



Biosciences Working Group

Jason Haga

AIST

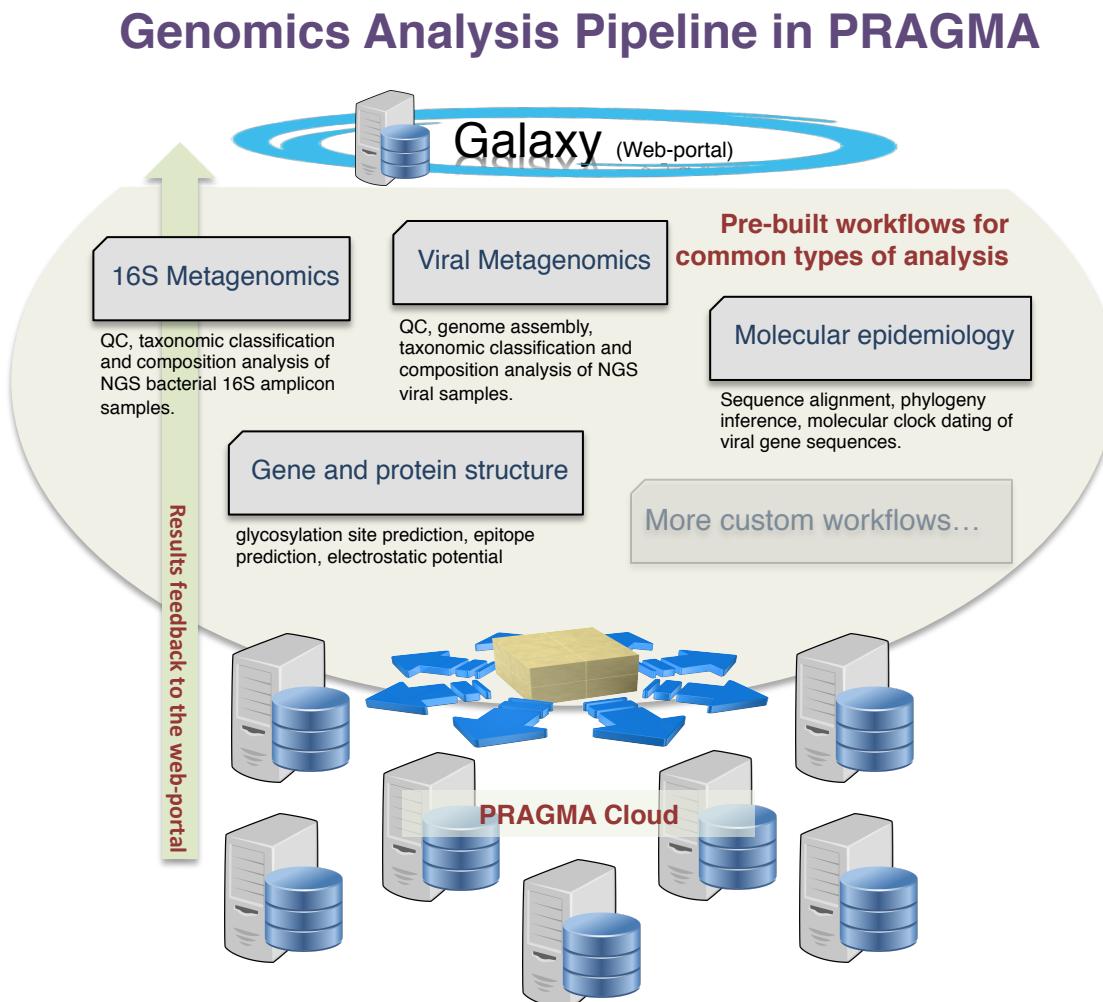
PRAGMA 30, Jan 28, 2016

Bioscience Activities

- Genomics Analysis Pipeline
 - Creating resources to enable genomics analysis
- Natural Products Discovery Portal
 - Create a unified platform for drug discovery
- Both driving applications of PRAGMA Cloud
 - Cross-working group interaction

Genomics Analysis Pipeline

Infectious disease as domain



Phase 1: Set up web-portal and computing infrastructure for Galaxy and genomics tools in PRAGMA resources.

Phase 2: Build workflows for common types of genomic analysis, initially for field of infectious disease research and then extend to other fields with the contributions of other PRAGMA participants.

Phase 3: Testing using case studies contributed by PRAGMA participants.

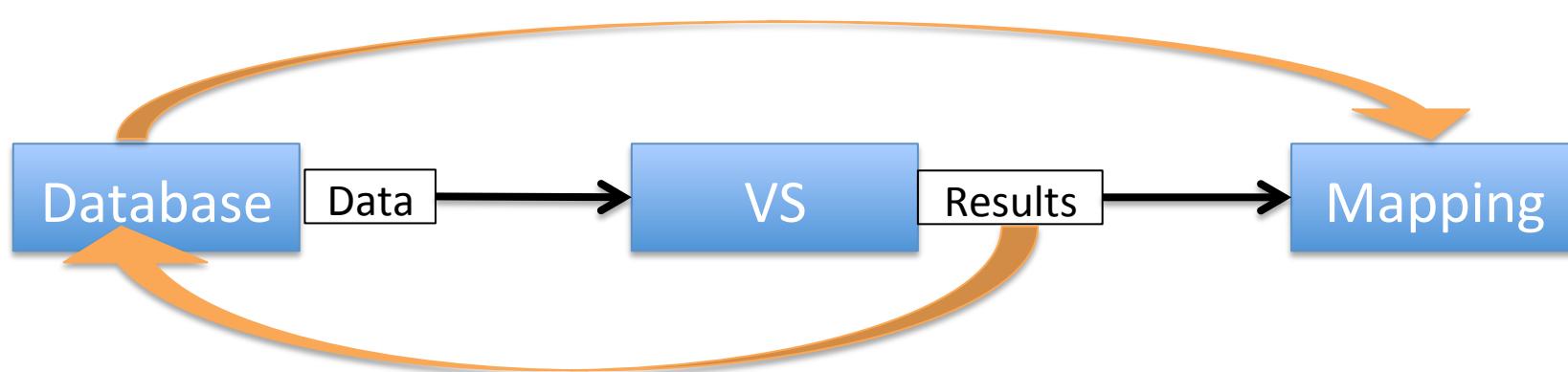
Phase 4: Education via training workshops and invite wider contributions to building common analysis workflows and application with PRAGMA participants' data.

Since PRAGMA29

- Genomics Analysis Pipeline
 - Created KVM-based VM with Biolinax (Kohei)
 - Preliminary testing of VM (Tommy)
 - VM is now on UCSD resources and in scheduler (Shava)
 - Starting to build up PRAGMA Cloud resources (AIST)

PRAGMA Natural Products Discovery Platform

- Components
 1. Natural products database
 2. Virtual screening
 3. Biodiversity mapping
- What are the infrastructure requirements for each component?
 1. Storage, data format
 2. Network, compute nodes
 3. Mapping tools, visualization



BASIS DATA TANAMAN OBAT INDONESIA

<http://herbaldb.farmasi.ui.ac.id/v3/index.php>

Beranda Daftar Spesies Daftar Senyawa FAQs Tentang Kami IND ENG

Database Senyawa Aktif Tanaman Obat Indonesia

Box of Search

Kategori Pencarian
Kunci Pencarian
Search

Members Login

Ingin Mendaftar? Login

Avocado (Alpukat)
Tumbuhan Avocado berasal dari Meksiko dan Amerika Tengah dan kini banyak dibudidayaan di Amerika Selatan dan Amerika Tengah sebagai tanaman perkebunan monokultur dan sebagai tanaman pekarangan di daerah-daerah tropika lainnya di dunia.

3 / 3

Sto

HerbalDB (Indonesia)

http://kanaya.naist.jp/KNAPSAcK_Family/

“KNAPSAcK” Family
Since 2008.07

NAIST

KNAPSAcK Metabolomics

3D Since 2012.11 **Core System** Since 2004.04 **Search Engine** Since 2008.12

Pocket Search for Functional Species

Food & Health	Crude Drug	Biology
Lunch Box 食用データベース Since 2008.07	WorldMap 世界の薬用植物 データベース Since 2009.06	Biological Activity Natural Activity Since 2011.06
DietNavi 病気予防データベース Since 2012.11	KAMPO 漢方薬 生薬 データベース Since 2008.09	Biological Activity Metabolite Activity Since 2013.01
FoodProcessor 加工食品データベース Since 2012.11	JAMU IndonesiaHerb データベース Since 2009.11	
DietDish 食べ合わせデータベース Since 2012.11	Tea Pot ハーブ データベース Since 2011.06	
MARCHÉ 毎データベース Since 2014.04	Skewered KNAPSAcK 串刺し検索 Since 2010.10	

Picnic Gene Annotation
Arabidopsis Since 2008.04

Strap Correlation Coefficient
Arabidopsis Since 2009.09

Bacillus Bacillus Since 2008.05

Pickaxe Metalloprotein Database
MetalMine Since 2009.09

Motorcycle Metabolic Pathway
代謝データベース Since 2011.08

Bicycle Algae Metabolic Pathway
代謝データベース Since 2013.09

[Instruction Manual\(Japanese\)](#)

[Instruction Manual\(English\)](#)

KNAPSAcK (Japan)

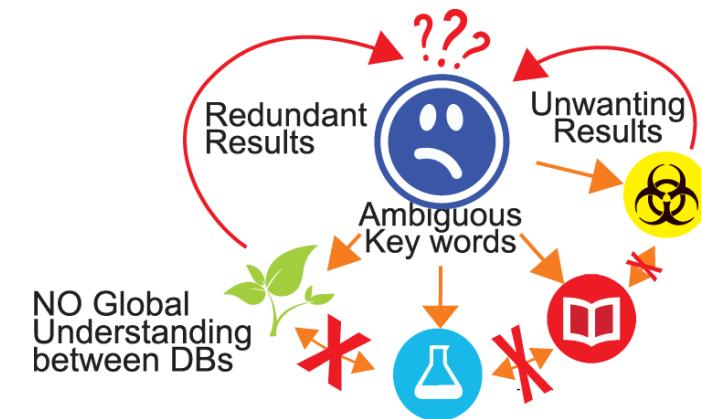
Since PRAGMA29

- Natural Products Discovery Portal
 - Database refinement/cleanup (Arry)
 - VietHerb database (Ly)
 - Autodock in Docker containers (Heru)
 - DOCK in Docker containers (Weicheng, Max)

Ontology-based Herb Database System for Pharmaceutical Science and Drug Development

Background and Challenge

- Huge but fragmented DBs for herbs, metabolites, and diseases
 - ➡ Big barrier for interdisciplinary studies
- Widely used DBs not semantically related
 - ➡ No knowledge sharing for analogous terminologies
- Terminologies ill-defined
 - ➡ Unwanted/redundant/incorrect results



Objectives and Contributions: VietHerb

- Integration of fragmented resources to unify and correlate Knowledge of Herbs, metabolites and diseases
- User friendly interface to assist both non-experts and interdisciplinary scientists for best results
- Updating mechanism for anonymous experts to collaboratively contribute to provide new knowledge

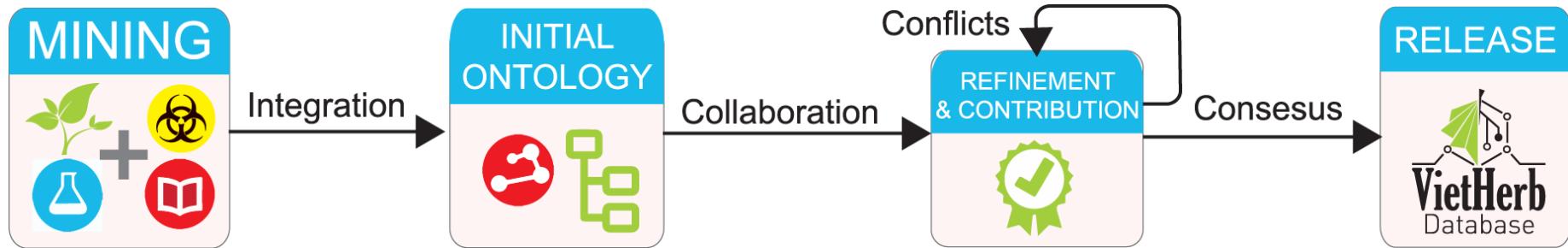
The screenshot shows the homepage of the VietHerb Database. It features a background image of various herbs and a central logo with the text "VietHerb Database". A banner on the right side reads "Most updates on Vietnamese Medical Herbal Database." and includes a link "[Find out more >>](#)". Below the banner is a navigation bar with links: HOME, DICTIONARY, SUGGESTIONS, FAQ, ABOUT, PARTNER, TIẾNG VIỆT. A search bar is located at the top right. The main content area displays a table with the following data:

ID	SCIENTIFIC NAME	Vn-NAME	Eng-NAME	METABOLITES	THERAPEUTIC EFFECTS
HID_00001	Aconitum napellus L.	Áu ô đầu	Monk's-hood [2] aconite wolfsbane, fuzi, monk's blood	Balanine G; T2-epi-Napelline; Songasmine	Animocleptic; Analgesic; Anti-epileptic; anti-convulsive; Anti-convulsive; Anti-shock; Anti-inflammatory; Anti-bacteria; Anti-funga;

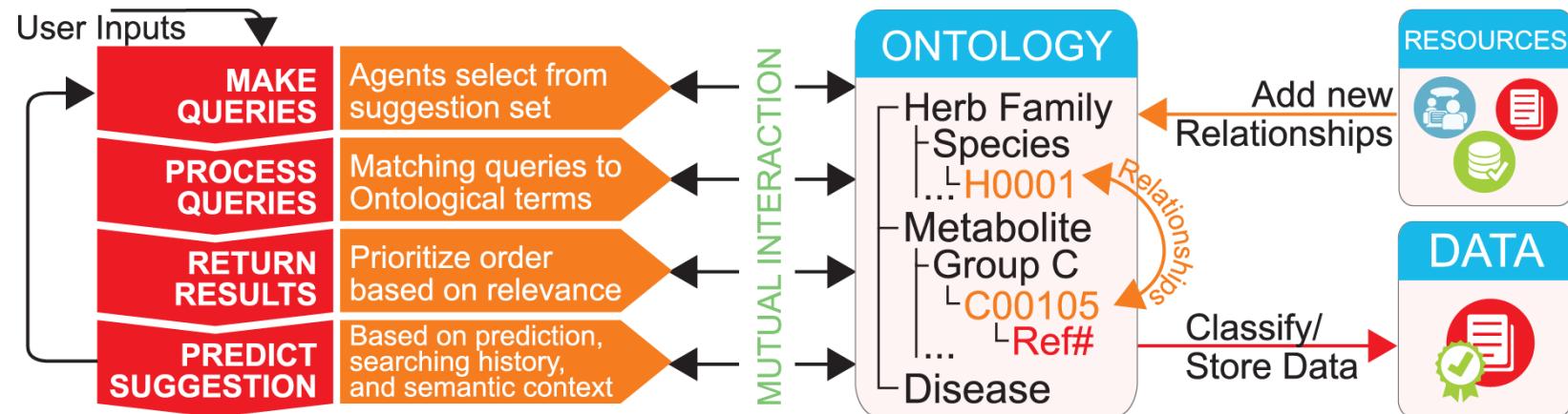
Ontology-based Herb Database System for Pharmaceutical Science and Drug Development

Key technologies

- Integration of different sources to build a sound ontology



- Use of ontology for semantic search and information exchange



PI: Ly Le (International Univ., Vietnam National University), FY15:30K, FY15:30K)



Biosciences WG Breakout

PRAGMA 30, Jan 28, 2015

Biosciences WG – Session 1

Proteomics

- Heru
 - Preliminary results using Docker and Autodock
 - Compare with AWS, physical cluster (Docker seems better)
 - For AMBER – no difference physical vs Docker
- Ricky Nellas
 - Schistosoma – looking for inhibitors
 - Maybe should try NAMD on GPU
- Kwan
 - System built and looking for users
 - NAMD users, GROMACS users
 - Restriction – need to collaborate with HKU researcher, then can use

Genomics

- Ramil Mauleon IRRI
- 3K rice genomes project
- Genetic approach to find adaptive traits already there
- 17TB of raw data, 120TB of data from analysis
- 500 compute cores 3-4 weeks – doable
- Breeding cycle for rice is 90days, thus need to do analysis in this time

Biosciences WG – Day 2

- Joint meeting with Resources WG
- Two driving applications
 - Genomics analysis VM
 - Genomics VM on SDSC resources
 - Eager to use PRAGMA resources for smaller-scale analysis
 - Possible to repeat PID service on rice genomics data analysis
 - Sending a copy to them so they can customize
 - Virtual screening VM
 - Already created and tested
 - To be placed on PRAGMA Cloud for use
- Installing DOCK on Docker containers
 - Need to synchronize efforts (e.g. NCHC, UF)

Tentative Timeline

- PRAGMA 31
 - DOCK VM on PRAGMA Cloud
 - Custom VM for rice genomics
 - Define the use of PID service in rice genomics
 - Need to define what this is
 - Attaching PIDs to VMs
 - Attaching PIDs to Galaxy workflows
- PRAGMA 32
 - PID service with genomics VMs
 - More complete testing of genomics VM
 - Test natural products database in virtual screen
- PRAGMA 33
 - Genomics analysis