Phytoplankton morpho-functional trait dataset from French water-bodies

Christophe Laplace-Treyture1,2✉, Jonathan Derot1,3, Emilie Prévost1, Anne Le Mat1 & Aurélien Jamoneau1,2

In lake ecosystems, phytoplankton communities can be studied by adopting taxonomic-based approaches. However, these approaches suffer from identification issues and are sometimes of limited ecosystem ecological value. The recent development of functional approaches may allow an evaluation of other aspects of ecosystem quality, functions and interactions with abiotic parameters or other communities. Here, our aim was to create a phytoplankton trait database at the French scale. This database will be relevant for the analysis of phytoplankton communities that will lead to a better understanding of phytoplankton functional ecology in lakes of France and other European countries possessing similar biological communities. We used a French national database of phytoplankton occurrences sampled from 384 lakes over the entire French metropolitan territory. A final list of 636 taxa was used to compile 53 morpho-functional traits associated with taxonomic information. The traits encompassed variables such as shape, biovolume, motility, toxin production and Reynolds groups. With this new database, we aim to provide data for morpho-functional analyses of phytoplankton assemblages from local to European scale.

Background & Summary
In water bodies, phytoplankton communities are key elements of ecosystems, regulating biogeochemical cycles and representing one of the most important photoautotrophic primary producers. Because of their rapid responses to water quality and anthropogenic pressures, phytoplankton communities are commonly used as trophic and ecological indicators in Europe and in the United States. Therefore, they are mainly studied using a classical taxonomic-based approach, i.e. abundance and biovolume of species. Taxonomic lists are used as a whole, with the sum of individual biovolume employed for total biomass metric assessment and at the species level for selecting indicator species. This selection is usually based upon the most common species of the phytoplankton community, for which information and ecological profiles can be established. Therefore, the assessment is only based on a part of the community. Rare species are not considered in such cases although they can be very informative (e.g. new invasive species with very specific response to a trophic or a polluted state). Data can be aggregated at higher taxonomic level (family or phylum) to take into account all available taxa. However, different taxa of the same phylum can indicate a very different ecological status while species from different phylum can reflect very similar ecological conditions. Moreover, the ecological significance can be neutralized at the phylum level by species indicative of opposite conditions.

During the last decades, approaches based on functional groups were created to overcome these drawbacks and allow evaluating other aspects of ecosystem quality, functioning and biotic interactions. As outlined in Salmaso et al. many studies use functional classifications and/or functional groups with the objective to better investigate links between phytoplankton communities and the main environmental factors. Some classifications mentioned by Salmaso et al. are based on one characteristic (size classes) or several (morphologically based classifications) while others use functional properties (functional or morpho-functional groups). Indices based on functional groups were also published for lakes and for rivers to be used for the implementation of Water Framework Directive. Nevertheless, as explained by Padišák et al. functional groups were not built using all encountered taxa. Ecologists have to adjust their groups manually or through an R code for the

1INRAE, UR EABX, 50 avenue de Verdun, F-33612, Cestas, Cedex, France. 2Pôle R&D Ecosystèmes Lacustres (ECLA), F-13100, Aix-en-Provence, France. 3Estuary Research Center, Shimane University, 1060 Nishikawatsu-cho, Matsue, Shimane, 690-8504, Japan. ✉ e-mail: christophe.laplace-treyture@inrae.fr
Reynolds Functional Groups\textsuperscript{23}. These additions, may not be adequate and lead to misinterpretations. According to Wentzky et al.\textsuperscript{24}, the use of functional-trait at the species level (and not at higher taxonomical level) can limit these issues without influencing any assumption about taxon functional group.

Thus, data collections of phytoplankton traits at the species level are required for a large number of taxa. Accordingly, different datasets were created considering the taxa size. Olenina et al.\textsuperscript{25} reported size-classes (PEG biovolume, http://ices.dk/data/vocabularies/Documents/Forms/AllItems.aspx) but mainly for marine waters. Kremer et al.\textsuperscript{26} established a dataset for freshwater and estuaries for a large number of taxa (>1200 species), which had the advantage that it considered cell size and natural unit size (colonial and filamentous growth forms). At the French level, Druart and Rimet created a trait dataset including only biovolumes\textsuperscript{27} and published, more recently, a phytoplankton trait dataset\textsuperscript{28} mainly based on Alpine lake surveys and including around 10 compiled traits.

To our knowledge, specific traits such as lorica, protuberance presence (and kind of protuberance), potential toxicity associated with potential toxin production by cyanobacteria, heterocyte and vacuole presence have never been merged together in one trait database. More generally, to our knowledge, no database covers a large number of ecological and morphological traits for phytoplankton communities at the French or European level. To be able to use many morpho-functional traits for water bodies in ecological analyses in France, the construction of such a dataset was initiated. This was based upon the survey database established for the WFD application on French water bodies since 2005. This dataset will be particularly helpful for studying the relationships between functional traits and environmental variables, as well as to develop new trait-based indicators for water quality assessment.

We aim in this paper to (1) compile morpho-functional traits of phytoplankton taxa in French water bodies (metropolitan territory) (2) offer information of the corresponding phylogeny (3) provide a dataset helpful for analyzing functional and morpho-functional phytoplankton assemblages from local to national scale (4) deliver a dataset allowing consistent data comparisons between European countries.

Methods

Site description. This work was based on the French national database consisting of phytoplankton communities surveyed in 384 lakes between 2005 and 2016 and used for applying the Water Framework Directive. As each lake has been sampled several times this database contains 2,987 phytoplankton samples. Sampled lakes cover the whole French territory (Fig. 1) and encompass all natural and artificial types of lakes encountered in France: from lowlands to alpine location, from small to large lakes (0.09 to 577.12 km\textsuperscript{2}), from shallow to deep lakes (mean depth from 0.30 to 153 meters) and also from oligotrophic to hypereutrophic environmental conditions.
Data acquisition and compilation. In order to create our dataset we firstly defined the taxa list recorded in the national database. The extracted list was firstly examined for taxonomic consistency. Taxa only determined at a high taxonomic level (order, class, phylum) were rejected from the list. Thus, genus and below this level (species, variety, form, etc.) were conserved to obtain an accurate taxonomical list for morpho-functional traits (e.g. biovolume, cell form). A taxa name in the national database can be different over time because of taxonomic changes and determination by different operators. To harmonize taxa denomination and to obtain one up-to-date taxa list, the original names used in the lake database were corrected following the most recent research reported on the AlgaeBase website (https://www.algaebase.org/ - last consultation on 31st July 2020) and in scientific papers. This will allow the traits table to be used for future studies. Thus, a list of 1,318 taxa, associated with abundance and biovolume data, was obtained from the database. In order to remove idiosyncratic species and any potential determination errors, we only kept taxa which occurred in more than 15 samples. A final taxa list of 636 taxa was obtained.

Secondly, we selected a list of 53 morpho-functional traits that we considered to be the most relevant for the description/analyses of functional phytoplankton communities (see Online-only Table 1 for detail of these traits). We compiled available trait data from published sources: we used general books (John et al.25,29; Reynolds3; Wehr et al.30) for all taxa and specific references listed in Table 1 for more precise information of some traits or taxa. The traits not available in the literature (biovolume, size class, carbon contents, etc.) were determined/calculated. All available synonyms of the taxa recorded in the literature were taken into account in order to consider morpho-functional traits taken from as a many taxonomic references as possible. Taxonomic information was essentially obtained from the AlgaeBase website.

From these publications we mainly obtain morphological characteristics such as cell form, presence of ornamentation on cell and number of chloroplasts. If the information was divergent between two references, the most recent is accepted.

To facilitate the use of the dataset in statistical analyses, taxa code (6 letters in upper case), used in the free computer software Phytobs31 (commonly used in France for freshwater phytoplankton counting), was added for each taxon. These codes were constructed with the first 3 letters for genus and the last 3 for species and infra specific denomination. In order to have a global clustering of phytoplankton taxa (sometimes useful for result presentations), more or less independent of phylogeny updates and thus stable in time, we included a phytoplankton group name. We classified phytoplankton taxa in 9 groups based on common clustering of algae and cyanobacteria such as chlorophytes (green-algae), diatoms, xanthophytes (brown-algae), chrysophytes (gold-algae). The trait life form was added to the dataset and is defined as the common living form of the taxon in its natural environment (cell, colony or filament). Because some taxa could exhibit a second life form in nature, this information was also indicated in the dataset.

We determined the cell and individual form traits for each taxon with the use of 14 simple geometric forms: sphere, cylinder, rotational ellipsoid, flattened ellipsoid, cymbelloid, staurastrum and ceratium form, oval cylinder, double cone, parallelepiped, prism on parallelogram base and on triangular base, cone with half sphere and tetrahedron as defined in the computer tool Phytobs31 and illustrated in Fig. 2. Choices were made with the help of book taxa illustrations and according to Hillebrand et al.32, Olenina et al.33, Druart and Rimet27, Hutorowicz34 and Padišak and Adrian35.

The minimum and maximum cell dimensions were taken from the literature and were used for the definition of the size class (see below).

We determined taxa biovolume following the European standard NF EN 1669535 with the use of the software Phytobs31 for the calculation (this counting software provides also a biovolume calculation tool). Thus, from real samples observed during the period of the study, 30 individuals of each taxon were measured, as recommended by the European standard, to calculate a mean cell biovolume. For taxa not observed in enough quantity, common minimum and maximum dimensions were taken from the literature in order to determine mean dimensions and calculate a corresponding mean cell biovolume.

We also specified the mean carbon content per cell in our dataset. We used the formulas described in Menden-Deuer and Lessard36 and reproduced in European standard NF EN 1669535 because they consider the decrease in specific carbon content with cell size. Thus, the general formula used for phytoplankton is Carbon [pg C] = 0.216 × Biovolume0.811 and the specific formula for diatom taxa is Carbon [pg C] = 0.288 × Biovolume0.828.

Three size class systems were also subsequently defined in the dataset. The first two use the Sieburth et al.37 class limits and are based on the cell dimensions. The third one uses the biovolume thresholds defined by

| Phylum          | References                                                                 |
|-----------------|---------------------------------------------------------------------------|
| chlorophytes    | Huber Pestalozzi and Thienemann52; Komárek et al.37 and for desmids group: Coesel and Meesters34,35 |
| xanthophytes    | Starmach38                                                                  |
| cyanobacteria   | Komárek and Anagnostidis76,77; Komárek78 and for the toxicity: Whitton ed.79; Dittman et al.80,81; Sanseverino et al.82,83; Meriläinen et al.84; Annes85 |
| diatoms         | Bey and Ector86; Core87; Druart and Straub88; Houk89; Houk and Klee90; Houk et al.91,92; Krammer and Lange-Bertalot93,94; Lange-Bertalot et al.95; Siver et al.96 |
| dinoflagellates | Popovsky and Pfiester97; Meenstrup and Calado98                             |
| euglenophytes   | Huber-Pestalozzi31                                                          |
| chlorophytes    | Ettl3; Rieth98                                                               |

Table 1. Specific books and references used for some traits and taxa for the different phyla.
Ignatiades. For Sieburth et al. class limits, two variables were defined: one using the minimal individual length and the second using the maximal individual length. In each case, taxa were classified within the following classes: picophytoplankton (0.2–2 µm), nanophytoplankton (2–20 µm), microphytoplankton (20–200 µm), mesophytoplankton (0.2 mm–20 mm). The Ignatiades classification was calculated on individual biovolume: cell biovolume for unicellular taxa and individual biovolume for colonial and filamentous taxa. The classes defined by Ignatiades were: nanoplankton (10–1,000 µm³), microplankton (1,000–1,000,000 µm³), macroplankton (> 1,000,000 µm³) plus a specific class, picoplankton (< 10 µm³), used for the very small taxa.

Different binary traits (1-presence; 0-absence) were added: motility, presence of flagella, aerotopes, contractile vacuoles, mucilage, akinetes, heterocytes, cysts, plasts, sheath, tractus, siliceous skeleton, loria, external plates, scales, ornamentations, protuberances, chlorophyll-b and c, xanthophyll, phycobilin and potential toxin production (with toxin family specified: microcystin, anatoxin-a, anatoxin-a(S), cylindrospermopsin, saxitoxin, Beta-Methylamino-L-Alanin). These characteristics were mainly extracted from the literature. The precise number of plastids and flagella were also added. Some distinctions were defined for the protuberance type, their size and their number onto the individuals. The protuberances were split into seven different types including granule, needle or bristle and the size was divided in two size classes (large and small).

The main nutrition mode was indicated for each taxon following the common definitions: autotroph, mixotroph and heterotroph.

The reproduction mode was specified more precisely when mentioned in the literature or just indicated as sexual or asexual when not.

Some taxa are known only from freshwater and others also from the marine environment. Thus, this information was indicated as a Water_Type in the trait dataset. The trophic status of the water where the taxon lives was
defined under "Water_Trophy" as from oligotrophy to hypereutrophy. In the case of a taxon with a large trophic
range, the main trophic status was indicated.

In some literature and mainly in Reynolds3, the tolerance to abiotic parameters are defined. We also indicated
this trait as a tolerance trait in the dataset (literal form).

We finally reported in the dataset the Reynolds functional groups, based on the original description of
Reynolds3 and developed further (with more taxa) by in Borics20 and Padisak et al.19,22.

**Data Records**

The phytoplankton morpho-functional traits dataset46 is in the form of a large table with taxa in rows and
their corresponding morpho-functional traits in columns. This trait table is contained in a formatted file with
semicolon-separated values (*.csv) and stored in a public repository available on https://data.inrae.fr/dataverse/eabx,
with the following DOI: https://doi.org/10.15454/GJGIAH. The name of the file was defined as FRENCH_
PHYTOPLANKTON_TRAITS.csv. This trait table is designed to be updated with new data especially from
French lakes (new surveys and new taxa occurrences) and/or new morpho-functional traits relevant for the study
of phytoplankton communities.

In the table some variables are numerical and continuous such as biovolumes or cell dimensions, other are
categorical such as life forms and other boolean as toxin production. Consequently, the table is a mixed format
with numeric and text. The different variables are listed with their attributes in Online-only Table 1 to facilitate
understanding and reuse of the data. Missing information was labelled "#NA".

In the dataset, the chlorophytes group is the most represented with 36.5% of the total (Fig. 3) and cryptophytes
is the least (2.2%). Other groups like cyanobacteria, diatoms and chrysophytes represent between 10 to 20% of the
dataset taxa. The individual life form of each taxon listed is mainly represented by the cell form (56.8%), followed
by the colonial form (32.9%) and, finally, the filamentous form of around 10% (Fig. 4).

**Technical Validation**

The technical validation of the data stored in the dataset was undertaken following the scheme outlined in Fig. 5.
All the information collected from the literature was double-checked and all sources are provided in the methods
section above. All measures (biovolume, carbon content, size classes attribution, etc.) were made by the same
individual in order to reduce error.

After that, all data included in the traits table were controlled by the same person (C.L.T.), an expert on
French phytoplankton. The dataset thus created was explored through preliminary analyses with box-plots and
histograms in order to check distribution and to highlight erroneous data (outliers, aberrant text, missing data, etc.). Many cross-checks were undertaken in order to highlight missing and inconsistent data (C.L.T.):

- motility set to “1” when aerotope, contractile vacuole or flagellum are present;
- number of flagella indicated with presence of flagellum;
- absence of flagellum for desmids, diatoms and cyanobacteria;
- compulsory flagellum for cryptophytes;
- number of plastids indicated with the presence of plastids;
- absence of plastids for cyanobacteria taxa;
- presence of akinete and heterocyte only for cyanobacteria taxa;
- compulsory cyst for chrysophytes and dinophytes taxa;
- siliceous skeleton only for diatoms;
- presence of external plates only for dinophytes;
- presence of lorica for chrysophytes and few chlorophytes and euglenophytes;

Fig. 5 Scheme for the technical validation of the dataset. Extraction of the taxa list from the national phytoplankton database, simplification with the synonymy stored in Phytobs software. Implementation of the traits by 3 different operators to obtain a beta version of the dataset, then double-check step and implementation of the calculated traits to obtain a first version of the dataset. Quality control of the data by the referent scientist through the use of graphs and cross-checks to obtain the final version of the dataset.
- protubercule type other than "None" and protuberance size not null with presence of protuberance;
- presence of chlorophyll-b and xanthophyll and also absence of chlorophyll-c and phycobilin for chlorophytes (equivalent cross-controls for the other phytoplankton groups were done);
- presence of toxin only for some cyanobacteria;
- the toxin family indicated with presence of toxin.

Usage Notes
The phytoplankton trait dataset is freely available and stored in https://data.inrae.fr/dataverse/eabx public repository and should be appropriately referenced by citing the present paper. A survey taxa list should be first checked for synonymy before being analyzed with the morpho-functional dataset. This will avoid the mismatch of taxa names. We strongly encourage users to use the synonymy database present in the Phytobs software.

Note that many traits were obtained from literature indicating general characteristics of species or genera and do not represent individual responses to environmental conditions. Some of them can be affected by different environmental conditions. For example, traits based on dimensions (min, max length and width), biovolume, carbon content and in some cases size-classes can be directly influenced by environmental conditions as well as the presence of heterocyte and akinete for some cyanobacteria taxa.

Code availability
The software Phytobs, used for some trait compilation or calculation (biovolumes), was created by our team in a first version in 2009 by Hadoux and Laplace-Treytoure for phytoplankton counting purposes. It was upgraded over the years to allow biovolume calculations and to integrate taxonomy and synonymy. This free software is now in version 3.2 (French, English and Spanish languages) and publicly available on https://hydrobio-dce.inrae.fr/phytobs-software/.

Received: 5 August 2020; Accepted: 17 December 2020;
Published online: 02 February 2021

References
1. Litchman, E. et al. Global biogeochemical impacts of phytoplankton: a trait-based perspective. J. Ecol. 103, 1384–1396 (2015).
2. De Senerpont Domis, L. N. et al. Plankton dynamics under different climatic conditions in space and time. Freshw. Biol. 58, 463–482 (2013).
3. Reynolds, C. Ecology of Phytoplankton. (Cambridge University Press, 2006).
4. Phillips, G. et al. Water Framework Directive Intercalibration: Central Baltic Lake Phytoplankton Ecological Assessment Methods. 189 (Jönköping Research Center, 2014).
5. Ptcnik, R., Solimini, A. & Brettum, P. Performance of a new phytoplankton composition metric along a eutrophication gradient in Nordic lakes. Hydrobiologia 633, 75–82 (2009).
6. Pollard, A. L. Hampton, S. E. & Leech, D. M. The promise and potential of continental-scale limnology using the U.S. Environmental Protection Agency’s National Lakes. Assessment. Limnol. Oceanogr. Bull. 27, 36–41 (2018).
7. de Hoyos, C. et al. Water Framework Directive Intercalibration: Mediterranean Lake Phytoplankton Ecological Assessment Methods. 189 (Jönköping Research Center, 2014).
8. Mischke, U., Riedmüller, U., Hoehn, E., Schönfelder, I. & Nixdorf, B. Description of the German System for Phytoplankton-Based Assessment of Lakes for Implementation of the EU Water Framework Directive (WFD). 31 (Univ. Cottbus, 2008).
9. Laplace-Treytoure, C. & Feret, T. Performance of the Phytoplankton Index for Lakes (IPLAC): A multimetric phytoplankton index to assess the ecological status of water bodies in France. Ecol. Indic. 69, 686–698 (2016).
10. Xue, Y. et al. Distinct patterns and processes of abundant and rare eukaryotic plankton communities following a reservoir cyanobacterial bloom. ISME J. 12, 2263–2277 (2018).
11. Barbe, J. et al. Actualisation de la Méthode de Diagnostic Rapide des Plans d’Eau: Analyse Critique des Indices de Qualité des Lacs et Propositions d’Indices de Fonctionnement de l’Écosystème Lacustre. 107 (Cemagref, 2003).
12. Marchetto, A., Padella, B., Mariani, M., Luglie, A. & Sechi, N. A numerical index for evaluating phytoplankton response to changes in nutrient levels in deep mediterranean reservoirs. J. Limnol. 68, 106–121 (2009).
13. Kruk, C., Mazzeo, N., Lacerot, G. & Reynolds, C. S. Classification schemes for phytoplankton: A local validation of a functional approach to the analysis of species temporal replacement. J. Plankton Res. 24, 901–912 (2002).
14. Reynolds, C. S. Phytoplankton designer – or how to predict compositional responses to trophic-state change. Hydrobiologia 424, 123–132 (2000).
15. Reynolds, C. S., Huszar, V., Kruk, C., Naselli-Flores, L. & Melo, S. Towards a functional classification of the freshwater phytoplankton. J. Plankton Res. 24, 417–428 (2002).
16. Mileitner, J. & Reichert, P. Modelling functional groups of phytoplankton in three lakes of different trophic state. Ecol. Model. 211, 279–291 (2008).
17. Rangel, L. M., Soares, M. C. S., Paiva, R. & Silva, L. H. S. Morphology-based functional groups as effective indicators of phytoplankton dynamics in a tropical cyanobacteria-dominated transitional river–reservoir system. Ecol. Indic. 64, 217–227 (2016).
18. Salmaso, N., Naselli-Flores, L. & Padasik, J. Functional classifications and their application in phytoplankton ecology. Freshw. Biol. 60, 603–619 (2015).
19. Padasik, J., Borics, G., Grigorszky, I. & Sórocázi-Pintér, É. Use of phytoplankton assemblages for monitoring ecological status of lakes within the water framework directive: The assemblage index. Hydrobiologia 553, 1–14 (2006).
20. Borics, G. et al. A new evaluation technique of potamo-plankton for the assessment of the ecological status of rivers. Large Rivers 17, 465–486 (2007).
21. European Parliament. Directive 2000/60/CE du Parlement Européen et du Conseil du 23 Octobre 2000 Établissant un Cadre pour une Politique Communautaire dans le Domaine de l’Eau. 72 (Communauté Européenne, 2000).
22. Padasik, J., Cressetti, L. O. & Naselli-Flores, L. Use and misuse in the application of the phytoplankton functional classification: A critical review with updates. Hydrobiologia 621, 1–19 (2009).
23. Kruk, C. et al. Classification of Reynolds phytoplankton functional groups using individual traits and machine learning techniques. Freshw. Biol. 62, 1681–1692 (2017).
24. Wentzyk, V. C., Tittel, J., Jäger, C. G., Bruggeman, J. & Rinke, K. Seasonal succession of functional traits in phytoplankton communities and their interaction with trophic state. J. Ecol. 108, 1649–1663 (2020).
Acknowledgements
We would like to thank the six water agencies and private labs for participating in the collection of biotic data. We are also grateful to the lake team of Irstea Aix-en-Provence for the lake database management. Many thanks are going to one of our team engineer, S. Boutry, for the map in Fig. 1. We kindly thank the anonymous referees for their useful comments on the previous version of this manuscript. We are also thankful to D.M. John for the English edits. This work was supported by the French Ministry of the ecological and solidarity transition (MTES) and funded by the “Office Français de la Biodiversité” (OFB) through the convention between INRAE and OFB between 2016 and 2018. It was also funded by the Labex COTE “Continental To coastal Ecosystems: evolution, adaptability and governance” within the Cluster of Excellence COTE (ANR-10-LABX-45) and the Adour-Garonne water agency for some personnel costs.

Author contributions
Christophe Laplace-Treyture (C.L.T.) was the coordinator of the project and the dataset was assembled by Jonathan Derot (J.D.), Anne Le Mat (A.L.M.), Emilie Prévost (E.P.) and Christophe Laplace-Treyture. For the manuscript: conceptualization, methodology, formal illustration and analyses by C.L.T. and A.J.; draft preparation mainly by C.L.T., A.J. with the help of J.D., A.L.M. and E.P. All authors have read and agreed to the published version of the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to C.L.-T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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/.

The Creative Commons Public Domain Dedication waiver http://creativecommons.org/publicdomain/zero/1.0/ applies to the metadata files associated with this article.

© The Author(s) 2021