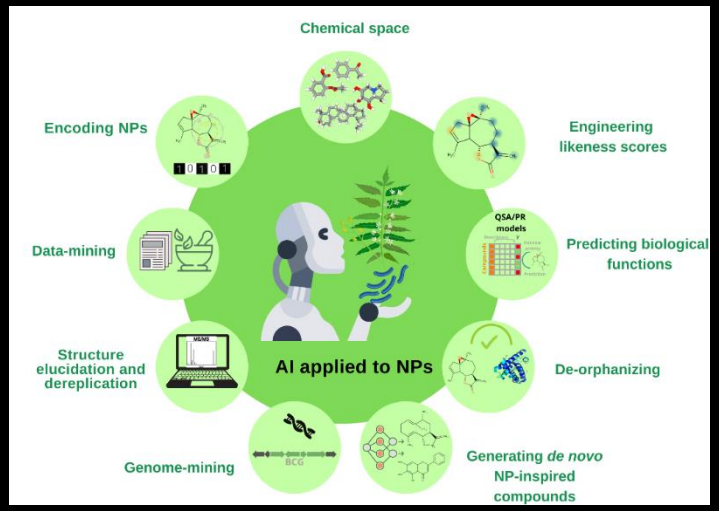
Expanding the chemical space and multiverse of natural products and food chemicals



José L. Medina-Franco

Alejandro Gómez-García, K. Eurídice Juárez-Mercado

DIFACQUIM Research Group, Department of Pharmacy, School of Chemistry, National Autonomous University of Mexico

XXIX Symposium on Bioinformatics and Computer-Aided Drug Discovery

20 September, 2023

Outline

- Chemical space and *chemical multiverse*
- Natural product databases in Latin America
  - LANaPDB: Latin American Natural Product Database
- Food chemicals and epigenetic targets
  - Building an (Epi) Food Chemical Database
- Summary

Chemical space

'An M-dimensional Cartesian space in which compounds are located by a set of M physicochemical and/or chemoinformatic descriptors'

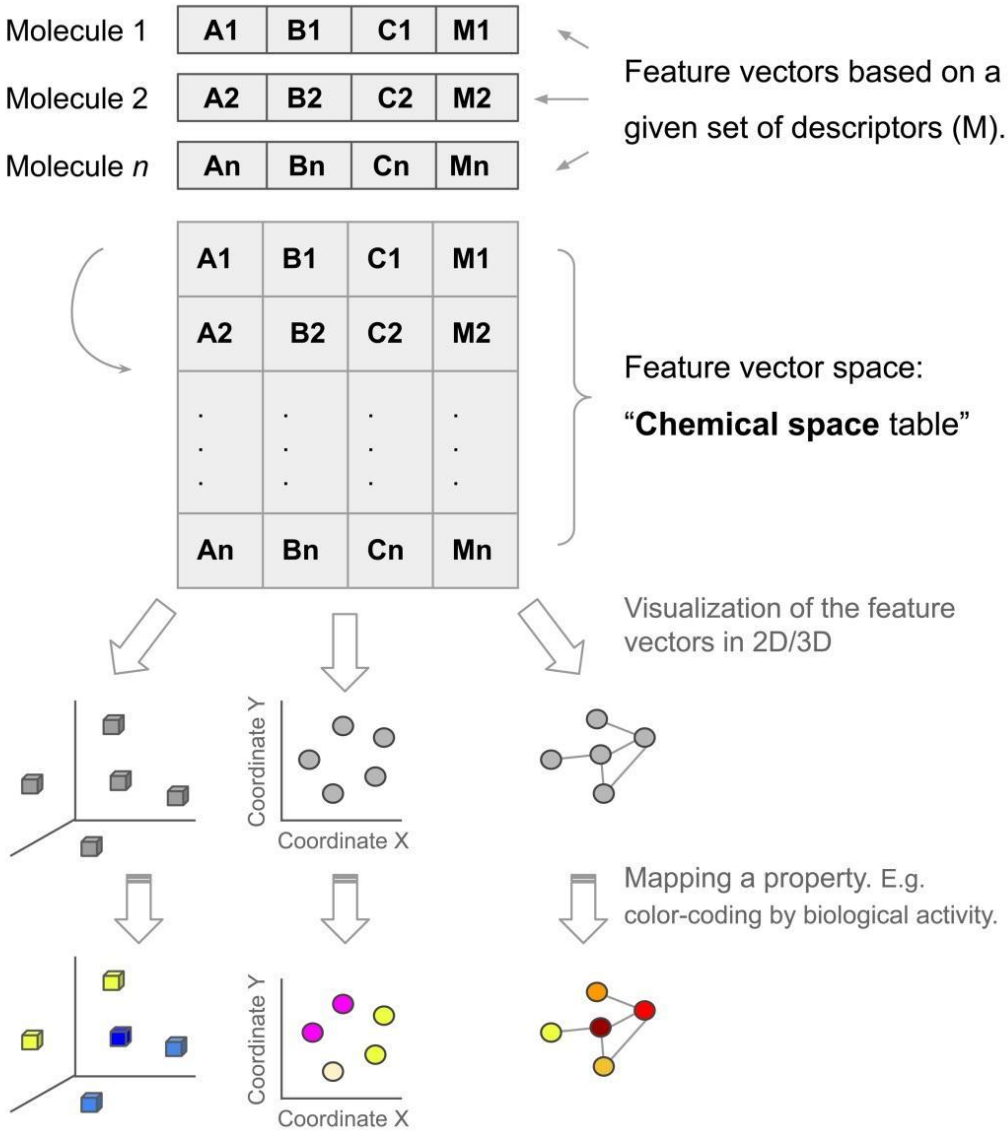
Virshup AM, Contreras-García J, Wipf P, et al. *J Am Chem Soc* 2013 135:7296-303

Chemical compound type	ID	Physicochemical properties		Topological descriptors		Molecular fragments		Similarities values based on shapes		Others
		D1	D2	D3	D4	D5	D6	D7	D8	Dm
Approved drugs	M1									...
Natural products	M2									...
Food chemicals	M3									...
Virtual compounds	M4									...
Synthesizable compounds	M5									...
Organometallic compounds	M6									...
Peptides	M7									...
Others	Mn	...	...	...	...	...	...	...	...	...

*Saldívar-Gonzalez et al. *Exp. Opin. Drug Discovery* 2022 17:789-798

Chemical space

Visual representation



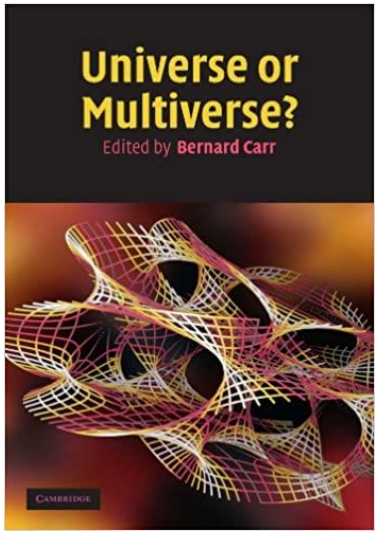
Applications:

- Diversity analysis.
- Data sets comparisons.
- Library design.
- Structure-property relationships, etc.

In Physics

Multiverse

- "A hypothetical collection of potentially diverse observable universes, each of which would comprise everything that is experimentally accessible by a connected community of observers."

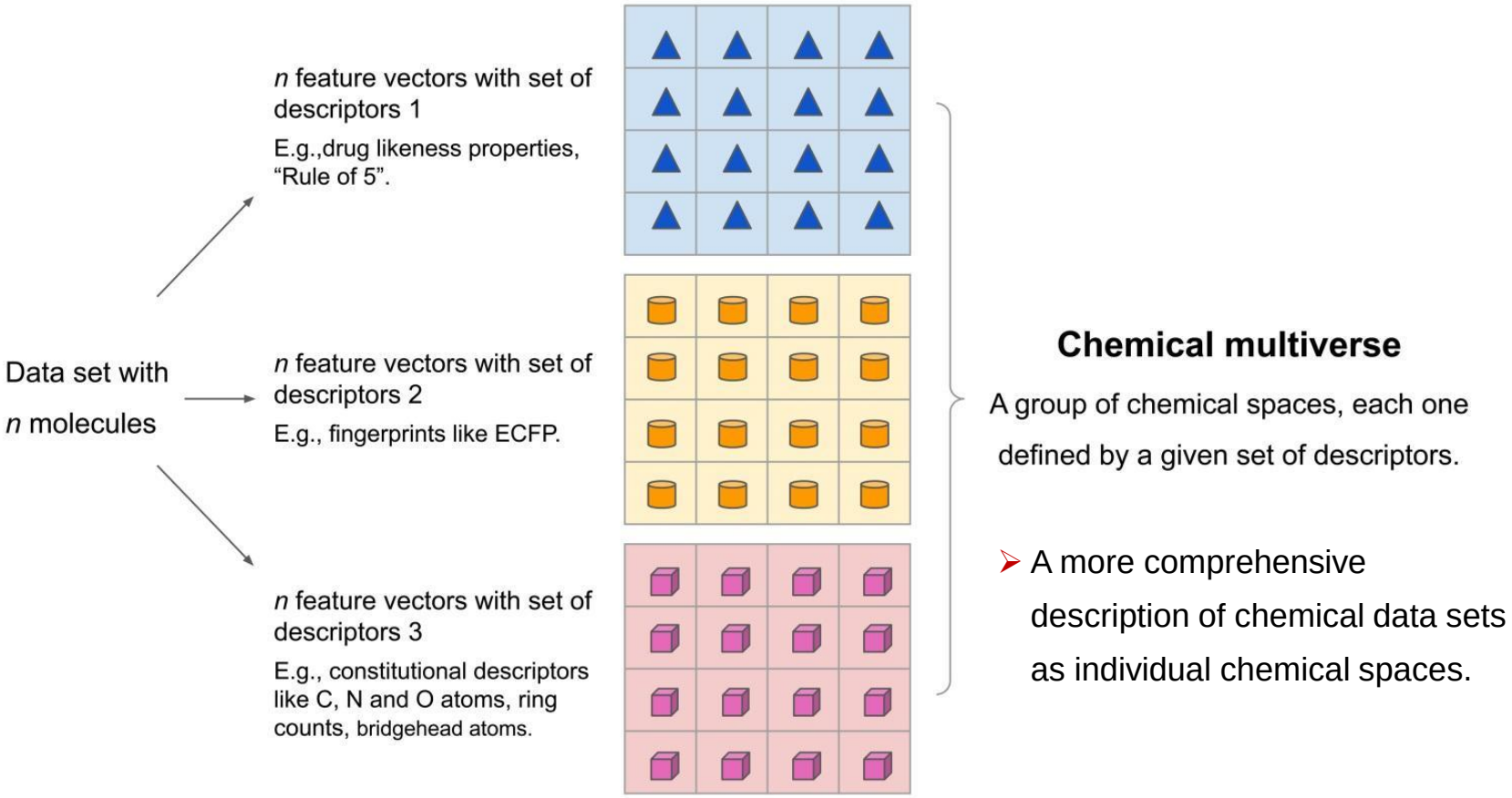


- "A hypothetical group of multiple universes," and regions in the universe detached from one another exhibit distinct properties.

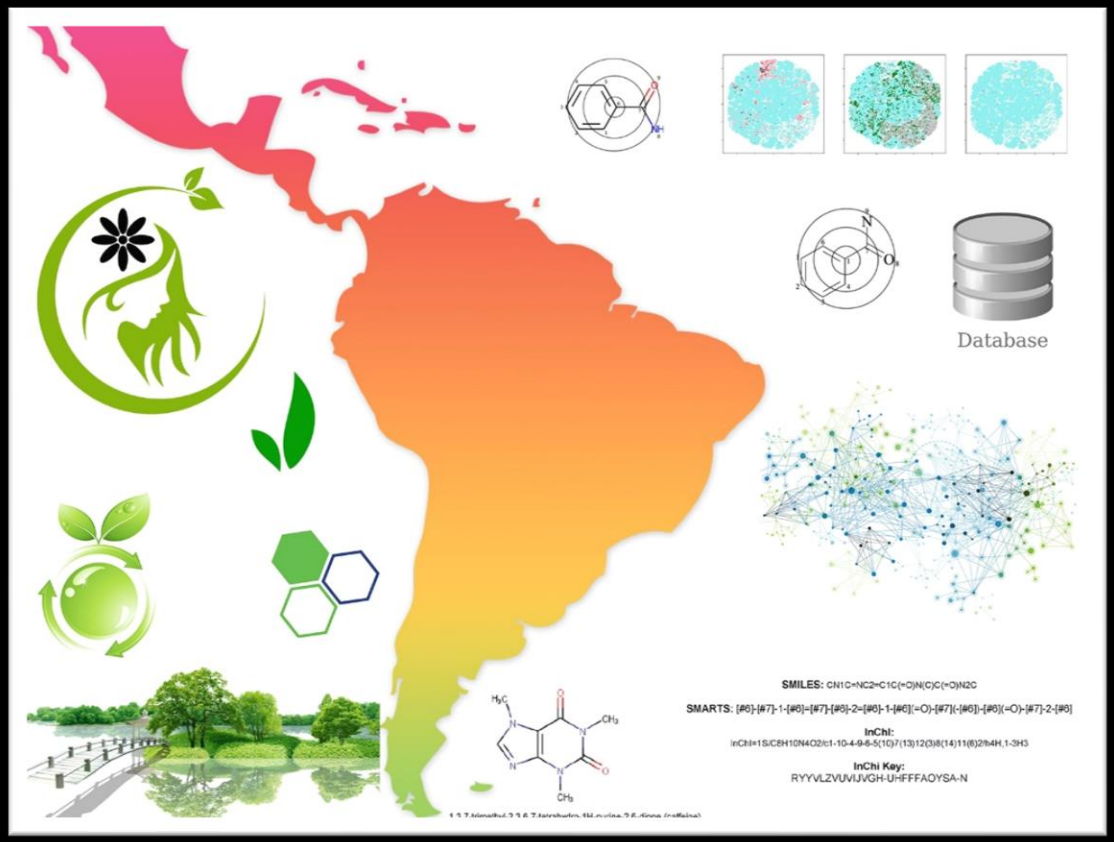
H. Everett, Hugh Everett Theory of the Universal Wavefunction, Thesis, Princeton University, 1957

Chemical multiverse

Group of numerical vectors that describe the chemical space differently from the same set of molecules.



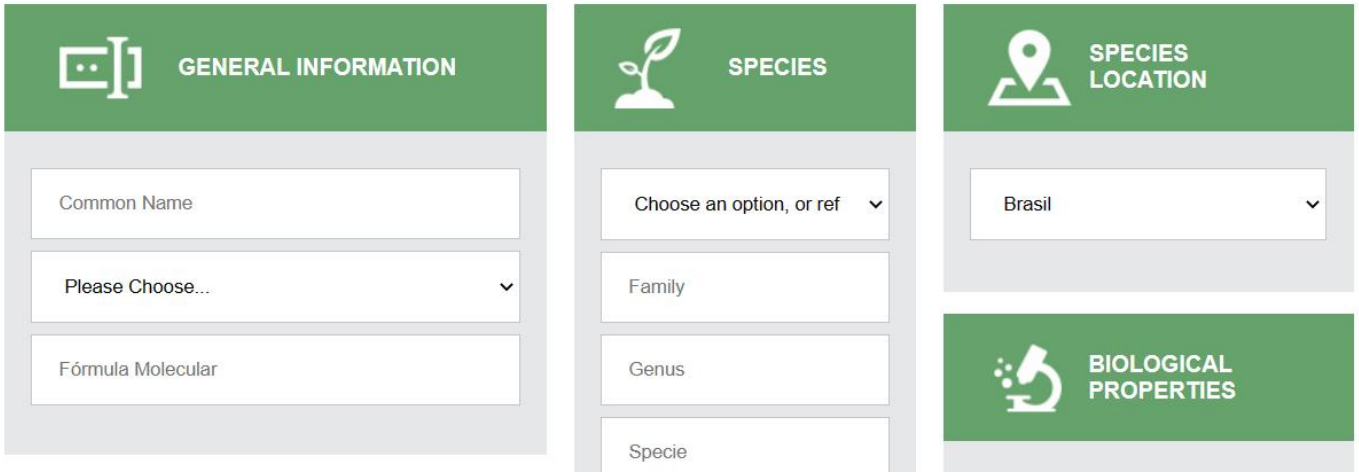
- A more comprehensive description of chemical data sets as individual chemical spaces.



Natural products databases

Contributions from Latin America

Nuclei of Bioassays, Biosynthesis and Ecophysiology of Natural Products (NuBBE)



- 2223 compounds
- Plants, animals; microorganisms
  - São Paulo State University; University of São Paulo

http://www.nubbe.iq.unesp.br/portal/nubbe-search.html

Valli M et al. Development of a natural products database from the biodiversity of Brazil. Journal of Natura Products 2013 76:439–444

Mexican research groups

Mexican biodiversity

Isolation and purification

www.difacquim.com/d-databases

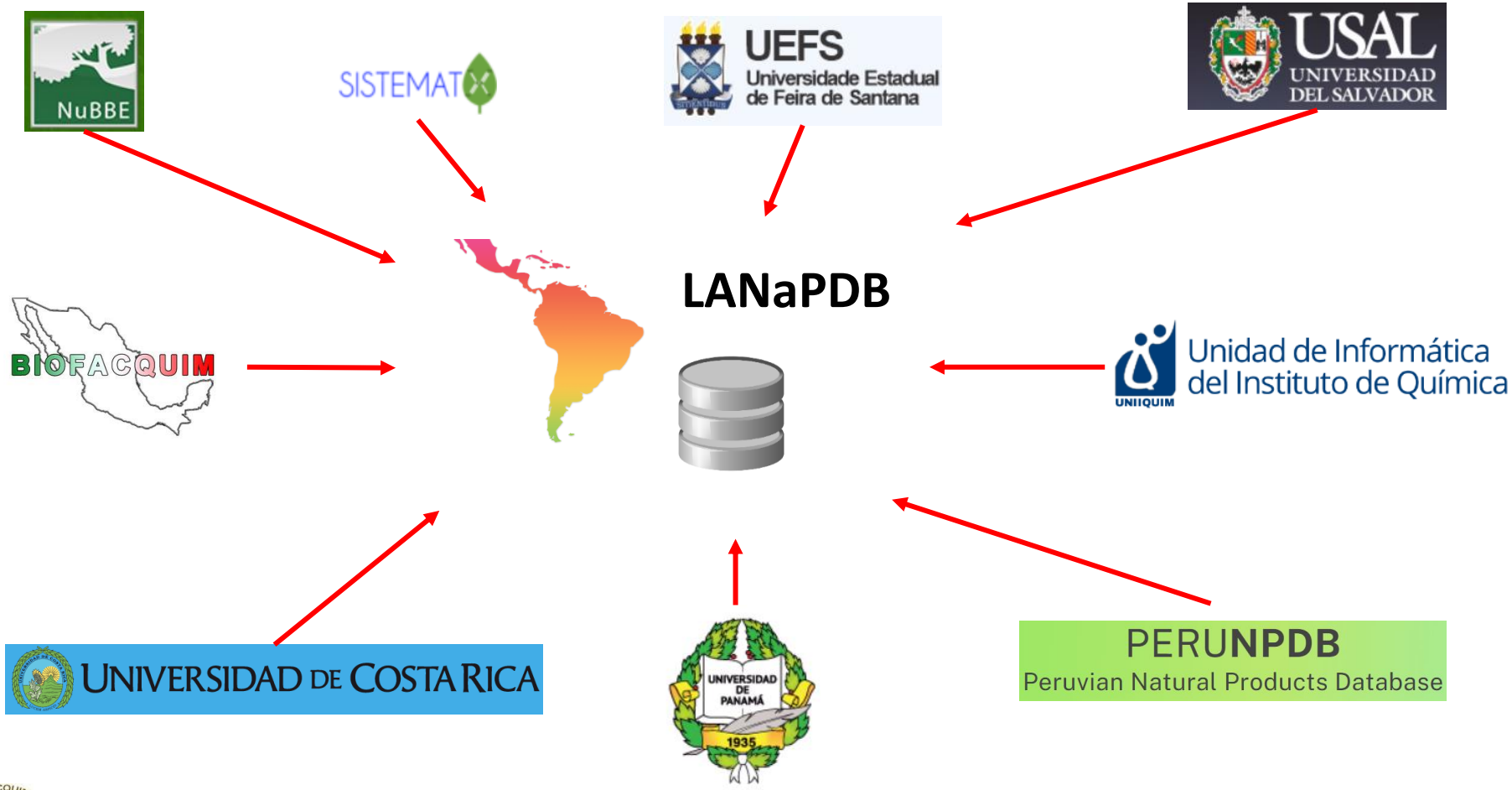
- 553 compounds
- Pants, fungi, marine animals
  - UNAM

http://zinc15.docking.org/catalogs/biofacquimnp/

Pilón-Jiménez BA et al BIOFACQUIM: A Mexican compound database of natural products. Biomolecules 2019 9:31

Latin American databases

Currently there is no database that unifies the content of the NPs of Latin America



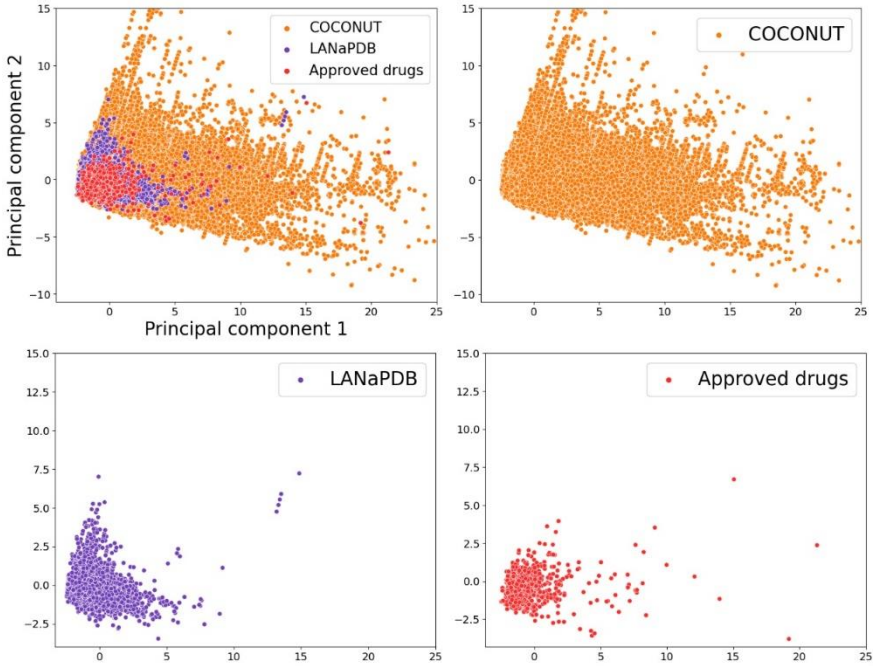
LANaPDB
Latin American Natural Products Database

Database	Size	Country	Accessibility	Source	Website	Year release
$NuBBE_{DB}$	2223	Brazil	Open-access	Plants Microorganisms Terrestrial animals Marine animals	http://nubbe.iq.unesp.br/portal/nubbe-search.html	2013, 2017
SistematX	9514	Brazil	Open-access	Plants	https://sistematx.ufpb.br/	2018, 2021
UEFS	503	Brazil	Open-access	Plants	http://zinc12.docking.org/catalogs/uefsnp	NA
NAPRORE-CR	359	Costa Rica	Access under request	Plants Microorganisms		Not published yet
LAIPNUDELSAV	214	El Salvador	Access under request			NA
CIFPMA	454	Panama	Access under request	Plants	No available. Structures available upon request.	2017
PeruNPDB	280	Peru	Open-access	Plants Animales	https://perunpdb.com.pe/	2023
UNIIQUIM	~1112	Mexico	Open-access	Plants	https://uniiquim.iquimica.unam.mx/	NA
BIOFACQUIM	553	Mexico	Open-access	Plants Fungi Propolis Marine animals	Version 1 https://biofacquim.herokuapp.com Version 2 https://figshare.com/articles/dataset/BIOFAQUIM_V2_sdf/11312702	2019, 2020

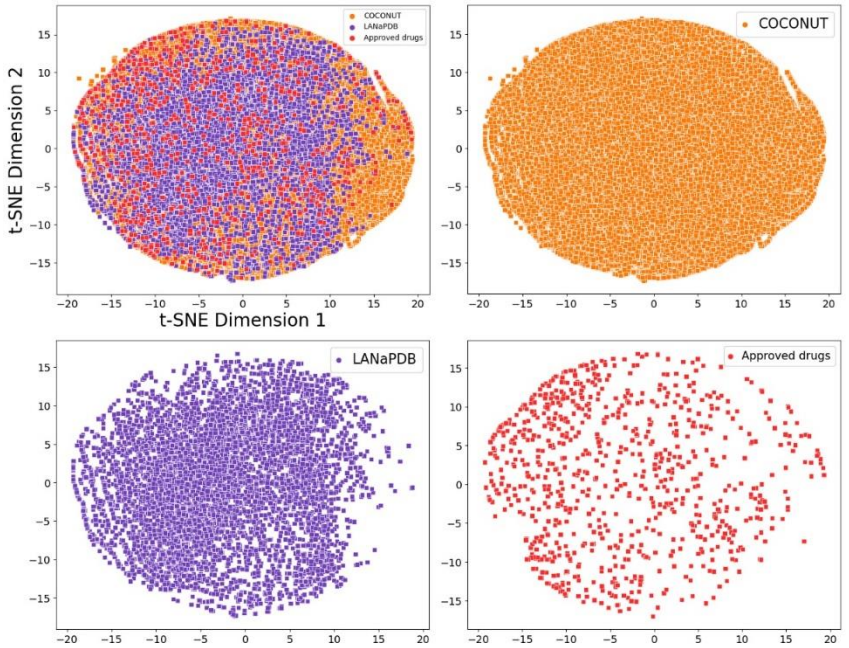
LANaPDB

Chemical space based on properties

Principal component analysis



t-Distributed stochastic neighbor embedding

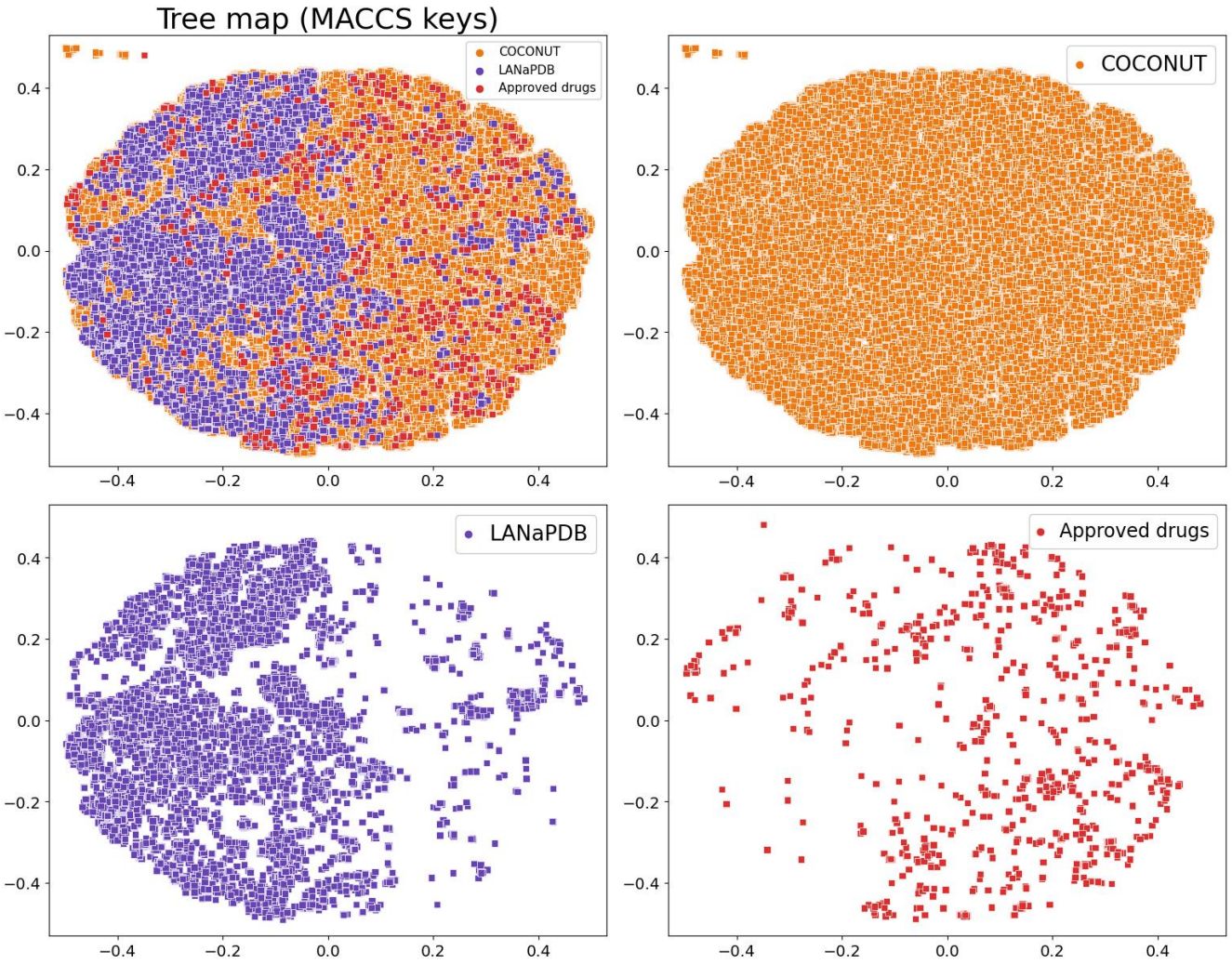


Chemical space described by 6 properties of pharmaceutical interest:

- **COCONUT** is the database that covers the largest area of the chemical space, followed by **LANaPDB** and approved drugs.
- The 3 databases overlap in a certain area of chemical space.

Chemical space based on fingerprints

MACCS keys

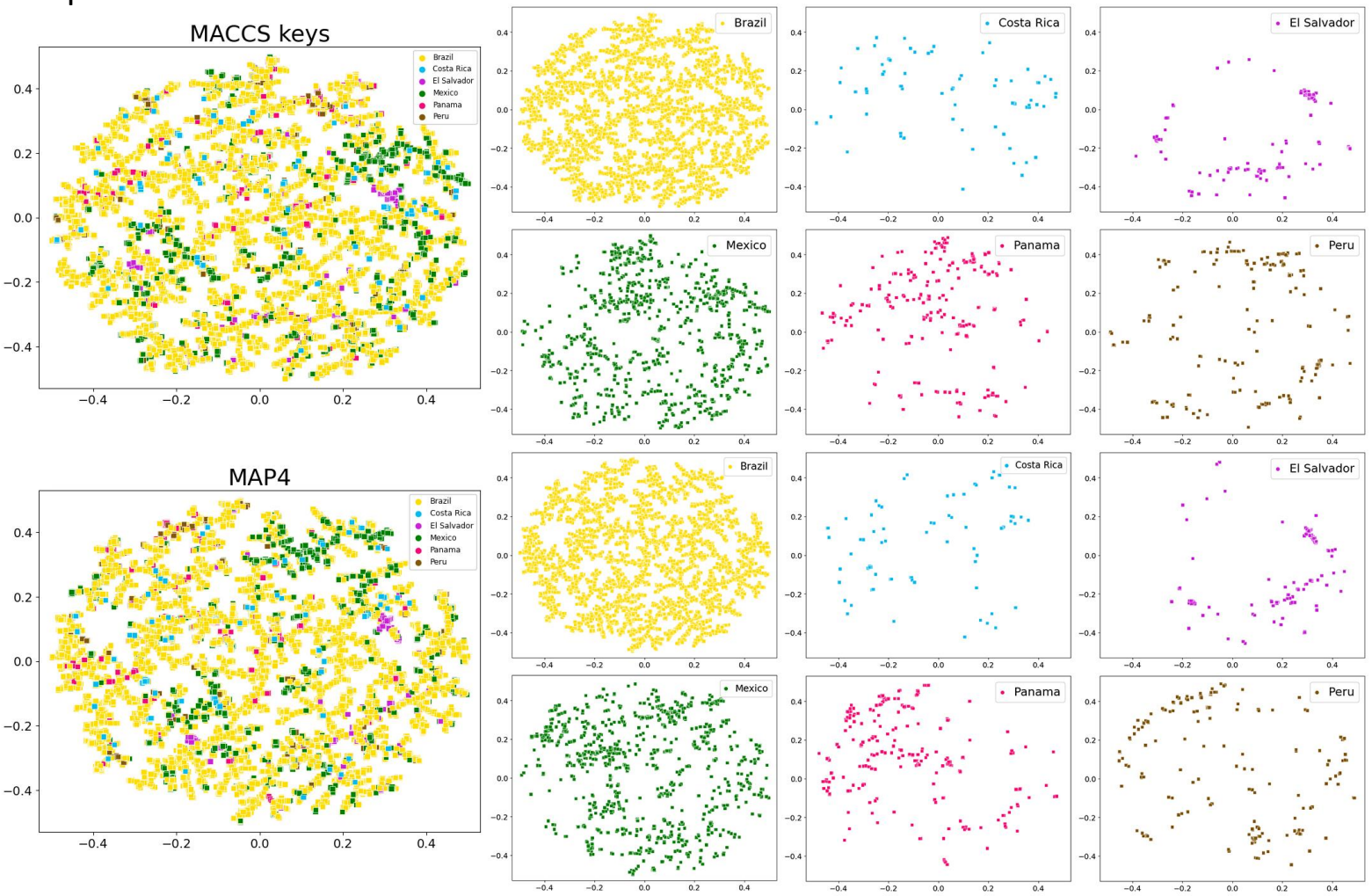


- LANaPDB totally overlaps with COCONUT.
- The approved drugs are more dispersed and some of them overlap with LANaPDB.

Chemical multiverse

MACCS keys and MAP4 fingerprints

Tree Maps



Natural products from Brazil and Mexico cover a broad region of the chemical multiverse of LANaPDB.

Latin American team of LANaPDB

Collaborators


Dr. José L. Medina Franco
MSc. Alejandro Gómez García

México
BIOFACQUIM

Dr. Willam Zamora

Costa Rica
NAPRORE-CR

Dr. Marvin José Núñez

El Salvador
LAIPNUDESAL

Dr. Dionisio A.Olmedo

Panamá
CIFPMA



Dra. Vanderlan da Silva Bolzani
Dr. Adriano D. Andricopulo
Dra. Marilia Valli

Brasil
NuBBE$_{DB}$

Dr. Héctor Fabio Cortés Hernández
MSc. Johny Roberto Rodríguez Pérez

Colombia

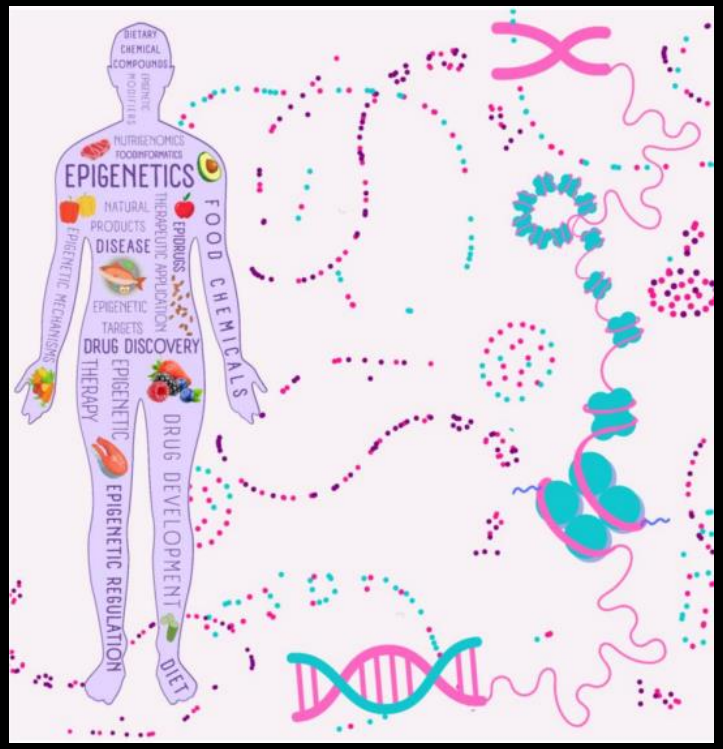
Dr. Miguel Ángel Chávez Fumagalli

Perú
PeruNPDB



Dra. Valeria Patricia Sülsen
Dra. Soledad Ravetti
Dra. Manuela Emilia García

Argentina



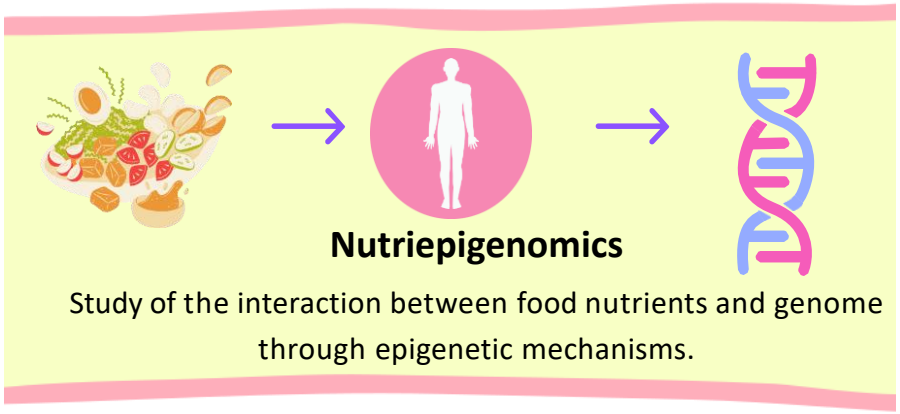
Food chemicals and epigenetic targets

Assembling an Epi Food Chemical Database

Motivation

EPIGENETICS

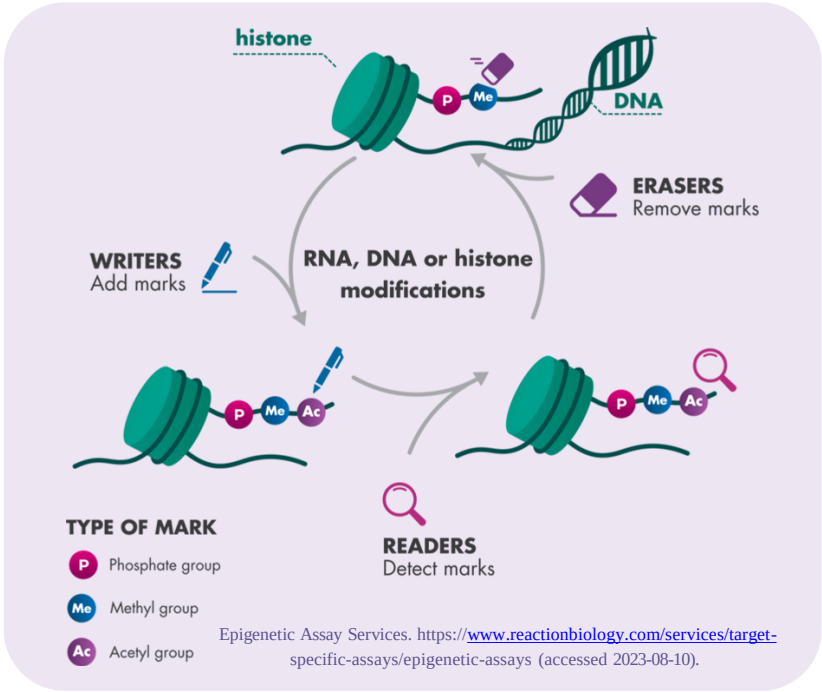
Is the study of heritable changes in genome function that are not associated with DNA sequence alterations.



Diseases associated with food and epigenetic targets:

- Type I and type II diabetes
- Liver fibrosis
- Nonalcoholic fatty liver disease
- Cancer

Epigenetic enzymes control the mechanics of genetic expression, acting as "on" and "off" switches for the human genome.

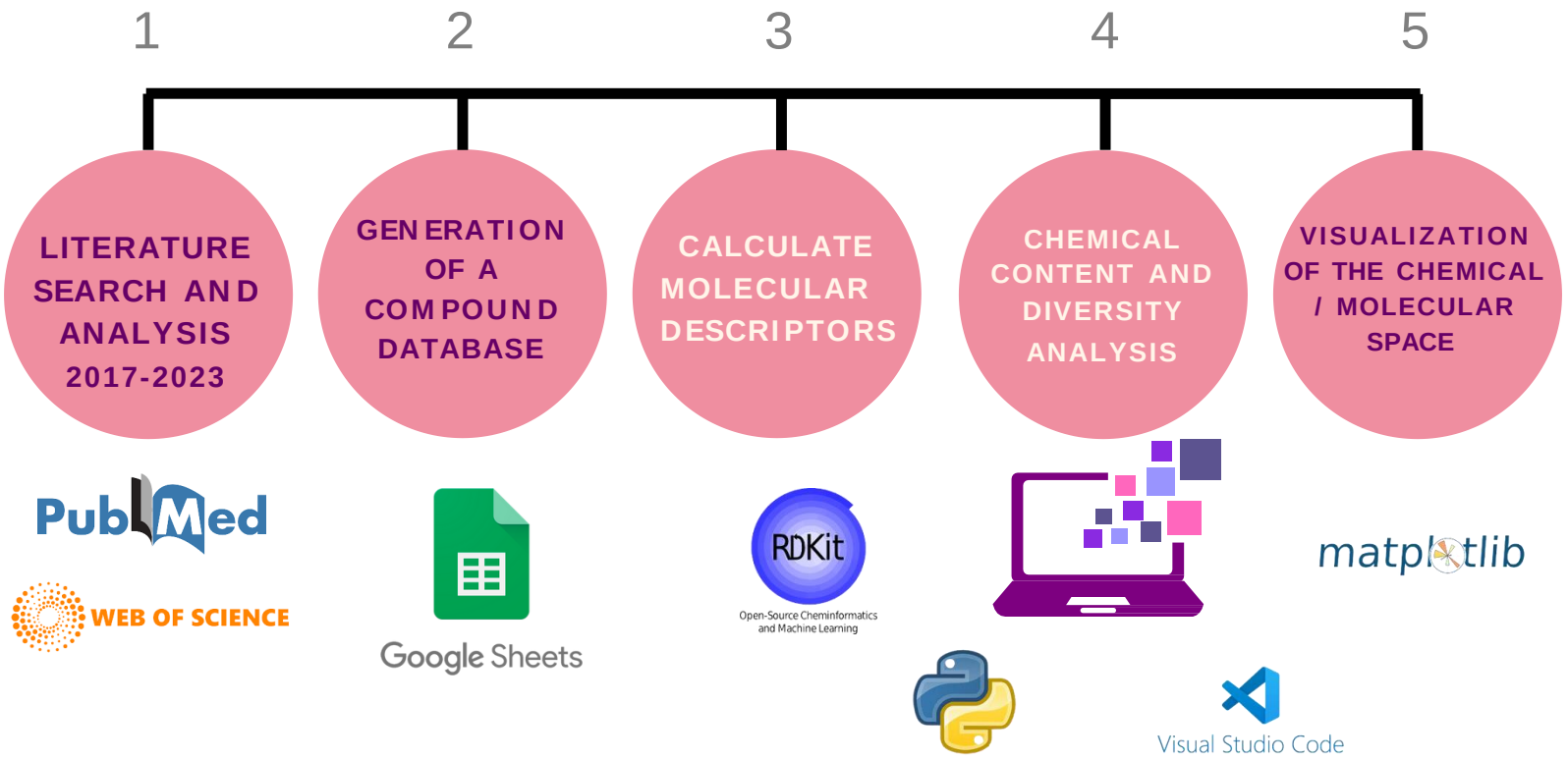


Epigenetic Assay Services. https://www.reactionbiology.com/services/target-specific-assays/epigenetic-assays (accessed 2023-08-10).

Casadesús, J.; Noyer-Weidner, M. Epigenetics. Brenner's Encyclopedia of Genetics 2013: 500–503

Goals and approach

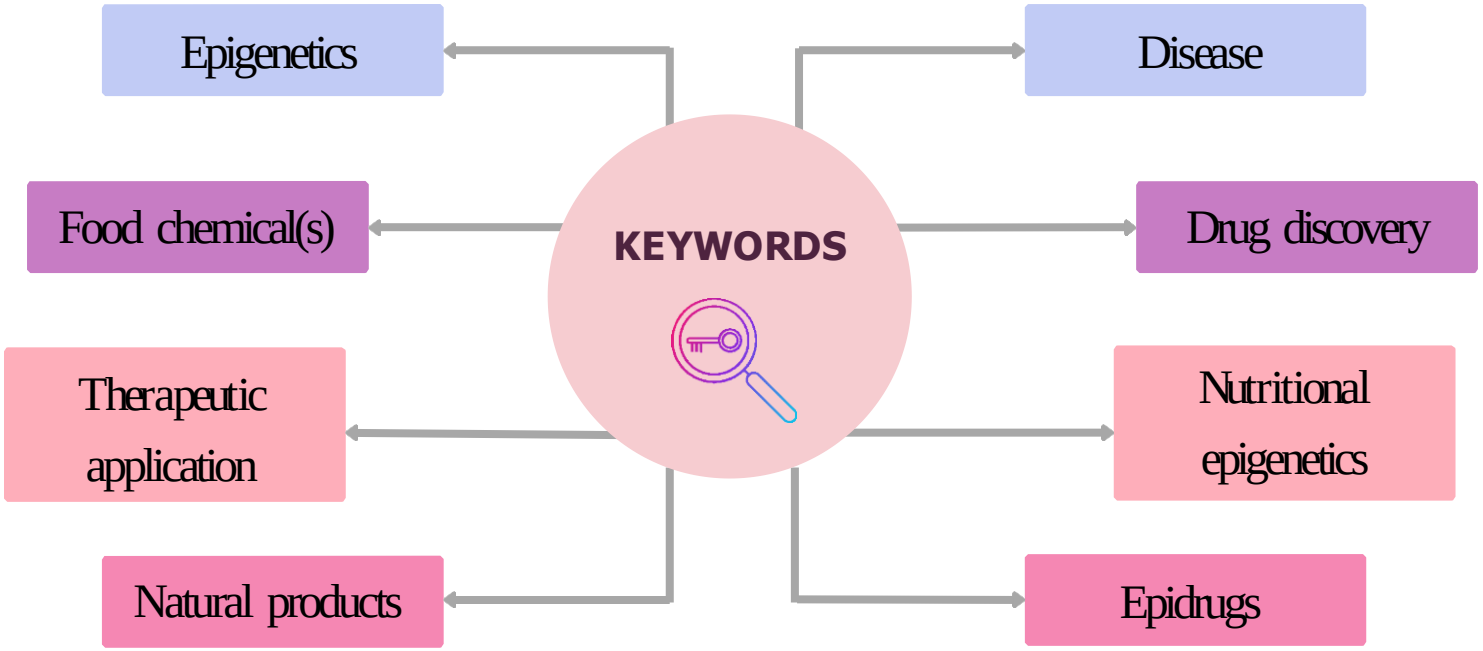
- Generate a **compound database** that **integrates** the information of the chemical structure of **food chemicals with epigenetic activity** reported in the literature.
- Analyze the molecular database using chemoinformatics and data visualization.



Results

Literature analysis

2017-2023



WEB OF SCIENCE

5,960 papers

PubMed

7,430 papers

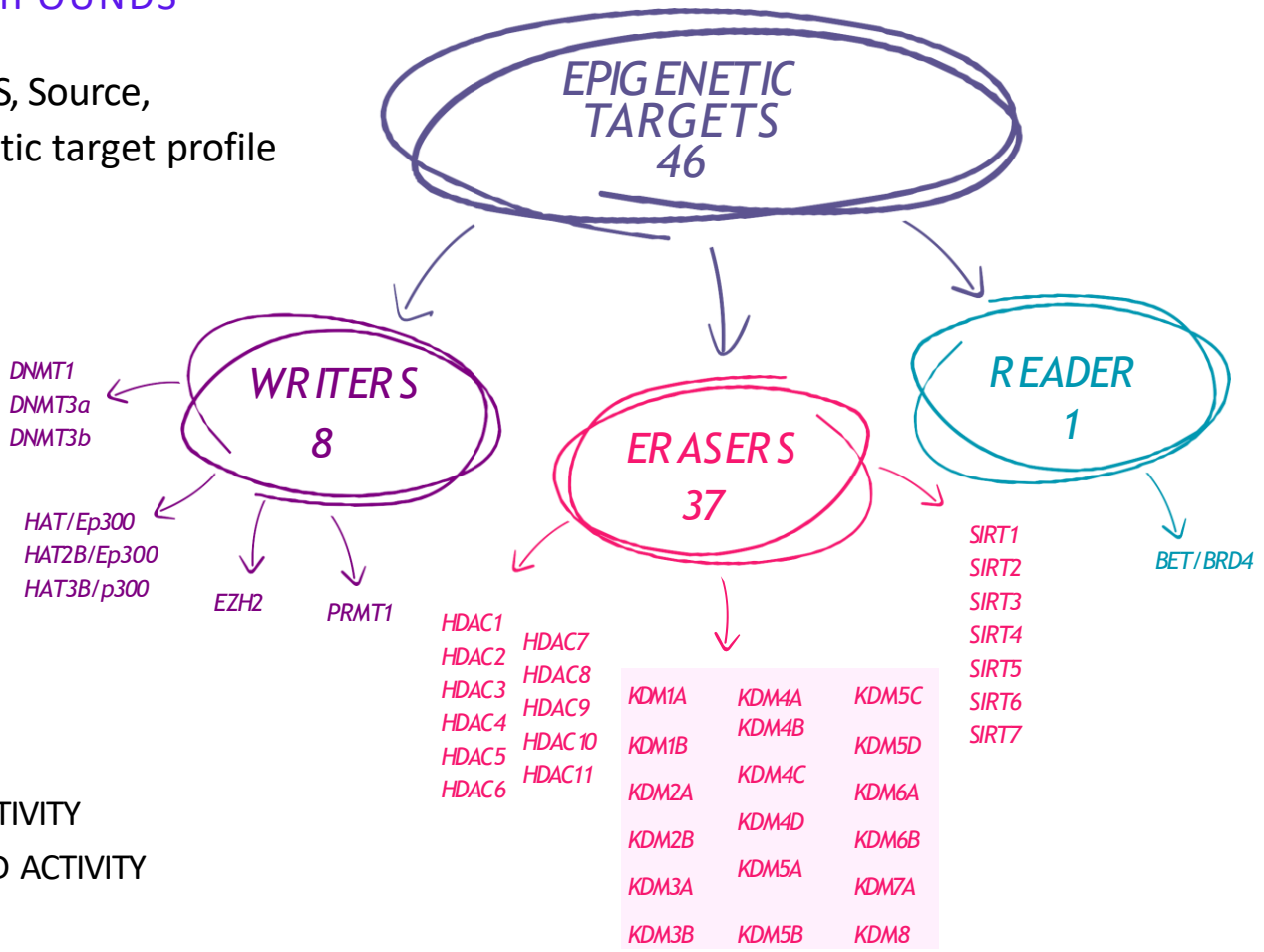
8,911 UNIQUE PAPERS

Results

Epi Food Chemical Database

184 UNIQUE COMPOUNDS

Compound name, SMILES, Source, Reference (DOI), Epigenetic target profile

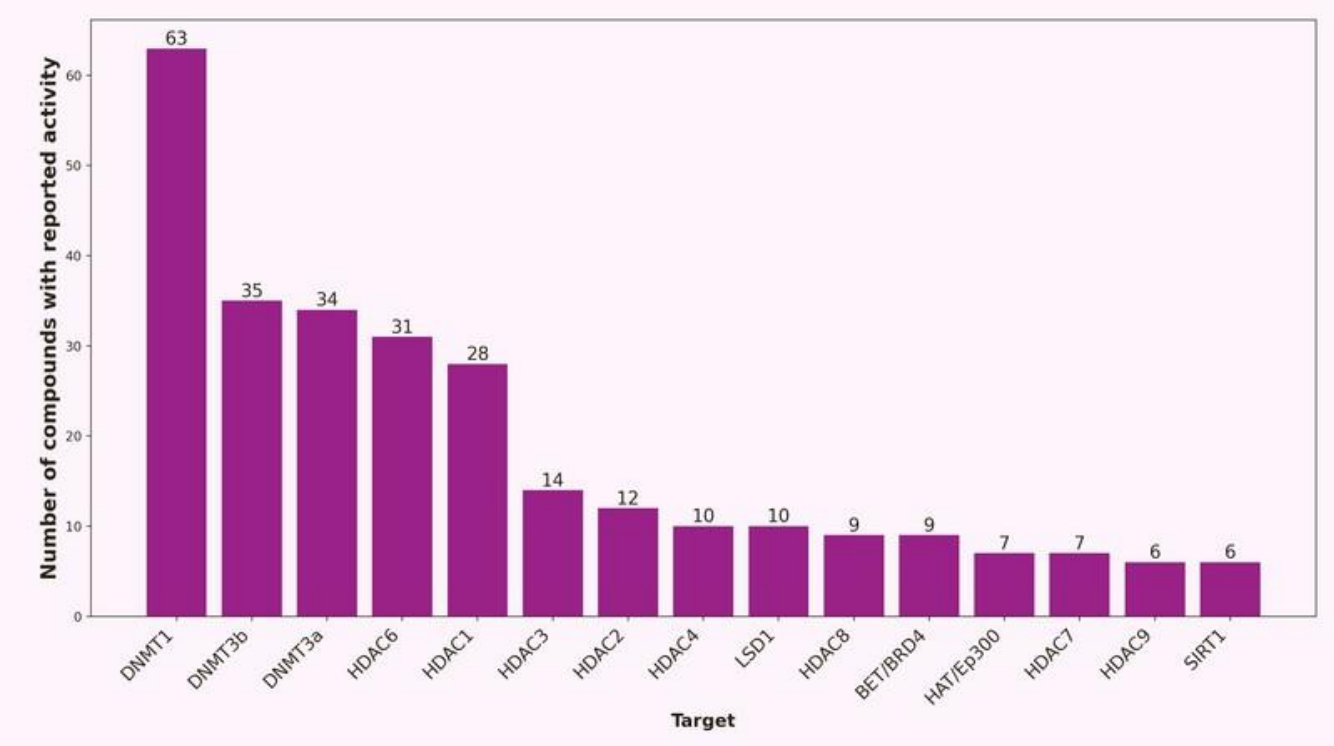


1 WITH REPORTED ACTIVITY

0 WITHOUT REPORTED ACTIVITY

Results

Target frequency



Associated diseases with **DNMT1**:

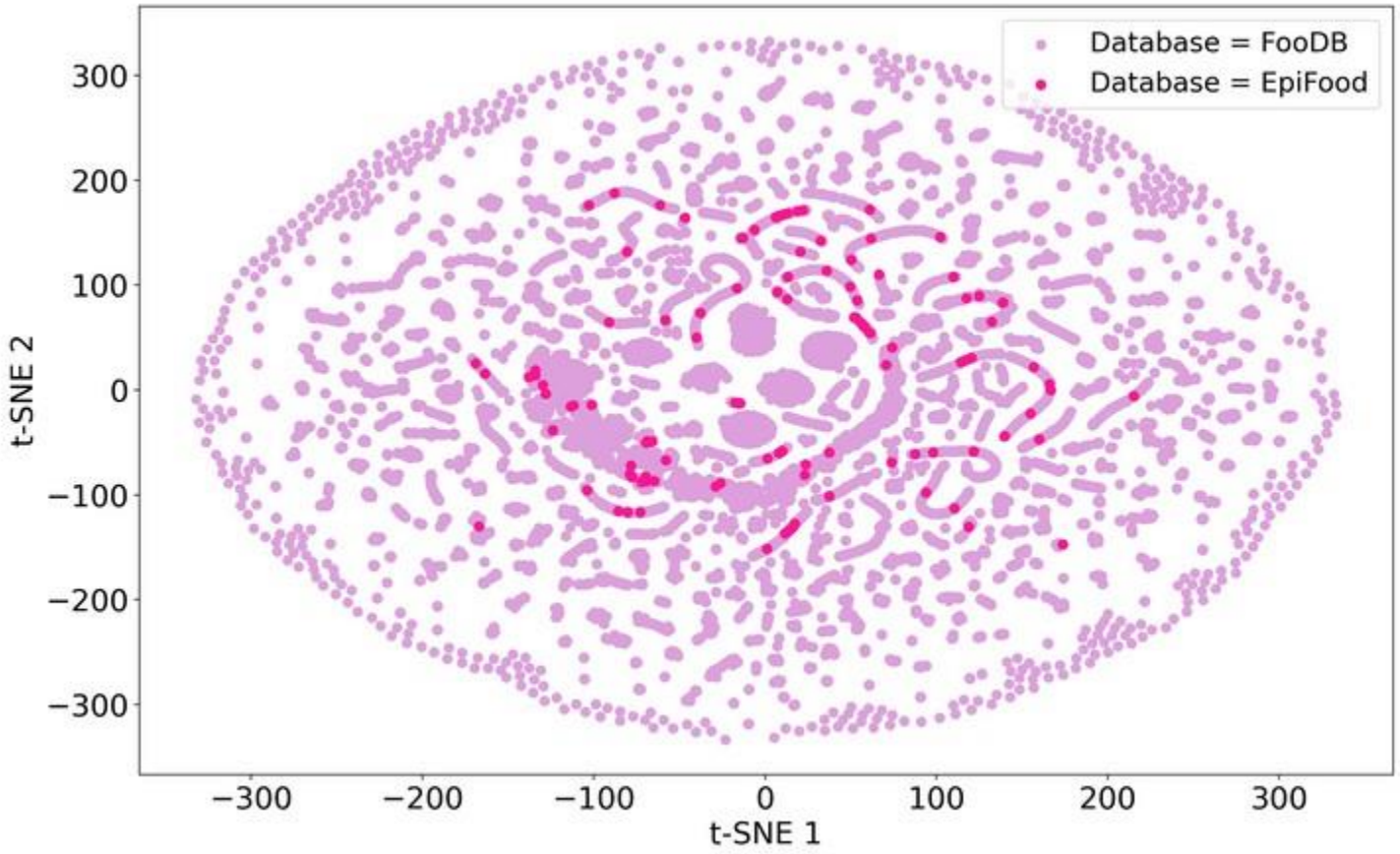
- Breast cancer
- Cervical cancer
  Prostate cancer
- Colon cancer
- Lung cancer

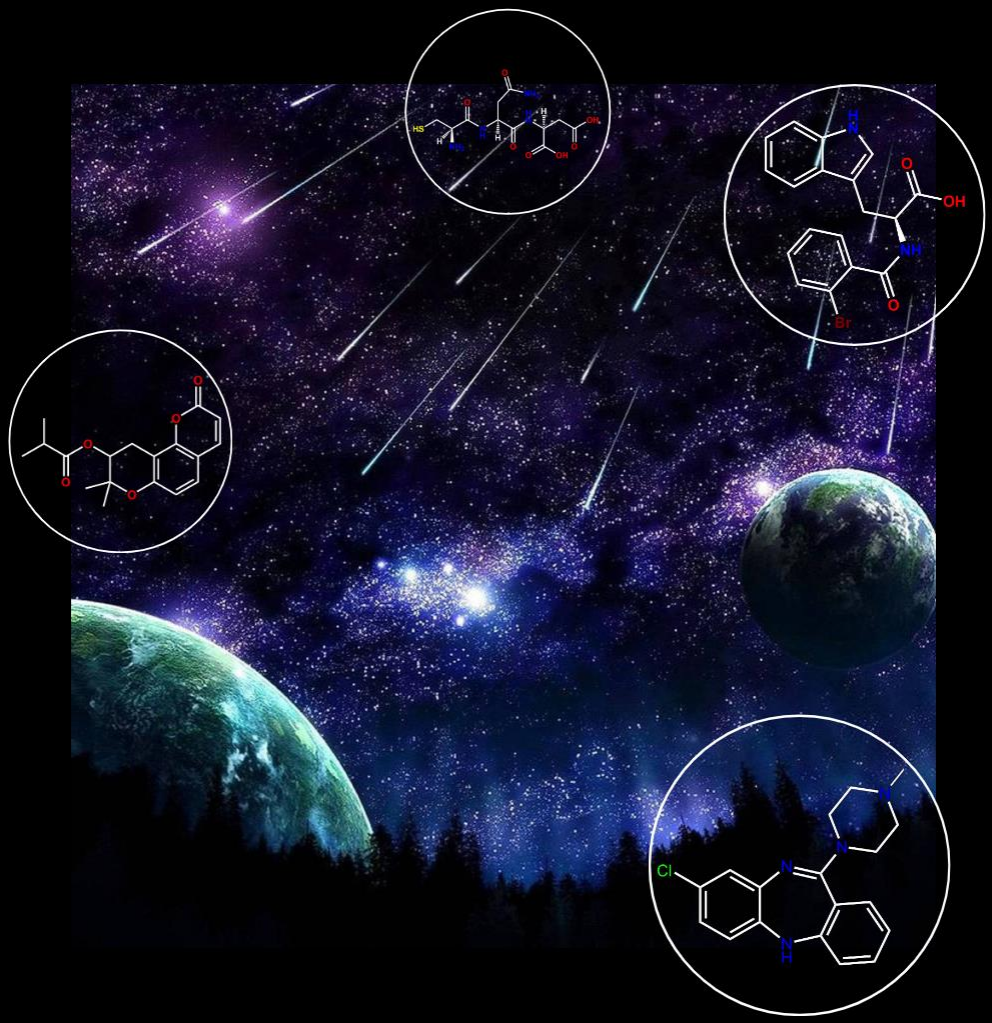
Montalvo-Casimiro M et al. Epidrug Repurposing: Discovering New Faces of Old Acquaintances in Cancer Therapy. Frontiers in Oncology 2020: 10

Results

Visualization of the chemical space

t-Distributed Stochastic Neighbor Embedding





Summary

Figure credit: Yazmin Percastre

Chemical universe

Conceptual framework of chemoinformatics with many applications.

Chemical multiverse

- A **group of chemical spaces**, each one defined by a given set of descriptors.
- Useful for a comprehensive analysis of chemical space.

Latin American Natural Product Database: LANaPDB

- Nine databases from Brasil, Colombia, Costa Rica, Mexico, Panama, Peru, El Salvador
- Current version: 12,959 compounds.

(Epi) Food Chemical Database

- A curated a compound database of 184 food and natural products with structural information and epigenetic target activity profile.
- Starting point to do structure-epigenetic activity relationships.

Biological and Medicinal Chemistry

Navigating the Chemical Space and Chemical Multiverse of a Unified Latin American Natural Product Database: LANaPDB

24 August 2023, Version 1

Working Paper

Alejandro Gómez-García, Daniel A. Acuña Jiménez, William J. Zamora, Haruna L. Barazorda-Ccahuana, Miguel Á. Chávez-Fumagalli, Marilia Valli, Adriano D. Andricopulo, Vanderlan da S. Bolzani, Dionisio A. Olmedo, Pablo N. Solís, Marvin J. Núñez, Johny R. Rodríguez Pérez, Hoover A. Valencia Sánchez, Héctor F. Cortés Hernández, José L. Medina-Franco

Show author details

Agriculture and Food Chemistry

Food Chemicals and Epigenetic Targets: Assembling an Epi Food Chemical Database

30 August 2023, Version 2

Working Paper

K. Eurídice Juárez-Mercado, Juan F. Avellaneda-Tamayo, Hassan Villegas-Quintero, Ana L. Chávez-Hernández, Claudia Daniela López-López, José L. Medina-Franco

Show author details

Research group DIFACQUIM
Computer-Aided Drug Design at School of Chemistry

BSc

Hassan Villegas

Daniela Gaytán

Jazmín Miranda

Atzi Valdez

Fernando Martínez

MSc

Eurídice Juárez

Brayan Raziel

Massyel S. Martínez

Juan F. Avellaneda

Carlos Daniel

PhD

Alejandro Gómez

Fernanda Saldivar

Diana Prado

Ana Chávez

C. Adrián González

Co-advisor: Dr. Enrique García

Edgar López (CINVESTAV)

Co-advisor: Dr. Carlos C. Rojas

National Autonomous University of Mexico, UNAM

Funding

CONACYT
dgapa Dirección General de Asuntos del Personal Académico
miztli
HUAWEI
SRE SECRETARÍA DE RELACIONES EXTERIORES
RMAS UNAM