

EssoilDB: A semantic phytochemical knowledgebase

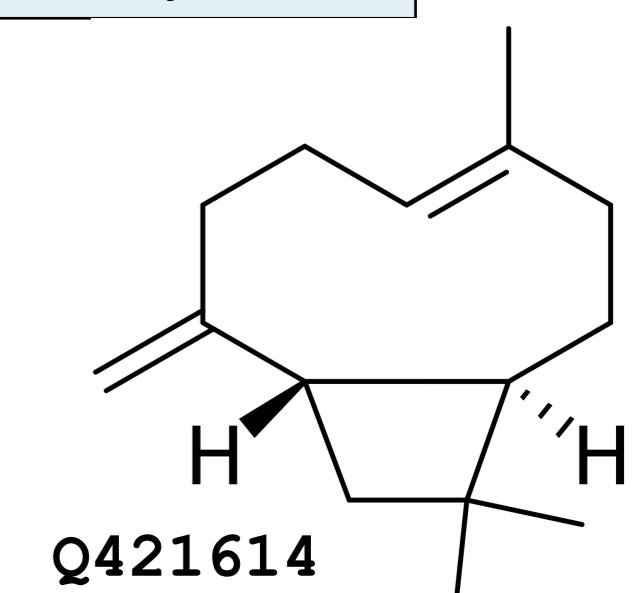
Vinita Lamba¹, Shruthi M¹, Manish Kumar¹, Ambarish Kumar¹, Gitanjali Yadav^{1,2}, Peter Murray-Rust^{2,3,a}

¹ National Institute of Plant Genome Research, New Delhi, IN; ² University of Cambridge, UK; ³ ContentMine Ltd, Cambridge, UK

OVERVIEW. Plants are capable of generating a diverse spectrum of volatile organic compounds for maintaining communication and fruitful adaptive interaction with their environment. These are proven to be important protective and signalling molecules & constitute a platform for plant–plant interaction. For 13 years, EssoilDB in NIPGR, has captured the plant literature on terpenes (class of phytochemicals).



BACKGROUND: In 2019-02 [1] GY, VL, AK collaborated with ContentMine/PMR to run a TIGR²ESS workshop on mining the plant literature. We deployed Open **text-mining** tools (ContentMine's **getpapers**, **AMI**) and used Wikipedia/Wikidata to create WikiFactMine **dictionaries** (e.g. pests, microorganisms, soiltypes, invasive species, spices...) and showed that Open semantic text-mining was robust and could be deployed in many areas of Plant Science.



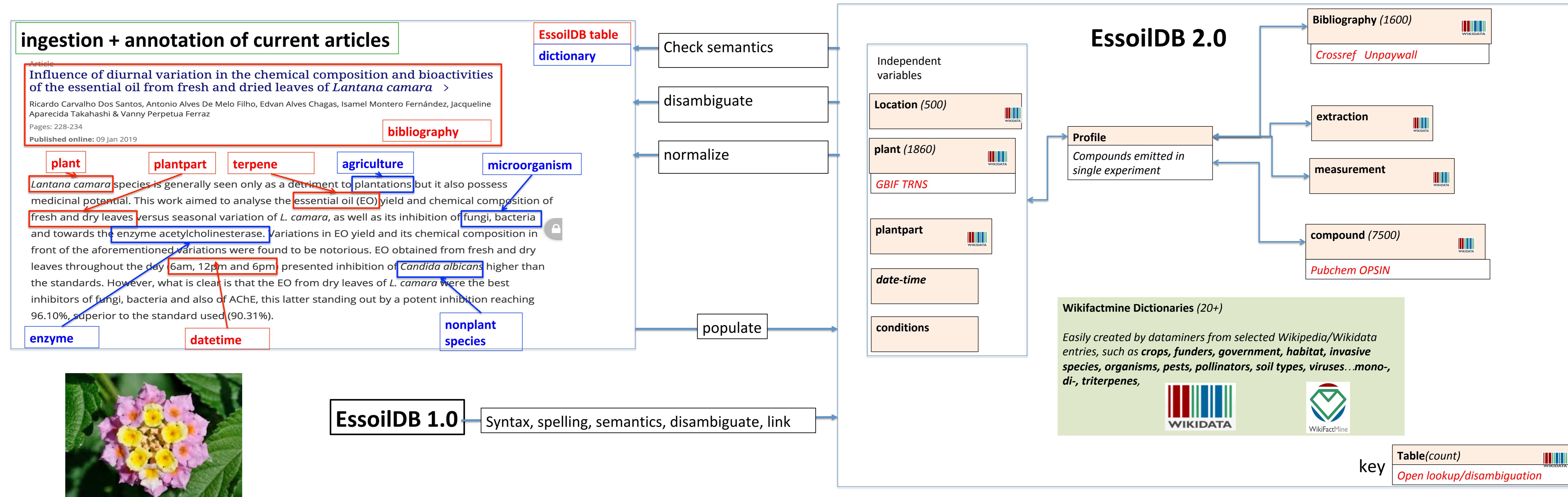
SEMANTICS: Wikidata (a Wikimedia project) has over 50 million **items** of Open semantic knowledge. **Items** have unique identifiers (**Q numbers**) related by **properties (P)**. Example: “the invasive species *Lantana camara* produces beta-caryophyllene” is represented by the triples:

Q332469 P5626 "56" (L. camara hasGISDId 56 – (GlobalInvasiveSpeciesDatabase))
Q421614 P1528 Q332469 (beta-caryophyllene isNaturalProductOf L. camara)

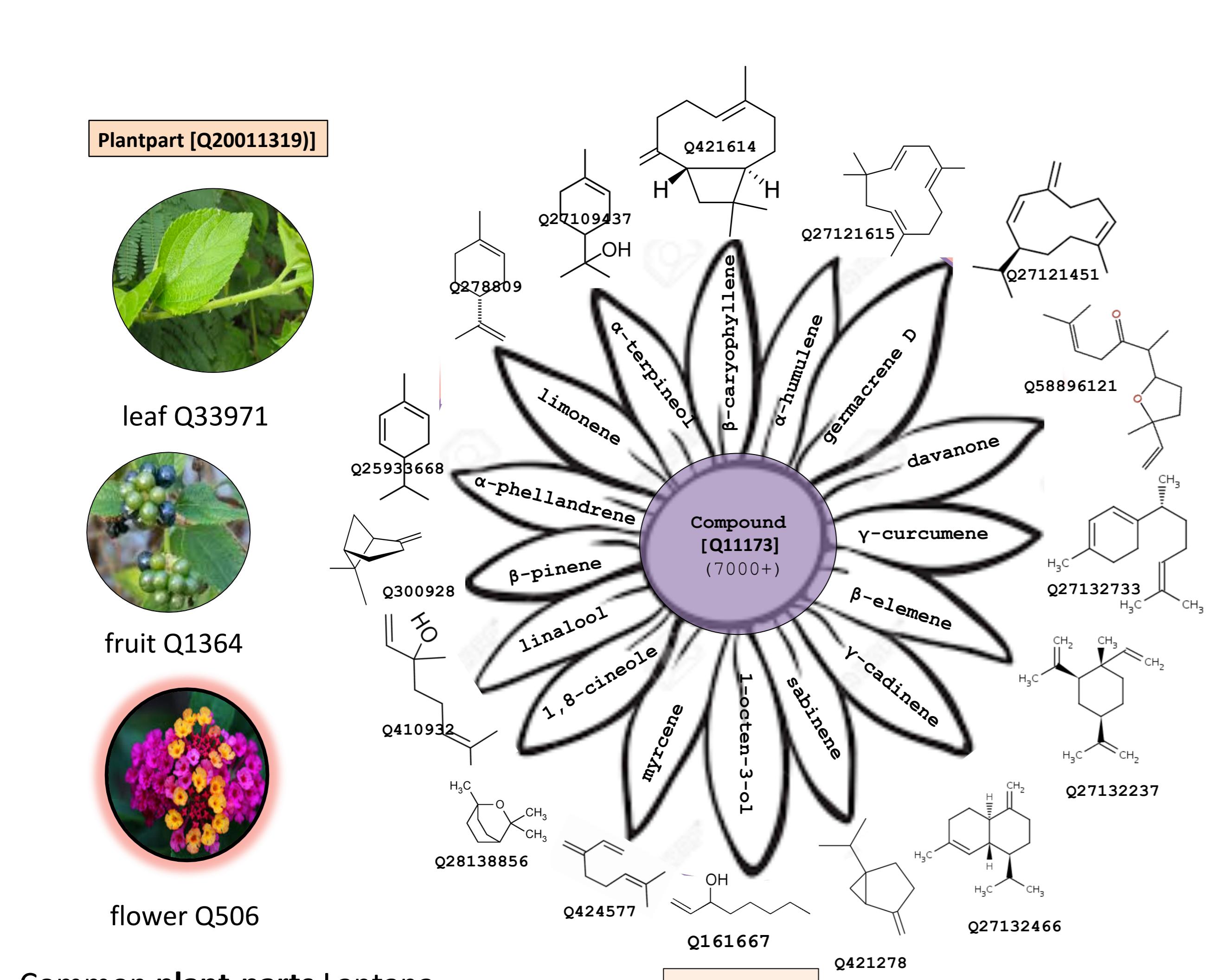
The new EssoilDB 2.0

We use EssoilDB2.0 to support text-mining ...

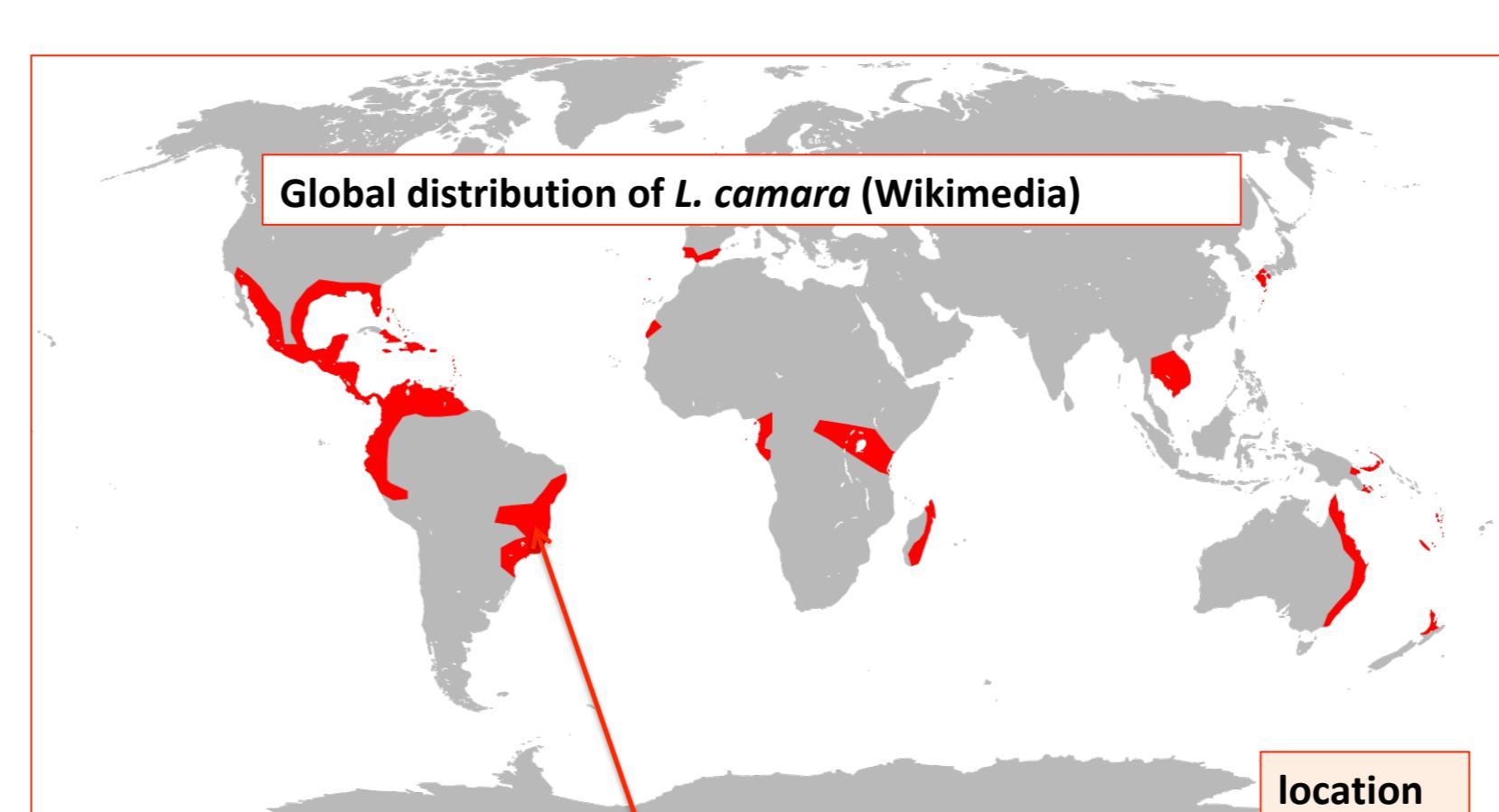
... and text-mining to populate EssoilDB2.0!



Knowledge Components in EssoilDB 2.0 (illustrated by *L. camara*)



Common plant-parts Lantana

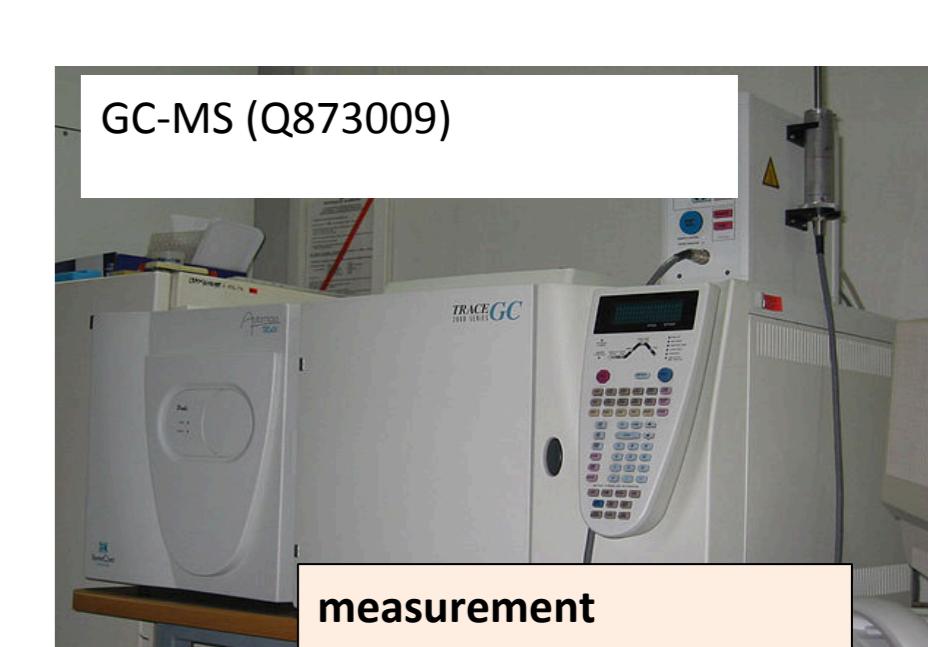
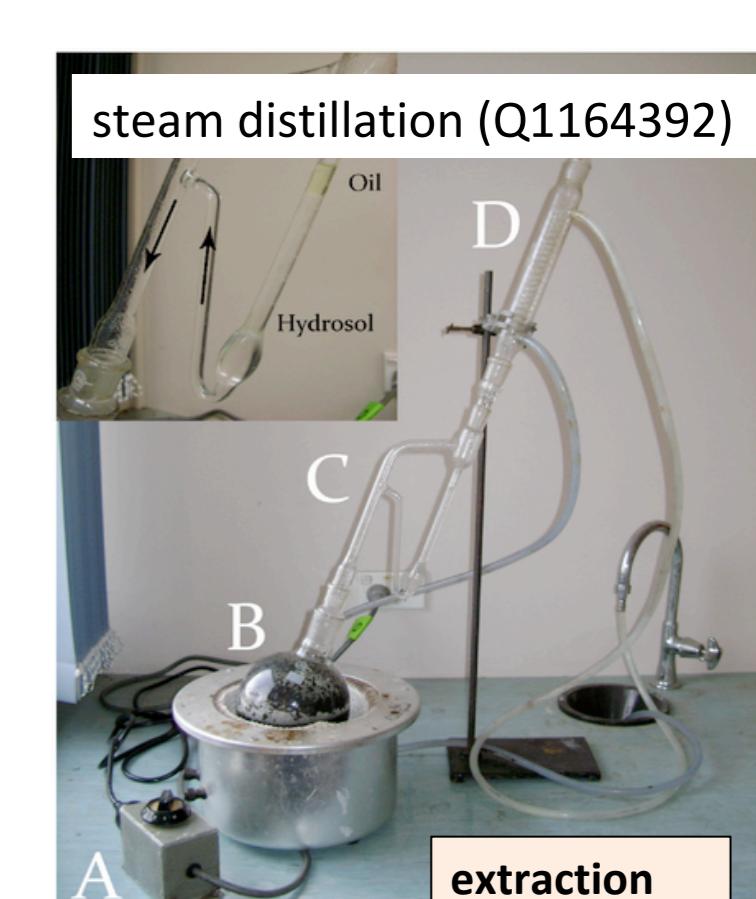


Essential Oil Composition of Two *Lantana* Species from Mountain Forests of Pernambuco (Northeast of Brazil)

José C. S. de Oliveira, Ilzenayde A. Neves, Claudio A. Q40942 & Manfred O. E. Schwartz
Pages 530-532 | Received 01 Mar 2007, Accepted 01 Jul 2007, 8 Dec 2011
Download citation | https://doi.org/10.1080/10412905.2008.9700080

bibliography

Profile	Fresh leaves (%)		Dried leaves (%)		
	6am	12pm	6pm	12pm	6pm
Sabinene	0.9	1.1	0.9	0.8	0.3
β-Pinene	0.2	0.2	0.2	0.2	-
β-Elemene	3.3	4.1	3.8	4.2	3.1
β-Caryophyllene	12.8	13.0	10.9	8.9	9.5
Humulene	2.9	3.2	2.9	2.8	2.4
Germacrene D	28.9	25.6	30.6	27.6	36.1



Queries that EssoilDB2.0/Wikidata can support

- Do invasive species produce Sesquiterpene toxins?
- Can Chemical fingerprints be identified for Invasives across Habitats?
- Who funds, or is likely to fund Terpene research?
- What new Chemicals have been reported in 2019 preprints?
- Identify plants that produce chemicals similar to known medicinal drugs.

References and Acknowledgements

- TIGR2ESS workshop : <https://tigr2ess.globalfood.cam.ac.uk/news/adventures-r-reflections-tigr2ess-workshop-r-genomics-and-data-mining>
- ContentMine.org software: <http://github.com/contentmine> and : <http://github.com/petermr/normami>
- <https://www.wikidata.org/wiki/WikiData:WikiFactMine>
- http://www.all4export.com/index.php?route=product/product&product_id=411
- Images from Wikimedia Commons; Licence CC0 and CC BY-SA
- GY, VL, AK, MK thank NIPGR for financial support.
- GY Thanks the DBT-Cambridge program for Funding (University of Cambridge, Dept of Biotechnology, Govt of India)