

How to create a custom database for SIRIUS with the compounds already identified for family/genus/species ?

Elnur Garayev

06/06/2022

Open LOTUS



Find natural products

Name, InChI, SMILES, formula, LOTUS id, Wikidata, chemical classification, ...

 Search

[Structure Search](#) | [Advanced Search](#)

[Home](#)

[Compound Browser](#) ▼

[Search](#) ▼

[Download](#)

[Documentation](#)

Natural Products Online is an open source project for Natural Products (NPs) storage, search and analysis. This page hosts LOTUS, the natural prOducTs occUrrence databaSe, one of the biggest and best annotated resources for NPs occurrences available free of charge and without any restriction. LOTUS is a living database which is hosted in parallel at [Wikidata](#) and here. The Wikidata version allows for community curation and addition of novel data. The current version allows a more user friendly experience (such as structural search, taxonomy oriented query, flat table and structures exports).

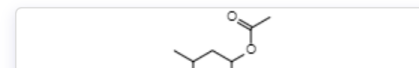
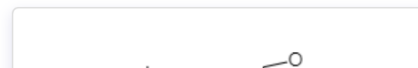
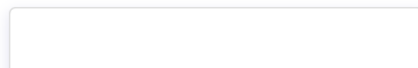
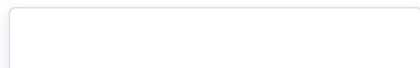
Component Browser

 Cards

 Table

There are 276 518 natural products in the database

« < 1 2 3 4 5 6 7 8 9 10 ... 11521 > »



Specify the genus or species and download SDF



Find natural products

Search

[Structure Search](#) | [Advanced Search](#)

[Home](#)

[Compound Browser](#) ▾

[Search](#) ▾

[Download](#)

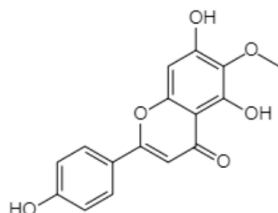
[Documentation](#)

Search Results

Your search for "Inula montana" returned 9 natural products.

Not what you were searching for? Try our [structure](#) or [advanced](#) search for more results

[Download SDF](#)



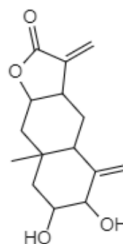
Q15410994

Hispidulin

Mol. formula C₁₆H₁₂O₆

Mol. weight 300.26

Tmp. LOTUS id LTS0135598

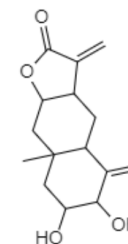


Q105113085

(3ar,4ar,6r,7r,8ar,9ar)-6,7-dihydroxy-8a-methyl-3,5-dimethylidene-octahydronaphtho[2,3-b]furan-2-one

Mol. formula C₁₅H₂₀O₄

Mol. weight 264.32



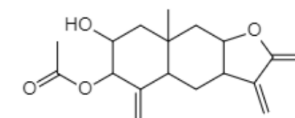
Q82060062

6,7-dihydroxy-8a-methyl-3,5-dimethylidene-octahydronaphtho[2,3-b]furan-2-one

Mol. formula C₁₅H₂₀O₄

Mol. weight 264.32

Tmp. LOTUS id LTS0087074



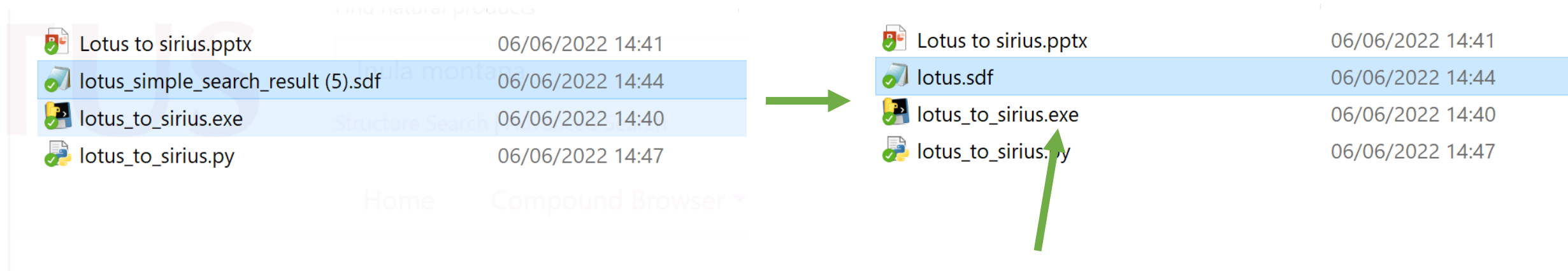
Q105255497

(3ar,4ar,6r,7r,8ar,9ar)-7-hydroxy-8a-methyl-3,5-dimethylidene-2-oxo-octahydronaphtho[2,3-b]furan-6-yl acetate






Mol. formula C₁₇H₂₂O₅

Mol. weight 306.35

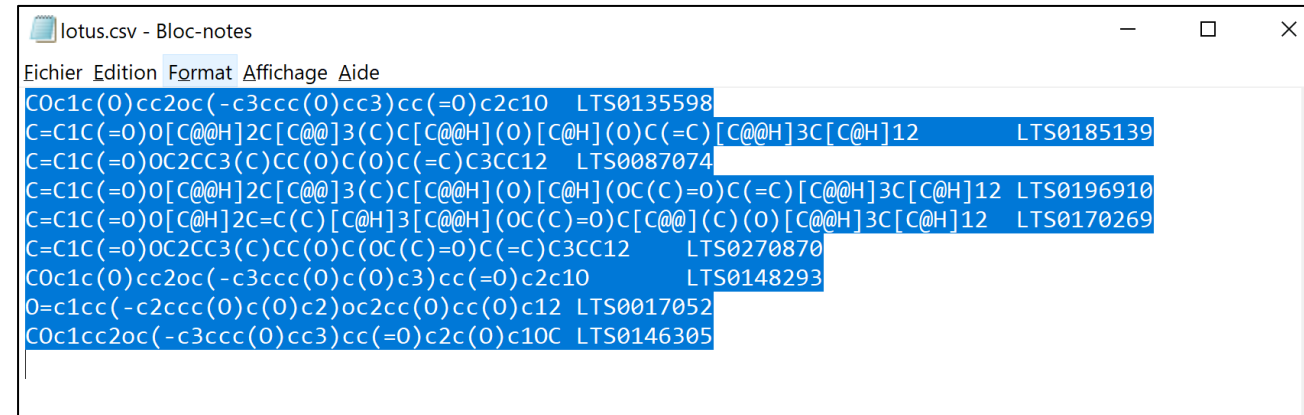
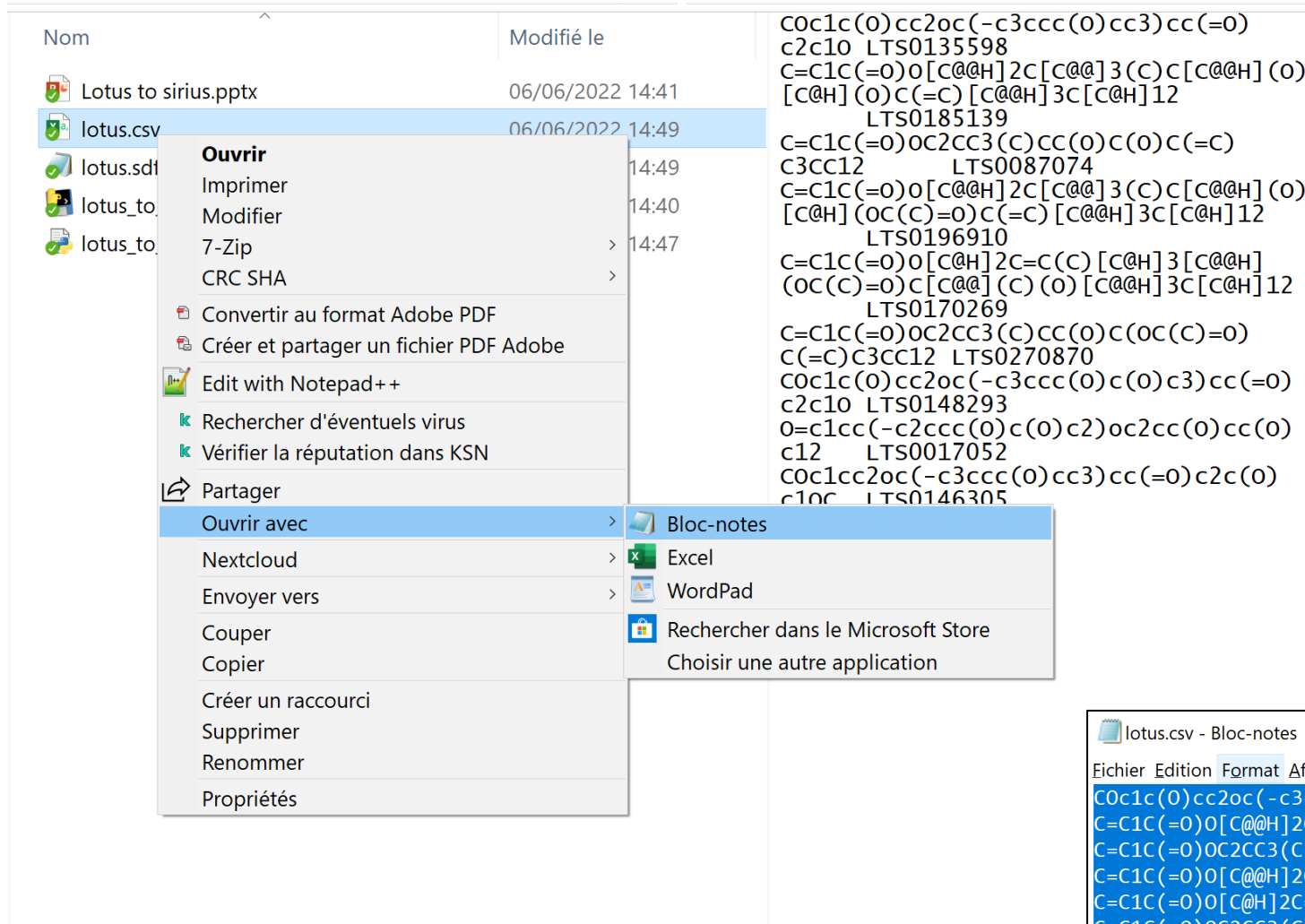
Rename the sdf file to lotus.sdf and open exe file



csv file with SMILES and lotus ID is generated

Nom	Modifié le	
 Lotus to sirius.pptx	06/06/2022 14:41	
 lotus.csv	06/06/2022 14:49	<chem>COC1C(O)CC2OC(-C3CCC(O)CC3)CC(=O)c2c1O</chem> LTS0135598
 lotus.sdf	06/06/2022 14:49	<chem>C=C1C(=O)O[C@@H]2C[C@@]3(C)C[C@@H][C@H](O)C(=C)[C@@H]3C[C@H]12</chem> LTS0185139
 lotus_to_sirius.exe	06/06/2022 14:40	<chem>C=C1C(=O)OC2CC3(C)CC(O)C(O)C(=O)C3CC12</chem> LTS0087074
 lotus_to_sirius.py	06/06/2022 14:47	<chem>C=C1C(=O)O[C@@H]2C[C@@]3(C)C[C@@H][C@H](OC(C)=O)C(=C)[C@@H]3C[C@H]12</chem> LTS0196910
		<chem>C=C1C(=O)O[C@@H]2C=C(C)[C@H]3[C@@H](OC(C)=O)C[C@@](C)(O)[C@@H]3C[C@H]2</chem> LTS0170269
		<chem>C=C1C(=O)OC2CC3(C)CC(O)C(OC(C)=O)C(=O)C3CC12</chem> LTS0270870
		<chem>COC1C(O)CC2OC(-C3CCC(O)C(O)C3)CC(=O)c2c1O</chem> LTS0148293
		<chem>O=C1CC(-C2CCC(O)C(O)C2)OC2CC(O)CC(C12)</chem> LTS0017052
		<chem>COC1CC2OC(-C3CCC(O)CC3)CC(=O)c2c(C1OC)</chem> LTS0146305

Open csv file with bloc-note and copy the content



Open SIRIUS and click on Databases

SIRIUS 5.5.3 on Project: 'C:\Users\Elnur\AppData\Local\Temp\sirius-tmp-project-94738263-fedd-4a06-89d8-5d323bd7d739'

Buttons: New, Open, Save As, Save Copy, Import Compound, Import, Summaries, FBMN Export, Compute All, **Databases**, Jobs, Log, Settings, Webservice, Account, Help, About

Filter: Hit enter to search...

LC-MS | Formulas | Spectra | Trees | Predicted Fingerprint | Structures | Substructure Annotation | Compound Classes

XLogP: 0,00 | Similarity: 100 | Filter: Hit enter to search

☐ additional ☐ ALOE ☐ Bio Database ☐ Biocyc ☐ CHEBI ☐ COCONUT ☐ EcoCyc Mine ☐ GNPS ☐ HMDB ☐ HSDB ☐ INULA ☐ KEGG ☐ KEGG Mine ☐ KnapSack ☐ Lipid ☐ Maconda ☐ MeSH ☐ Natural Products ☐ NORMAN ☐ Plantcyc ☐ PubChem ☐ PubMed ☐ Training Set ☐ YMDB ☐ YMDB Mine ☐ ZINC bio

Rank	Name	SMILES	Molecular Formula	Adduct	CSI:FingerID Score	Similarity	#PubMed IDs	XLogP	InChIKey	Lipid Class
1	Caffeate	<chem>C1=CC(=C(C=C1)C=CC(=O)O</chem>	C ₉ H ₈ O ₄	[M + H] ⁺	-2.348	100,000%	84,1280712,1368360,13...	0,970	QAIPRVGONGVQAS	Lipid Class

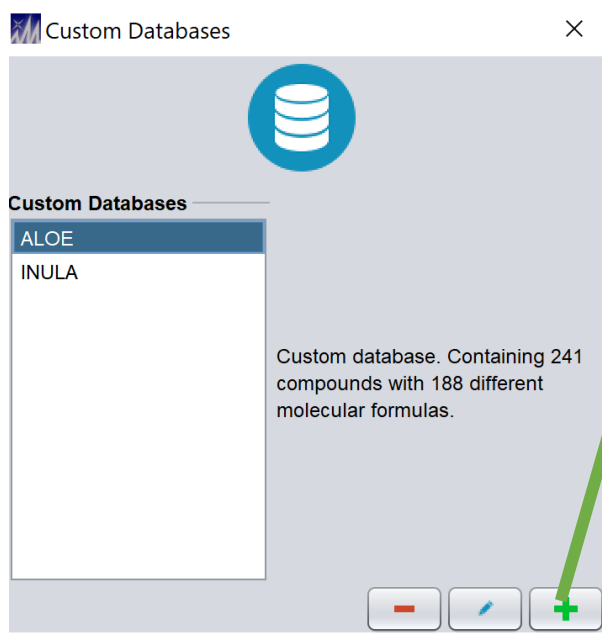
Mode: MS2 | merged

Relative Intensity vs m/z

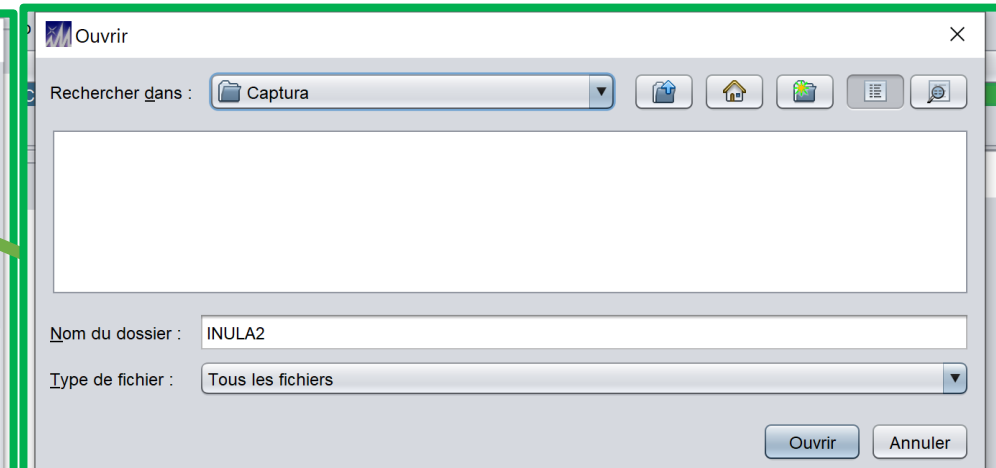
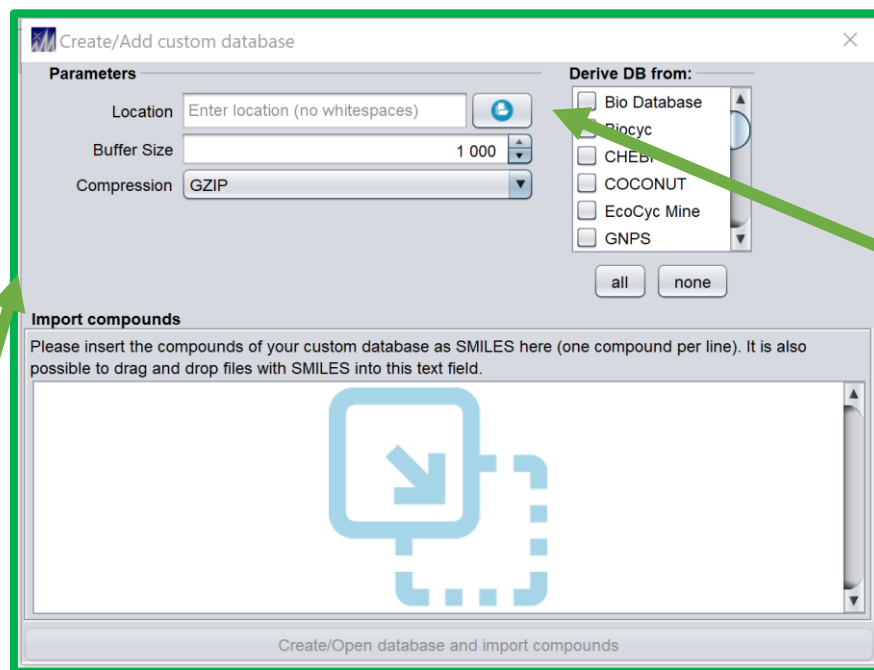
Chemical Structure: Caffeate

Formula: C₉H₈O₄ + H⁺
Intensity: 0.0000
m/z: 181.0493
Mass deviation: +0.1346 mDa (+0.7432 ppm)
Total fragmenter score: 72.83
Fragment score: 5.95

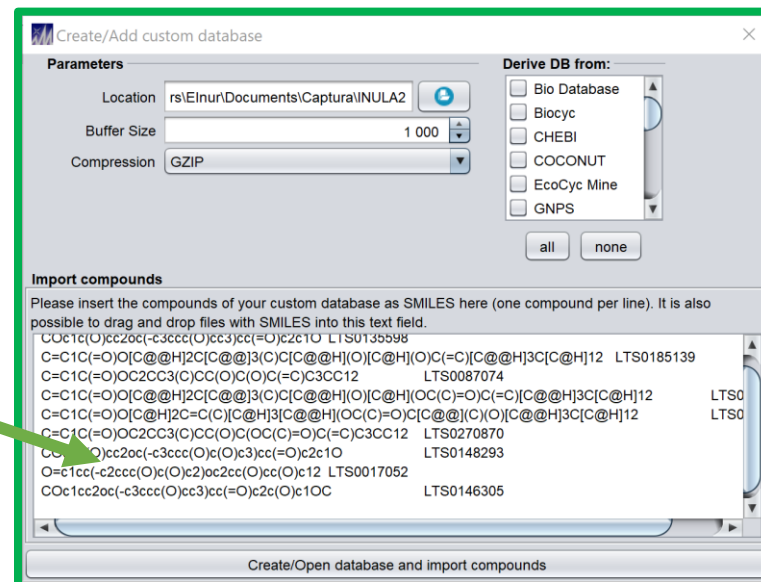
1. Add new database



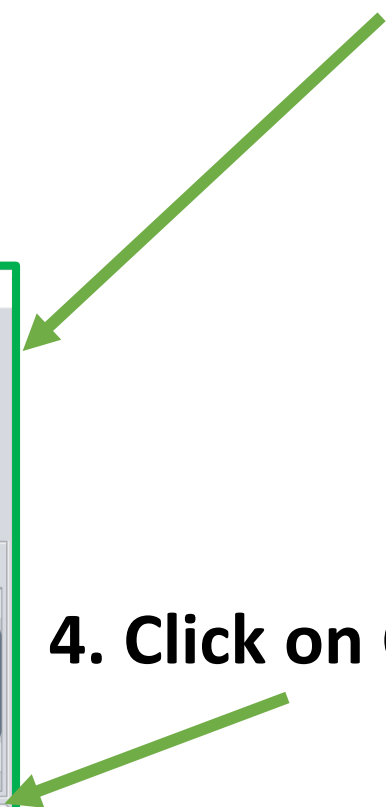
2. Specify the location without spaces in the filename



3. Paste the content of csv file into the window



4. Click on Create



Database is created. Enjoy ;)

