



GENERAL INTRODUCTION

our vision

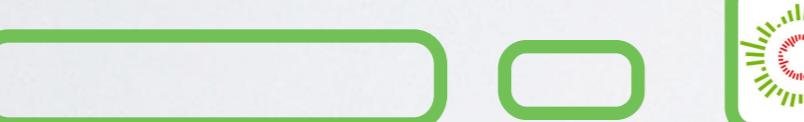
our team

projects and collaborations

CONTENT MINE

OPEN KNOWLEDGE

many researchers want their work to be read and re-used by others



~80% of academic research is publicly funded



we believe that charitably and publicly funded research should be available to all at no additional cost



THE RIGHT TO READ IS
THE RIGHT TO MINE

*The Hargreaves report (UK)
allowing limitations and
exceptions for non-commercial
content mining research*



OUR MISSION

“make 100,000,000 facts
from the scholarly literature
open, accessible and reusable”



THE SCALE OF THE TASK

- ~ 27,000 peer reviewed journals*
- > 5,000 publishers
- ~ 3,000 new papers per day

*Ulrich's database: <http://ulrichsweb.serialssolutions.com/login>



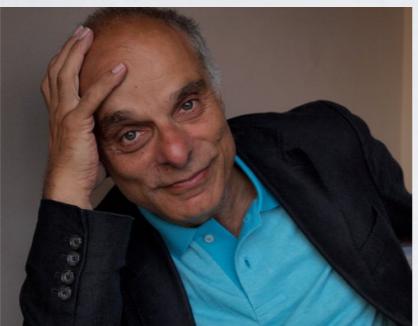
STRUCTURED INFORMATION

- chemical names and structures
- species
- metabolism
- phylogenetic trees

CONTENT MINE

OUR TEAM

Charles Oppenheim



@CharlesOppenh

Ross Mounce



@rmounce

Graham Steel



@McDawg

Mark MacGillivray



@cottagelabs

Richard Smith-Unna



@blahah404

Jenny Molloy



@jenny_molloy

Peter Murray-Rust



@petermurrayrust

Stephanie Smith-Unna



@treblesteph



COLLABORATIONS

- Mint phylogeny working group
- Phyloinformatic Literature Unlocking Tools (PLUTO)
- EBI - MetabоЛights
- OpenFarm
- OpenOil / OpenCorporates

MINT PHYLOGENY WORKING GROUP

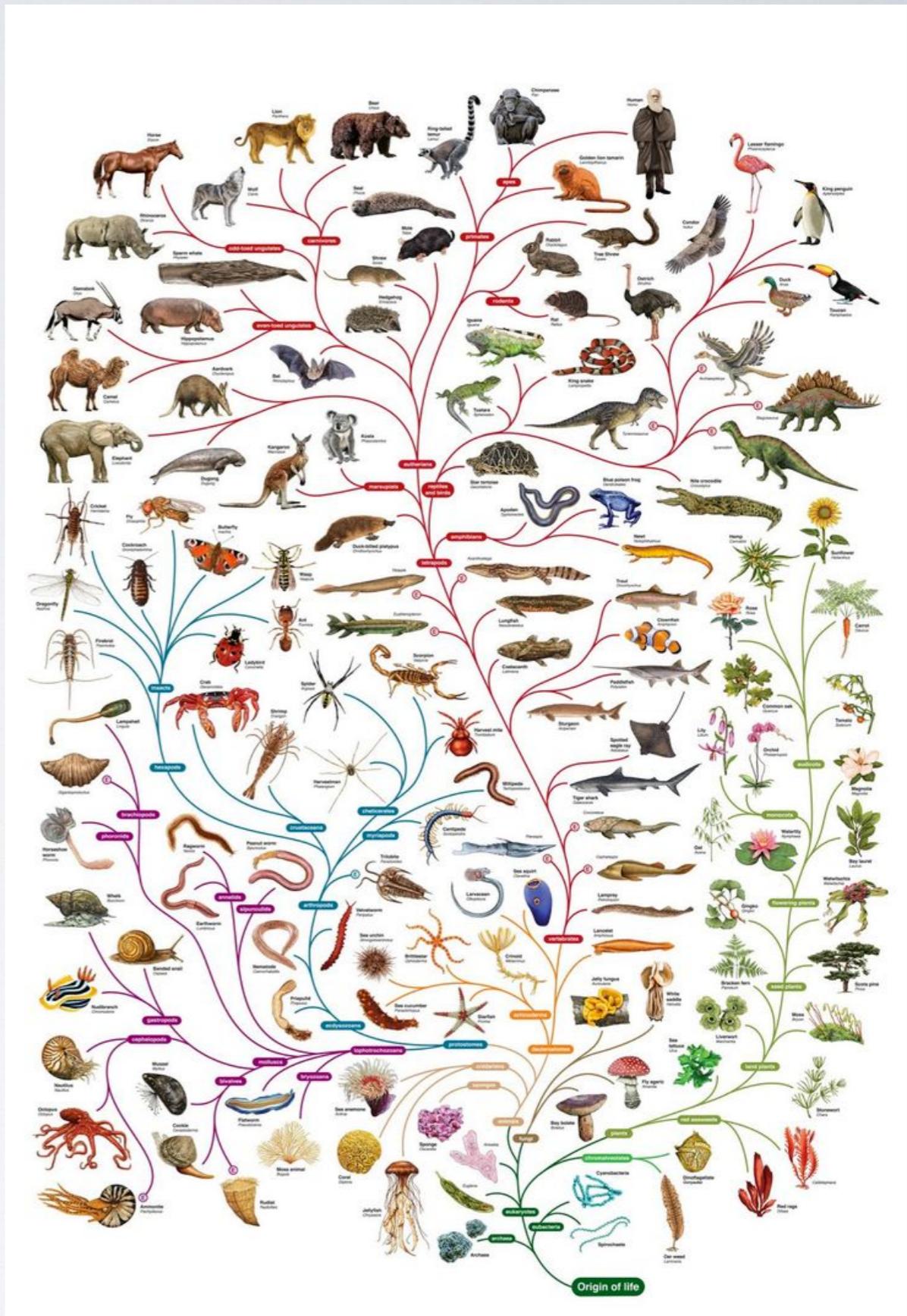


- Sixth largest family of flowering plants
- Major economic and cultural importance
- Studying coevolution of their chemical compounds and genomes
- We are helping to extract these phytochemicals using AMI-species
- Bryan Drew, Douglas Soltis, Pamela Soltis

CONTENT MINE

PLUTO

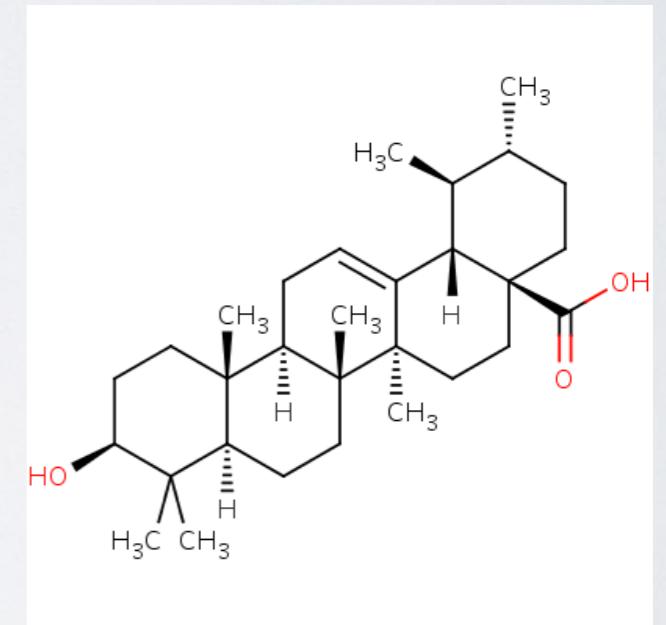
- Phyloinformatic Literature Unlocking Tools (PLUTo)
- Phylogenetic information is scattered across >100,000 papers
- Hard to synthesise manually
- PDF >> extract phylogenetic trees
>> reusable data in standardised format (Newick trees)





METABOLIGHTS

- European Bioinformatics Institute
- database for metabolomics experiments and derived information
- cross-species, cross-technique, structures, biological roles, locations, concentrations
- <http://www.ebi.ac.uk/metabolights/>





OPENFARM

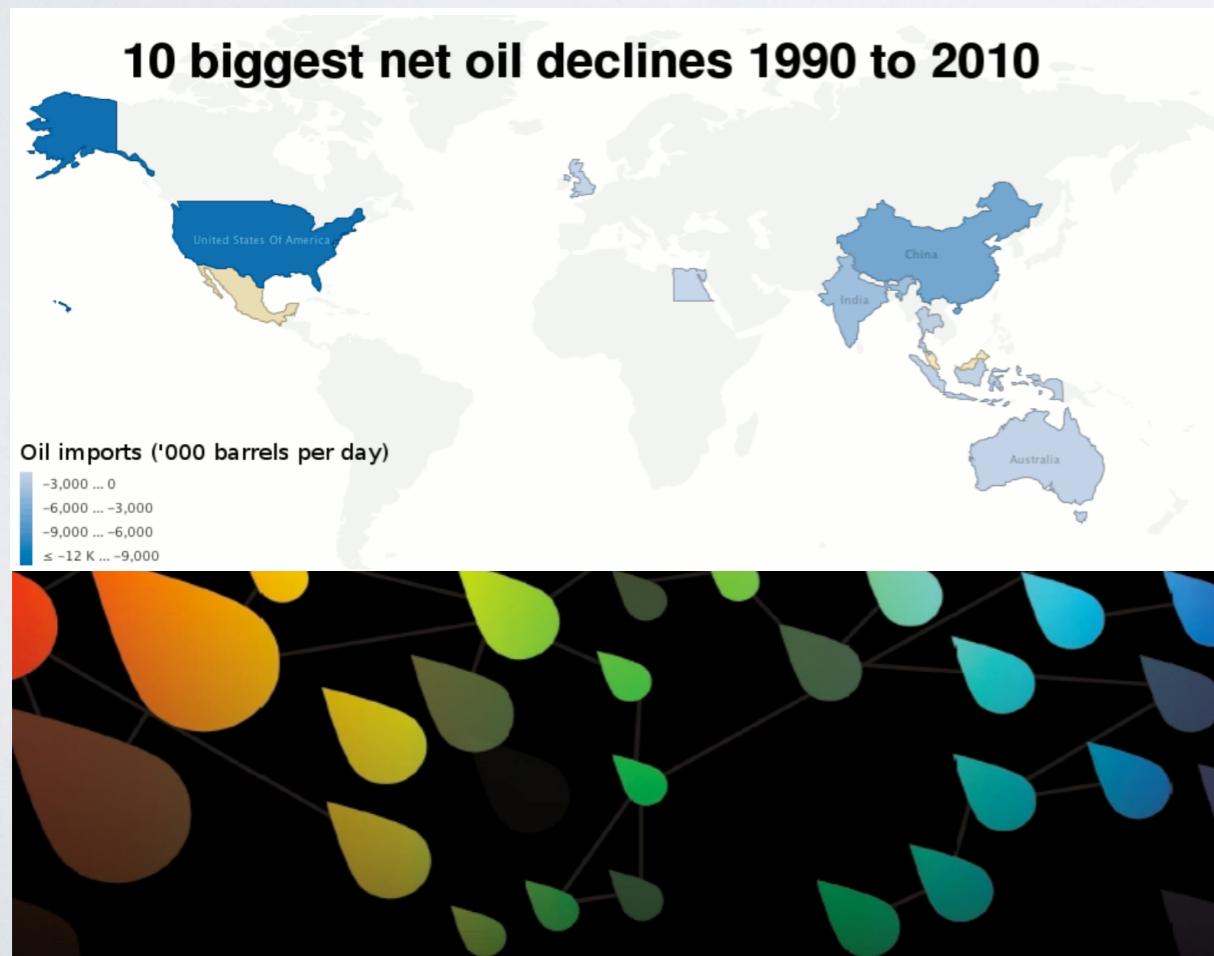
- community-driven open knowledge base of plant information related to agriculture and horticulture
- we provide technical and social infrastructure for high throughput fact extraction
- <https://openfarm.cc/>

openfarm



CONTENT MINE

OPENOIL / OPENCORPORATES



- creative, practical, socially progressive policy-making
- corporate network mapping
- we provide the crawl/scrape/normalise/extract pipeline
- <http://openoil.net/>

