

## SYKE-plankton\_IFCB\_2022

### Authors:

Kaisa Kraft, Otso Velhonoja, Jukka Seppälä, Heidi Hällfors, Sanna Suikkanen, Pasi Ylöstalo, Sílvia Anglès, Sami Kielosto, Harri Kuosa, Sirpa Lehtinen, Johanna Oja, Timo Tamminen

Contact: [kaisa.kraft@syke.fi](mailto:kaisa.kraft@syke.fi), [jukka.seppala@syke.fi](mailto:jukka.seppala@syke.fi), [finmari-data@syke.fi](mailto:finmari-data@syke.fi)

### Marine Research Centre of the Finnish Environment Institute (SYKE)

Agnes Sjöbergin katu 2,  
00790 Helsinki,  
Finland

The data set available here is published with article “Kraft et al. (2022). Towards operational phytoplankton recognition with automated high-throughput imaging, near real-time data processing, and convolutional neural networks. *Front Mar. Sci.*” (submitted on 2.2.2022, status not updated). It contains approximately 63 000 images belonging to 50 different classes, consisting mainly of phytoplankton. The images can be used to e.g. train a classifier to identify phytoplankton images.

The images were collected with an Imaging FlowCytobot (IFCB, McLane Research Laboratories, Inc., U.S., [Olson and Sosik, 2007](#)) from different locations in the Baltic Sea. In 2017 and 2018 the data were collected from a continuous deployment at the [Utö Atmospheric and Marine Research Station](#) (59°46.84' N, 21°22.13' E; [Laakso et al., 2018](#); [Kraft et al., 2021](#)) operated by Finnish Environment Institute and Finnish Meteorological Institute (n=62). In 2016 and 2019 water samples were collected using the Alg@line ferrybox systems of M/S Finnmaid and Silja Serenade ([Ruokanen et al., 2003](#); [Kaitala et al., 2014](#)) and manually ran in the laboratory (n=52). The images were manually annotated by expert taxonomists. The class list and labeled image set is a continuous work in progress, thus there may be a need for revision in future. The data set available with this doi will not be revised. More detailed explanation and example images can be found from the publication Kraft et al. 2022.

The zipped folder contains 50 different folders, and the images are located in the class-specific folders. The image names may refer to an old class (e.g. folder Cryptophyceae-Teleaulax contains images with names Cryptophyceae\_drop, Cryptophyceae\_small, Teleaulax sp.) that has been joined with another one / revised otherwise.

*The work utilized SYKE and FMI marine research infrastructure as a part of the national FINMARI RI consortium. The work was partly funded by Tiina and Antti Herlin Foundation (personal grant for KK), Academy of Finland project FASTVISION (grant no. 321980), JERICO-S3 project, funded by the European Commission's H2020 Framework Programme under grant agreement No. 871153, and PHIDIAS project, funded by the European Union's Connecting Europe Facility under grant agreement INEA/CEF/ICT/A2018/1810854.*

### References:

Olson, R. J., and Sosik, H. M. (2007). A submersible imaging-in-flow instrument to analyze nano-and microplankton: Imaging FlowCytobot. *Limnol. Oceanogr. Methods* 5, 195-203. doi: 10.4319/lom.2007.5.195.

Kaitala, S., Kettunen, J., and Seppälä, J. (2014). Introduction to Special Issue: 5th ferrybox workshop—Celebrating 20 years of the Alg@ line. *J. Mar. Syst.* 140, 1–3. doi: 10.1016/j.jmarsys.2014.10.001

Kraft, K., Seppälä, J., Hällfors, H., Suikkanen, S., Ylöstalo, P., Anglès, S., et al. (2021). First application of IFCB high-frequency imaging-in-flow cytometry to investigate bloom-forming filamentous cyanobacteria in the Baltic Sea. *Front. Mar. Sci.* 8:282. doi: 10.3389/fmars.2021.594144

Laakso, L., Mikkonen, S., Drebs, A., Karjalainen, A., Pirinen, P., and Alenius, P. (2018). 100 years of atmospheric and marine observations at the Finnish Utö Island in the Baltic Sea. *Ocean Sci.* 14:617-632. doi: 10.5194/os-14-617-2018

Ruokanen, L., Kaitala, S., Fleming, V., and Maunula, P. (2003). "Alg@line–joint operational unattended phytoplankton monitoring in the Baltic Sea" in *Elsevier Oceanogr. Ser.* 519–522. doi: 10.1016/S0422-9894(03)80083-1

The list of classes and the number of images in a class:

Class / taxonomic group	Samples	Class / taxonomic group	Samples
<b>Cyanophyceae</b>		<b>Diatomophyceae</b>	
<i>Dolichospermum</i> sp. / <i>Anabaenopsis</i> sp.	12280	<i>Skeletonema marinoi</i>	4128
<i>Aphanizomenon flosaquae</i>	6989	<i>Thalassiosira levanderi</i>	2537
Oscillatoriales	4440	<i>Chaetoceros</i> sp. chain	1382
<i>Snowella</i> sp. / <i>Woronichinia</i> sp.	2950	Pennales thin	781
<i>Dolichospermum</i> sp. / <i>Anabaenopsis</i> sp. coiled	2504	Centrales	480
<i>Chroococcus</i> sp.	827	<i>Chaetoceros</i> sp. single	213
<i>Nodularia spumigena</i>	169	Pennales thick	210
Chroococcales	142	<i>Pauliella taeniata</i>	119
<i>Merismopedia</i> sp.	98	<i>Cyclotella choctawhatcheeana</i>	102
<i>Aphanothece paralleliformis</i>	29	<i>Licmophora</i> sp.	74
		<i>Nitzschia paleacea</i>	65
<b>Cryptophyceae</b>		<i>Ceratoneis closterium</i>	45
Cryptophyceae / <i>Teleaulax</i> sp.	6830	<i>Melosira arctica</i>	43
Cryptomonadales	713		
		<b>Chrysophyceae</b>	
<b>Euglenophyceae</b>		<i>Uroglenopsis</i> sp.	516
<i>Eutreptiella</i> sp.	2247	<i>Pseudopedinella</i> sp.	379
Euglenophyceae	102		
		<b>Chlorophyta</b>	
<b>Dinophyceae</b>		<i>Pyramimonas</i> sp.	1224
<i>Heterocapsa triquetra</i>	3276	<i>Oocystis</i> sp.	842
Dinophyceae	1433	<i>Monoraphidium contortum</i>	327
<i>Peridiniella catenata</i> single	899	<i>Cymbomonas tetramitiformis</i>	199
<i>Heterocapsa rotundata</i>	614	Chlorococcales	95
<i>Prorocentrum cordatum</i>	276		
<i>Dinophysis acuminata</i>	217	<b>Ciliates</b>	
<i>Peridiniella catenata</i> chain	193	<i>Mesodinium rubrum</i>	1132
<i>Gymnodinium</i> like cells	158	Ciliata	243
Gymnodiniales	69		
<i>Gonyaulax verior</i>	22	<b>Additional classes</b>	
<i>Amylax triacantha</i>	19	heterocyte	263
		beads	125
<i>Katablepharis remigera</i>	54		

Below, example images representing the 50 classes: 1) *Aphanizomenon flosaquae*, 2) *Dolichospermum* sp. / *Anabaenopsis* sp. coiled, 3) *Nodularia spumigena*, 4) *Dolichospermum* sp. / *Anabaenopsis* sp., 5) *Snowella* sp. / *Woronichinia* sp., 6) Chroococcales, 7) *Merismopedia* sp., 8) Oscillatoriales, 9) *Aphanothece paralleliformis*, 10) *Chroococcus* sp., 11) *Eutreptiella* sp., 12) Euglenophyceae, 13) Cryptomonadales, 14) Cryptophyceae / *Teleaulax* sp., 15) *Katablepharis remigera*, 16) *Pseudopedinella* sp., 17) *Pyramimonas* sp., 18) *Ceratoneis closterium*, 19) *Uroglenopsis* sp., 20) *Cymbomonas tetramitiformis*, 21) Chlorococcales, 22) *Monoraphidium contortum*, 23) *Oocystis* sp., 24) Pennales thin, 25) Pennales thick, 26) Centrales, 27) *Thalassiosira levanderi*, 28) *Cyclotella choctawhatcheeana*, 29) *Chaetoceros* sp. single, 30) *Melosira arctica*, 31) *Skeletonema marinoi*, 32) *Nitzschia paleacea*, 33) *Licmophora* sp., 34) *Chaetoceros* sp., 35) *Pauliella taeniata*, 36) *Peridiniella catenata* chain, 37) *Peridiniella catenata* single, 38) Gymnodiniales, 39)

Gymnodinium like cells, 40) *Heterocapsa triquetra*, 41) *Heterocapsa rotundata*, 42) *Prorocentrum cordatum*, 43) *Gonyaulax verior*, 44) *Amylax triacantha*, 45) Dinophyceae, 46) *Dinophysis acuminata*, 47) *Mesodinium rubrum*, 48) Ciliata, 49) Beads, 50) Heterocyte

