

Thomas Rutten Projects

Date : 28 October 2025
By : Thomas Rutten
Ref.nr : TRP 2025.029

Building a database in EasyClus, several options, same goal

Introduction

The EasyClus database contains optical data and images and is a very useful tool for recognizing species (phytoplankton) in water samples. A database can also consist of functional type groups, clustered on basis of similar optical properties. The EasyClus supervised clustering methods RULES, LASSO and HYBRID use the database to recognize and classify flow cytometry data. There are particles that match the indicated groups or species in the database and other particles that do not match (not recognized). A good database ensures that about 95% or more particles are recognized.

Filling a database can be done in different ways:

1. Clusters are added to the database after unsupervised clustering via a tool – the name is given or edited in the database per each group or cluster
2. Extract images and copy images to database names folders – update button EC database
3. Import directly from CytoClus isolated data (e.g. .xml sets)
4. Build automatically from Hybrid method (and rename – optimize afterwards)
5. Build automatically from ‘autonamed’ sorted images (e.g. NanoFw3Sw4Re8Or3Ye4)

Tip 1: Start collecting many files with images and merge

It can be useful to create a database based on many files, because then the biological variation, instrument variation and other variation of measurements are all included in the data of the database. If a database is to be created with the combination of flow cytometry data and images only, it can be useful to create a composite file, which is based on collecting all images from all available files and using this 'merged' composite file as the basis to create the database. This merged file also provides insight into the more and less dominant particles. Also, at low numbers of species, clusters of these

species will be formed by merging more files, making them easier to add to the database as a cluster.

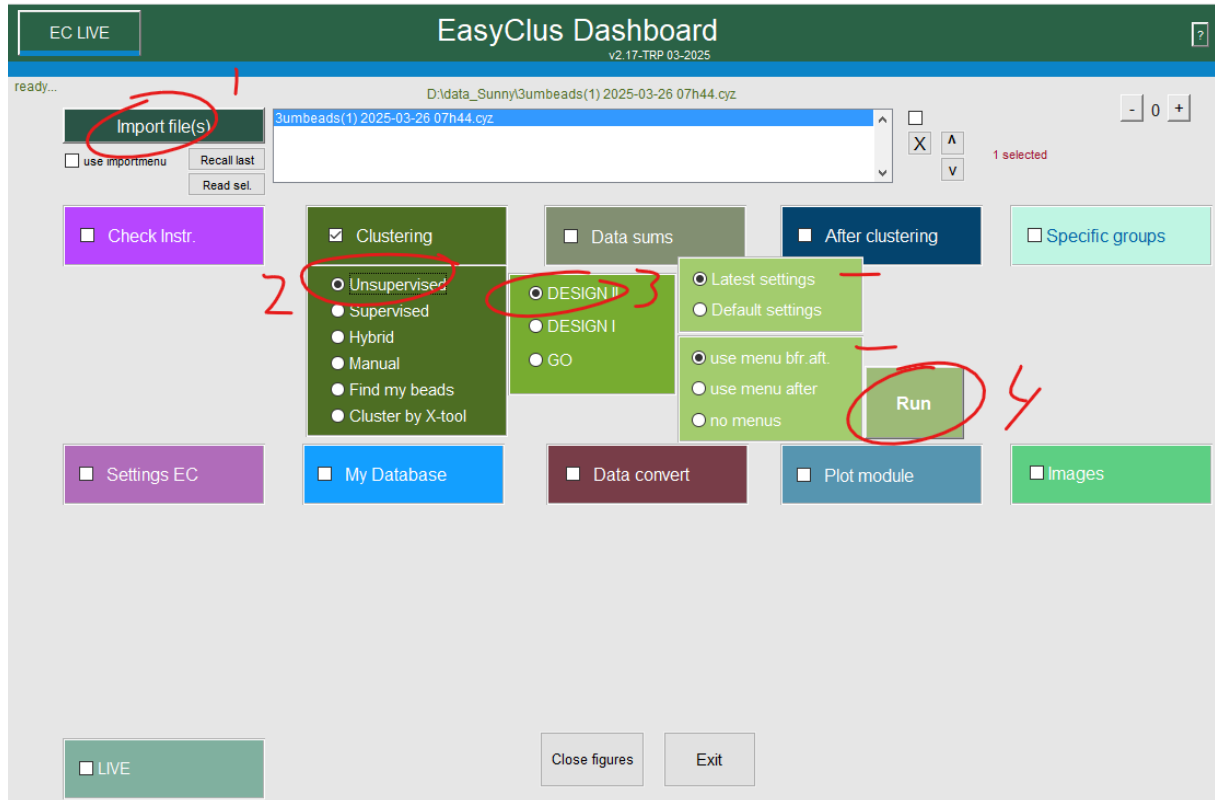
Tip 2: How about smaller particles without an image?

It is usually only useful to make images of particles larger than 3-5 μm , because small particles are difficult to recognize due to a combination of magnification and pixel size. Usually, the equipment is also set to only photograph particles from a certain size. This means that if a database is created consisting of only flow cytometry data with an image, these small particles are missing for the database and can never be recognized. Because small particles must also be recognized in every analysis, it is necessary to add small particles without a photo to the database as well. In addition, it is also possible to add small particles with a virtual 'image' to the database. This virtual image is based on the signal characteristics.

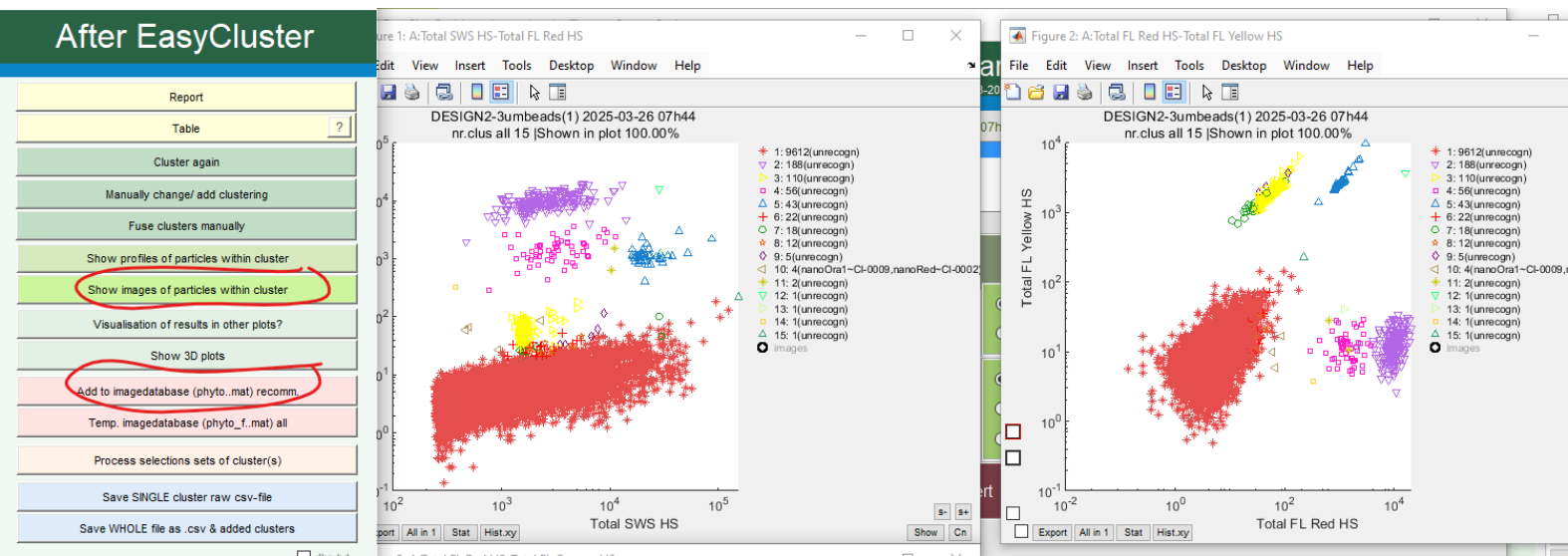
Thomas Rutten Projects

1. Clusters are added to the database after unsupervised clustering via a tool – the name is given per cluster

Import file & cluster by EasyClus



Look at cluster result in combination with images in clusters 'Show images ...'



Thomas Rutten Projects

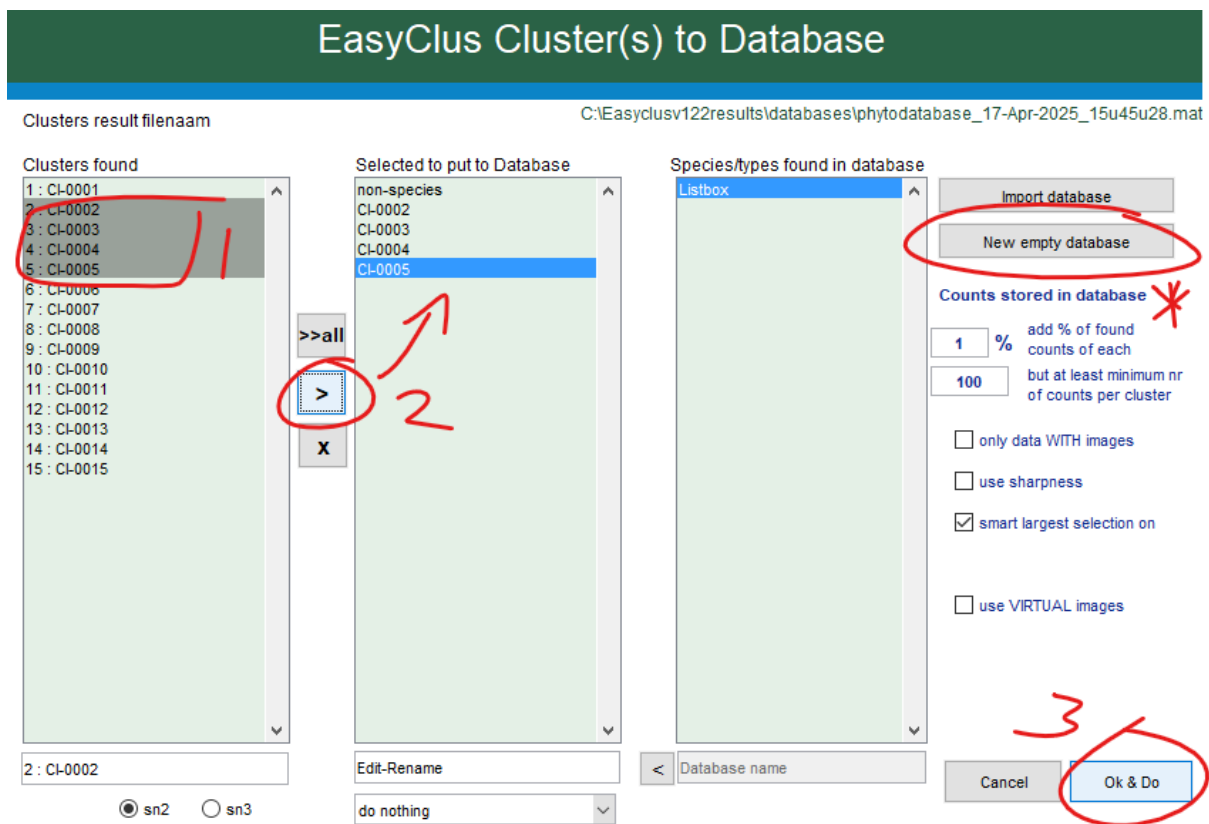
Add clusters of interest to the database by 'Add to imagedatabase'

If you want to start a new empty database press 'New empty database' see *

Select clusters of interest (1)

Add to list in the middle by '>' button (2)

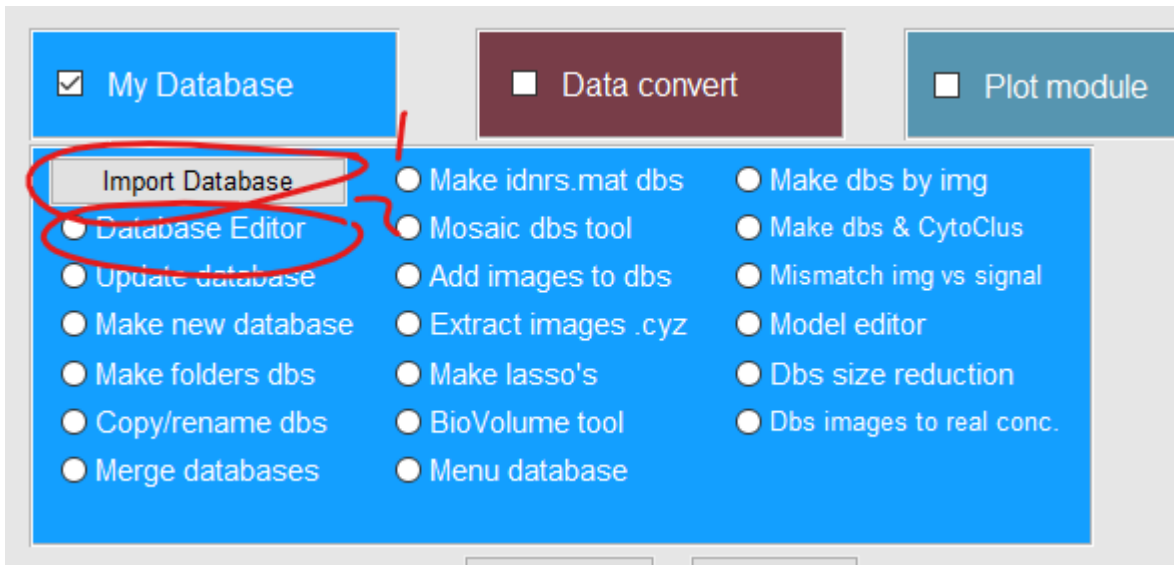
Press 'Ok& Do' (3)



Look in the database editor and rename

Thomas Rutten Projects

Import database (1) then Database Editor (2)

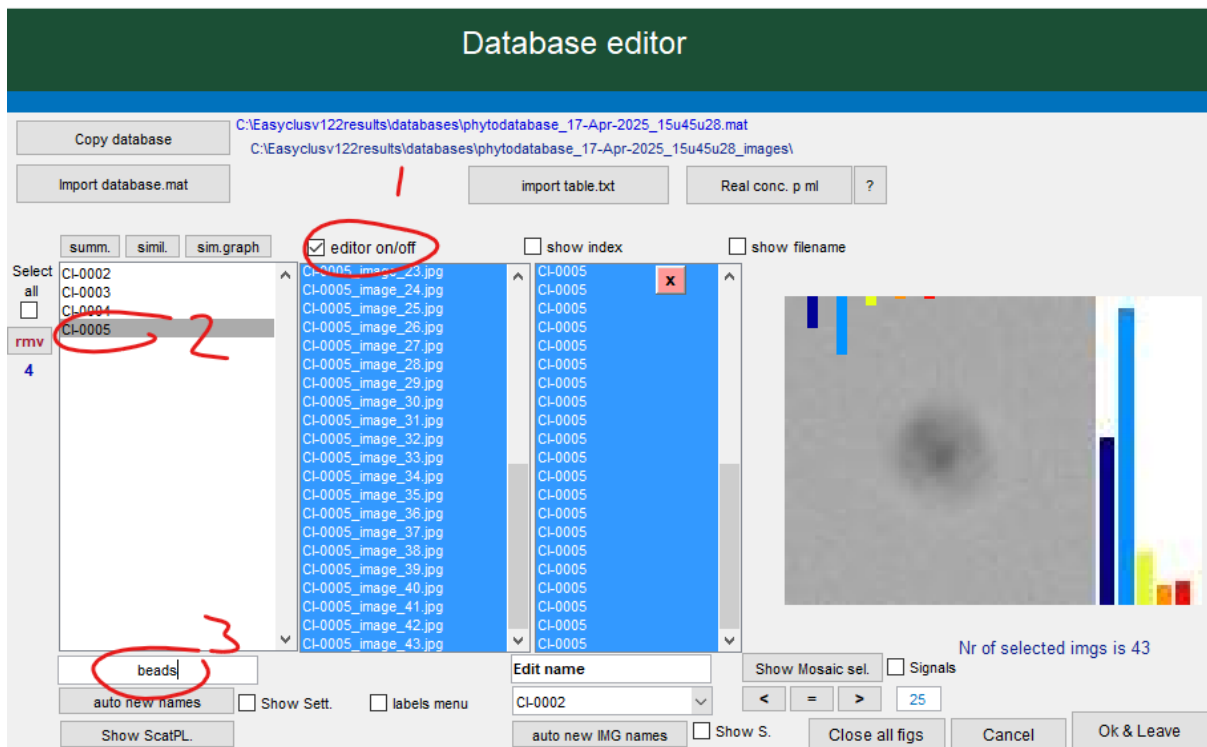


Look in the database editor and rename by:

Click 'editor on' (1)

Click on requested cluster (2)

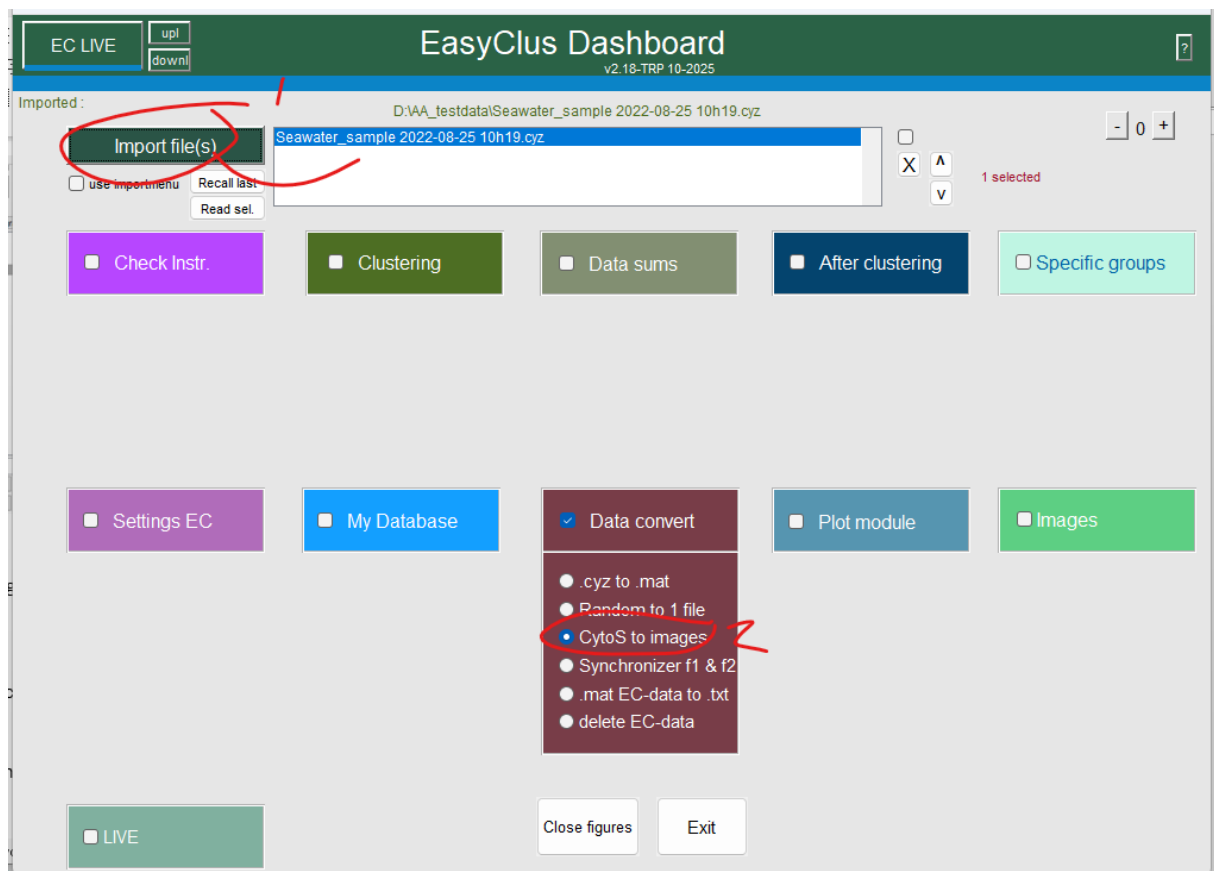
Rename (3) and 'Enter'



2. Extract images and copy images to database names folders – update button EC database

This method uses photos of algae that are put in folders corresponding to the species name on the basis of the species names. The database is then updated via an update button. The original cyz file belonging to the photos is needed to read the flow cytometric data. This original cyz-file(s) should be located in the EasyClusvxxxresults\datafiles\ folder.

The method starts with reading a file and then we export all images to .jpg files

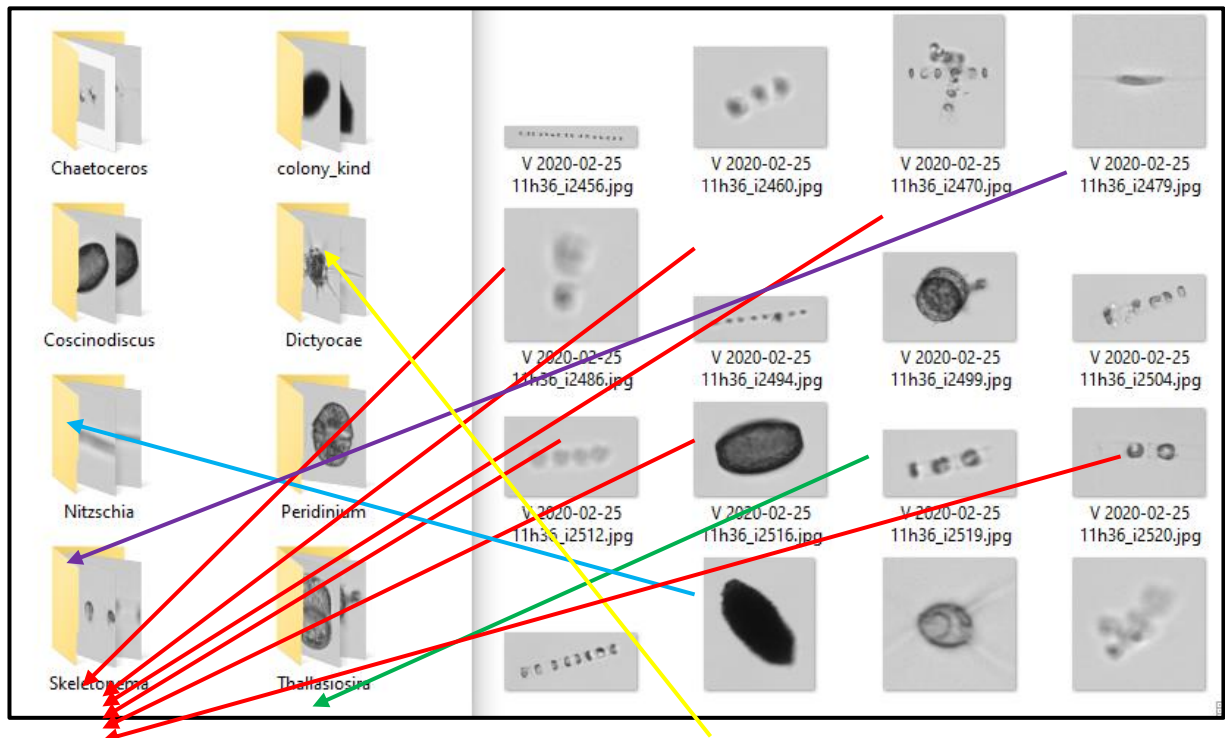


Put or move the photos of algae in (self-created) folders with the folder name of the phytoplankton species to which the photo belongs. The name of the photo does not need to be changed.

Note: It is also possible to use this option for the database of photos, which are in the 'database name_images' folder. These folders contain all the photos of the species and species groups in the database, where the species name is included in the file name (of the .jpg file). Photos can also be copied from here into a new folder with a different species name, after which the 'update' button makes the adjustments automatically.

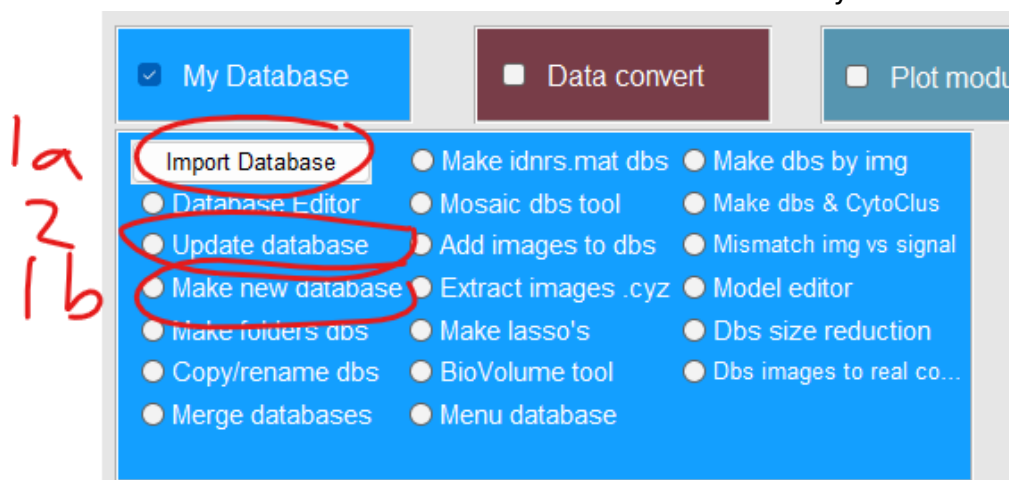
Thomas Rutten Projects

The photos (.jpg) are copied (in the case of the database) or moved (or copied) in cyz-files to the folder with the correct desired species name or species type.



In EasyClus :

Import database (1a) or if necessary create a new empty database (1b) and 'Update the database' (2) and be sure that if the images are not yet in the database already but newly added that the original cyz-file(s) should be located in the EasyClusvxxxresults\datafiles\ folder. It is often necessary to re-import a database afterwards to ensure that the most recent database is loaded in EasyClus.



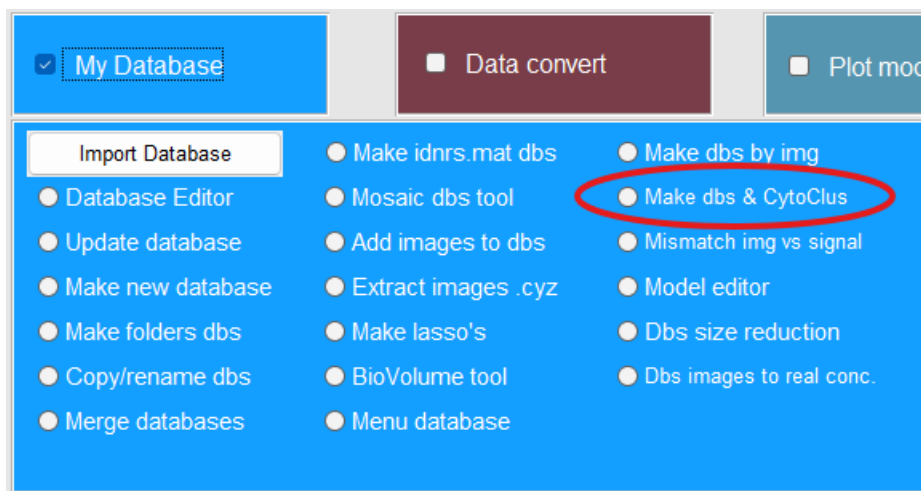
Thomas Rutten Projects

3. Import directly from CytoClus isolated data (e.g. .xml sets)

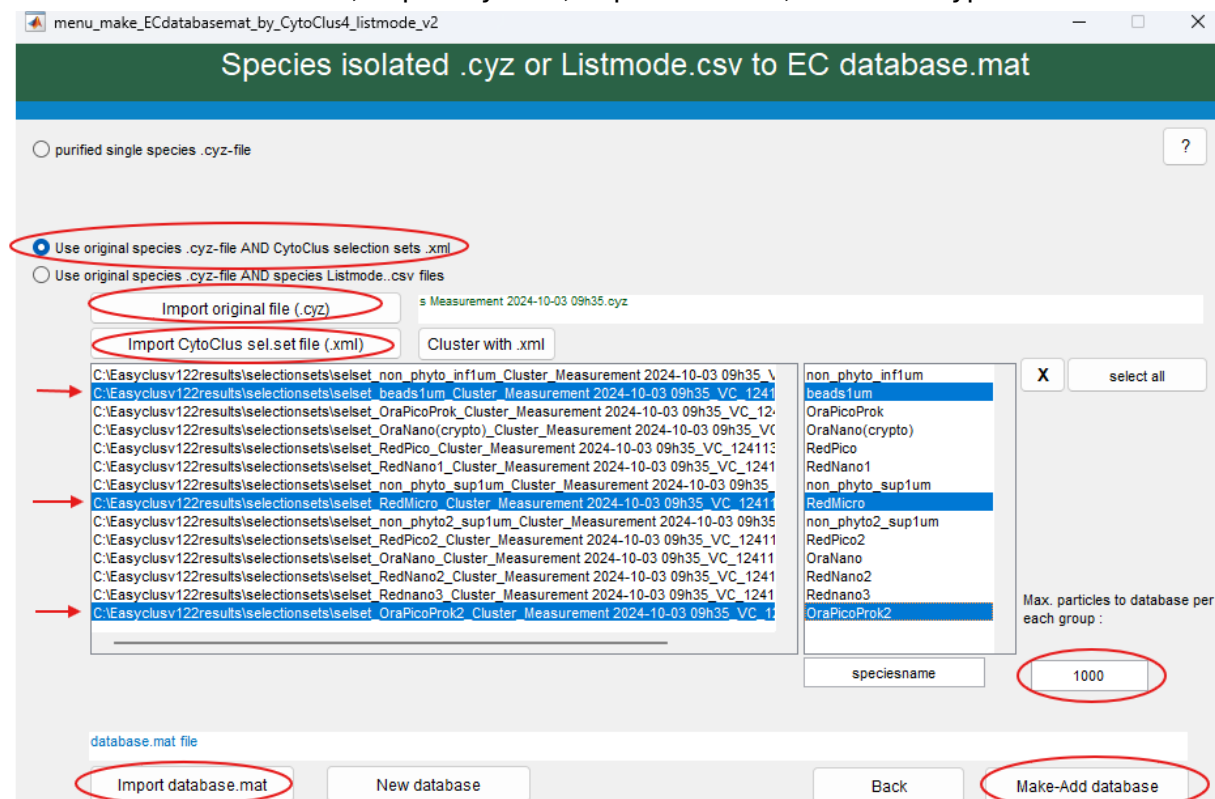
3a By using the database-module: Import cyz-file AND import selection set or lasso xlm.file made in CytoClus from isolated species type together, and add to a database

3b By using the EasyClus lasso clustering method: Import cyz-file AND import selection set or lasso xlm.file made in CytoClus from isolated species type together, and add to a database

3a : Database + .xml



Select the choices below, import cyz-file, import .xml file, select the types

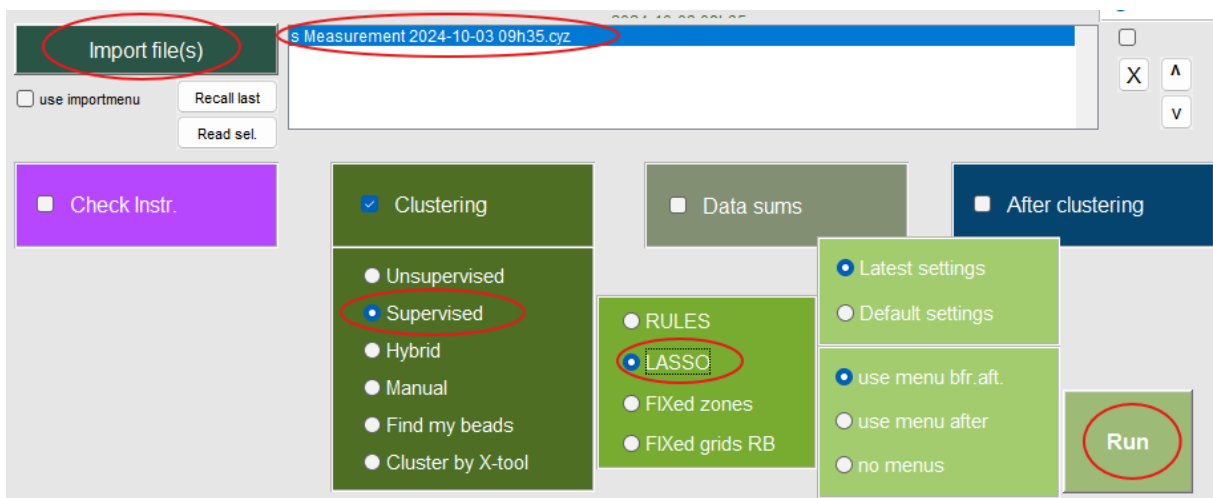


Thomas Rutten Projects

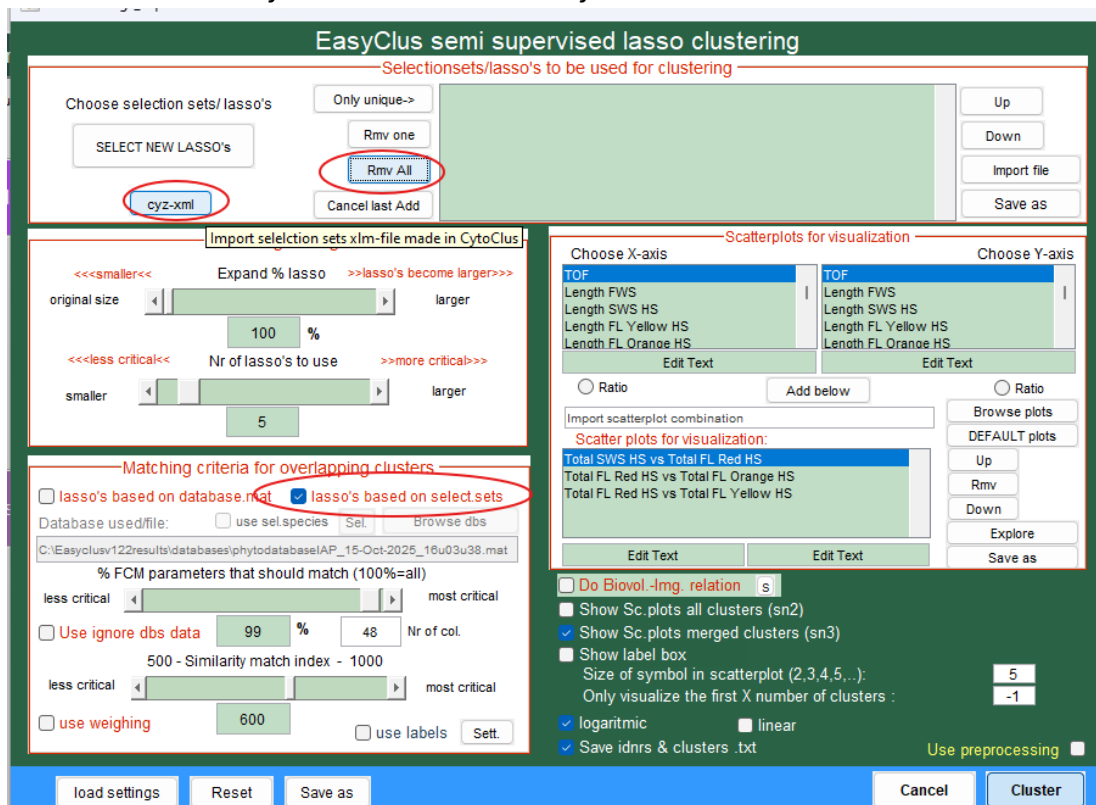
that need to go to the database (in the example there are three), the (maximum) number that is stored in the database per type and add it to an imported or new database.

Via the same module, there are also other options for importing data processed via CytoClus into the database i.e. by importing CytoClus selected cluster(s) or selection set(s) as a 'purified' saved cyz-file(s) or by importing the Listmode.csv file(s) of clusters or selection sets.

3b via EasyClus LASSO clustering



Import the cyz-file, start the supervised LASSO clustering (with menu). Press cyz-xml button to convert the CytoClus .xml file to EasyClus lasso's selection sets.



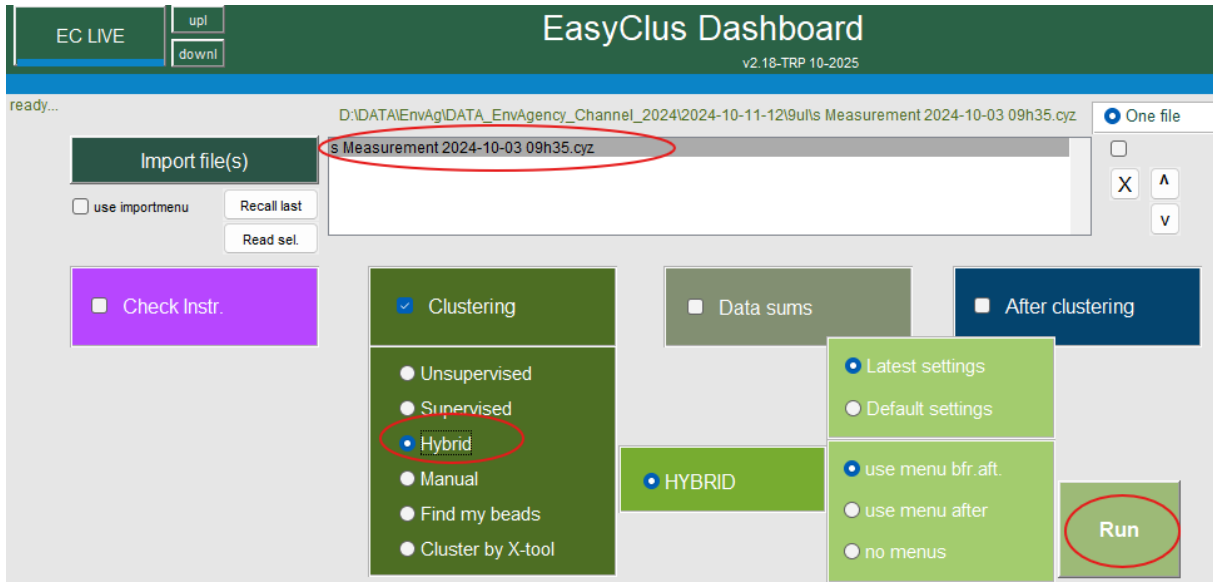
Thomas Rutten Projects

Remove all previously loaded lasso's by pressing the 'Rmv-All' button. Import the newly created lasso selection sets with 'Select New LASSO's' button (three types have been chosen here) and start the LASSO clustering ('Cluster' button)

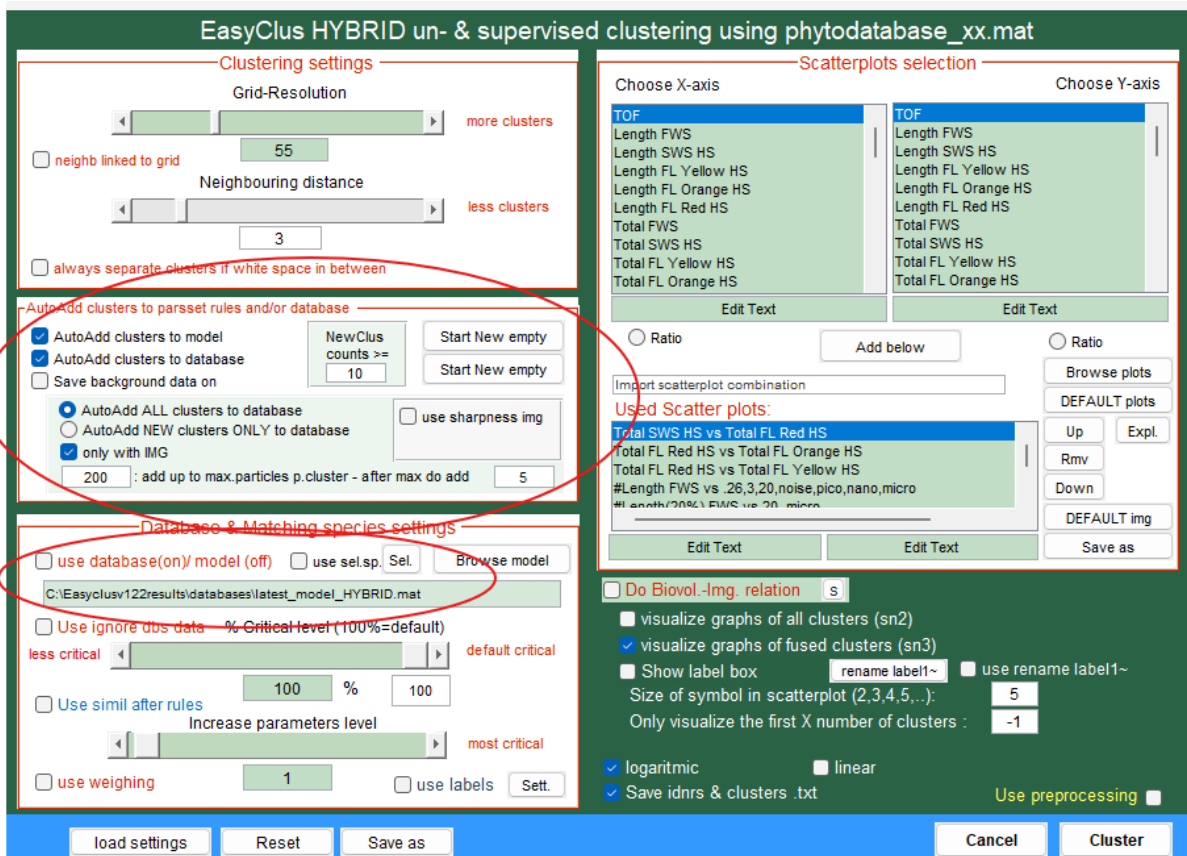
Add the found clusters to the database via the 'Add to imagedatabase' button

4. Build automatically from Hybrid method (and rename – optimize afterwards)

Using the HYBRID method, unsupervised DESIGNII is clustered and then clusters and particles from clusters are recognized via the database. Unknown clusters are automatically added to the database. This database can be modified later via the database editor if necessary.



With HYBRID, a model is built, and a database can be built simultaneously. The model



Thomas Rutten Projects

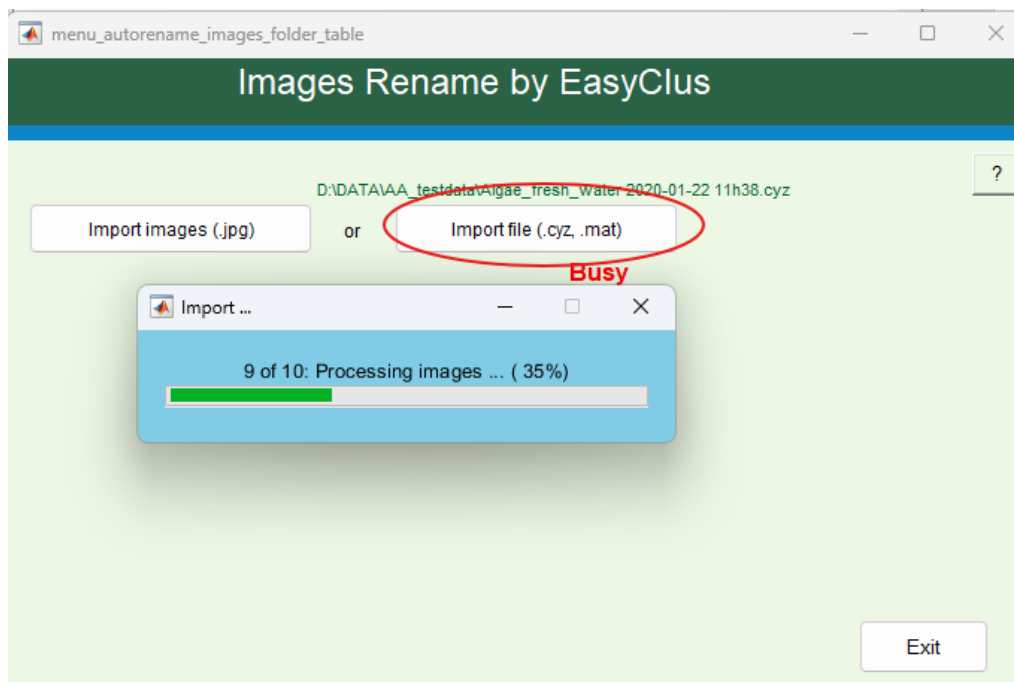
(RULES method) is used for matching and recognition, the database is only used in this example to add the species that have been found (with an image) to a database. This method is especially useful if it is used over a larger number of files, because data is added to the database from each file.

5. Build automatically from 'autonamed' sorted images (e.g. NanoFw3Sw4Re8Or3Ye4)

Images of phytoplankton particles are automatically given names that are constructed according to a 'coding' based on the (optical) signal heights of that particle. For example, Re1 equates to a low signal in the red fluorescence and Re10 to a high signal in the red FL. These names can be used automatically directly in the database, but also on individual photos of particles created via EasyClus. The images with the same name are particles that are optically related to each other. It provides pre-sorted photos based on optical properties, which can be used directly for the database

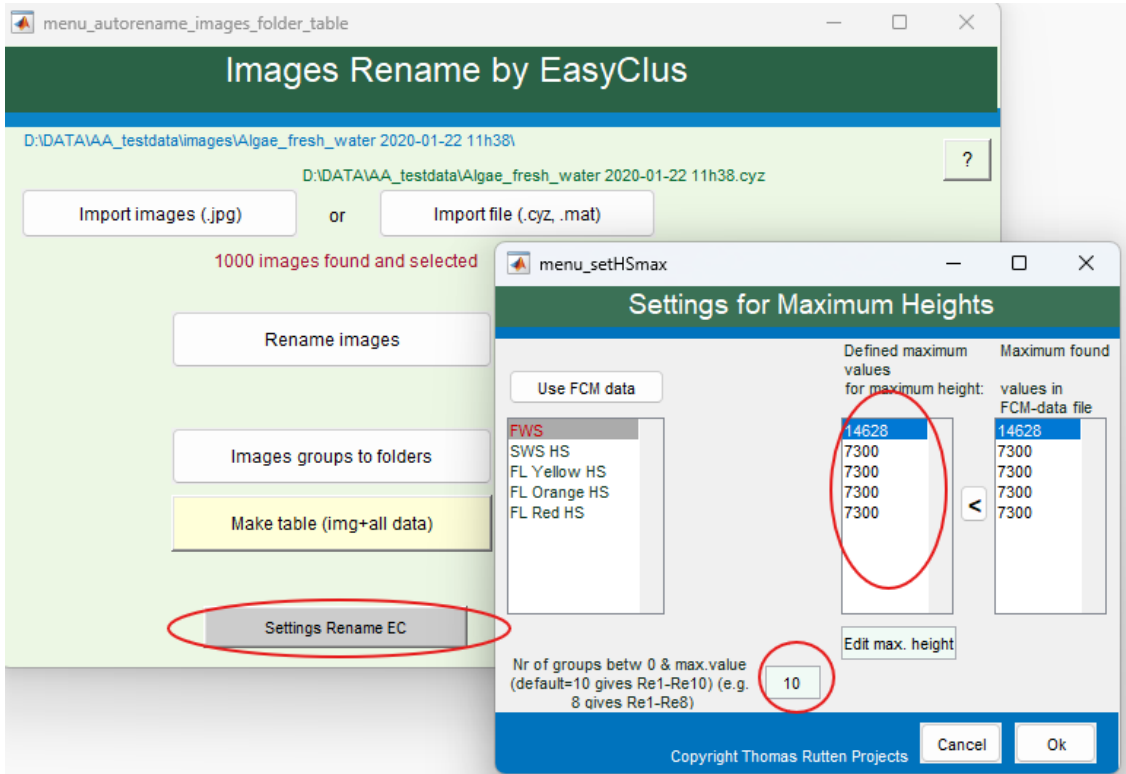
5a Re-naming of photos and then adding to database

Import a cyz-file (or previously exported .jpg photos), photos are read.

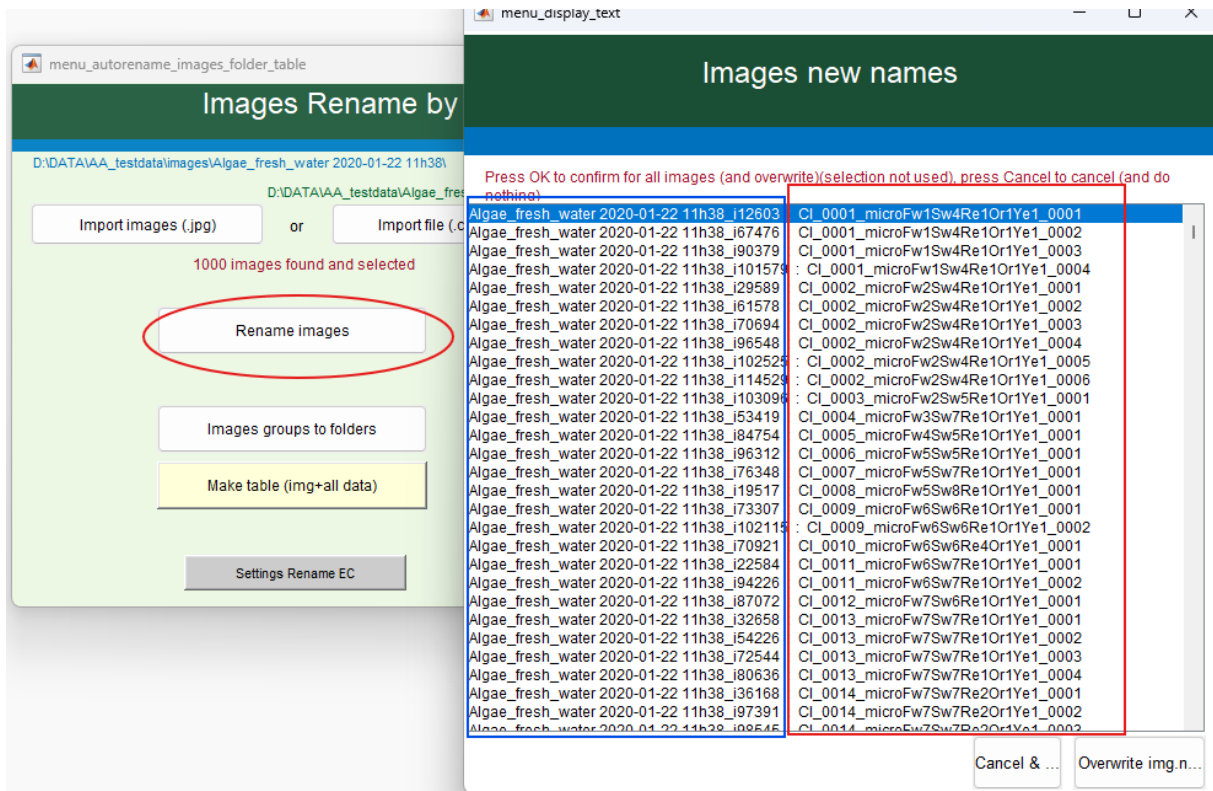


Check settings to see which optical characteristics are used and set the 'limits' or maximum possible values per detector:

Thomas Rutten Projects

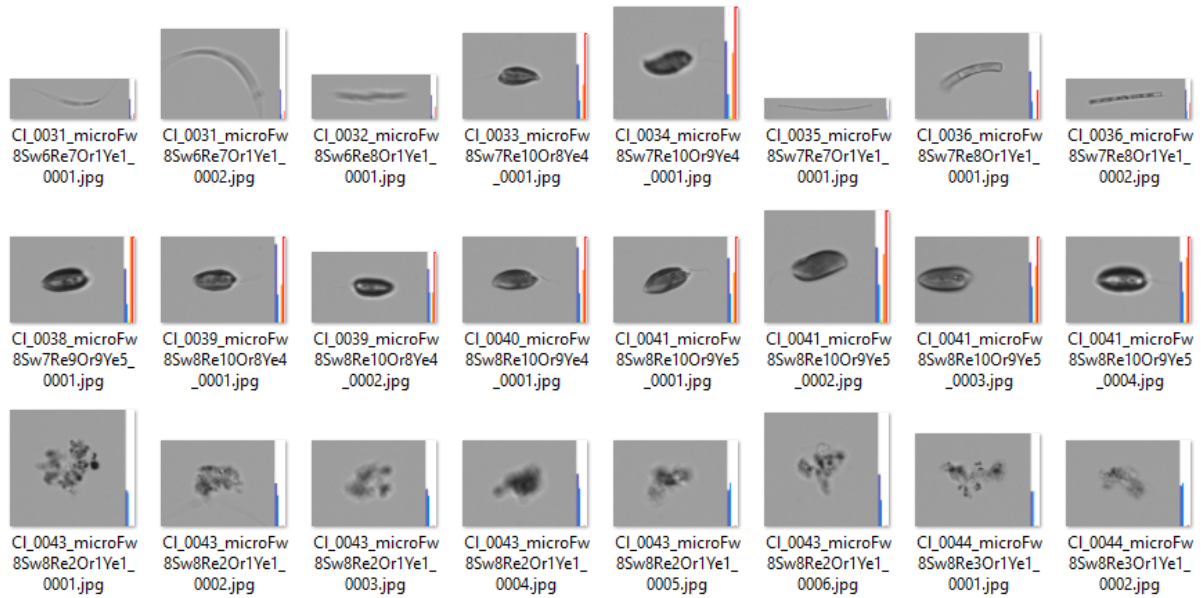


Original names are in blue, the new names in red. The 1000 photos are



Thomas Rutten Projects

divided into new names, which are numbered. An example of the above file with photos is indicated below. Types with the same name (optical characteristics) are put in one group with a cluster number and the name behind it. From the photos it can be concluded that the same types of photos are put together.



Via the option, images to folder, the photos are moved by the same name in the folders of the same name.

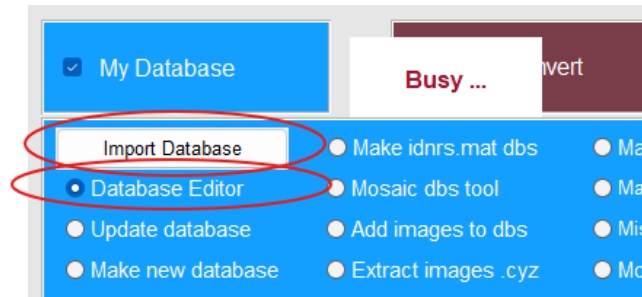
Via the option table, the types with the same name can be further studied

IMGcl	UniqNames	Cntr_all	Conc_all_p_ml	Rel_contr	Counts	img_counts	img_vs_C	SzGrp	Length_FWS	Nr	NmoSw	NmoFwSw	Fw	Sw	Re
microFw10Sw10Re7Or7Ye6		1	1.4710	8.5053e-06		1	0	0	micro	388.5283	1010776	10776	776	10	10
microFw11Sw4Re1Or1Ye1		2	14.7080	8.5053e-05		10	0	0	micro	27.8985	14111	1111	111	1	4
microFw2Sw4Re1Or1Ye1		3	19.1200	1.1057e-04		13	0	0	micro	33.3925	24111	2111	111	2	4
microFw2Sw5Re1Or1Ye1		4	7.3540	4.2526e-05		5	0	0	micro	36.4735	25111	2111	111	1	5
microFw2Sw6Re1Or1Ye1		5	1.4710	8.5053e-06		1	0	0	micro	28.5423	26111	2111	111	2	6
microFw3Sw5Re1Or1Ye1		6	5.8830	3.4021e-05		4	0	0	micro	25.9344	35111	3111	111	3	5
microFw3Sw7Re1Or1Ye1		7	1.4710	8.5053e-06		1	0	0	micro	37.6247	37111	3111	111	3	7
microFw4Sw5Re1Or1Ye1		8	4.4120	2.5516e-05		3	0	0	micro	36.8035	45111	4111	111	4	5
microFw4Sw6Re1Or1Ye1		9	1.4710	8.5053e-06		1	0	0	micro	31.7349	46111	4111	111	4	6
microFw5Sw5Re1Or1Ye1		10	1.4710	8.5053e-06		1	0	0	micro	28.2744	55111	5111	111	5	5
microFw5Sw6Re1Or1Ye1		11	1.4710	8.5053e-06		1	0	0	micro	53.7755	56111	5111	111	5	6
microFw5Sw7Re1Or1Ye1		12	1.4710	8.5053e-06		1	0	0	micro	42.3738	57111	5111	111	5	7
microFw5Sw8Re1Or1Ye1		13	1.4710	8.5053e-06		1	0	0	micro	50.0818	58111	5111	111	5	8
microFw6Sw5Re2Or1Ye1		14	1.4710	8.5053e-06		1	0	0	micro	48.6186	65211	6211	211	6	5
microFw6Sw5Re3Or1Ye1		15	1.4710	8.5053e-06		1	0	0	micro	29.1727	65311	6311	311	6	5
microFw6Sw6Re1Or1Ye1		16	8.8250	5.1032e-05		6	0	0	micro	22.1819	66111	6111	111	6	6
microFw6Sw6Re4Or1Ye1		17	2.9420	1.7011e-05		2	0	0	micro	43.6424	66411	6411	411	6	6
microFw6Sw7Re1Or1Ye1		18	8.8250	5.1032e-05		6	0	0	micro	29.2727	67111	6111	111	6	7
microFw6Sw7Re2Or1Ye1		19	5.8830	3.4021e-05		4	0	0	micro	46.0131	67211	6211	211	6	7
microFw6Sw7Re3Or1Ye1		20	1.4710	8.5053e-06		1	0	0	micro	48.1175	67311	6311	311	6	7
microFw6Sw6Re2Or1Ye1		21	1.4710	8.5053e-06		1	0	0	micro	40.8679	68211	6211	211	6	8
microFw7Sw4Re2Or1Ye1		22	1.4710	8.5053e-06		1	0	0	micro	21.8786	74211	7211	211	7	4
microFw7Sw5Re2Or1Ye1		23	1.4710	8.5053e-06		1	0	0	micro	47.3881	75211	7211	211	7	5

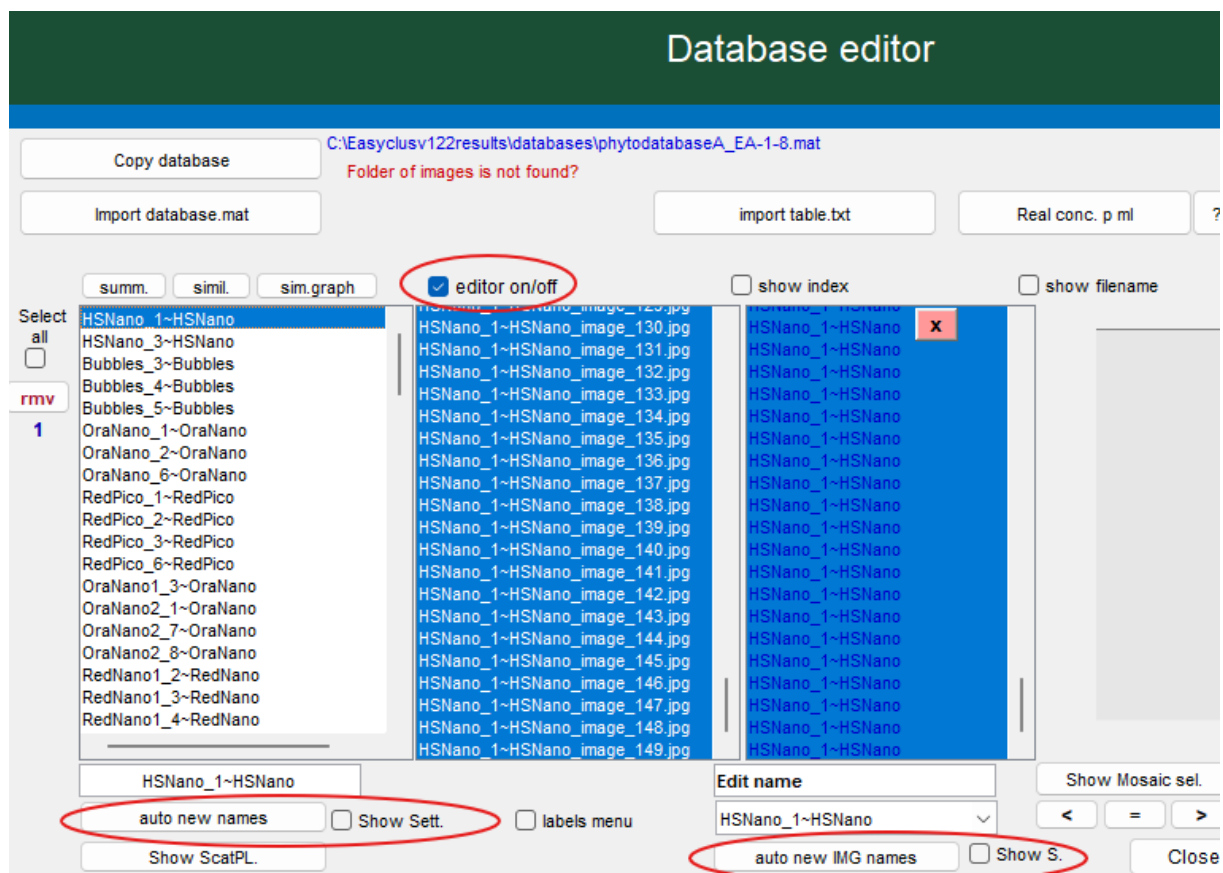
Thomas Rutten Projects

5b Re-naming of species in the database

Import the database and open the database editor



Click 'editor on/off' to 'on'

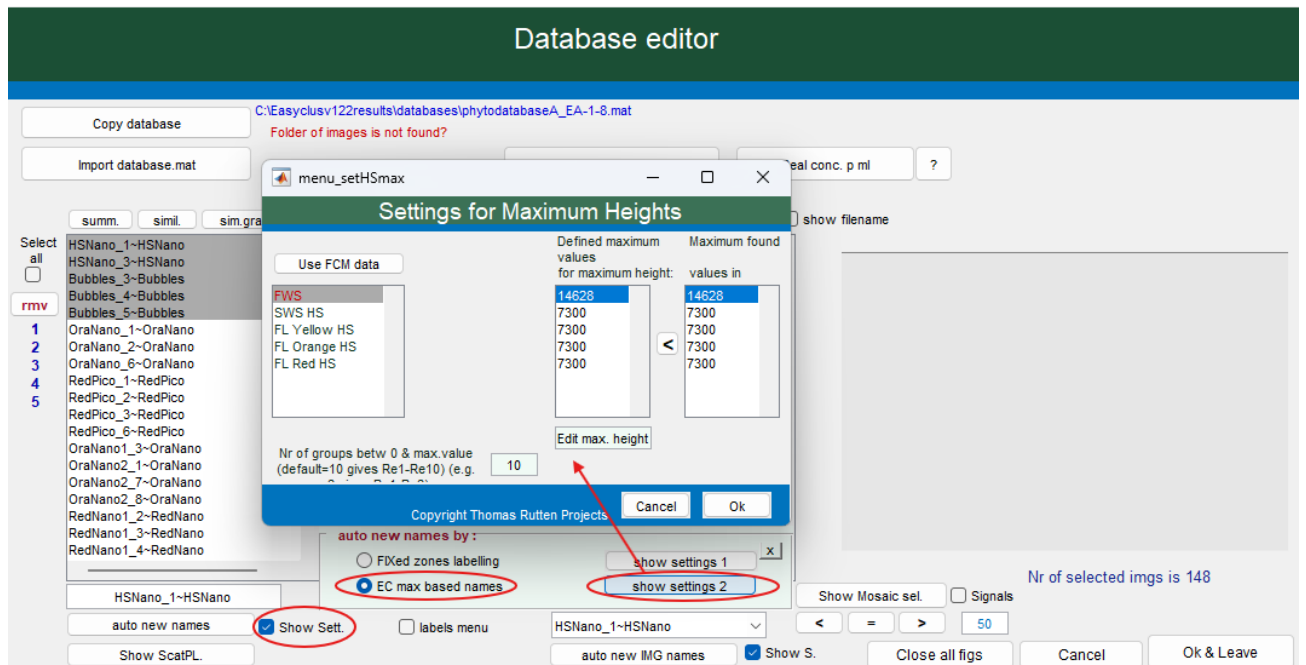


There are two options in the database editor: on the left you can rename the entire (selected) database name group (and all the particles in it) or on the right you can rename all the individual (selected) particles and rename each of them.

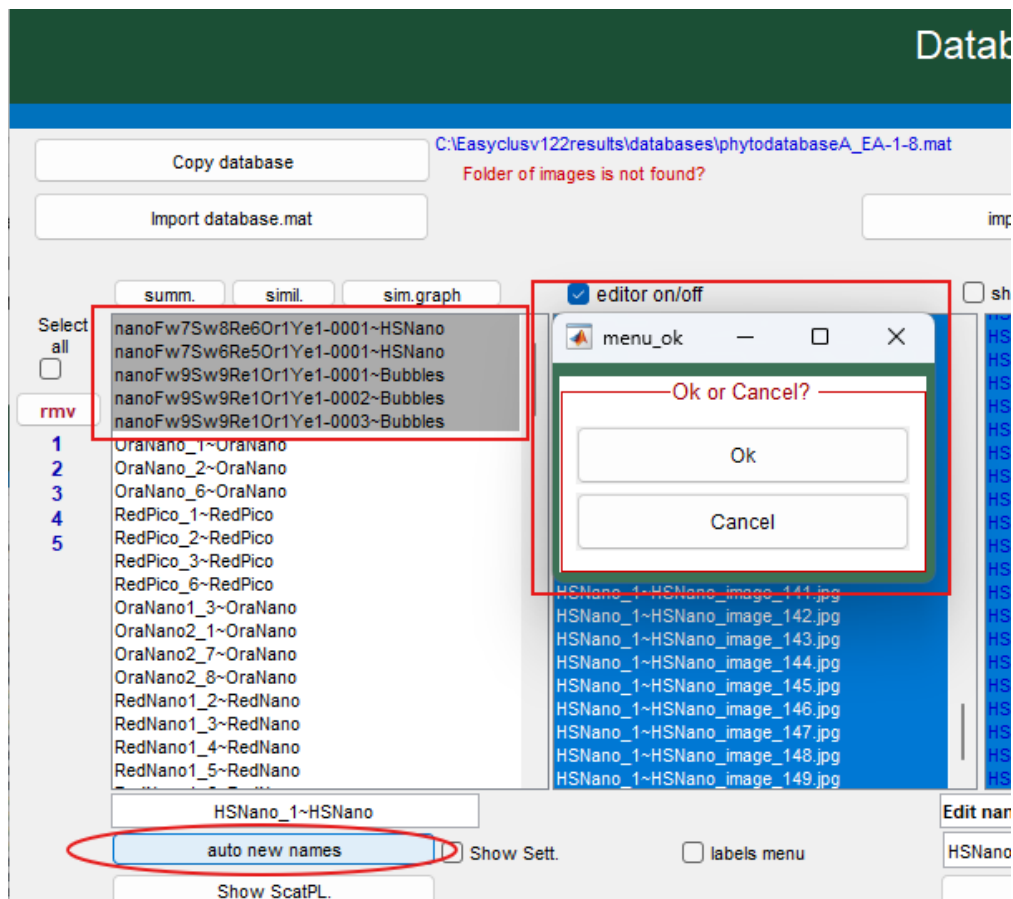
The renaming is like the above and can be set via the settings (although an additional FIX option is offered here). The renaming of the entire (selected) database group (left) is done based on the average of all particles within this group, while for the right setting

Thomas Rutten Projects

this is done per individual particle. This individual approach usually results in the creation of more groups within one species.

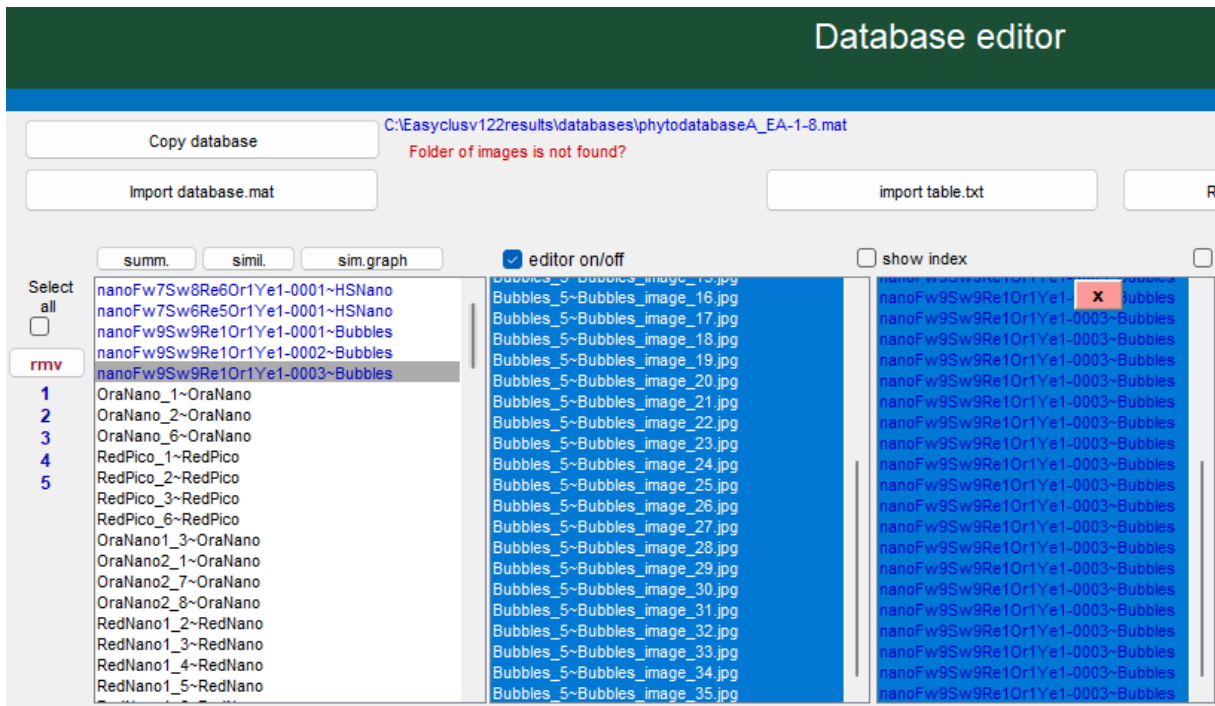


After autonames (left), the new and old name is presented



Thomas Rutten Projects

After confirmation, old and new names are kept next to each other. All individual particles are given this new and old name.



If right 'autonames' option is used, new names are determined for each particle.

For example, the first group is split into 18 groups, with 18 group names:

