EGEE Conference 2006

swissPIT
a usecase for proteomic software
on the SwissBioGrid
infrastructure

Andreas Quandt
Swiss Institute for Bioinformatics in Geneva
andreas.quandt@isb-sib.ch

Project overview

swissPIT = Swiss Protein Identification Toolbox

...tries to analyze mass spectrometry data with different programs.

(more details by Patricia Hernandez on Thursday)

Project overview
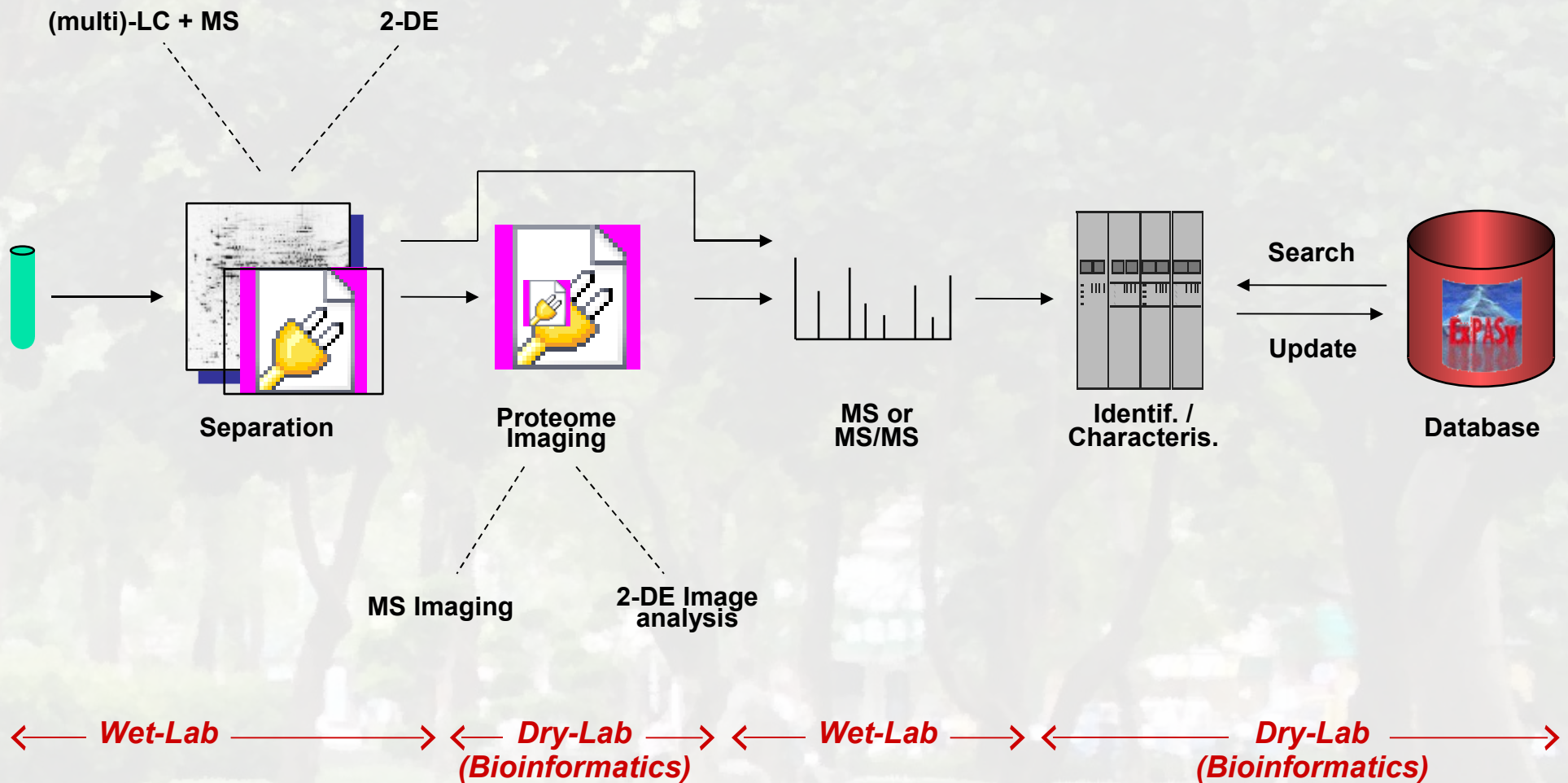
swissPIT = Swiss Protein Identification Toolbox

...tries to analyze mass spectrometry data with different programs.

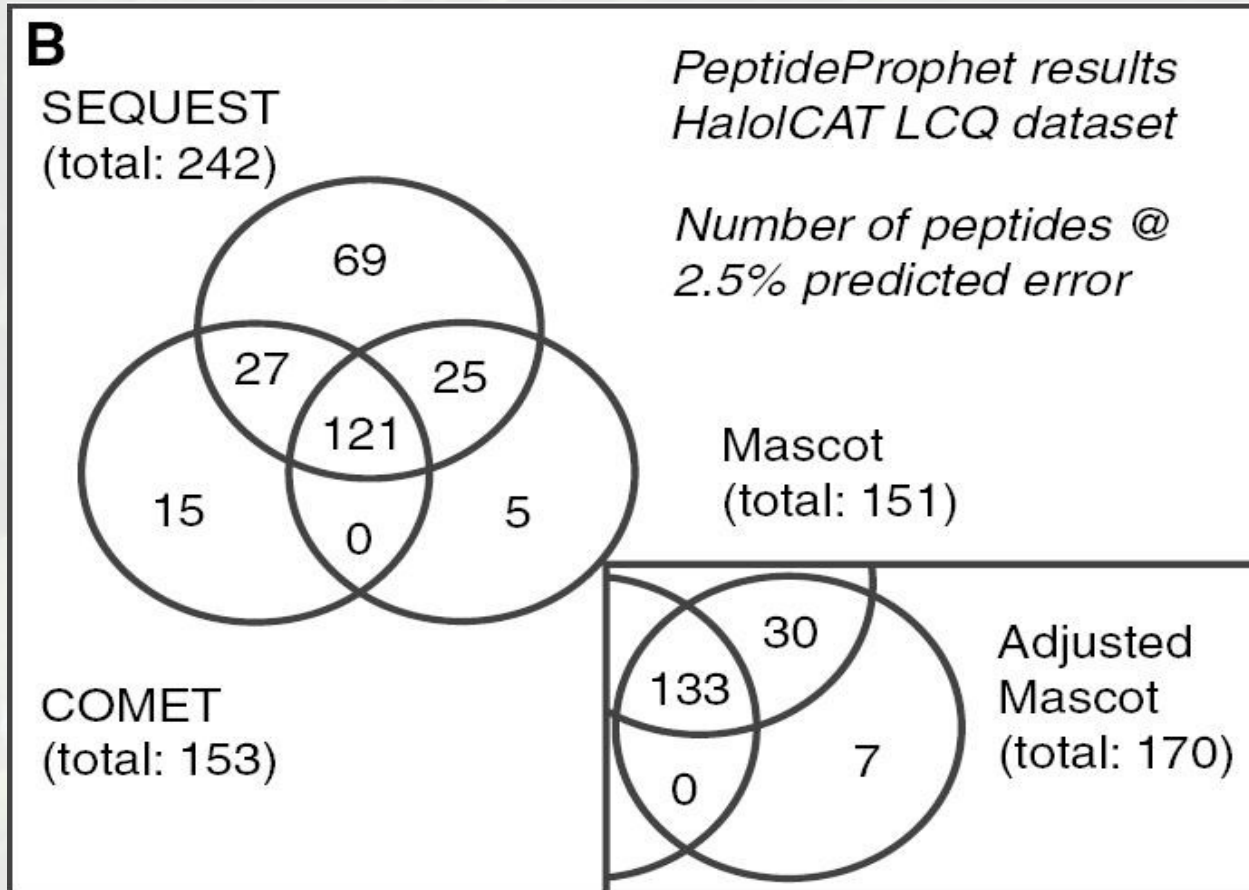*(more details by Patricia Hernandez on Thursday)*

I am sure some of you tink:

"What does this guy want? This is the EGEE Conference 2006 and not some classes in biology...?!?"

Proteomics workflow

(multi)-LC + MS   2-DE

Separation → Proteome Imaging → MS or MS/MS → Identif. / Characteris. → Database

Search

Update

MS Imaging   2-DE Image analysis

← Wet-Lab → ← *Dry-Lab (Bioinformatics)* → ← *Wet-Lab* → ← *Dry-Lab (Bioinformatics)* →

by Ron D. Appel

4

A.Keller et al. (2005)

Scope of the user

- The Biologist:
  - Semi-automatic system to analyze MS-MS data
    - Easy to understand and simple to use
    - The results if possible yesterday

- The Bioinformatician/Statistician/experienced Biologist:
  - Benchmarks
    - How to improve the identification process ?
      - Which strategy/program for which data?
      - Which parameter set?
  - Data analysis across projects and groups
    - How meaningful are the results ?
      - Comparison and Validation?

swissPIT and the big computers

Can I install swissPIT on my own computer?

...yes, but you do not become happy with it, because...

- Pierre-Alain Binz (GeneBio SA):
  - Data files up to 300MB packed binary ~ 2GB raw data
  - Computing time up to 32h (depending on parameter + CPU's)

- Paul Jenoe (Biozentrum Basel):
  - Sequest is 10x slower in comparison to Phenyx (GeneBio SA)

Goals of swissPIT

1. Compare available software
2. Implementation of analysis workflows
3. Unified interface for different programs

Where to start first:

- Web interface
  - Data management
  - Parameter sharing + standardization between programs
  - Unified result visualization

- Automatic system to execute programs in parallel

What do we need from the community

- Automatic certificate handling
- Distributed DB's and version controlling
- User space + data sharing
- Data backup
- New licensing models for commercial software
- Automatic workflow control
- Job monitoring
- Support to optimize software for distributed computing

End of the talk

Thank you for your attention!