

(11) Henniges MC, Powell RF, Mian S et al (2022). A taxonomic, genetic and ecological data resource for the vascular plants of Britain and Ireland. Sci Data 9, 1. doi: 10.1038/s41597-021-01104-5

(12) World Checklist of Vascular Plants, version 2.0 (2020) Royal Botanic Gardens, Kew. <http://wcvp.science.kew.org/>. Accessed 14 Feb 2022

(13) Grime JP (1974) Vegetation classification by reference to strategies. Nature 250(5461): 26–31

(14) Raunkiær C (1934) The life forms of plants and statistical plant geography being the collected papers of C. Raunkiær. Oxford at the Clarendon Press

(15) Whitney KD et al. (2010). Patterns of hybridization in plants. Perspectives in Plant Ecology, Evolution and Systematics 12(3), 175–182

(16) Wickham H, Hester J, Chang W, Bryan J (2021). devtools: Tools to make developing R packages easier. R package version 2.4.3. <https://CRAN.R-project.org/package=devtools>. Accessed 14 Feb 2022

(17) Wickham H, François R, Henry L, Müller K (2021) dplyr: A Grammar of Data Manipulation. R package version 1.0.7. <https://CRAN.R-project.org/package=dplyr>. Accessed 14 Feb 2022

(18) Wickham H (2016) ggplot2: Elegant graphics for data analysis. Springer-Verlag New York. R package version 3.3.5. <https://cran.r-project.org/web/packages/ggplot2/index.html>. Accessed 14 Feb 2022

(19) Tennekes M (2021) treemap: Treemap visualization. R package version 2.4-3. <https://CRAN.R-project.org/package=treemap>. Accessed 14 Feb 2022

(20) Becker RA, Wilks AR. Original S code. Brownrigg R. R version. Minka TP, Deckmyn A. Enhancements (2021). maps: Draw geographical maps. R package version 3.4.0. <https://CRAN.R-project.org/package=maps>. Accessed 14 Feb 2022

Figure captions

Fig. 1 Overview of information categories included within the resource (reproduced from ref. 11).

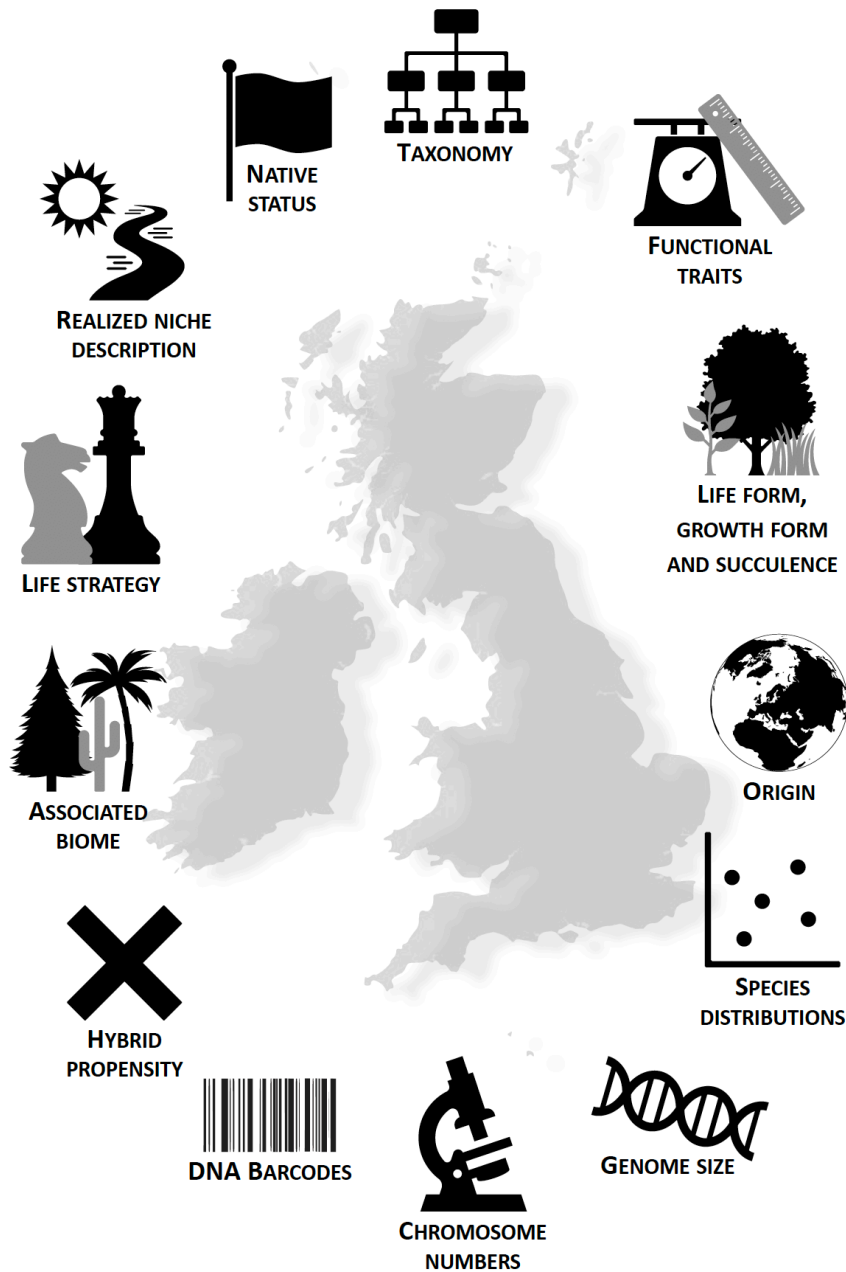


Fig. 2 Logo of the 'BIFloraExplorer' R package.

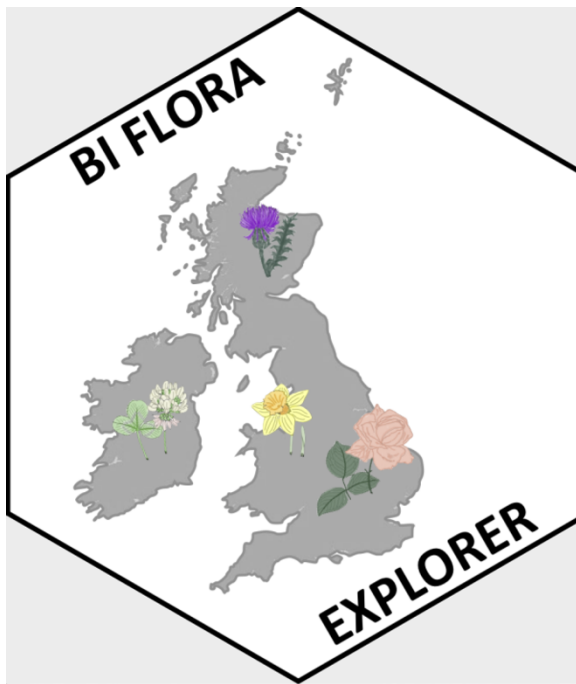


Fig. 3 Examples of visualizations that can be generated with the dataset. (a) Treemap of status distributions across Britain and Ireland. (b) Histogram of genome size data in megabasepairs [Mbp] per haploid genome [1C], which exist for 66% of the native and alien flora of Britain and Ireland. The plants overlaid over the histogram are located close to their respective genome size and are, from left to right: *Linnaea borealis* L. (791 Mbp/1C), *Botrychium lunaria* (L.) Sw. (11,858 Mbp/1C), *Erythronium dens-canis* L. (24,490 Mbp/1C), *Fritillaria meleagris* L. (46,354 Mbp/1C), *Tulipa sylvestris* L. (56,724 Mbp/1C), *Viscum album* L. (86,944 Mbp/1C). (c) Bubble plot visualizing the origin of alien species within the flora. Bubble size and numbers show how many species were introduced to Britain and Ireland from each continent. (d) Violin plots illustrating the distribution of genome sizes [log(Mbp/1C)] of species associated with different biome types.

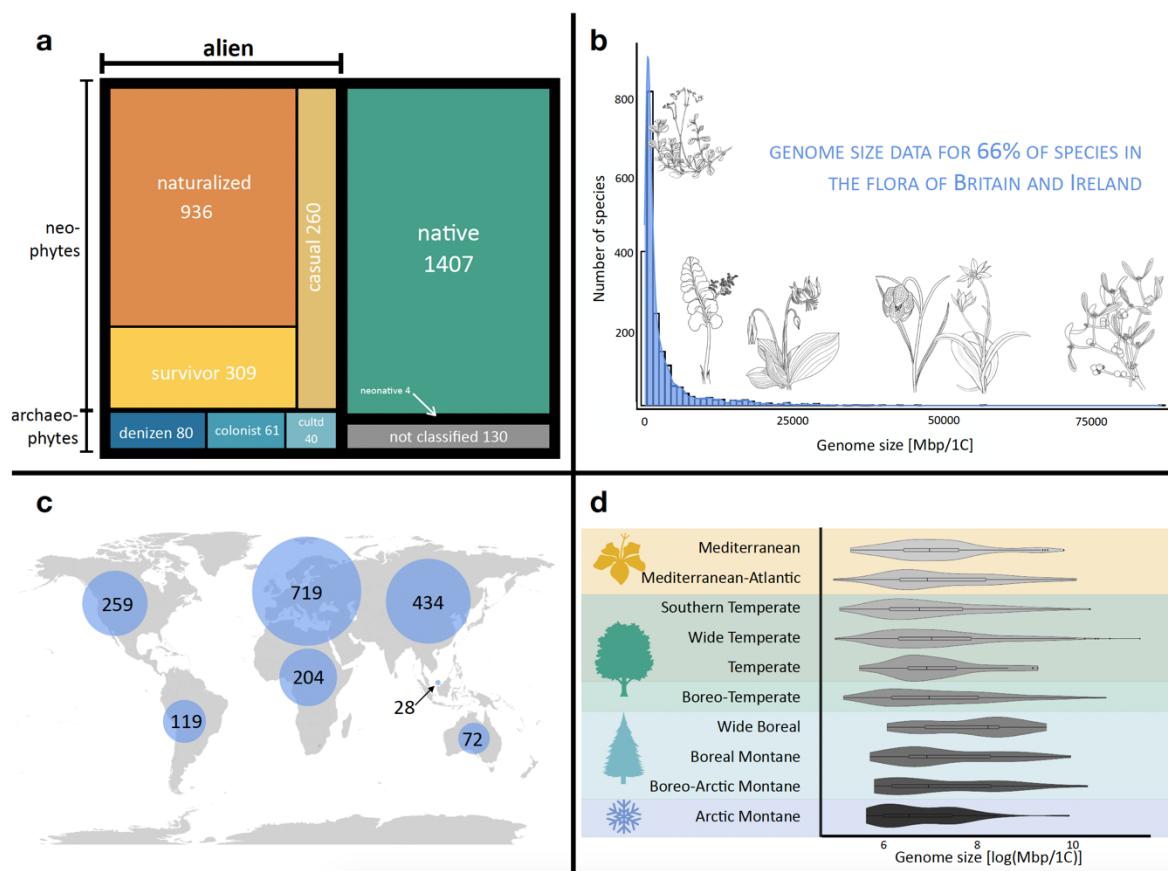


Figure log

Figure No.	Copyright holder	Copyright holder's required wording	Notes
1	The authors (ourselves) as per the copyright policy of Nature portfolio journals	The journal requires credit with first publication, no additional requirements	From previous publication
2	NA	NA	Not within prior publication
3 (a-d)	NA	NA	Generated for this chapter