

The GCP Platform: integrating data sources and facilitating analysis

Generation Challenge Programme
Annual Review Meeting
Benoni, South Africa
12-16th September 2007

Graham McLaren
IRRI-CIMMYT Crop Research Informatics Laboratory

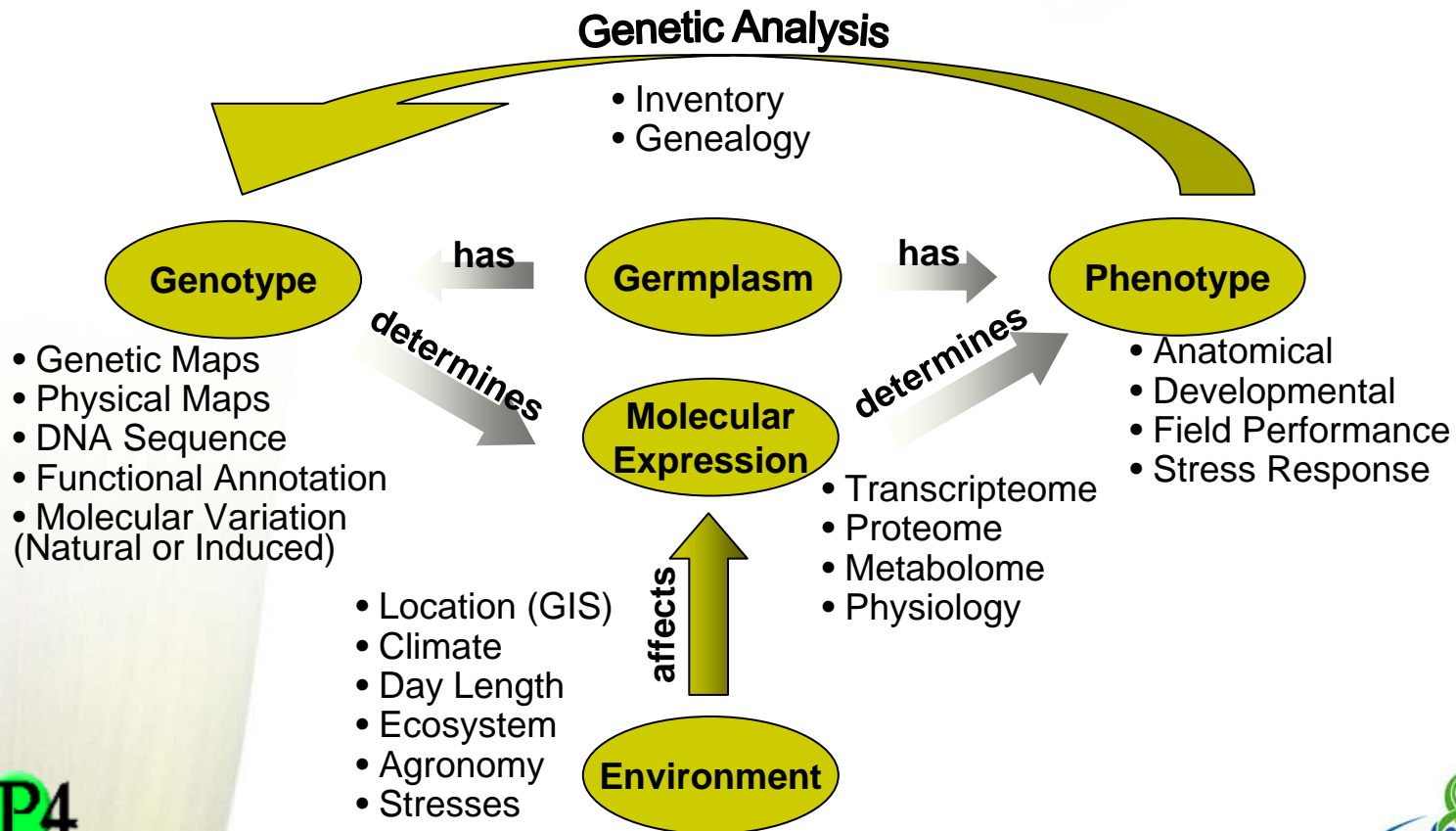


Rationale

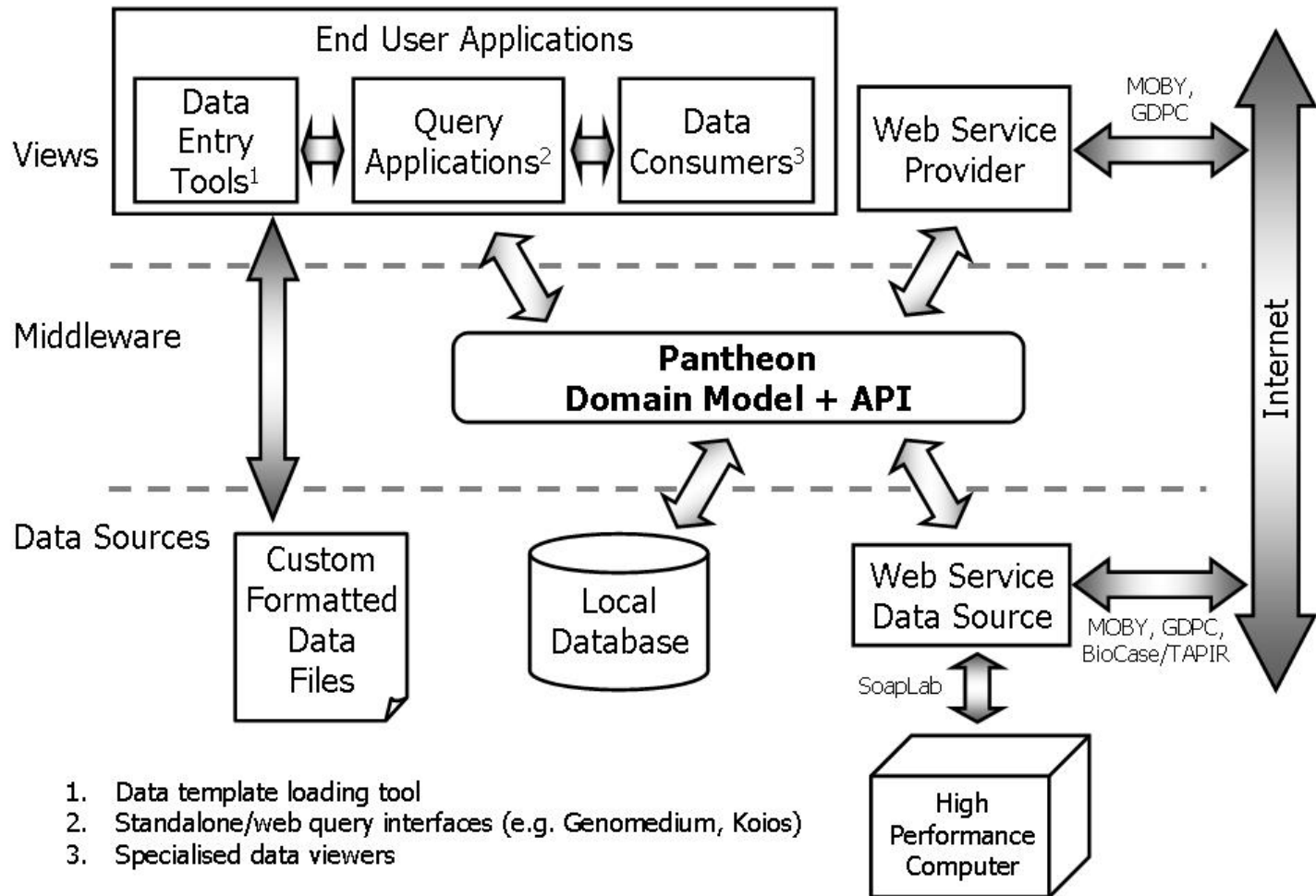
- ❖ Vast amount of germplasm and genomic data generated by CP partners
- ❖ Informatics requirements for data:
 - acquisition
 - storage
 - analysis
 - sharing
- ❖ Challenge: heterogeneous information systems and data sources dispersed across the CP institutes

Information Systems Platform & Network Design

- Crop biological concepts, relationships and data types



Generation CP Informatics Platform



2007 Team Members and Activities

- Principal Investigator: Graham McLaren - IRRI
 - SP1 Use Case: Manuel Ruiz – CIRAD
 - Brigitte Courtois, Claire Billot -CIRAD
 - SP2 Use Case: Shoshi Kikuchi, Koji Doi – NIAS
 - SP2 Use Case: Richard Bruskiewich – IRRI
 - Ramil Mauleon – IRRI, Marcos Costa - EMBRAPA
 - SP3 Use Case: Guy Davenport – CIMMYT
 - Kyle Brack - CIMMYT
 - Martin Senger – EBI and Mathieu Rouard – INIBAP
- Activities:
 - http://cropwiki.irri.org/gcp/index.php/Platform_Development_Activities_for_2007

2007 Activities: SP1 Use case

Analysis of biological diversity for GCP data sets.

- The first tool will be a sub-sampling tool.
- The second tool will allow to convert data for easy input into a range of software packages.
- A plug-in allowing to adapt the tool to new packages will be developed.
- Automatic linkage with the high power computing facilities.

List of available genotyping studies (by species and databases)

Studies for Barley and GCP Central Repository

- gcpcr_file655

Studies for Chickpea and GCP Central Repository

- gcpcr_file654

Studies for Coconut and GCP Central Repository

- gcpcr_file666
- gcpcr_file666

Studies for Rice and Tropgene

- Mini_GB_Isozymes

Studies for Sorghum and Tropgene

- GCP_Sorghum_3400accessions_39SSR

Studies for Chickpea and Tropgene

No study available

Studies for Coconut and Tropgene

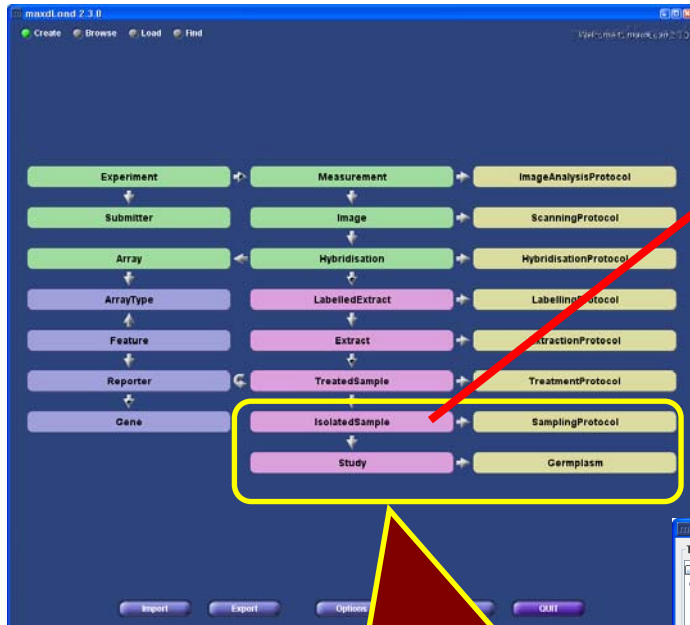
- Fingerprinting data of 1215 coconut accessions with 30 markers

2007 Activities: SP2 Use case

Analysis of Microarray Data.

- MAXD-based gene expression database constructed description of experimental conditions adhering to the MIAME/Plant standard.
- Probe data can be incorporated through the array converter system.
- New, but existing analysis tools, such as the TMEV suite will be integrated into the platform.
- The cis-element search and analysis application will be enhanced to accommodate new sources of sequence data and new types and locations of regulatory elements.

GCP Platform Support for Task 32 (MAXD Customization for GCP Microarray Data)



The screenshot shows the 'CREATE ISOLATEDSAMPLE' form in MAXD. The form includes fields for Study, ObservationUnit, Germplasm, SamplingProtocol, Name, and Attributes. A yellow callout box points to the 'Study' field.

"IsolatedSample" for microarray experiment is just a Germplasm Identifier recorded in a GCP Study as an "Observation Unit"

MAXD customized for crop data (MIAME-Plant) by connection to Germplasm & Study domain model compliant GCP Data Source and to GCP ontology database

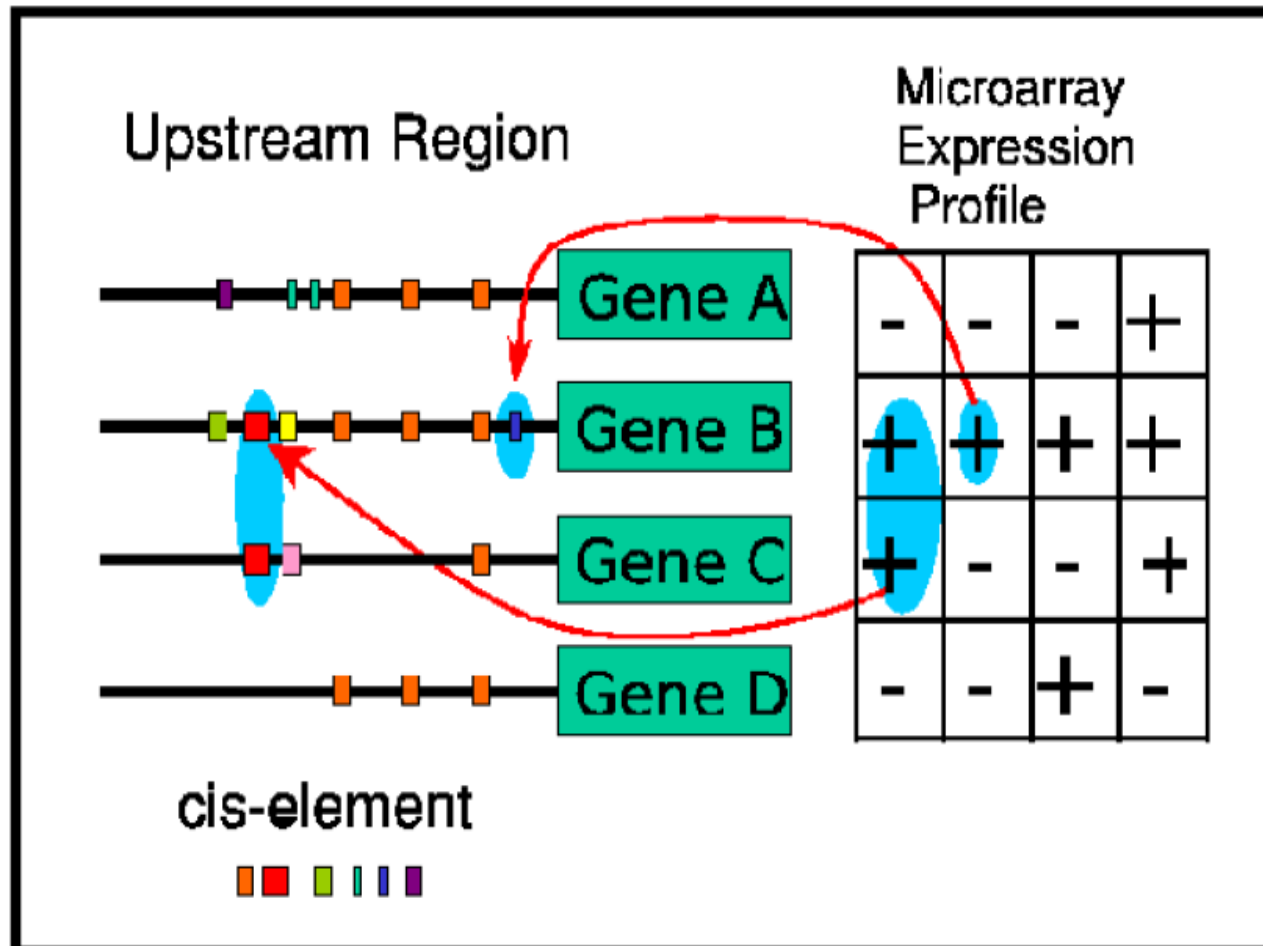
The screenshot shows the 'Ontology terms' dialog box. It features a 'Terms Tree' on the left with 'clone_start' and 'gap' selected. The 'Search' section is also visible, showing a search for 'junction'.

MIAME-Plant ontologies imported into MAXD through an "interactive" ontology database GCP DataSource



Basic Strategy for Search and Evaluation of *Cis*-Element Candidates

If motifs overrepresented in the upstream region of the focused genes, they could play specific roles on expression regulation of the genes.



2007 Activities: SP3 Use case

Tools for Marker assisted backcrossing (MABC)

- Protocols for routine management of genotyping and evaluation data for molecular breeding in platform compliant databases will be developed.
- A suite of applications including graphical visualization tools for MABC and display tools for background or target genotype selection from high throughput genotyping studies.

2007 Results - presentations

- **Combining decentralised genotyping data for diversity analysis (C. Billot)**
 - **GenDiversity:** a Web application for querying and combining different genotyping studies and delivering different output formats.
 - **GenMapping:** a Web application for getting genetic and physical mapping data related to the markers used in genotyping studies.
- **GCP Informatics Platform Demonstration**
 - **Standalone tools for genomics analysis (R. Mauleon)** cross-integration of microarray and mapping data with MAXDLoad, TMEV and Cytoscape and with GCP genomic DataSources (using GMOD and Chado) and GCP germplasm and phenotype DataSources.
 - **Dayhoff:** Comparative stress gene catalogue (S. Wanchana) uses GCP data sources in Chado and data consumer technology to integrate ATV and JalView as well as web services to communicate with Greenphyl.
 - **Genoma:** Management of EST data (M. Costa) a GCP partner tool adapted to use GCP data source and consumer technology to link third party tools such as Apollo
 - **Koios Querybuilder:** GCP domain model and API compliant web query engine and interface (R. Bruskiwich)

http://koios.generationcp.org

Generation Challenge Programme Koios Search Engine - Mozilla Firefox

http://koios.generationcp.org/Koios/views/main.jsp

Koios Search Engine

Search for Germplasm

Submit

All Data Sources
Selected Data Sources:

IRIS

Data Type Selection

- GCP Data Type
 - Entity
 - Study
 - Ontology Term
 - Germplasm**
 - Map Feature

Search Result Synopsis

Retrieve Result List

| # | Data Type | Value | Data Source | Hits |
|---|-----------|-------|--------------------------|------|
| 1 | Germplasm | IR64 | http://www.iris.irri.org | 1 |
| 2 | Germplasm | IR8 | http://www.iris.irri.org | 4 |

Deselect All Clear All Delete

Retrieve Item Details

| # | Identifier |
|---|------------|
| 1 | YS 528 |
| 2 | IR 8 |
| 3 | IR 8 |
| 4 | IR 8 |

Detailed Information

Data Sources: http://www.iris.irri.org

Unique Identifier: iris.irri.org|Germplasm|603539

Name: IR 8

| Attribute | Value |
|---------------|--|
| Annotations | [] |
| AuditTrail | [] |
| Citations | INT'L RICE GENE BANK |
| CropName | 603539=RICE |
| Relationships | [] |
| Sources | [] |
| Studies | [] |
| Synonyms | [603539 = IR 8, 603539 = CRI 1133, 603539 = IRGC 83082, 603539 = IR8, 603539 = O. SATIVA, 603539 = 1133] |

start

BIAC IV - Microsoft In... Microsoft Outlook We... Generation Challenge... Microsoft PowerPoint ...

6:59 PM

2007 Results - posters

- 2.1 GCP Informatics Platform Support for Crop Informatics (R. Bruskiwich)
- 2.2 The development of a novel data mining tool to find cis-elements in rice gene promoter regions (S. Kikuchi)
- 2.4 Comparative stress gene catalogue and ortholog function display (S. Wanchana)
- 3.4 The many faces of Genomedium (K. Braak)
- 4.13 The GCP microarray data integration and analysis system (R. Mauleon)

Official Generation CP project middleware and core applications

- ◆ Category: [GCP - Generation Challenge Program](#)
- ◆ Development Status: [2 - Alpha](#)
- ◆ Environment: [Other Environment](#)
- ◆ License: [OSI Approved](#)
- ◆ Operating System: [OS Independent](#)
- ◆ Programming Language: [Java](#)

Registered: 2006-02-23 12:30

Activity Percentile: 100%

View project activity [statistics](#).

View list of [RSS feeds](#) available for this project

Data format and release

Source code at

<http://CropForge.org/>

Project: GCP Pantheon

Pantheon (source code)

- 4 administrators, 47 developers, 17 months
- 27,000 source code contributions from 29 developers located in 6 institutions worldwide

Developer Info

Project Admins:

[Guy Davenport](#)

[Manuel Ruiz](#)

[Martin Senger](#)

[Richard Bruskiwich](#)

Developers:

[Akinola Akintunde](#)

[Alex Cosico](#)

[Alexis Dereeper](#)

[Andrew Farmer](#)

[Arlet Portugal](#)

[Ayton Meintjes](#)

[B Jayashree](#)

[Barry Peralta](#)

[Edwin Rojas](#)

[Gautier Sarah](#)

[Genevieve Mae Aquino](#)

[Georgios Pappas, Jr.](#)

[Isaiah Mukema](#)

[James Wagner](#)

[Jan Michael Yap](#)

[Jasmin Baladad](#)

[Jeffrey Morales](#)

[Jon Mendoza](#)

[Joseph Anthony Hermocilla](#)

[Kevin Manansala](#)

[Koji Suzuki](#)

[Kyle Braak](#)

[Lord Hendrix Barboza](#)

[Luis Avila](#)



GCP Pantheon

Table of Contents

[GCP Models](#)

[Domain Model](#)

[BioMoby Data Types](#)

[XML Schemas](#)

[GCP Ontologies](#)

[GCP Use Cases](#)

[GCP Platform Components](#)

[GCP Data Sources](#)

[GCP Data Consumers](#)

[GCP Data Transformers](#)

[GCP Platform Compliant Applications](#)

[Software Releases for Pantheon Developers](#)

[Tutorials](#)

[Tutorial: How to use Pantheon](#)

[Tutorial: How to start a new project](#)

[Tutorial: GCP Data Sources](#)

[Tutorial: GCP Data Consumers](#)

[Tutorial: GCP Data Transformers](#)

[Tutorial: GCP Components Validator](#)

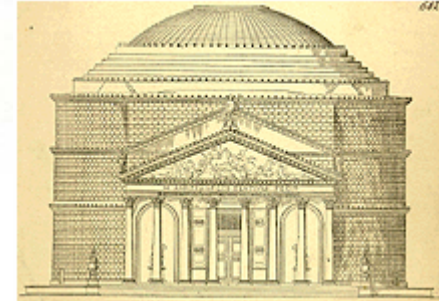
[Tutorial: Developing with Pantheon Eclipse](#)

[Tutorial: Using Maven from Ant](#)

[Tutorial: Using EMF to create the GCP domain model](#)

[Best Practices](#)

[Koios - the Pantheon Search Engine and Web Interface](#)



Documentation and Collaboration at

<http://pantheon.generationcp.org/>

Links with other projects

❖ SP1

- ❖ Data management for genotyping studies
- ❖ Analysis of genetic diversity

❖ Sp2

- ❖ Data management and analysis of microarray data
- ❖ Identification and documentation of orthologous stress genes
- ❖ Data mining for regulatory elements

❖ SP3

- ❖ Data management for pedigree breeding

❖ SP4 Data management for markers, maps and qtl.

Product delivery and impact

- **References**

Bruskiewich R, Davenport G, Hazekamp T, Metz T, Ruiz M, Simon R, Takeya M, Lee J, Senger M, McLaren CG and Van Hintum T. (2006) **Generation Challenge Programme (GCP): Standards for Crop Data**. OMICS: The Journal of Integrative Biology. 10(2):215-219

- **GenDiversity tool**

- <http://sat.cirad.fr/sat/gendiversity/servlet/ChoixInitiale>

- **Rice CIS element Searcher**

- <http://hpc.irri.cgiar.org/tool/nias/ces>

- **Koios Web-based query builder**

- http://koios.generationcp.org/Koios/index_mozilla.jsp

- Genomedium stand alone workbench

- <http://www.genomedium.org>

- Microarray repository and analysis tools

- <http://cropforge.org/projects/gcpmicroarray>

SP4