

EssoilDB: A semantic phytochemical knowledgebase

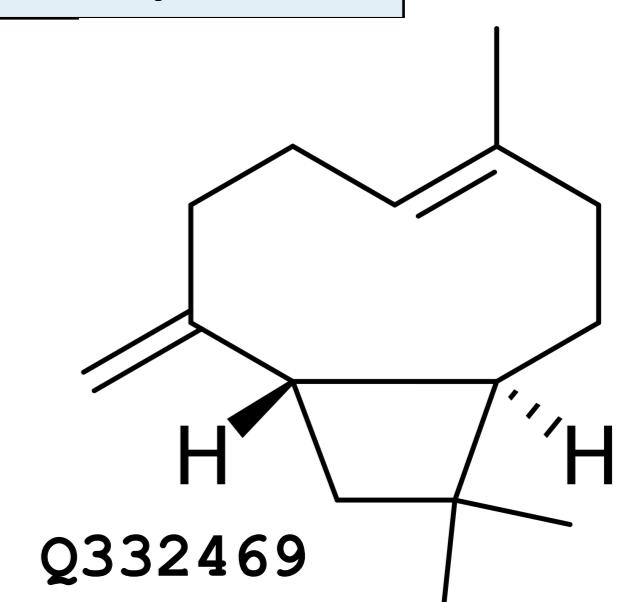
Vinita Lamba¹, Shruthi M¹, Manish Kumar¹, Ambarish Kumar¹, Gitanjali Yadav^{1,2}, Peter Murray-Rust^{2,3, a}

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



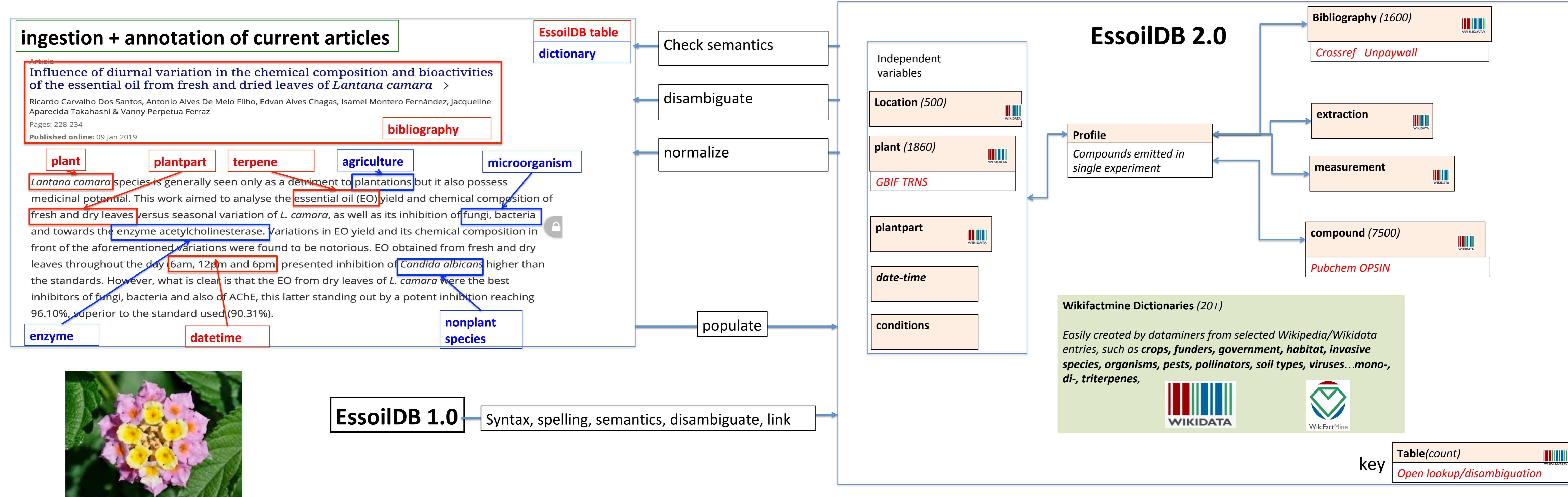
SEMANTICS: Wikidata (a Wikimedia project) has over 50 million **items** of Open semantic knowledge. **Items** have unique identifiers (**Q numbers**) related by **properties (P)**. Example: “the invasive species *Lantana camara* produces beta-caryophyllene” is represented by the triples:

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

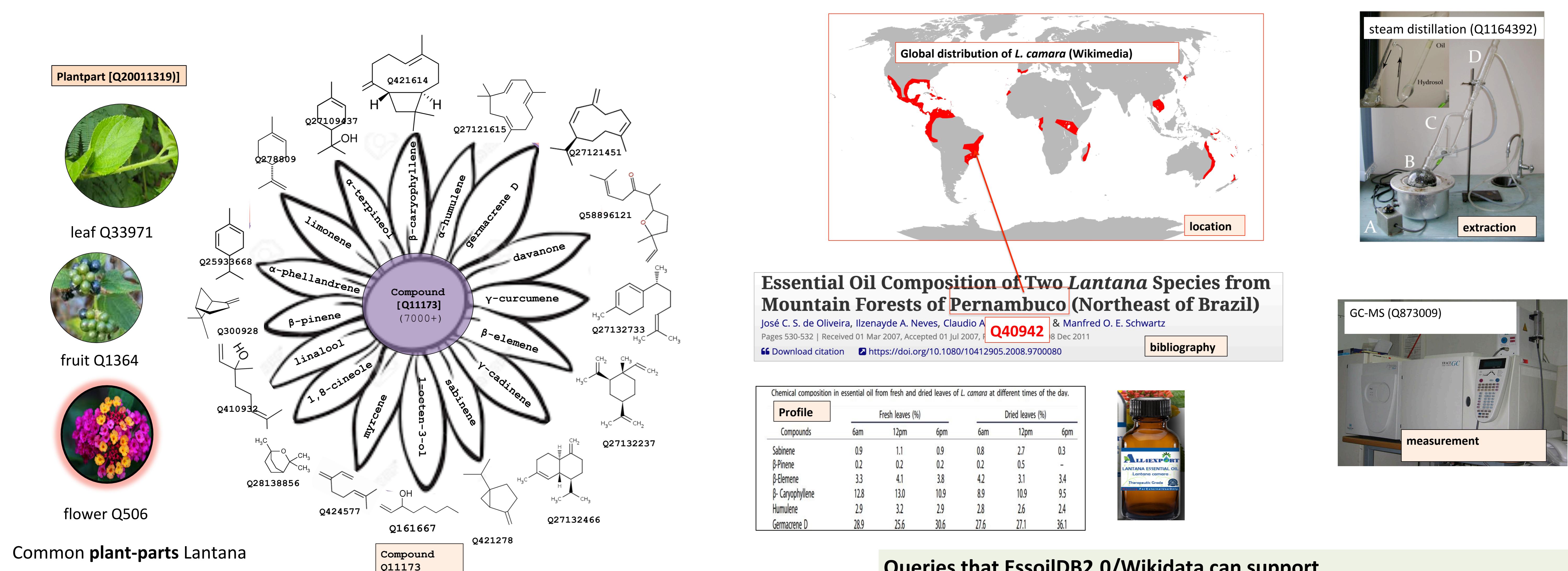
The new EssoilDB 2.0

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*)



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?