New Functions in ALMA Science Archive (ASA)

Line Search and Similar Image

Pei-Ying Hsieh (NAOJ, ASA EA cognizant lead)

2025 Major Updates

CARTA 5

- Viewing channel map is available before downloading the cubes.

Fastatronomy

- Spectral Line Identification.
- Similar Images.

Fastastronomy

find me all the CO2-1 line data with 0.5 km/s velocity resolution of M51 in the ALMA archive

okay, here is the table of the data and the scripts you request, do you want to export them to csv file?

Edit in a page

find me all the stellar sources exihibiting outflow structures

Smart

Spectral Line Database in ASA – Current and Future

- To help users quickly explore and mine the science content of ALMA data cubes, now we offer,
  - detecting spectral lines
  - computing moment maps

- **ADMIT (ALMA Data Mining Toolkit)** is a Python-based software. It was developed by The ADMIT Team at UMD in 2016.

- Most the ALMA archive data have ADMIT products.

https://casaguides.nrao.edu/index.php/ADMIT_Products_and_Usage_CASA_6
https://admit.astro.umd.edu/examplescripts.html#example-1-ingest-through-lineid-and-moment
https://help.almascience.org/kb/articles/where-can-i-get-additional-information-for-my-na-added-value-data-products

ADMIT Pipeline of Line Identification (brief version)

[CubeSum] Sum all cube along all channels of the frequency axis to make a 2-D map.

[SFind2D] Creates a list of sources found in a 2-D image brighter than 5 sigma.

[CubeSpectrum] Extract a spectrum through the cube **at a specified point (?).**

[LineID] Identify any spectral lines present in the data.

[LineCube] Create a cutout cube for each spectral lines identified.

[Moment] Compute moment 0,1,2 maps, and spectra for each cutout cube

https://science.nrao.edu/facilities/alma/science_sustainability/Teuben_study7_report.pdf

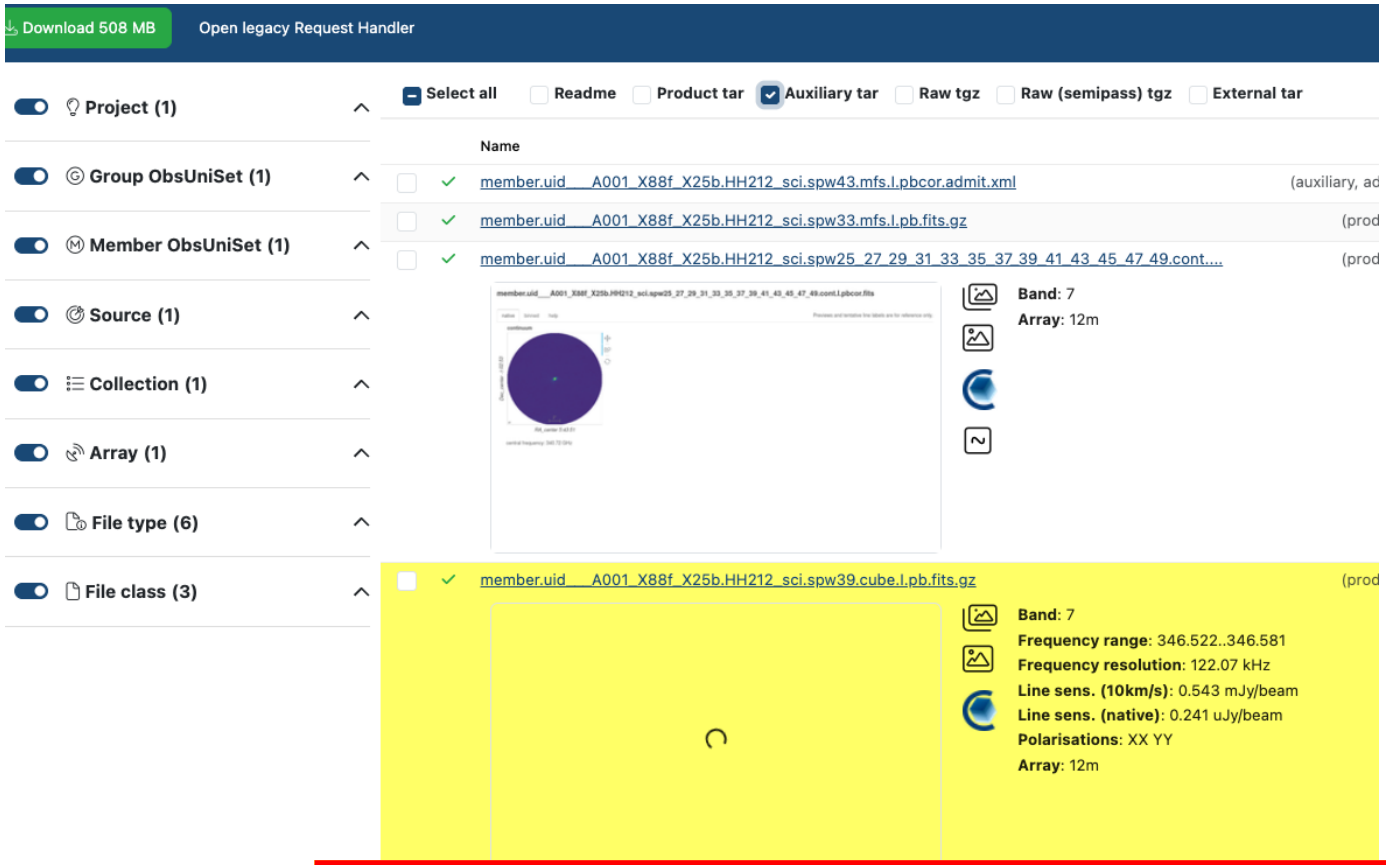
SFind2D (peaks > 5 sigma at the sum image)

ASA: Extract the spectrum of the first source, and do line ID. The source table is available in the ADMIT product.

Table of source locations and sizes (not deconvolved)

Name	RA	DEC	Flux	Peak	Major	Minor	PA
[]	[]	[]	[Jy]	[Jy/beam]	[arcsec]	[arcsec]	[deg]
1.000E+00	17:46:46.380	-28.32.09.94	1.409E+00	2.856E-01	2.636E+00	1.544E+00	1.349E+02
2.000E+00	17:46:46.396	-28.32.10.24	2.613E+00	2.651E-01	4.540E+00	1.805E+00	1.633E+02
3.000E+00	17:46:46.379	-28.32.09.86	3.375E+00	2.779E-01	5.847E+00	1.763E+00	1.538E+02
4.000E+00	17:46:47.058	-28.32.07.16	2.326E-01	2.511E-01	9.480E-01	7.653E-01	1.249E+02
5.000E+00	17:46:46.373	-28.32.09.33	5.231E+00	2.325E-01	8.747E+00	1.813E+00	1.629E+02
6.000E+00	17:46:46.696	-28.31.56.11	6.481E-01	2.213E-01	1.811E+00	1.457E+00	2.297E+01
7.000E+00	17:46:47.180	-28.32.14.31	7.134E-01	2.183E-01	1.982E+00	1.467E+00	3.658E+01
8.000E+00	17:46:46.312	-28.32.09.34	1.252E-01	2.159E-01	7.000E-01	7.000E-01	5.659E-04

Example: Download Admit Product in ASA



Download and untar *.admit.*.tgz

member.uid___A001_X88f_X25b.HH212_sci.spw33.cube.l.pbcor.admit.native.tgz

Use the ADMIT Weblog to View the Results

Open "**index.html**" with chrome, firefox etc., and review each steps.

Click each step to review the analysis.

LineID_AT: table of identified lines

LineID_AT (taskid=8) vlsr=1e-05 numsigma=5.0 minchan=4 maxgap=

Identified Spectral Lines

frequency	uid	formula	name	transition	velocity	El	Eu
[GHz]	[]	[]	[]	[]	[km/s]	[K]	[K]
3.473E+02	SiO_347.33058	SiO	Silicon Monoxide	8-7	-8.860E+00	5.835E+01	7.502E-

Future in the ASA Search (Some time after 2026)

Lines

Rest Frequency

Line

Redshift

VLSR

SNR

FWHM

Peak Flux

Publications (4544) | **Lines** (45060)

ALMA Source Name	Molecule (formula)	Transition	FWHM	Redshift	VLSR (radio)	Rest frequency	Rest Frequer
ATLAS_C2024_G3	HC3N	J=25-24	2.4535	0.0000	11.938	227.419	224.293..23
ATLAS_C2024_G3	13CS	5-4	2.6737	-0.0000	-0.213	231.221	227.814..23
ATLAS_C2024_G3	SiO	5-4	1.7376	-0.0000	-0.214	217.105	214.891..219
ATLAS_C2024_G3	H2CO	3(1,2)-2(1,1)	1.9478	0.0000	11.938	225.698	223.216..22
ATLAS_C2024_G3	CH3CHOvt=1	21(-1,20)-21(1,21)E	10.0000	0.0000	11.938	226.309	213.569..23
ATLAS_C2024_G3	CH213CHCN	24(13,11)-23(13,10)	1.2901	0.0000	11.938	227.009	225.365..22
ATLAS_C2024_G3	OCSv=0	20-19	2.5891	-0.0000	-0.597	243.218	239.919..24
ATLAS_C2024_G3	HC3N	J=24-23	2.5224	-0.0000	-0.214	218.325	215.111..221
ATLAS_C2024_G3	HDO	3(1,2)-2(2,1)	3.7895	0.0000	11.938	225.897	221.069..23
ATLAS_C2024_G3	SO	6(5)-5(4)	11.7741	-0.0000	-0.213	219.949	204.949..23
ATLAS_C2024_G3	c-HCC13CH	18(11,7)-18(10,8)	1.5856	-0.0000	-0.597	244.254	242.234..24
ATLAS_C2024_G3	CS	5-4	2.7791	-0.0000	-0.597	244.936	241.395..24
ATLAS_C2024_G3	CO	2-1	11.7741	-0.0000	-0.616	230.538	215.538..24
ATLAS_C2024_G3	U_220.2829		10.0000	-0.0000	-0.213	220.283	207.543..23

What is the Current Source of Redshift Used by ADMIT?

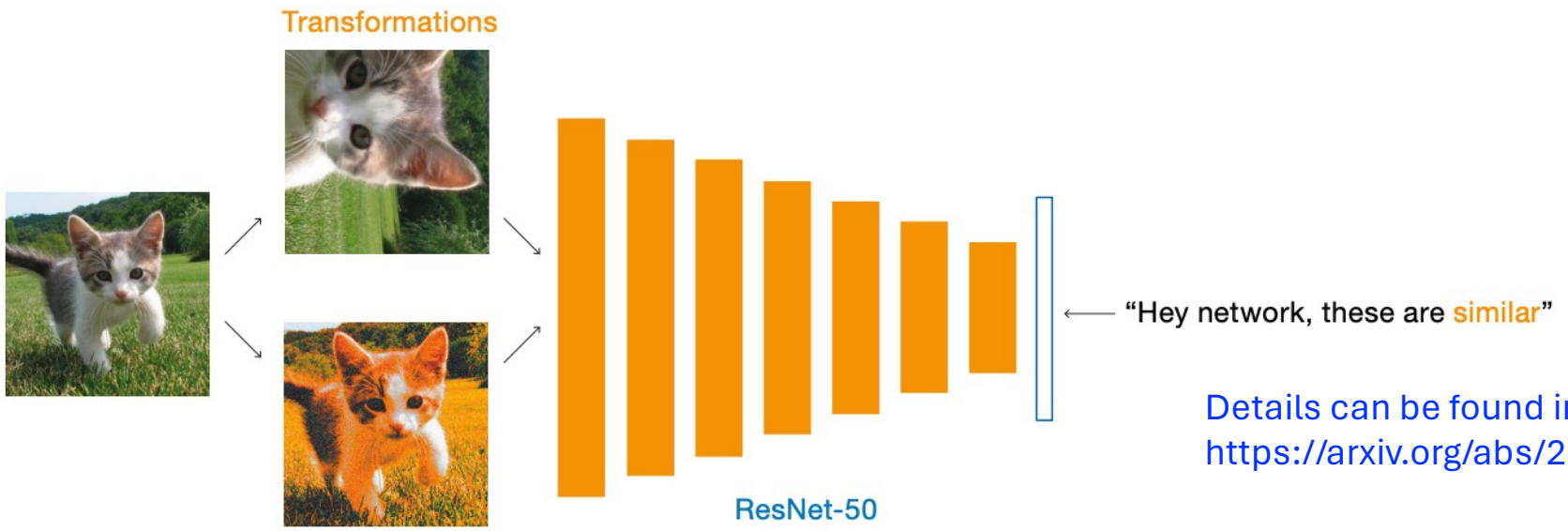
1. We use the redshift given by the PI.
2. If none is given, we go to SIMBAD and NED and find the 'best' matching source.
3. If none is found we try to use the lines that people wanted to observe (if given) and say that this would be 'in the center'.
4. If none is given, we use 0.

- This sometimes works reasonably well. Sometimes gives wrong results. E.g. if the PI put 0 for a high-z source.
- **Please note that uncertainties in VLSR may affect the identification, so we recommend reviewing the results carefully, especially for complex spectra.**

ALMA Similar Images – Model Training

ResNet-50 is a deep neural network that turns an image into a numerical summary of its visual structure.

Teach a model (ResNet-50): Unsupervised machine learning trains ResNet-50 using rotated and cropped images, producing the fingerprints **(**2048-number vector) that describes each image's morphology.



Details can be found in (Stoehr, F. 2025), https://arxiv.org/abs/2511.17061

ALMA Similar Images – Model Training

- **Input images** are **64 × 64 pixel cut-outs**, centered on the **brightest pixel** when the signal-to-noise ratio is greater than 5, and on the **image center** otherwise.
- **Pixel values are normalized.**
- The model is trained on **all continuum preview images (public)** in the ALMA Science Archive.
- A **separate training** is performed on **peak-flux (moment 8) images** from spectral-line cubes.
- During similarity inference, results are **restricted to scientifically meaningful categories**, preventing nonsensical matches (e.g., galactic discs vs. protoplanetary discs).

Details can be found in
https://arxiv.org/abs/2511.17061

Step 1: Model Training

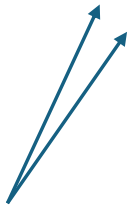
- **1.8 hours → Model training**
- **What:** Training the ResNet-50 model using SimCLR
- **Data:** ~100,000 images
- **Epochs:** 100
- **Hardware: 3 × NVIDIA TITAN RTX GPUs (in ESO)**
- **Purpose:**
  - Teach the model how to represent image morphology
  - Learn invariance to rotation, brightness, cropping, etc.
  - This step **is only done once**.

Step 2: Similarity Database for the Archive Images

Using the ResNet-50 Model…

1.7 hours → Similarity computation for the archive

- **What:** Computing similarity relationships **after the model is already trained**
- **Data:** All public archive images converted to vectors
- **Hardware: 1 × NVIDIA TITAN RTX GPU (in ESO)**
- **Purpose:**
  - Convert images into 1-D vectors
  - Find nearest neighbors to overcome $O(N^2)$ scaling
  - Build the similarity database (1000 images)



It is done **offline** and repeated only when new data are added. Image similarity is measured using a vector distance (cosine distance).

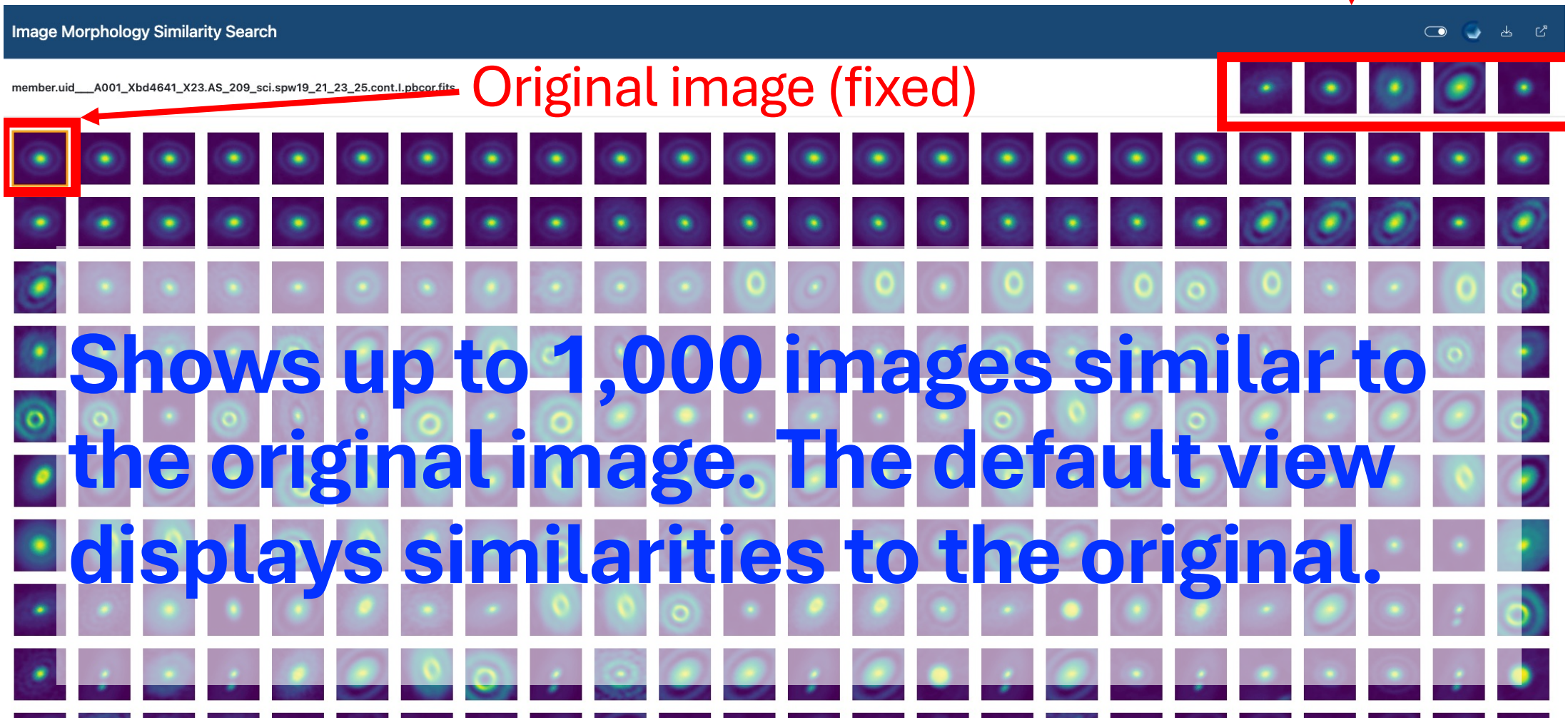
ALMA Similar Images in ASA

- **When you find a data in ASA, click the icon within preview** 

- 1000 similar images will return. In addition, we **group the similar images into 5 major clusters** to show the main types of shapes found.

The current similar image fucntion offers Continuum Maps for now.

Interface of Similar Images in ASA



Five Similar Groups by the Model (quick select)

- It groups the similar images into five morphology.
- These five clusters are based on the selected image's morphology vector.

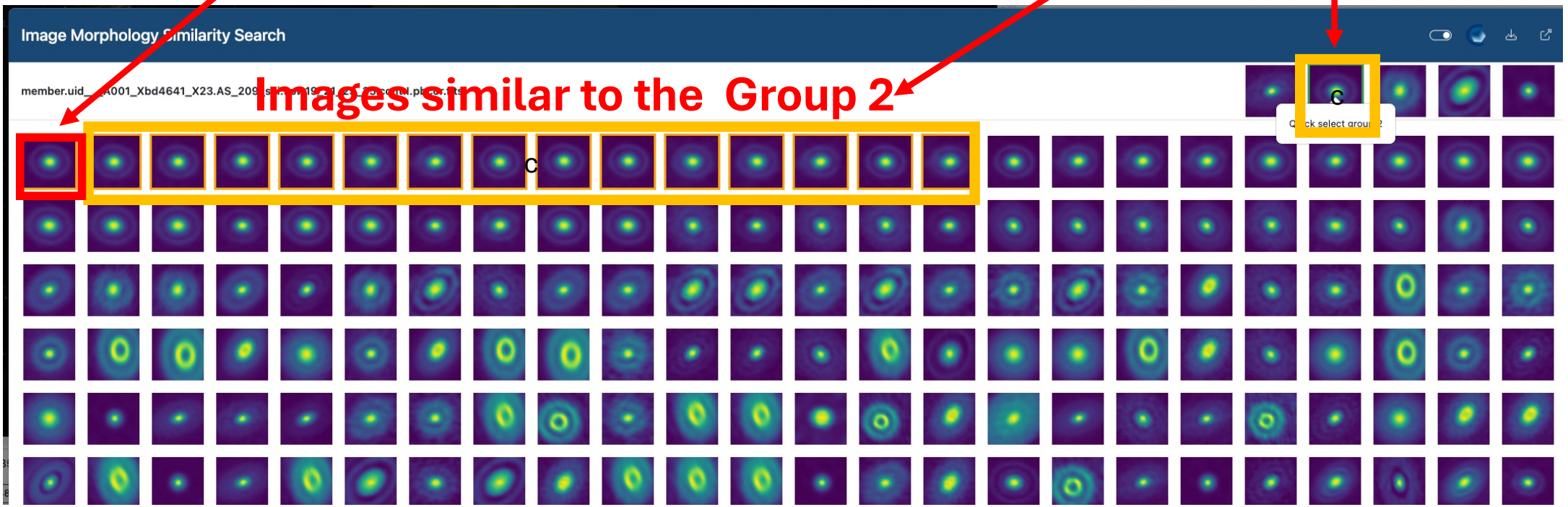
Example 1: Select One Group

- When you select Group 2, the original image and the 14 similar images in that group are highlighted with an orange frame.

Caveats: there is no unique answer

Original image

Click Group 2



Example 2: Users Selectable Conditions

- Instead of selecting the groups, you can select the similar images yourself. The results will update based on the new parameters.