Biosciences WG Updates

Jason Haga
AIST
PRAGMA 29, Oct 8, 2015

Since PRAGMA28

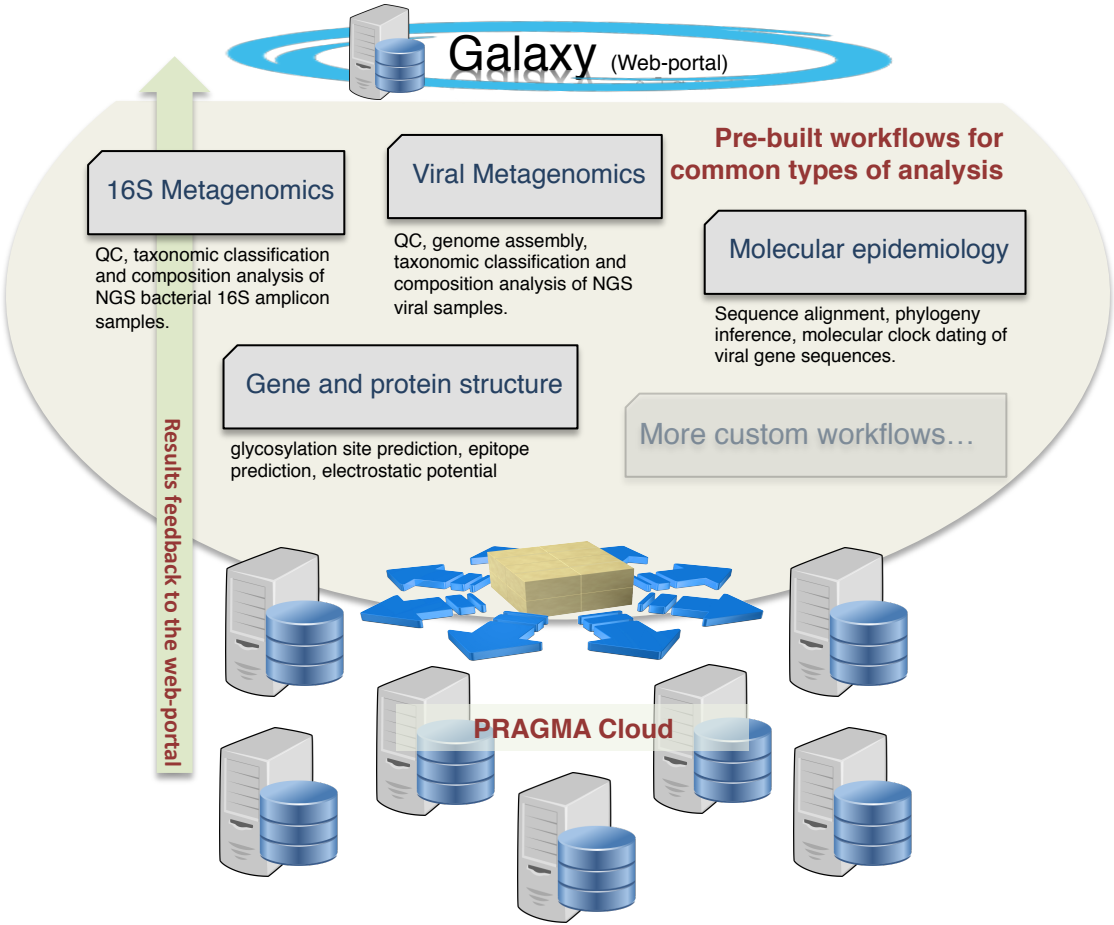
- Regular conference calls
- Two areas of activity: genomics and proteomics
  - Genomics Analysis Pipeline
    - Established a roadmap (Tommy)
    - Selected Biolinux (Jason/Tommy)
    - Created KVM-based VM with Biolinux (Kohei)
    - Preliminary testing of VM (Tommy)
  - Natural Products Discovery Portal
    - Database refinement/cleanup (Arry)
    - VietHerb database (Ly)

Genomics Analysis Pipeline

- Infectious disease as domain
- Regular Skype conference calls
- Draft a “roadmap” (Tommy)

- Action items
  - Select and install genomics tools (Tommy, Jason)
    - Galaxy, iPlant
  - Influenza A virus and MRSA as small test for assembly and annotation (Ly)
    - Define file size, tools used, etc.
  - Hadoop support (Andrea)

Genomics Analysis Pipeline in PRAGMA



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

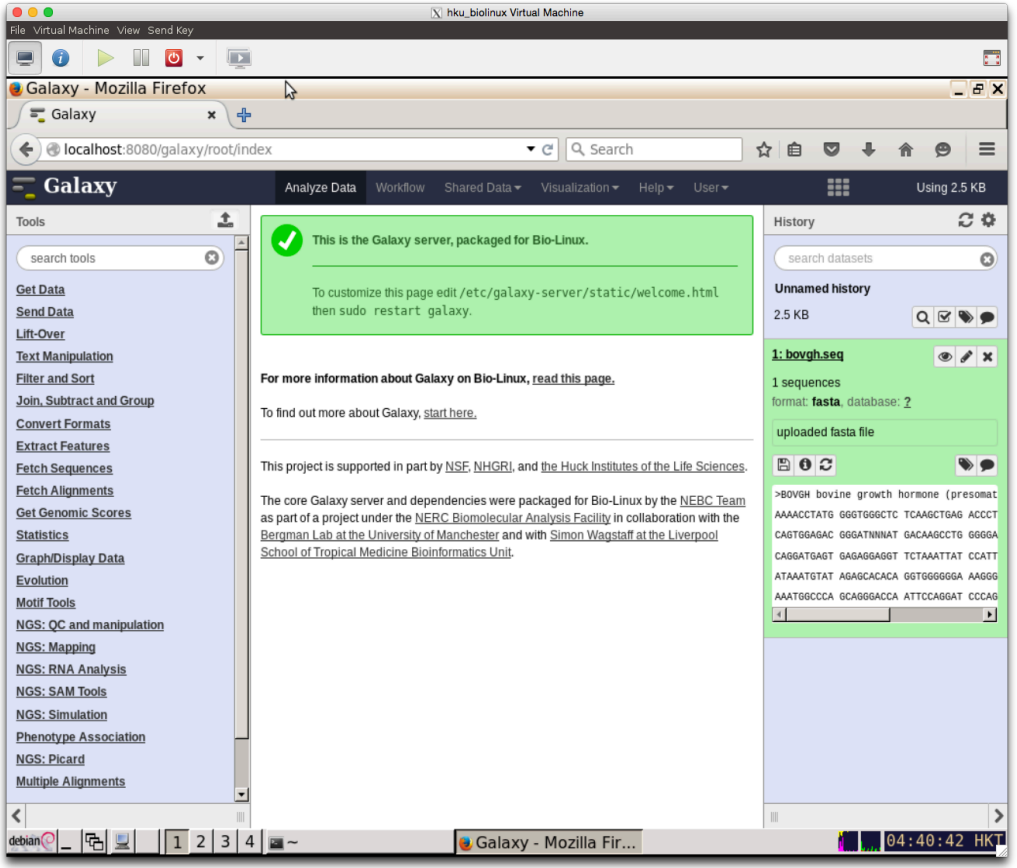
Phase 2: Build workflows for common types of genomic analysis, initially for field of infectious disease research and then extend to other fields with the contributions of other PRAGMA participants.

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.

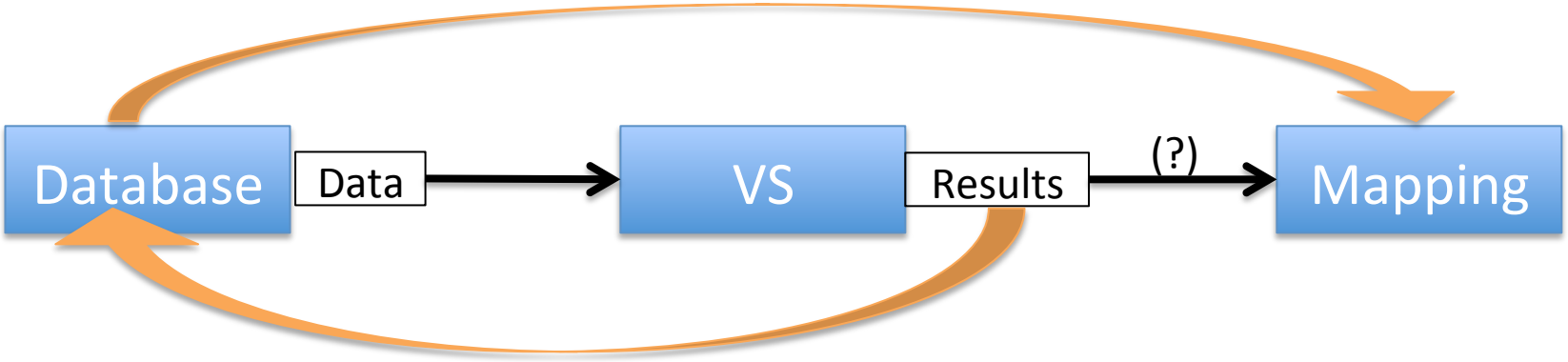
PHASE 1: Installation of GALAXY in VM

- GALAXY was successfully installed in the PRAGMA-compatible VM
  - with *Biolinux8*
  - with window manager *icewm*
- But this GALAXY inside the VM is supposed to be isolated from the world outside the VM, so we
  - need to create a central web-portal to accept computational jobs
  - need to allow connection between VM-inside GALAXY and the web-portal GALAXY
- Other tasks in PHASE 1
  - Identify minimum package of Biolinux8 to reduce transfer load.
  - [Do you have any ideas?]



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



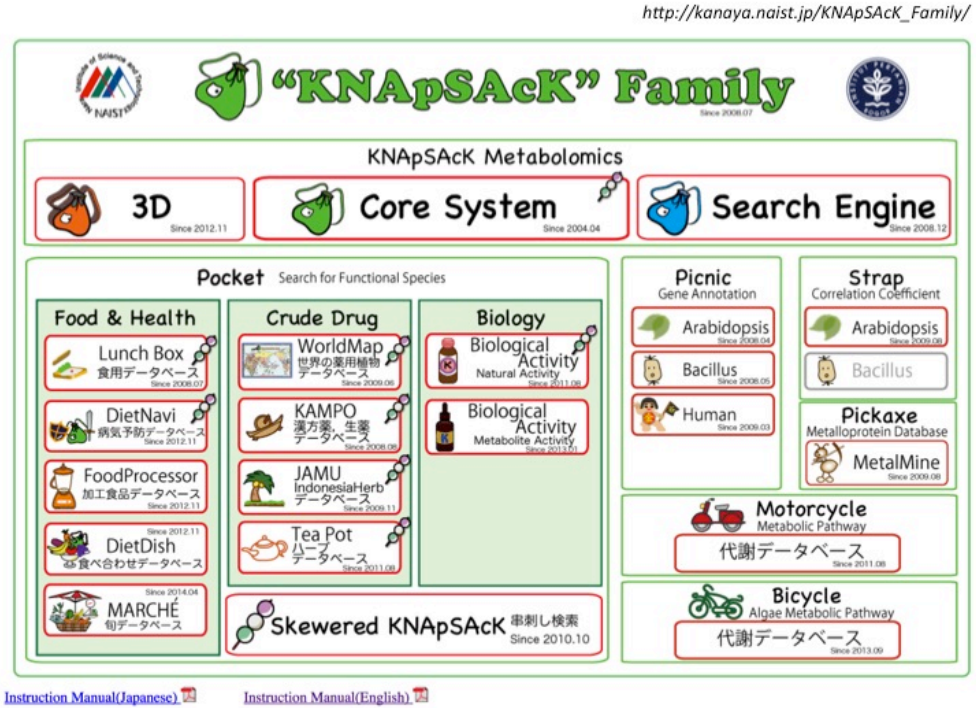
Natural Products Discovery Platform

- Regular Skype conference calls

- Action Items
  - Define ideal computing environment, applications, databases, size of VM (Arry, Ly)
  - Continued discussion with resources WG to create web services (Nadya, Ari)
    - Based around NBCR efforts e.g. PyRx
  - Start combining the natural products databases i.e. remove duplicates and combine (Arry, Kanaya, Ly)

- Benefit Vietnam (limited computational resources), Indonesia (move to virtualized system)

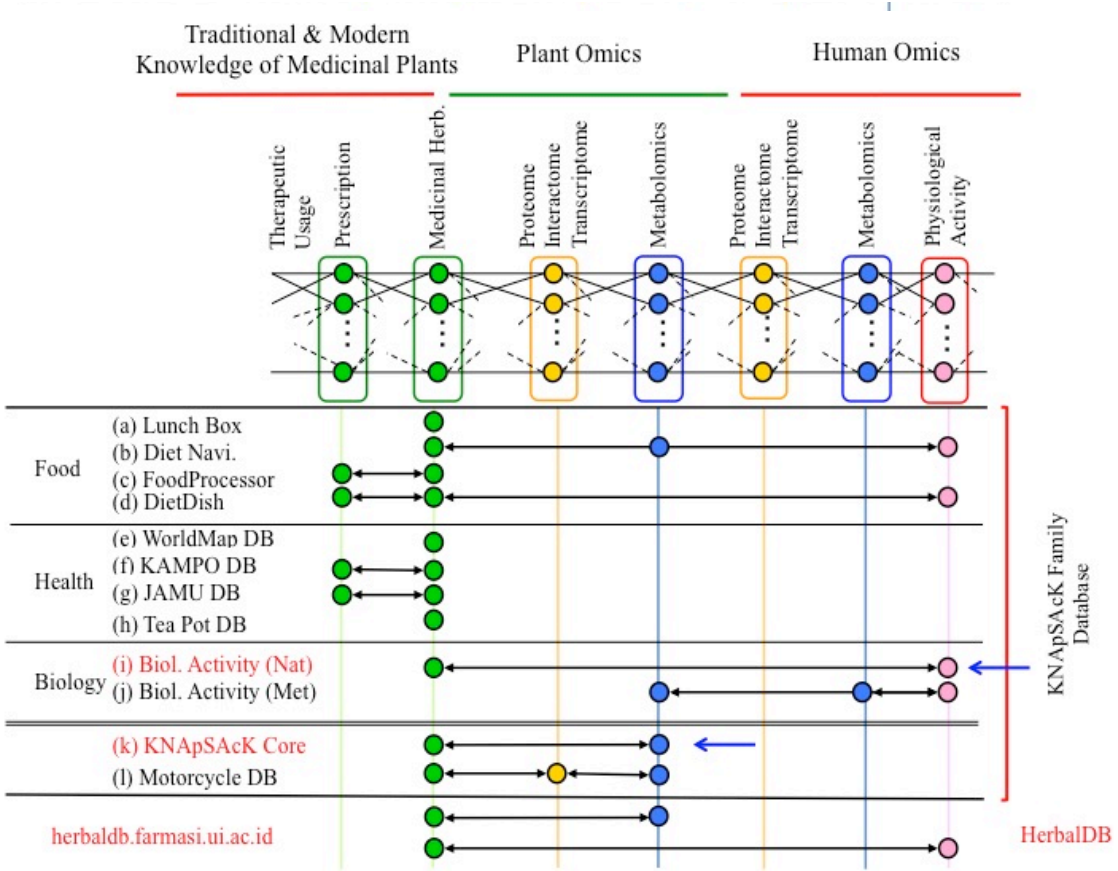
HerbalDB (Indonesia)



KNApSAcK (Japan)

HerbalDB & KNApSAcK intersection

- To find possibility of data sharing between HerbalDB and KNApSAcK DB



VietHerb Database

- Vietnam has an extremely rich sources of medicinal plant
- Detailed information about phytochemicals has not been standardized and digitized
- This database inspired by KNApSAcK database system (Japan) and Jamu database system (Indonesia)
- Create an interactive, electronic herbal database
  - Impartial, evidence-based information resource
  - Public site provides free access
  - Use of herbs for health