pagepress

Journal of Limnology

Journal of Limnology 2026; 85(s1):2254
doi:10.4081/jlimnol.2026.2254

DataLake: a georeferenced dataset of vascular plants from freshwater lakes of central-southern Italy

Lorenzo Pinzani,[1,2]* Dario Di Lernia,[1] Simona Ceschin[1,2]

[1]Department of Science, University of Roma Tre, Rome; [2]NBFC-National Biodiversity Future Center, Palermo, Italy

Abstract

Freshwater lakes are fundamental ecosystems that provide essential ecosystem services for humans and support high plant diversity, often including specialized and rare species. Nevertheless, in Italy, botanical knowledge of lake ecosystems remains fragmented, and floristic data are scattered across heterogeneous sources. *DataLake* is a dataset designed to integrate and standardize existing floristic knowledge on aquatic and riparian vascular plants occurring in 30 natural freshwater lakes of central-southern Italy. Records included in *DataLake* derive from both literature sources and original field surveys conducted by the authors. Overall, *DataLake* comprises 5140 georeferenced floristic records referring to 213 taxa, 97 genera and 43 families. The dataset highlights a markedly uneven distribution of taxa and records among lakes, with large tectonic and volcanic basins accounting for the highest floristic richness and number of records. The taxonomic composition is dominated by Cyperaceae, Potamogetonaceae and Juncaceae, while at the species level *Myriophyllum spicatum* L., *Ceratophyllum demersum* L., *Phragmites australis* (Cav.) Trin. ex Steud., *Potamogeton perfoliatus* L. and *Stuckenia pectinata* (L.) Börner are the most frequently recorded taxa. A substantial proportion of taxa is included in regional (39% of the dataset) and national (10.8%) conservation risk categories, underlining the role of lacustrine ecosystems as important reservoirs for plant species of conservation concern. Alien species were recorded in 14 lakes, and among these, 12 taxa are classified as invasive, including *Paspalum distichum* L., *Elodea canadensis* Michx. and *Bidens frondosa* L.. Biological and chorological spectra reflect the strong ecological dependence of lacustrine plant species on hydrological conditions, with dominance of hydrophytes and helophytes and multizonal taxa. By integrating heterogeneous floristic data into a single, coherent and openly accessible resource, *DataLake* provides a solid reference base for future floristic research, as well as for supporting long-term monitoring activities and actions aimed at conserving natural Mediterranean freshwater lake ecosystems.

Key words: freshwater lakes; floristic dataset; Mediterranean region; biodiversity data; aquatic plants; riparian plants.

Correspondence to: lorenzo.pinzani@uniroma3.it

Introduction

Inland water ecosystems, such as freshwater lakes, are unique and fundamental ecosystems as they provide essential ecosystem services for humans and preserve high biodiversity (Wetzel, 2001; Kalff, 2002). Despite their importance, lake ecosystems are highly vulnerable environments, and the conservation of their biodiversity is seriously threatened by hydrological alteration, eutrophication, water pollution, bank modification and biological invasions (Jeppesen *et al*., 2010; Bornette and Puijalon, 2011).

Despite the numerous botanical studies conducted on Italian natural freshwater lakes, floristic data on aquatic and riparian plants of these lakes remains fragmentary and unevenly accessible, particularly in the central and southern Italian regions. The available data, often characterized by low spatial accuracy, are scattered among historical publications, dated records, local studies, grey literature and taxonomic treatments targeting specific taxa. As a result, floristic knowledge is often limited to species lists that simply indicate the presence of a species in a lake without specifying its distribution and spatial frequency. Recent studies have attempted to collect and update scattered floristic knowledge into single contributions focusing on a subset of natural lakes in Italy, particularly Italian volcanic lakes (Pinzani *et al*., 2025a, 2025b).

In this context, the elaboration of a georeferenced floristic dataset that includes all available floristic data on natural freshwater lakes of different origins in central and southern Italy would represent a fundamental step towards a comprehensive botanical knowledge of these lakes. Recent efforts have similarly focused on integrating vegetation data from Italian freshwater systems, such as the PONDY database on vegetation of Italian ponds (Cannucci *et al*., 2025), highlighting the growing importance of structured and accessible datasets on biodiversity of inland water ecosystems. Within this framework, *DataLake* was developed as a database containing all available floristic data on aquatic and riparian vascular plants from 30 natural freshwater lakes in central-southern Italy. *DataLake* was created by integrating original field data with records extracted from available botanical literature. Particular attention was paid to spatial data relating to single species recorded, georeferencing them to define as accurately as possible the distribution of each species at each lake considered, without being limited to indicating their presence/absence in each lake.

METHODS

Study area

The dataset includes floristic records from 30 freshwater lakes located in central-southern Italy (Fig. 1). The study area spans a wide latitudinal range encompassing peninsular Italy and includes lakes of different origins: volcanic, karst, glacial, alluvial, tectonic and landslide-dammed lakes (Tab. 1). Only natural lakes were considered, excluding artificial reservoirs and heavily modified water bodies, paying to attention on systems where the plant record reflects the natural ecological gradients rather than recent engineering interventions.

Lake selection was primarily driven by the availability of floristic data in botanical literature. For each lake, all accessible floristic sources referring to aquatic and riparian vascular plants surveyed were screened, with no temporal restriction, allowing the inclusion of both historical and recent records for each lake, in addition to original data collected in the field by the authors.

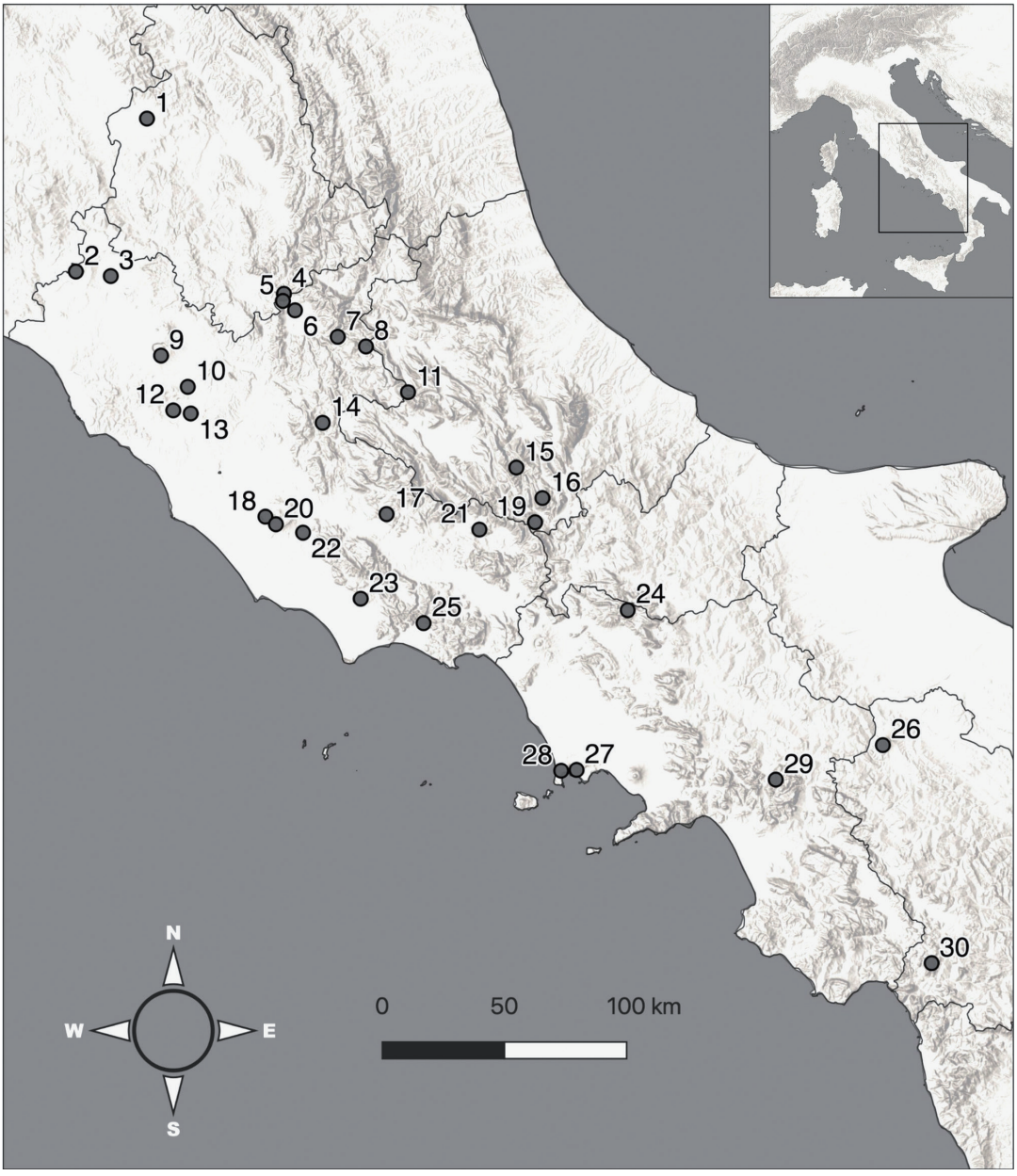

Fig. 1. Location of the 30 natural freshwater lakes included in *DataLake* across central-southern Italy. Lake numbering follows a north-south latitudinal gradient: 1. Trasimeno; 2. Mezzano; 3. Bolsena; 4. Piediluco; 5. Ventina; 6. Lungo-Ripasottile; 7. Paterno; 8. Rascino; 9. Vico; 10. Monterosi; 11. Duchessa; 12. Bracciano; 13. Martignano; 14. Percile; 15. Scanno; 16. Pantaniello; 17. Canterno; 18. Albano; 19. Vivo; 20. Nemi; 21. Posta Fibreno; 22. Giulianello; 23. Vescovo; 24. Matese; 25. Sette Cannelle; 26. Monticchio; 27. Astroni; 28. Averno; 29. Laceno; 30. Laudemio.

Data sources and selection

To ensure that the dataset included only vascular plant species strictly linked to lake ecosystem, Ellenberg autoecological index values for humidity (I_H) (Ellenberg *et al*., 1992) were applied as species selection criterion. The values adapted to the Italian vascular flora by Guarino and La Rosa (2019) were considered. On a scale of increasing hydrophily ranging from 1 to 12, only taxa with $I_H \geq 8$ were retained, corresponding to species ecologically linked to aquatic, semi-aquatic or riparian habitats. This threshold excludes facultative or moderately hygrophilous plants, focusing instead on taxa whose distribution and persistence depend directly on permanent water or high soil humidity.

Taxonomic nomenclature of the species followed Bartolucci *et al*. (2024) and Galasso *et al*. (2024), with updates from the Portal of the Flora of Italy (2025). Synonyms and outdated names reported in the original sources were harmonized accordingly, allowing the integration of historical and recent records under a unified taxonomic framework thereby ensuring consistency among heterogeneous sources. Life forms and chorotypes were attributed to each species following Pignatti *et al*. (2017-2019). In addition, conservation status was attributed according to the National Red Lists (Rossi *et al*., 2013, Orsenigo *et al*., 2020) and Regional Red Lists (Conti, 1997), while endemicity status was assigned following Bartolucci *et al*. (2024).

Georeferencing

All species records included in *DataLake* were georeferenced to ensure spatial consistency and usability for spatial analyses. Original field data were georeferenced in situ using GPS devices, providing high spatial accuracy. Bibliographic records were georeferenced based on locality descriptions reported in the original sources. In the case of historical data, when possible, locality information was cross-validated with contemporary maps to reduce spatial uncertainty and to assign coordinates consistent with current geographic references.

Georeferencing was carried out with particular attention to spatial accuracy. For each record, an explicit estimate of spatial uncertainty was assigned and stored, allowing differences in positional precision among records to be documented and accounted for in subsequent analyses. Seven spatial accuracy classes were defined (100, 500, 1000, 2000, 5000, 6000 and 7000 m), reflecting increasing levels of uncertainty associated with the original data source and the level of detail provided in the locality description.

When records referred generically to a lake without further spatial details (e.g., Lake Bracciano), geographic coordinates were assigned to the centroid of the lake polygon. In these cases, spatial accuracy was estimated using the mean radius of the lake calculat-

Tab. 1. For each of the 30 natural freshwater Italian lakes included in *DataLake*, geographic information (Italian region, latitude and longitude), geological origin, elevation, surface and floristic richness are reported. The floristic richness refers to the number of taxa and records of vascular plants.

Lake	Region	Coordinates	Origin	Elevation (m)	Surface (km²)	Taxa (n)	Record (n)	Reference (n)
Trasimeno	Umbria	43.144, 12.107	Tectonic	257	128.0	144	1577	41
Piediluco	Umbria	42.533, 12.758	Alluvial	375	1.58	54	67	2
Mezzano	Lazio	42.611, 11.769	Volcanic	452	0.9	52	154	10
Bolsena	Lazio	42.595, 11.935	Volcanic	305	113.0	64	296	13
Ventina	Lazio	42.508, 12.753	Alluvial	378	0.12	26	49	12
Lungo - Ripasottile	Lazio	42.475, 12.811	Alluvial	370	1.4	43	52	5
Paterno	Lazio	42.382, 13.014	Karst	430	0.03	1	2	2
Rascino	Lazio	42.348, 13.148	Karst	1146	0.1	24	65	7
Vico	Lazio	42.316, 12.174	Volcanic	510	12.9	84	668	14
Monterosi	Lazio	42.206, 12.301	Volcanic	243	0.3	51	126	8
Duchessa	Lazio	42.187, 13.348	Glacial	1788	0.04	5	9	0
Bracciano	Lazio	42.123, 12.232	Volcanic	164	57.5	85	473	19
Martignano	Lazio	42.112, 12.315	Volcanic	207	2.26	35	257	11
Percile	Lazio	42.079, 12.942	Karst	650	9	20	35	4
Canterno	Lazio	41.756, 13.246	Karst	541	1.6	31	64	9
Albano	Lazio	41.747, 12.670	Volcanic	293	5.9	47	138	9
Nemi	Lazio	41.720, 12.720	Volcanic	316	1.7	39	200	10
Posta Fibreno	Lazio	41.701, 13.687	Karst	288	0.3	71	204	14
Giulianello	Lazio	41.690, 12.849	Volcanic	235	0.12	26	51	4
Vescovo	Lazio	41.455, 13.123	Karst	30	0.05	15	30	5
Sette Cannelle	Lazio	41.368, 13.422	Karst	115	0.04	11	14	3
Matese	Campania	41.414, 14.392	Karst	1014	5.0	47	218	3
Astroni	Campania	40.842, 14.149	Volcanic	116	0.3	15	32	7
Averno	Campania	40.839, 14.076	Volcanic	2	0.55	13	22	7
Laceno	Campania	40.807, 15.096	Karst	1050	0.2	5	5	0
Monticchio	Basilicata	40.931, 15.605	Volcanic	823	0.54	71	307	9
Laudemio	Basilicata	40.143, 15.837	Glacial	1525	0.02	5	9	2
Scanno	Abruzzo	41.921, 13.863	Landslide	922	0.93	4	5	2
Pantaniello	Abruzzo	41.813, 13.987	Glacial	1818	0.03	8	8	2
Vivo	Abruzzo	41.727, 13.952	Glacial	1591	0.05	3	3	2

ed from its surface area, in order to represent the maximum distance between the assigned point and any potential original observation site along the lake perimeter. This approach was applied consistently to all those plant records whose original source did not provide more precise spatial information. For large lakes (e.g., Lake Bolsena and Lake Trasimeno), this resulted in spatial uncertainty values of several kilometers.

DataLake structure

DataLake is organized as a tabular database developed in Microsoft Excel (Office 365), in which each row represents a single georeferenced record of a vascular plant taxon at a given lake. In this context, a record is defined as a single georeferenced report of a plant taxon associated with metadata (e.g., source, date, spatial uncertainty). Records derive from both bibliographic sources and original field surveys and are integrated within a unified structure that preserves information on data provenance.

DataLake includes two main Excel sheets: record and metadata sheets. The "record sheet" contains floristic records accompanied by information related to taxonomic identity, spatial coordinates, collection time, record source and some characteristics on the taxa (life form, chorotype, Ellenberg indicator value for humidity, alien status, IUCN categories). The "metadata sheet" includes meaning, units and reference sources of all database fields reported in the "record sheet".

This database structure allows for the integration of floristic data recorded at different times and with different spatial accuracy within the same coherent information framework. It also supports the possibility of filtering, aggregating and analyzing data at multiple levels. The full matrix with all collected data is deposited in the Zenodo repository (10.5281/zenodo.18630469) (Pinzani *et al.*, 2026).

RESULTS AND DISCUSSION

DataLake comprises 5140 georeferenced records of vascular plants recorded in 30 natural freshwater lakes of central-southern Italy (Fig. 1, Tab. 1). Of these, 767 records (15%) derive from unpublished field data collected by the authors, whereas the remaining part (85%) were extracted and georeferenced from 135 bibliographic references (*List S1*). The temporal distribution of records shows a marked increase in sampling intensity from the second half of the 20th century onwards, with a peak in the most recent decades (Fig. 2). Richness of taxa and records is unevenly distributed among the investigated lakes and, more clearly, among lake types. Volcanic lakes account for the highest number of records (n=2724), followed by tectonic lakes (n=1577), whereas karst (n=637) and alluvial lakes (n=168) contribute substantially fewer records. Glacial (n=29) and landslide-dammed lakes (n=5) are represented by very limited data overall (Tab. 1). This pattern primarily reflects differences in data availability among lake origins. In addition, records for tectonic lakes are largely concentrated

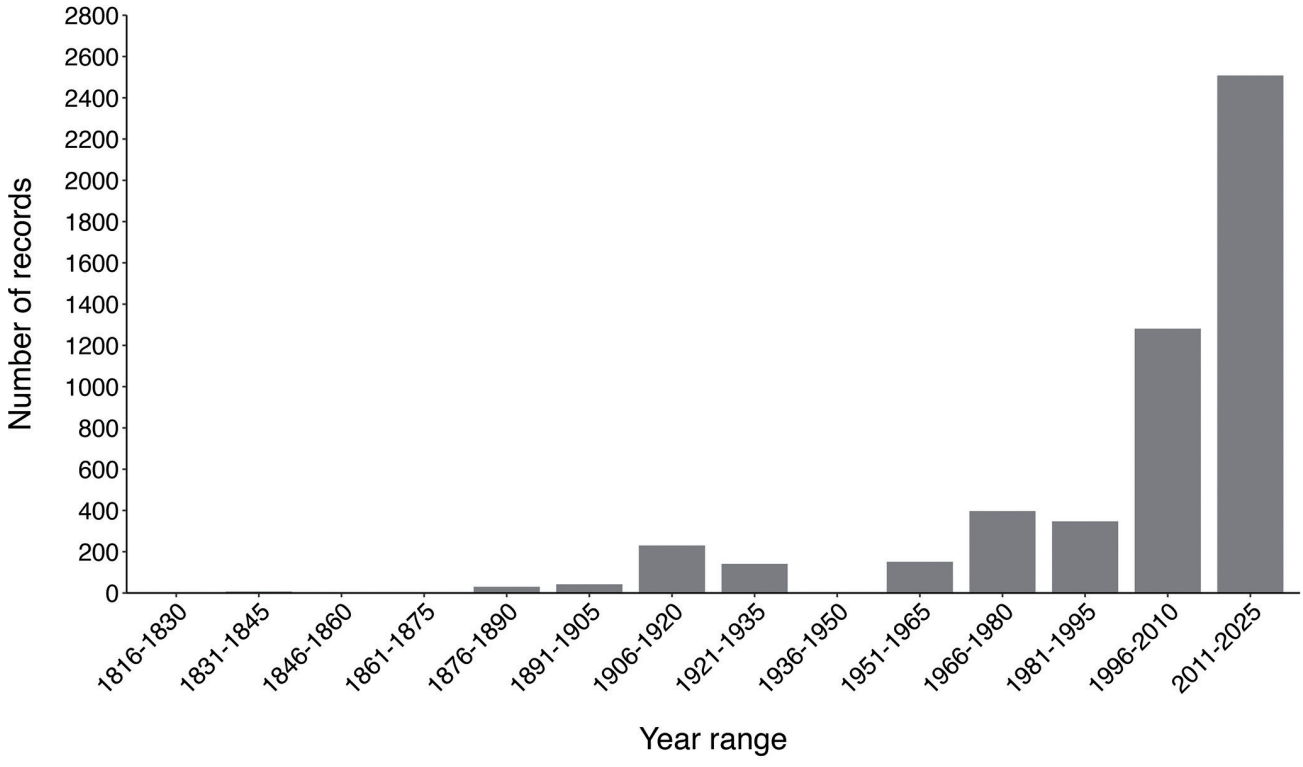

Fig. 2. Temporal distribution of records relating to vascular plants included in *DataLake*. Records are grouped by 15-year intervals from the first reports in the 19th century to the current year. The bars indicate the number of georeferenced records for each temporal range.

in a single major basin (Lake Trasimeno), while volcanic lakes are represented by multiple water bodies with substantial data coverage (Tab. 1, Fig. S1). Consequently, patterns observed at the lake-type level should be interpreted with caution, given the uneven representation of records across geological origins.

The spatial accuracy of the records is high overall. In fact, of the 5140 records included in *DataLake*, 2907 records (56.5%) are associated with a low spatial uncertainty of 100 m, based on GPS data or very detailed information on the floristic sampling site. Records with intermediate spatial accuracy include 327 records (6.3%) with an uncertainty of 2000 m, generally corresponding to sites defined on a sub-lake or municipal scale. Less precise records were primarily associated with generic lake-level locality descriptions. The least accurate records, falling into the 5000 m class and above, count 1050 records (20.4%), and are mainly associated with generic descriptions of sites around the largest lakes.

The 5140 georeferenced records included in *DataLake* correspond to 213 taxa, 97 genera and 43 families. Consistent with the aquatic and riparian environments considered, the taxonomic composition is dominated by families typical of these habitats, with Potamogetonaceae, Cyperaceae, Poaceae, Juncaceae, Hydrocharitaceae, Haloragaceae and Salicaceae accounting for the highest number of taxa and records (Fig. 3). Although dominance was defined at the overall dataset level, the contribution of the most representative families varies across lake types (Fig. S1). In tectonic lakes, record richness is strongly concentrated in a limited number of dominant families, largely reflecting the weight of a single basin. In contrast, volcanic lakes show a more even distribution of records among the dominant families across multiple systems. Karst and alluvial lakes display intermediate patterns, whereas glacial and landslide-dammed lakes, due to their very low record numbers, do not allow robust family-level comparisons.

At the genus level, *Potamogeton*, *Myriophyllum* and *Ceratophyllum* are the dominant genera in the aquatic habitats, while *Juncus* and *Carex* along the riparian ones. At the species level, *Myriophyllum spicatum*, *Ceratophyllum demersum*, *Phragmites australis*, *Potamogeton perfoliatus* and *Stuckenia pectinata* are the most frequently recorded species.

From a conservation perspective, the dataset is particularly relevant. A total of 84 taxa (39% of the dataset) are included in regional risk categories according to Conti *et al*. (1997), corresponding to 1366 records (27%). In addition, 23 taxa (10.6% of the total) are listed in national IUCN risk categories according to Orsenigo *et al*. (2020), accounting for 268 records (Tab. S1). These results highlight the vulnerability of the aquatic and riparian flora associated with Italian freshwater lakes and emphasize the role of lacustrine ecosystems as important reservoirs for plant species of conservation concern. By documenting the spatial distribution and frequency of threatened taxa across multiple lake systems, *DataLake* provides a valuable base for planning conservation actions, monitoring activities and future assessments of floristic changes.

Alien species were recorded in 14 lakes (47%) and account for 19 taxa (8.9% of the total flora), represented by 249 records (4.8%). Among these, 12 taxa are classified as invasive, including *Paspalum distichum*, *Elodea canadensis* and *Bidens frondosa*. Their presence across multiple lake systems underlines the relevance of the dataset for documenting biological invasions in freshwater environments and for supporting long-term monitoring activities.

The biological spectrum is mainly characterized by hydrophytes and helophytes, with hemicryptophytes also contributing substantially, while therophytes, geophytes and phanero-

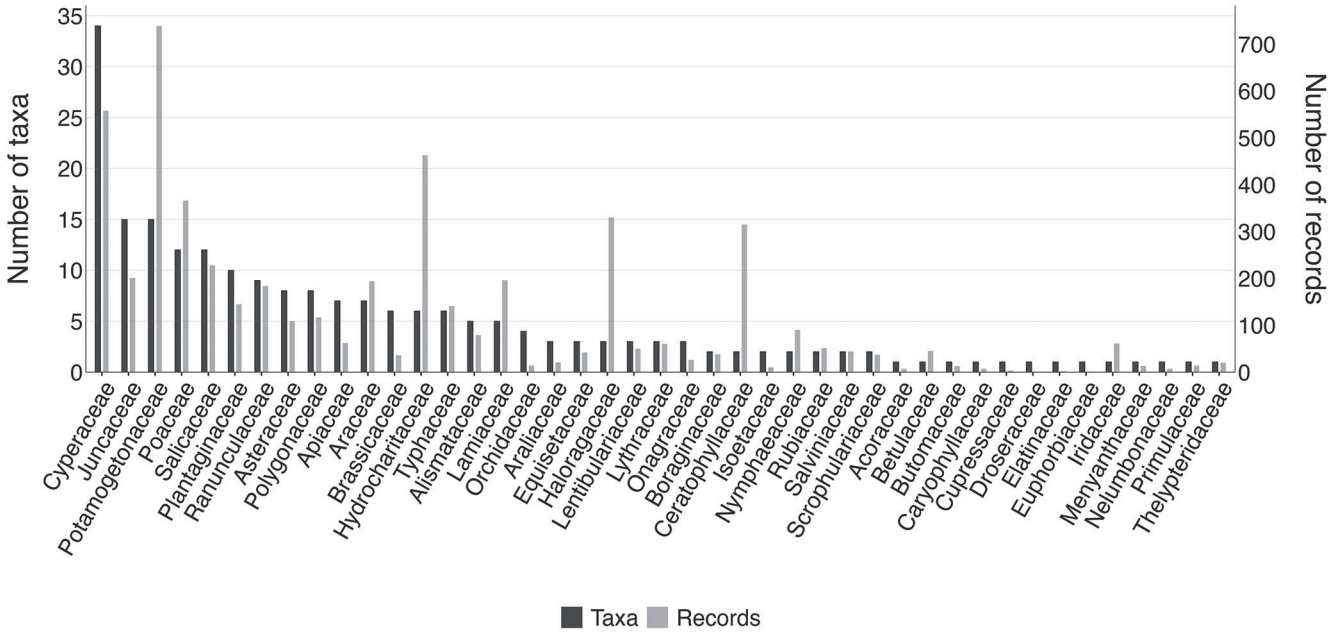

Fig. 3. Distribution of taxa and records among vascular plant families in *DataLake*. Families represented by very low numbers of records may appear absent at the adopted graphical scale due to their minimal contribution.

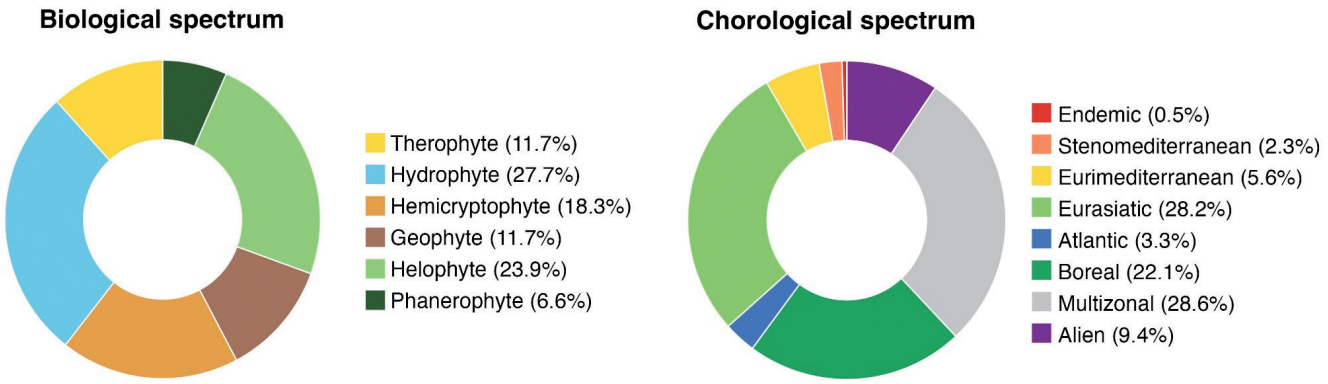

Fig. 4. Biological and chorological spectra of plant taxa included in *DataLake*.

phytes occur less frequently (Fig. 4). The chorological spectrum is dominated by Multizonal, Eurasiatic and Boreal taxa, followed by Aliens, Euromediterraneans, Atlantics, Stenomediterraneans and Endemics (Fig. 4). The observed biological and chorological patterns are consistent with the strong ecological dependence of aquatic and riparian plant species on hydrological conditions rather than on regional climatic constraints. In freshwater lake environments, as commonly observed in other inland water bodies, the buffering effect of water availability tends to reduce the influence of macroclimatic factors, favoring the prevalence of multizonal taxa over strictly Mediterranean ones (Wetzel, 2001; Bornette and Puijalon, 2011).

CONCLUSIONS

DataLake provides a georeferenced dataset on aquatic and riparian vascular plants from 30 natural freshwater lakes in central-southern Italy. Thanks to the comprehensive floristic data available in *DataLake*, it was possible to document how the aquatic and riparian flora of these lakes is rich in species but also vulnerable, as evidenced by a significant percentage of species of conservation interest, as well as of invasive alien species that pose a real local threat to the conservation of the native plant biodiversity.

It should be noted that the importance of this digital database lies in summarizing in a unique, coherent, standardized and easily accessible resource, floristic data from heterogeneous sources across time and space for the lakes considered.

Despite differences in the representativeness of the various types of lakes investigated, the dataset provides a comprehensive and well-documented overview of the diversity of vascular plants associated with natural Mediterranean freshwater lakes in central and southern Italy. Therefore, *DataLake* can provide a solid reference base for future floristic research, as well as for supporting long-term monitoring activities and actions aimed at conserving natural lake ecosystems of the Mediterranean.

REFERENCES

Bartolucci F, Peruzzi L, Galasso G, Alessandrini A, Ardenghi NMG, Bacchetta G, et al., 2024. A second update to the checklist of the vascular flora native to Italy. Plant Biosyst 158:219-296.

Bornette G, Puijalon S, 2011. Response of aquatic plants to abiotic factors: a review. Aquati Sci 73:1-14.

Cannucci S, Bolpagni R, Bonari G, Candini F, Dalla Vecchia A, Fanfarillo E, et al., 2025. Dive into the Italian PONDY dataset: pond vegetation data and water physico-chemical parameters. Veg Ecol Divers 62:e176891.

Conti F, Manzi A, Pedrotti F, 1997. [Liste Rosse Regionali delle Piante d'Italia] [in Italian]. Camerino, WWF Italia, Società Botanica Italiana.

Ellenberg H, Weber HE, Düll R, Wirth V, Werner W, Paulissen D, 1992. [Zeigerwerte von Pflanzen in Mitteleuropa] [Book in German]. Scripta Geobotanica 18. Göttingen, Erich Goltze Verlag.

Galasso G, Conti F, Peruzzi L, Alessandrini A, Ardenghi NMG, Bacchetta G, et al., 2024. A second update to the checklist of the vascular flora alien to Italy. Plant Biosyst 158:297-340.

Guarino R, La Rosa M, 2019. [Flora d'Italia digitale].[in Italian]. In: Pignatti S, Guarino R, La Rosa M (eds.), Flora d'Italia. Bologna, Edagricole.

Jeppesen E, Moss B, Bennion H, Carvalho L, DeMeester L, Feuchtmayr H, et al., 2010. Interaction of climate change and eutrophication, p. 119-151. In: Kernan M, Battarbee R, Moss B (eds.), Climate change impacts on freshwater ecosystems. Oxford, Wiley-Blackwell.

Kalff J, 2002. Limnology: inland water ecosystems. Upper Saddle River, Prentice Hall.

Orsenigo S, Fenu G, Gargano D, Montagnani C, Abeli T, Alessandrini A, et al., 2020. Red list of threatened vascular plants in Italy. Plant Biosyst 155:310-335.

Pignatti S, 2017-2019. [Flora d'Italia] [in Italian]. Vol. 1-3. Bologna, Edagricole.

Pinzani L, Di Lernia D, Pelella E, Ceschin S, 2025a. The vascular flora of Italian volcanic lake calderas: a comprehensive floristic study. Environments 12:327.

Pinzani L, Di Lernia D, Ceschin S, 2026. DataLake: a georeferenced dataset of vascular plants from freshwater lakes of central-southern Italy [Dataset]. Zenodo 18630469.

Pinzani L, Pelella E, Azzella MM, Ceschin S, 2025b. A bibliographic review on vascular flora of Italian volcanic lakes. Inland Waters 15:2475684.

Portal to the Flora of Italy, 2025. Version 2025.2. Accessed: 1 Dec 2025. Available from: https://dryades.units.it/floritaly/

Rossi G, Montagnani C, Gargano D, Peruzzi L, Abeli T, Ravera S, et al., 2013. [Lista Rossa della Flora Italiana. 1. Policy Species e altre specie minacciate].[in Italian]. Rome, Comitato Italiano IUCN e Ministero dell'Ambiente e della Tutela del Territorio e del Mare.

Wetzel RG, 2001. Limnology: lake and river ecosystems. San Diego, Academic Press.

Online supplementary material:

List S1. Complete list of bibliographic references consulted for the extraction and georeferencing of records included in DataLake.

Fig. S1. Percentage of records for the ten most represented vascular plant families in DataLake.

Tab. S1. Vascular plant taxa in DataLake included in national and/or regional Red Lists.

Received: 3 January 2026; Accepted: 26 February 2026.

Contributions: Lorenzo Pinzani, conceptualization, data curation, formal analysis, investigation, methodology, visualization, writing - original draft. Dario Di Lernia, investigation, data curation. Simona Ceschin, conceptualization, supervision, investigation, writing - review & editing. All authors read and approved the final version of the manuscript.

Conflict of interest: the authors declare no conflict of interest.

Data availability: DataLake: a georeferenced dataset of vascular plants from freshwater lakes of central-southern Italy is openly available on Zenodo at: https://doi.org/10.5281/zenodo.18630469

Acknowledgements: the authors acknowledge the support of NBFC to Department of Science-University of Roma Tre, funded by the Italian Ministry of University and Research, PNRR, Missione 4 Componente 2, "Dalla Ricerca all'Impresa", Investimento 1.4, Project CN00000033.