Supplemental Methods

Data collection and aggregation steps

On 1 July 2014 we began advertising (https://www.nceas.ucsb.edu/underice) the goals of the project and inviting broad collaboration by global researchers. We used an initial information-gathering survey to assess interest and characteristics of potential data. A total of 133 researchers expressed interest. The data collection template and data aggregation instructions were prepared in September 2014, informed by the initial survey responses and additional communications with respondents, and beta tested prior to release in November 2014 (Doc S1).

Observations were paired, with at least one 'pair' of observations - ice-on and ice-free - for each of the 135 variables, at a given lake, station, or year (Table S1; Fig. S1). There was one exception: we accepted "winter" data for Lake Erie in 2012, although the lake was almost entirely ice-free. These data were removed from all analyses, but retained for use in the public published dataset. We requested aggregates for the data fields, typically means or coefficients of variation, rather than researchers' raw data. Submitted aggregate data was assessed for completeness using a data validation R package created for this project (Woo 2015). Metadata fields were designed to provide information on lake- or season-specific factors that might contribute to variation in physical, chemical, and biological data.

When a lake had multiple sampling stations, the stations were generally treated independently. Exceptions were cases where researchers indicated two or more stations were functionally similar, and could be pooled in aggregate. After aggregation of these functionally similar stations, there were 135 stations and the majority of lakes (84 of the 101 lakes) were represented as a single sampling station. The remaining 17 lakes had two or more functionally distinct sampling stations. This means that stations separated by tens or hundreds of kilometers on the same large lake, or in distinct bays, were usually treated as independent systems in the analysis. There were 51 stations on these 17 lakes (38% of all stations), and most were located on large lakes such as the Laurentian Great Lakes.

In addition, we compiled under-ice data for primary producers using a literature survey. We searched the primary literature (between February 23 and June 23, 2015) using the search terms "lakes", "under-ice", "phytoplankton", "production", "winter", and "chlorophyll" in ISI Web of Science and Google Scholar. From the 14 papers we found with under-ice biological data that were comparable to those requested in the data aggregation template described above, we compiled data from 17 lakes (Fig. 1) taken extracted from text, tables, or from figures. using WebPlotDigitizer (http://arohatgi.info/WebPlotDigitizer/). For the literature review effort, we were unable to compare ice-on and ice-off data, as only 7 of the lakes in these papers also included biological data during the ice-off season (Cloern et al. 1992, Felip & Catalan 2000, Straskrabova et al. 2005, Lenard 2013). All compiled literature data are presented in Table S6).

Project collaboration and data sharing policies

We defined three data collation phases: 1) data integration, involving compilation of all submitted data into a master database, data quality assurance checking (submitted data available only to the core data management team, January 2015-April 2015); 2) analysis and group data sharing, including initial analyses, meetings, drafting manuscripts, and sub-project proposals (data available to all project collaborators via a password-protected website); and 3) manuscript submission and data sharing, during which initial manuscripts are submitted and data made publicly available [master database made public in the Knowledge Network for Biocomplexity, a member node of DataONE (Hampton et al. 2016]. All project collaborators agreed to the detailed data and coauthorship policy.

References

Cloern, J., Alpine, A., Cole, B. & Heller, T. (1992). Seasonal-Changes in the Spatial-Distribution of Phytoplankton in Small, Temperate-Zone Lakes. *J. Plankton Res.* 14, 1017–1024

Felip, M. & Catalan, J. (2000). The relationship between phytoplankton biovolume and chlorophyll in a deep oligotrophic lake: decoupling in their spatial and temporal maxima. *J. Plankton Res.* 22, 91–106

Hampton et al. (2016). Ice-on vs ice-free physical, chemical, and biological parameters of globally-distributed freshwater lakes. The Knowledge Network.

Lenard, T. (2013). Effect of Environmental Conditions on the Structure of Phytoplankton Community of Lake Piaseczno During Two Different Winter Periods. *Teka Komisji Ochrony i Ksztaltowania Srodowiska Przyrodniczego* 10, 223–231

Straskrabova, V. et al. (2005). Primary production and microbial activity in the euphotic zone of Lake Baikal (Southern Basin) during late winter. *Glob. Planet. Change* 46, 57–73

Woo, Kara (2015). icetest: Test data for under-ice ecology synthesis. R package version 1.0. https://github.com/karawoo/icetest