A community-driven paired data platform to accelerate natural product mining

Justin J.J. van der Hooft et al.

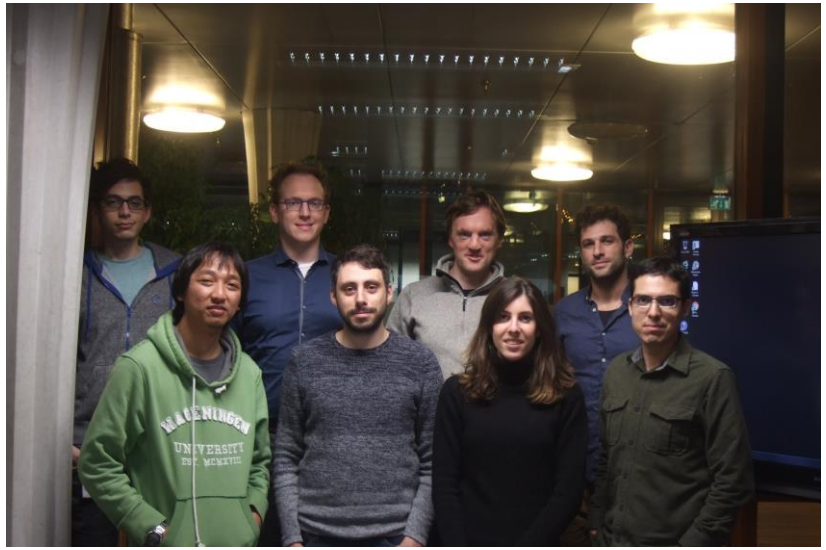
Bioinformatics Group – Wageningen University, The Netherlands

Wageningen, 12 December 2018

WAGENINGEN
UNIVERSITY & RESEARCH

100years

Team work! ☺

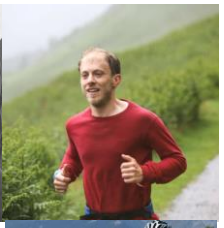


Medema lab - Wageningen UR, NL

NL eScience Center

Dorrestein lab – San Diego, USA

Glasgow Polyomics – University of Glasgow, UK



WAGENINGEN UNIVERSITY & RESEARCH

€€ Funding €€ NWO eScience center

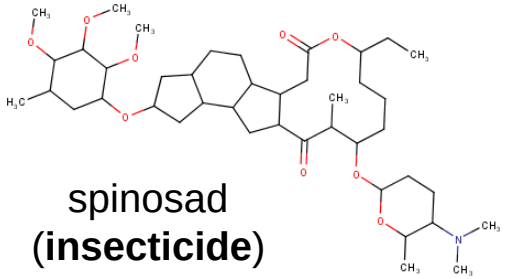
Glasgow Polyomics
www.glasgow.ac.uk/polyomics

University of Glasgow

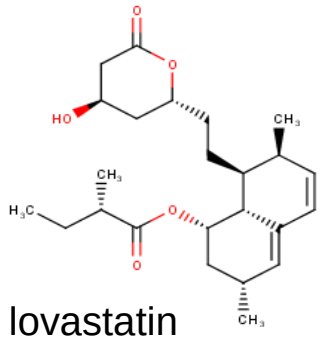
University of Strathclyde Glasgow

The challenge in metabolomics….

Nature produces a large & diverse arsenal of high-value molecules:

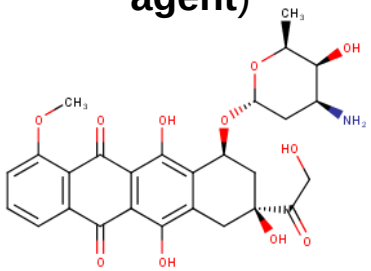


spinosad (**insecticide**)

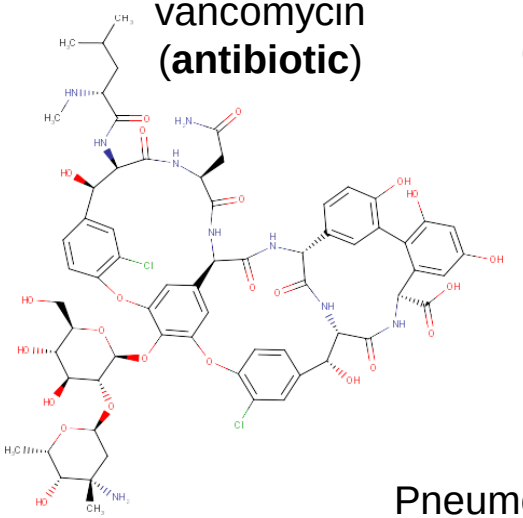


lovastatin (**cholesterol lowering agent**)

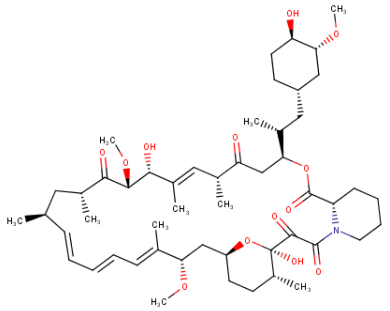
doxorubicin (**chemotherapeutic agent**)

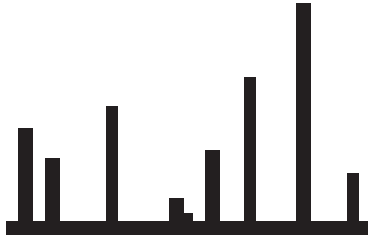


vancomycin (**antibiotic**)



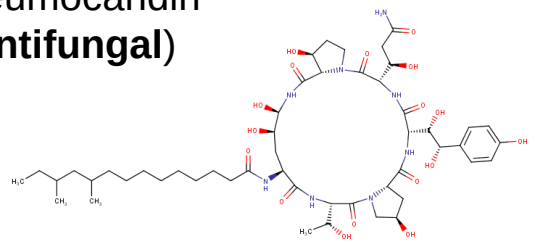
rapamycin (**immunosuppressant**)





Mass spectrometry fragmentation spectrum

Pneumocandin (**antifungal**)

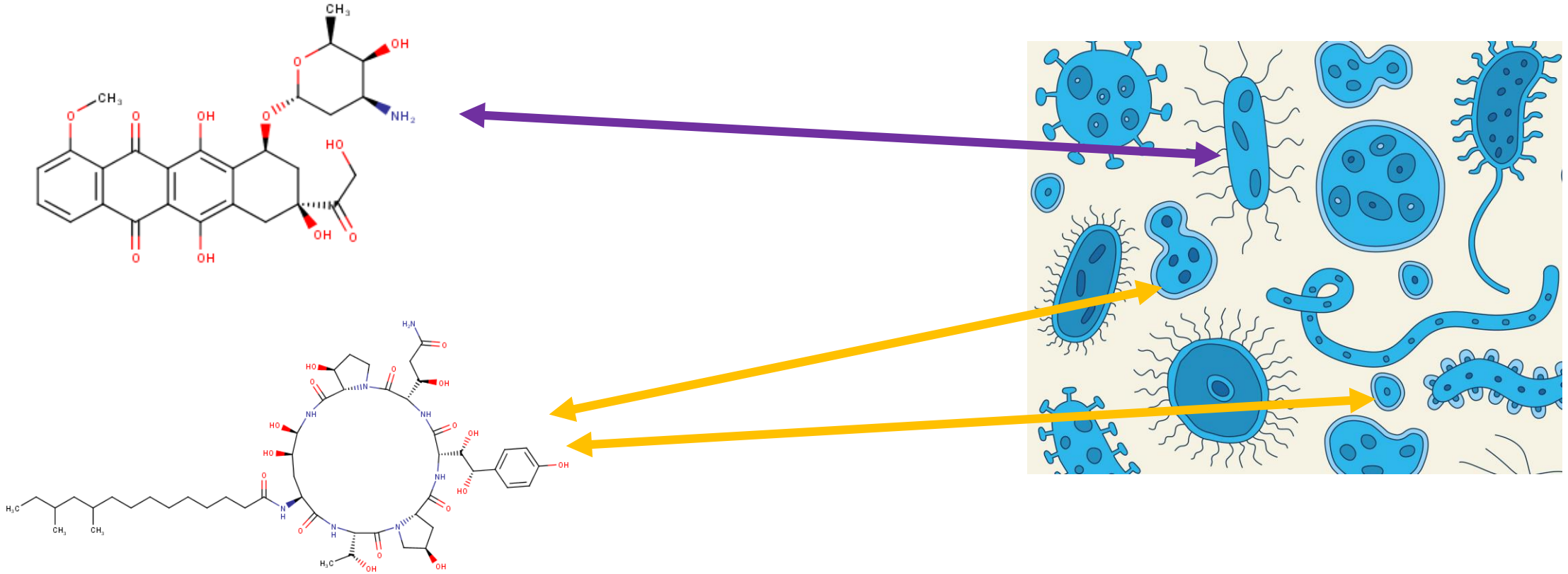


….is **large-scale coupling of spectral data to molecular structures**

of known & especially **novel** natural products molecules

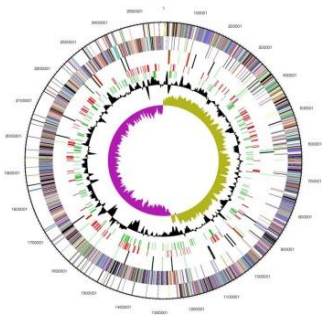
WAGENINGEN UNIVERSITY & RESEARCH

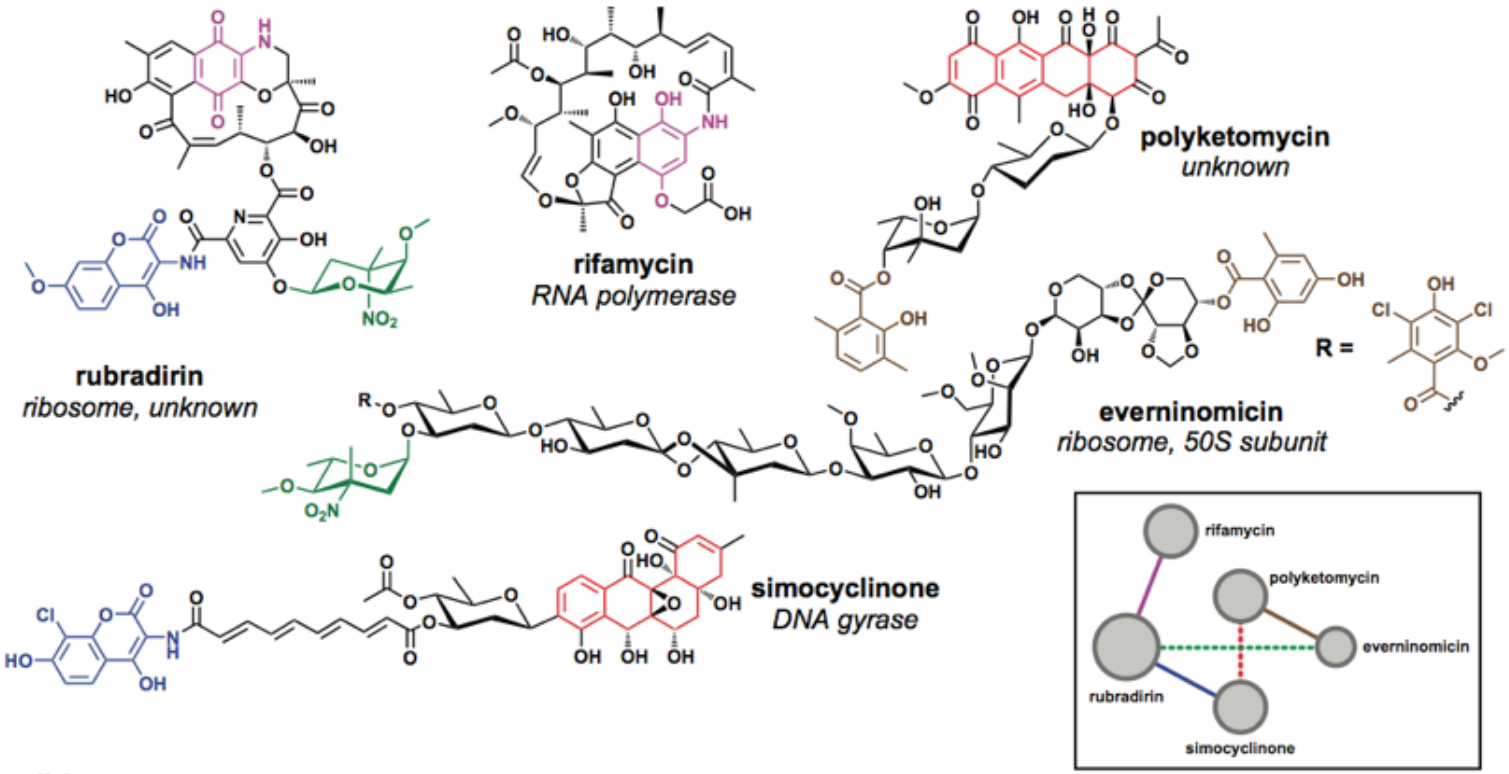
iOMEGA project:
integrated omics for metabolomics and genomics annotation

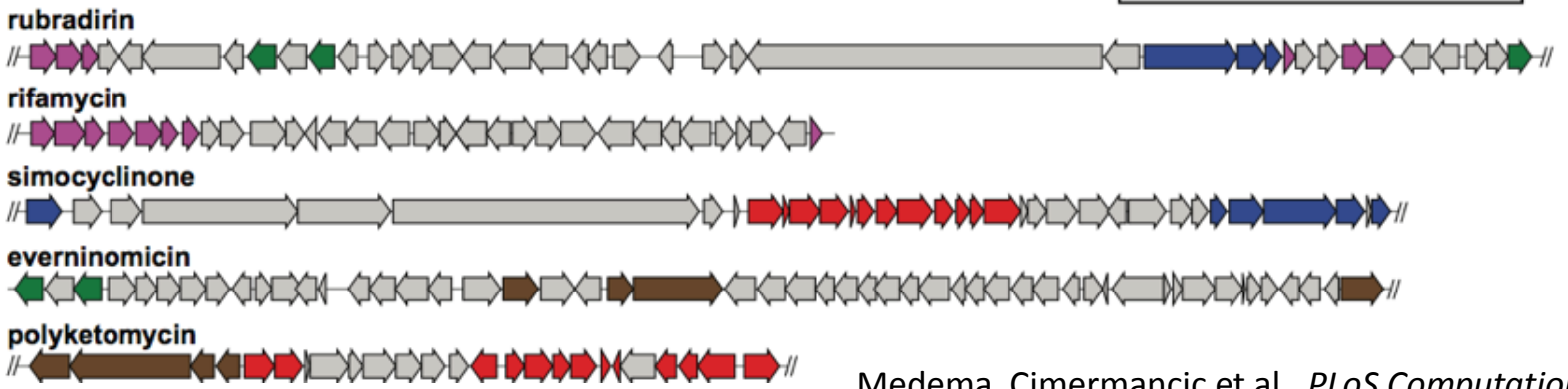


WAGENINGEN
UNIVERSITY & RESEARCH

Biosynthetic gene clusters: key to mining genomes for chemistry



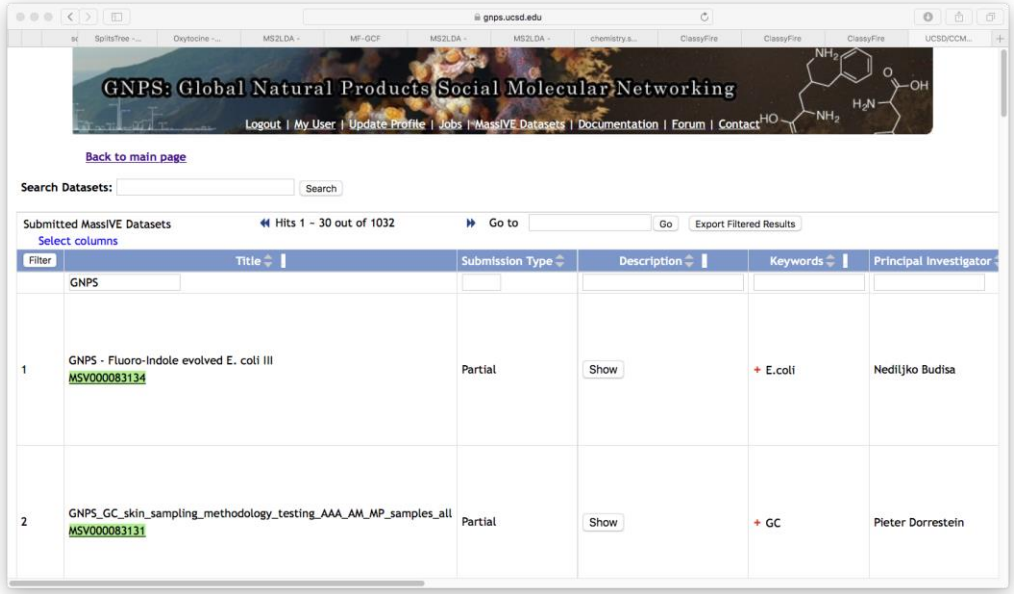




WAGENINGEN
UNIVERSITY & RESEARCH

Medema, Cimermancic et al., *PLoS Computational Biology* 10: e1004016.

State of the art….



Public genome repositories

Public metabolome repositories

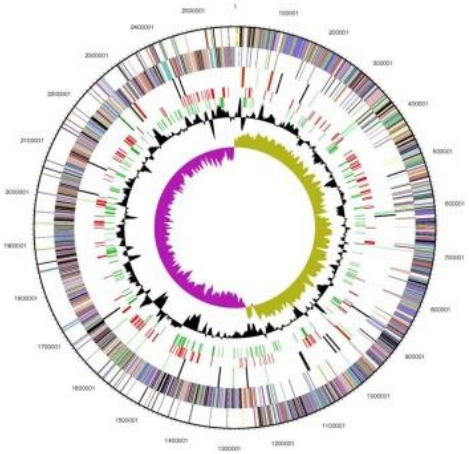
Libraries of validated biosynthetic gene clusters

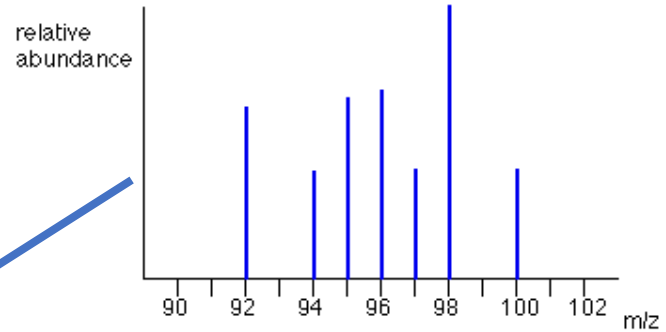
Libraries of annotated and identified molecular spectra

How to link it all?

WAGENINGEN
UNIVERSITY & RESEARCH

Paired Data Platform

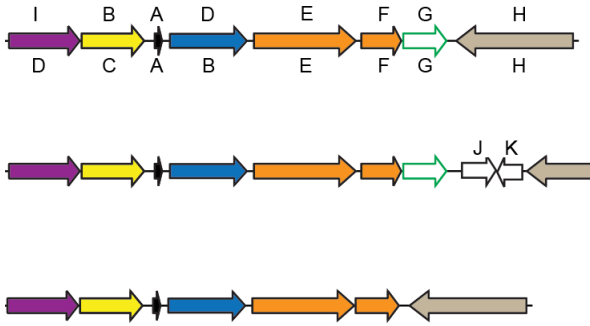




Paired Data Platform

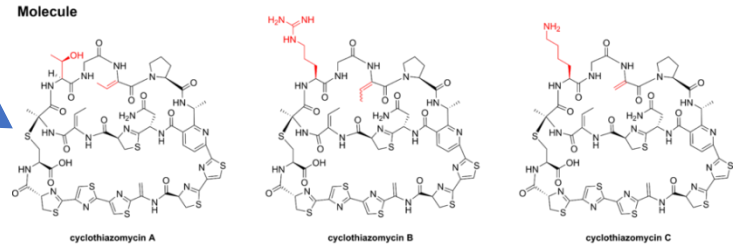
Genome	MS/MS
Name	MASSive /
NCBI Acc. #	GNPS link
Taxonomy ID	Experimental
antiSMASH	Parameters

Dr. Michelle Schorn, WUR



Gene clusters
Gene cluster families

LINK!



Molecules
Molecular families

WAGENINGEN
UNIVERSITY & RESEARCH

Cox et al., *ACS Chemical Biology* 10.1021/cb500324n.

Paired Data Platform

iOMEGA paired data platform schema JuNo

This is the JSON schema for paired genomic / metabolomic data.

version *

1

Personal data*

Name of contact for correspondence

This person will be the point of contact for any communication related to this entry.

Academic institution or company name

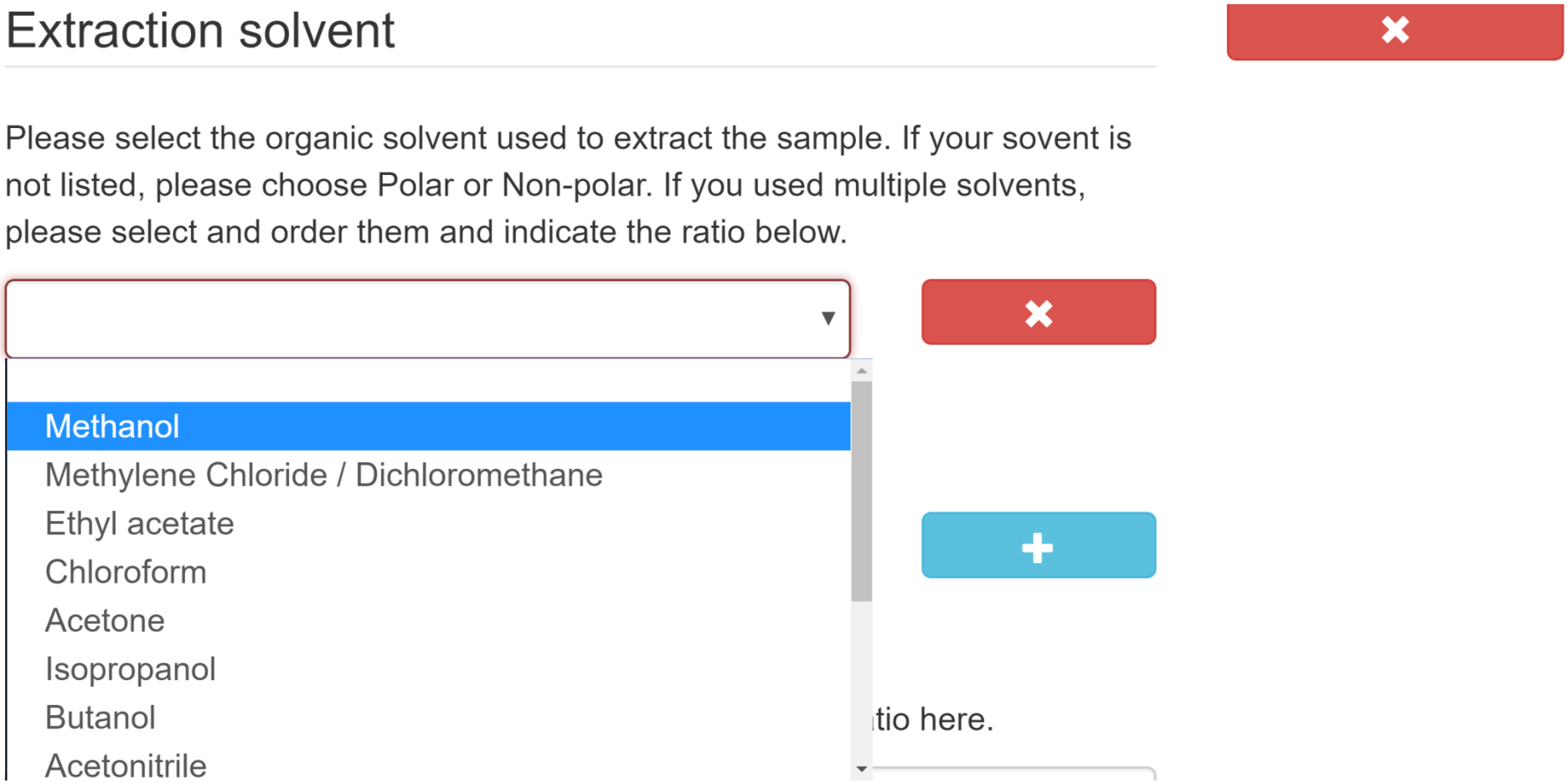
Please use the full, official name of your institute in English. E.g., 'Harvard University'.

Submitter contact e-mail address

Name of the principal investigator of the submitter

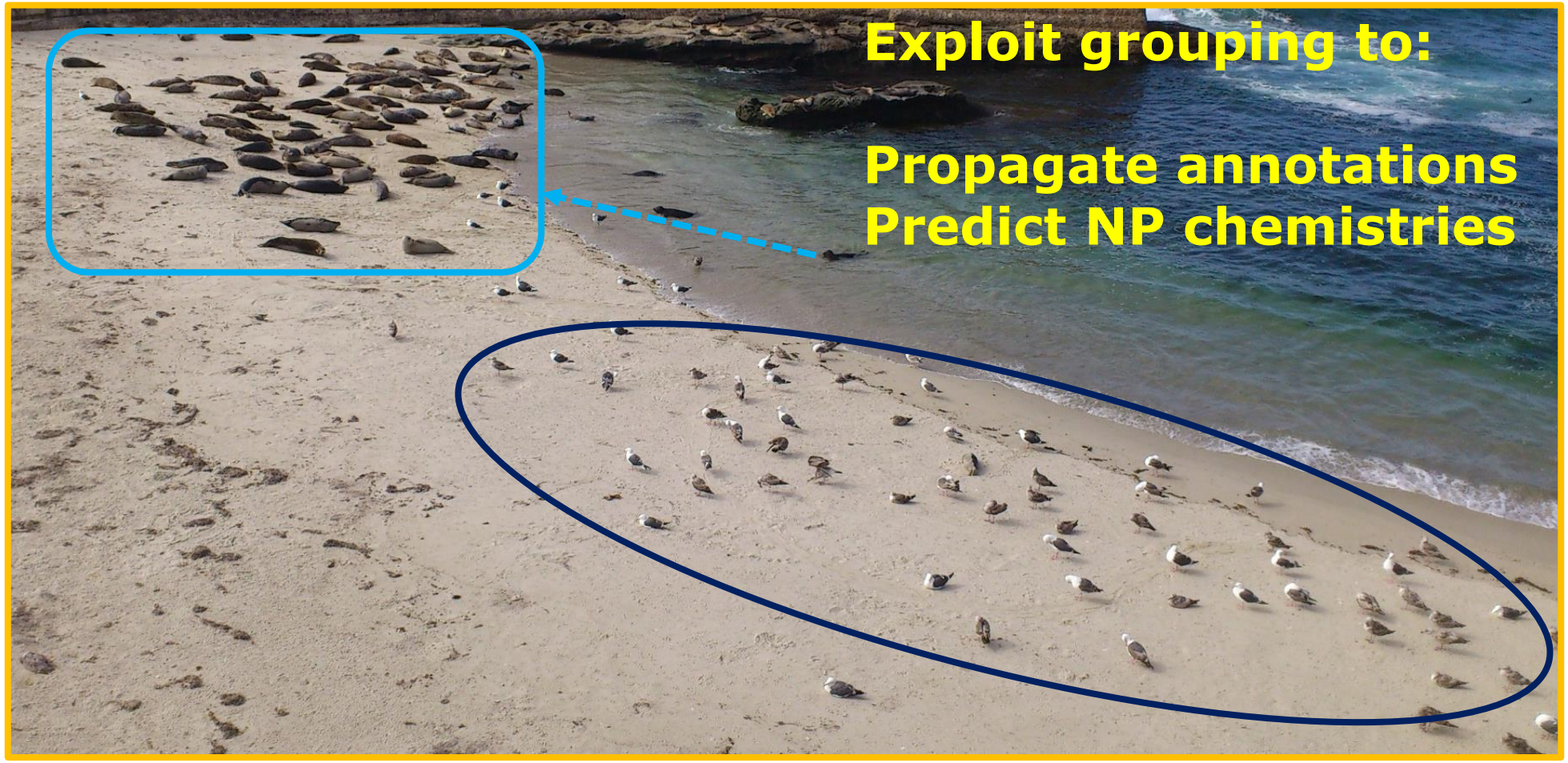
This person is contacted in case the submitter has moved institution

Paired Data Platform



WAGENINGEN
UNIVERSITY & RESEARCH

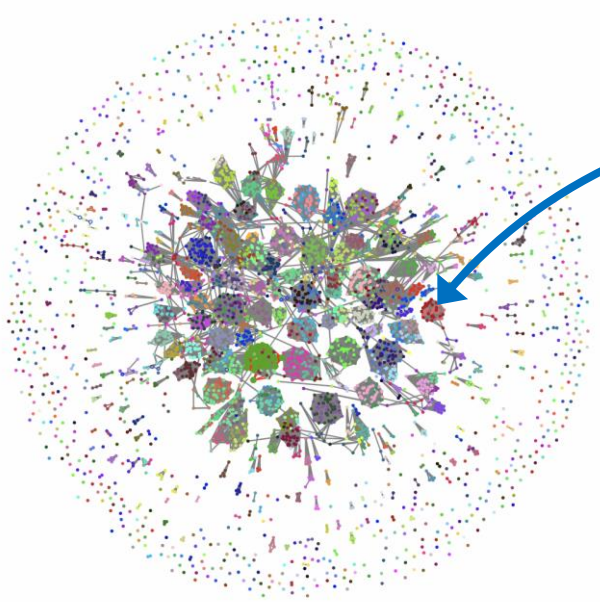
Improved annotation power by pattern mining and networking



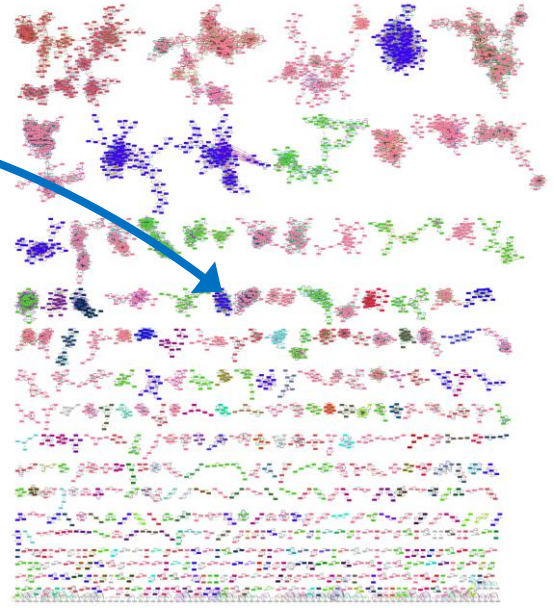
Linking substructures to genetic elements

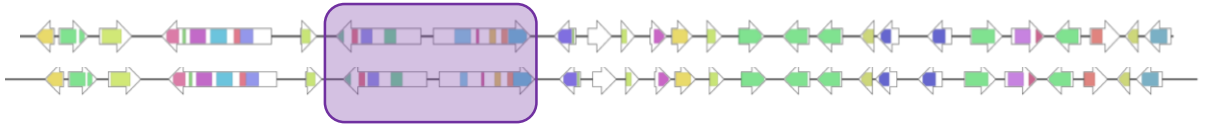
iOMEGA: Integrated Omics for MEtabolomics and Genomics Annotation

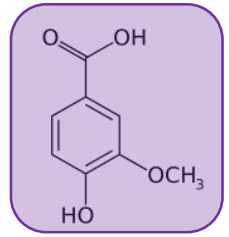
Gene Cluster Families & Metabolite Families

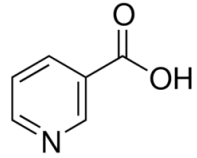


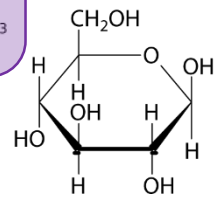
<--->





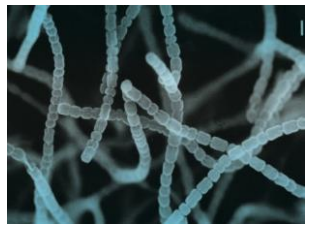


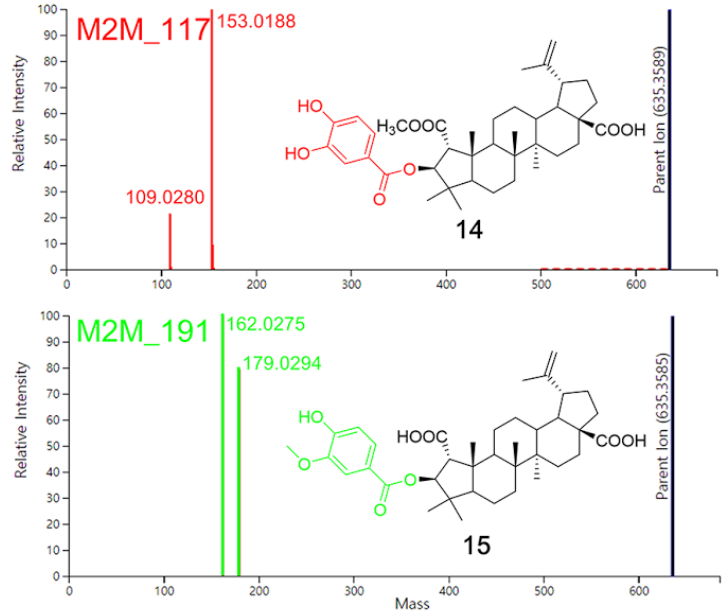


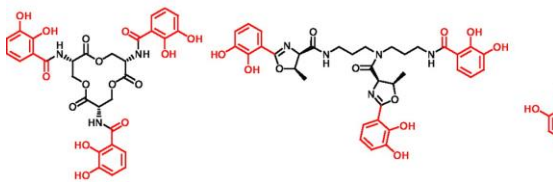


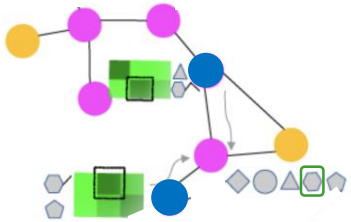
WAGENINGEN
UNIVERSITY & RESEARCH

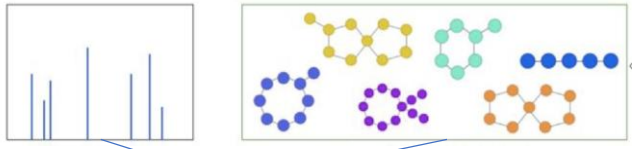
Integrated metabolomics workflow



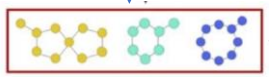





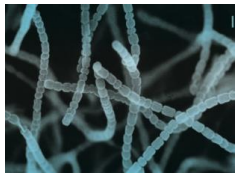


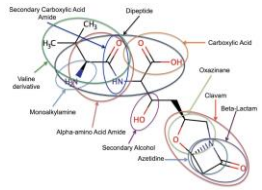


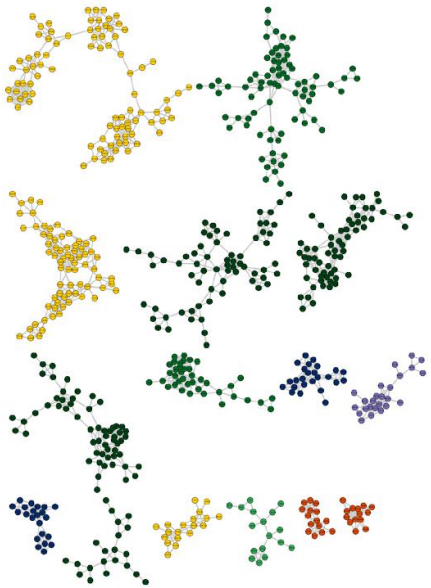
Spectrum



Peptide database





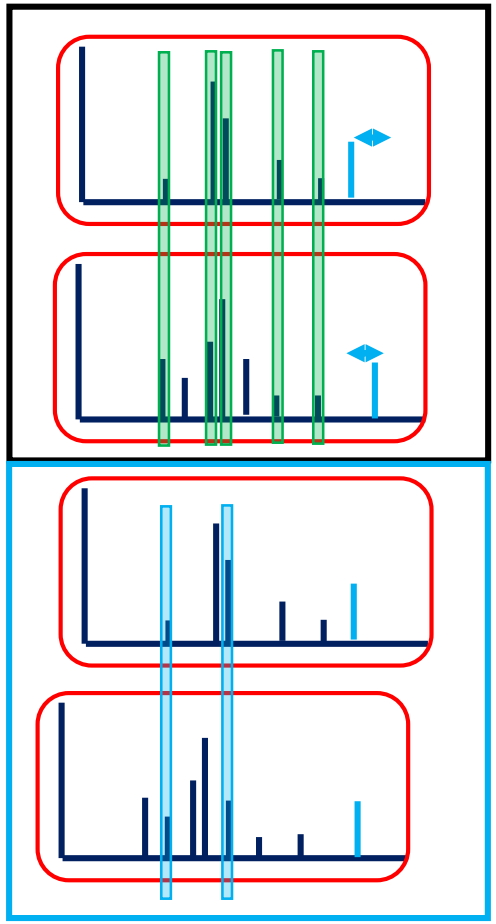


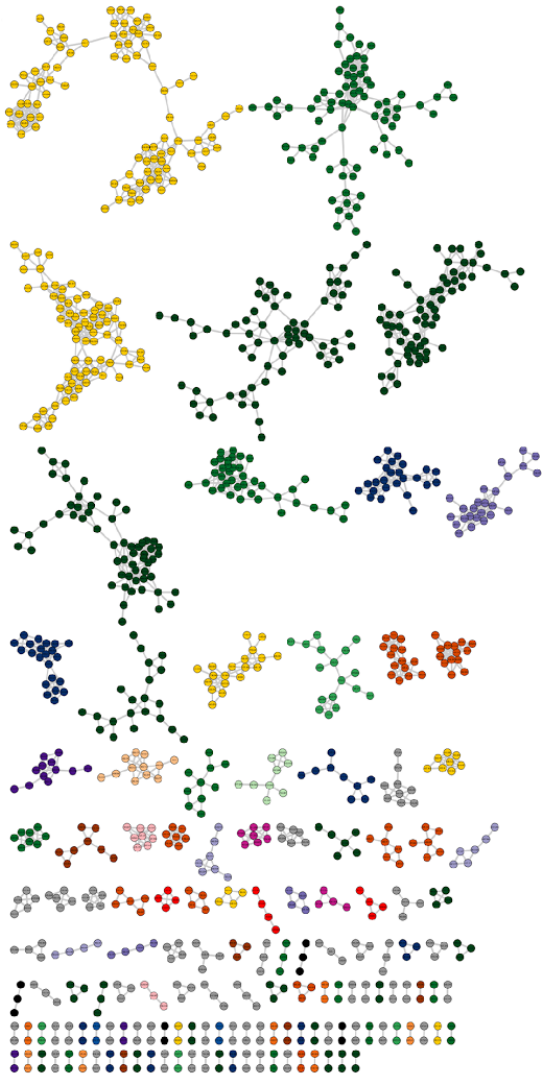
Dr Madeleine Ernst, UCSD

WAGENINGEN
UNIVERSITY & RESEARCH

Illuminating the Rhamnaceae plant chemistry

Molecular Networking





- triterpenoids
- triterpenoidal saponins
- steroidal saponins
- flavonoid 3-*O*-glycosides
- flavonoid 7-*O*-glycosides
- flavonoid *O*-glycosides
- flavones, flavonones, flavonols
- 8-methylated flavonoids
- (hydroxyl)anthraquinones
- xanthones
- hydrolyzable tannins
- phenolic glycosides
- coumaric acid and derivatives
- lignan glycosides
- iridoid glycosides
- peptide
- hybrid peptide
- oligosaccharides
- long-chain fatty acids
- others / no matches

plant related classifications:

different flavonoids

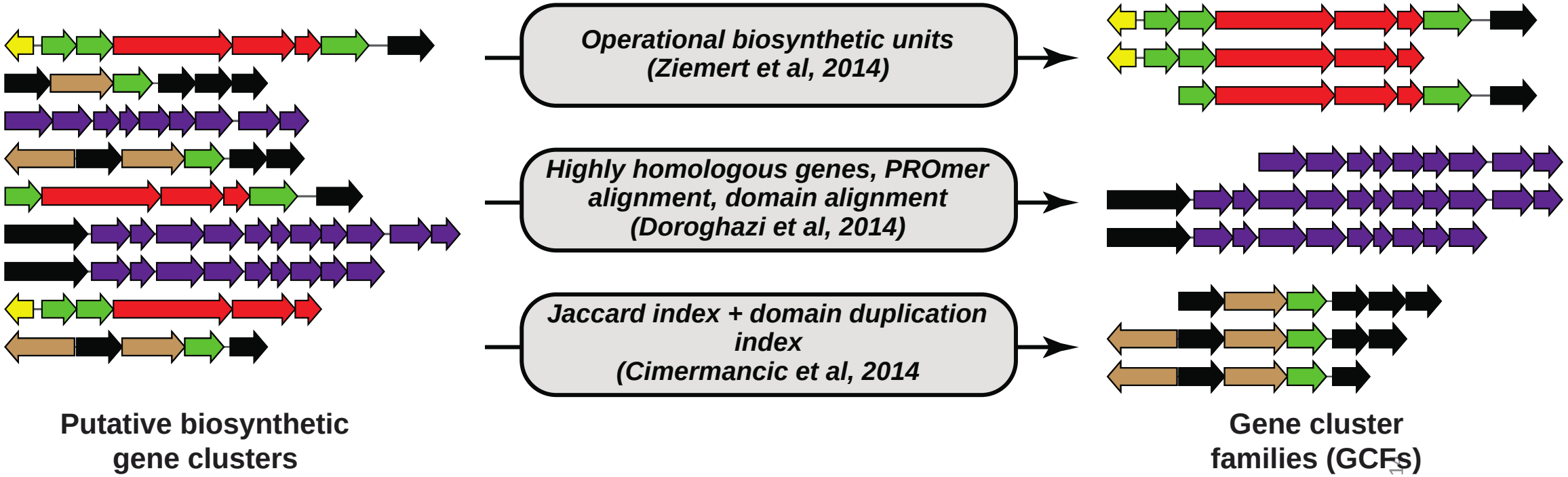
phenolic glycosides

triterpenoids

Dr Kyo Bin Kang, UCSD

WAGENINGEN
UNIVERSITY & RESEARCH

Grouping Biosynthetic Gene Clusters into Families

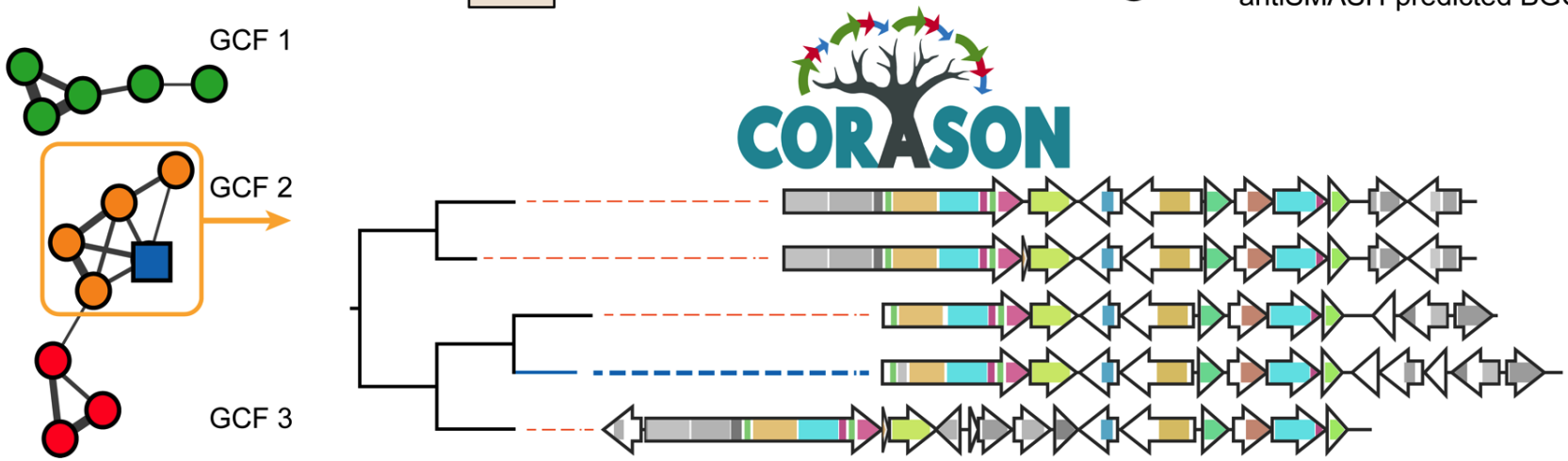


WAGENINGEN
UNIVERSITY & RESEARCH

Medema & Fischbach (2015) *Nat. Chem. Biol.* 11: 639-648.

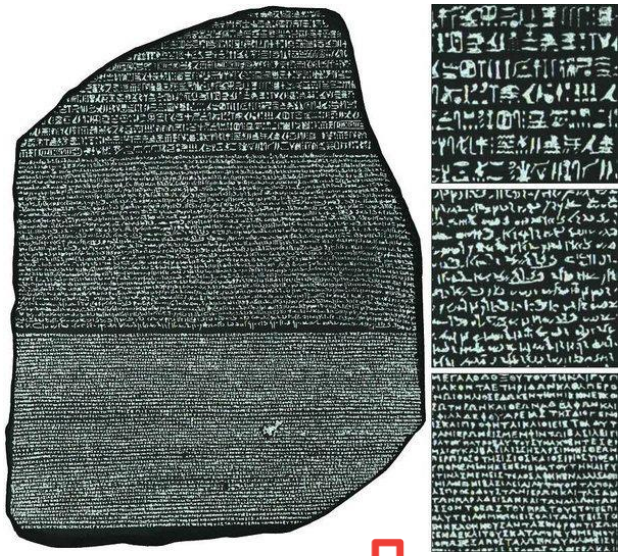
Automated reconstruction and phylogenomic analysis

Dr. Jorge Navarrio Muñoz



Nelly Selem-Mojica

A FAIR Future Outlook



Lots of potential but no computer-readable formal linking options available

Community-driven platform:

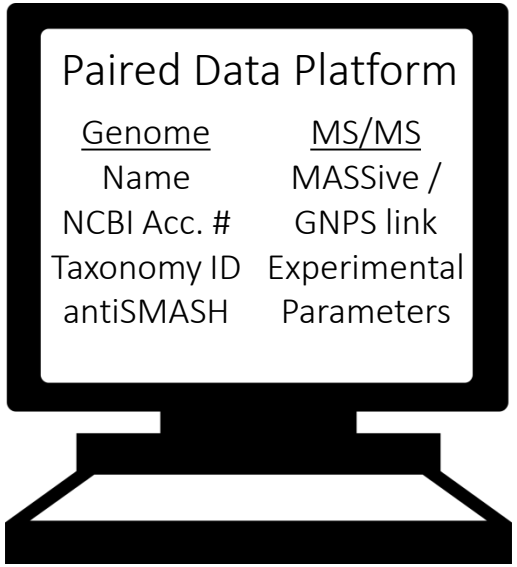
Collecting first feedback by close collaborators

Minimum information needed versus time commitment

The next step:

In collaboration with Glasgow we work on NPlinker:

Systematic exploration of linking algorithms



WAGENINGEN
UNIVERSITY & RESEARCH