



A Molecular Breeding Platform

Portal, Services, Information System,
and Toolbox

Graham McLaren

05/03/09





Rationale

- Revolutions in molecular biology and information technology offer opportunities for enhancing plant breeding
- Successfully deployed in the private sector but not generally used in the public sector and hardly ever in developing countries
- It is time to provide the appropriate support tools, services and infrastructure to offer breeding programmes in developing countries access to modern marker breeding technologies

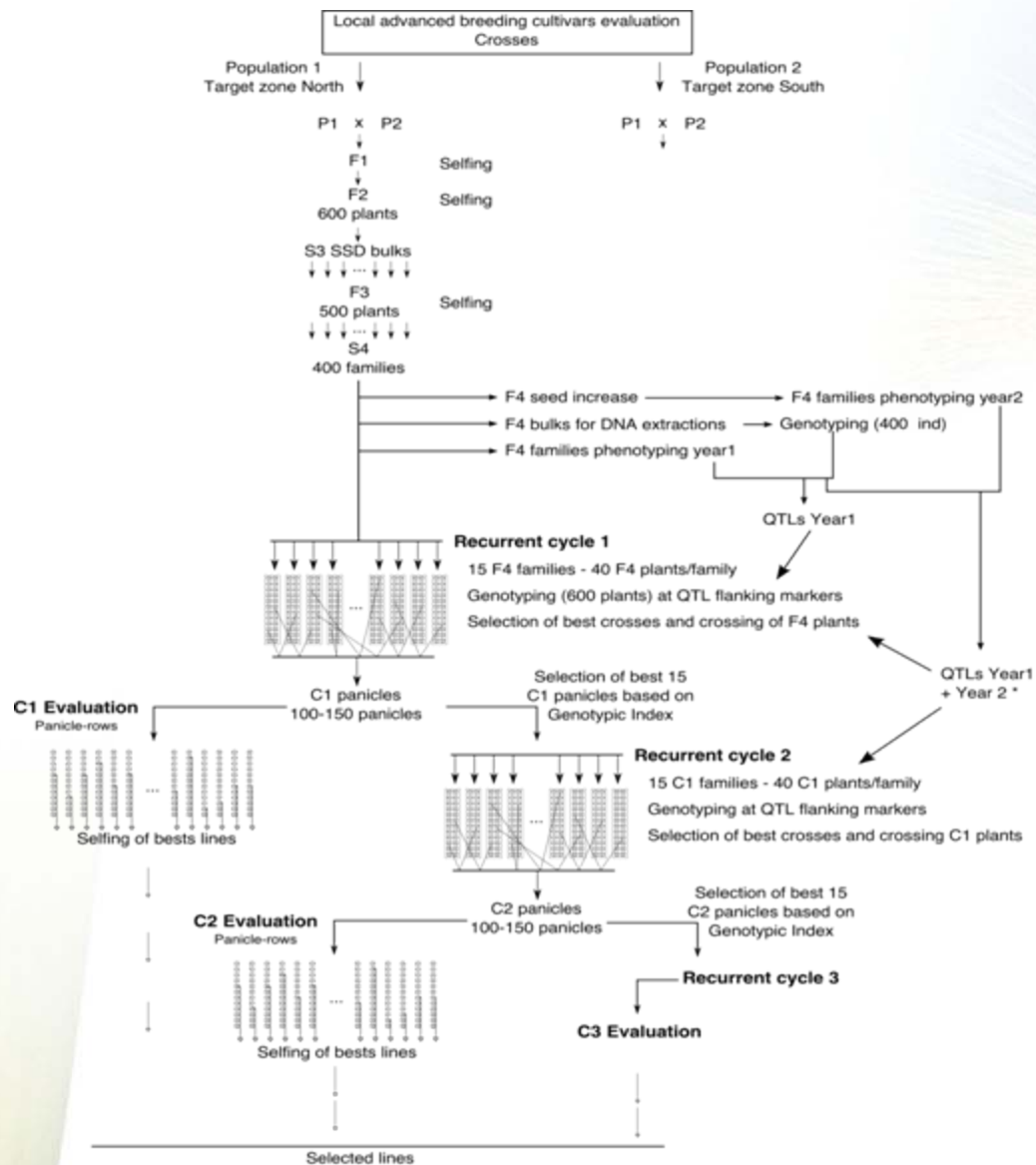
Project Objectives

- Develop and deploy a functional and sustainable molecular breeding platform
 - Providing access to molecular breeding services,
 - an information system and a
 - toolbox of analysis and decision support applications

Use case projects



- **Beans:** Improving tropical legume productivity for marginal environments in sub-Saharan Africa (Tropical Legumes I, Gates Foundation–GCP project, Mathew Blair, CIAT)
- **Chickpeas:** Improving tropical legume productivity for marginal environments in sub-Saharan Africa (Tropical Legumes I, Gates Foundation–GCP project, Pooran Gaur, ICRISAT)
- **Cowpeas:** Improving tropical legume productivity for marginal environments in sub-Saharan Africa (Tropical Legumes I, Gates Foundation–GCP project, Jeff Ehlers, University of California–Riverside)
- **Maize:** Drought-tolerant maize for Africa, (Gates Foundation project, Gary Atlin, CIMMYT)
- **Maize:** Drought-tolerant maize for Asia (GCP project, Pervez Zaidi, CIMMYT)
- **Rice:** Stress-tolerant rice for poor farmers in Africa and South Asia (Gates Foundation project, David Mackill, IRRI)
- **Rice:** Drought-tolerant rice for China (Gates Foundation project, Zhikang Li, CAAS–IRRI)
- **Sorghum:** Yield improvement of sorghum in Africa through marker-assisted recurrent selection (GCP project, Jean-François Rami, CIRAD)
- **Wheat:** Molecular marker technologies for faster wheat breeding in India (ACIAR project, Richard Trethowan, Sydney University)
- **Wheat:** Durable rust resistance in wheat (Gates Foundation project, Mike Pumphrey, USDA/Cornell Univeristy)



(*) : the QTL detection will be repeated for a second year.
 The consolidated QTL results including two years of phenotyping will be used either to refine the choice of target regions in C2 recurrent cycle or start another C1 recurrent cycle from F4 plants or a combination of F4 and C1 plants

Major outputs



- **A Molecular Breeding Portal**
 - A web-based interface, providing access to the platform
 - A help desk providing assistance to portal users
- **An Information System**
 - A modular information system
 - A middleware layer of software which integrates data
 - Accumulation of breeding data into Crop Databases
- **Services**
 - Breeding Services
 - Support Services



Project objectives

Objective 1.1: Establish and manage the Molecular Breeding Platform

Platform elements

- **1. Molecular Breeding Portal and Helpdesk**
- Platform management
 - Molecular breeding portal
 - Helpdesk



Objective 2.1: Make existing tools for data management and breeding logistics available to molecular breeding projects through the MBP

Objective 2.2: Develop a suite of analysis, prediction and simulation tools for MAB

Objective 2.3: Develop an information network, decision support tools and a workflow management system for molecular breeding

2. Information System

→ Logistics and Data management

- Pedigree Information Management
- Field Data Management
- Laboratory Information Management

→ Analysis and Decision Support

- Statistical and genetic analysis
- Cross prediction and selection
- Modeling and simulation tools
- Visualisation and decision support

→ Information Network and Workflow System

- Information network infrastructure
- Public crop information
- Configurable workflow system



Objective 3.1: Provide access to critical molecular breeding services

Objective 3.2: Provide assistance with a range of molecular breeding support services

3. Services

Breeding Services

- Genetic Resource Support Service
- Marker service laboratories
- Trait service laboratories

Support Services

- Business Plan development
- Information management & data curation
- Design and analysis
- Phenotyping sites & screening protocols
- Genotyping Support Service
- IP and policy

Project Implementation



- Build on infrastructure and expertise that already exists
- Boost individual initiatives and to link people with complementary expertise
- Work with partners with common interests in building and using a public Molecular Breeding Platform
- short-term: demonstrate that we can provide support to boost molecular breeding – starting with the selected use cases
- Facilitate automation and quality assurance
- Prioritize modules and services to meet use-case needs

Existing GCP Technology and Services



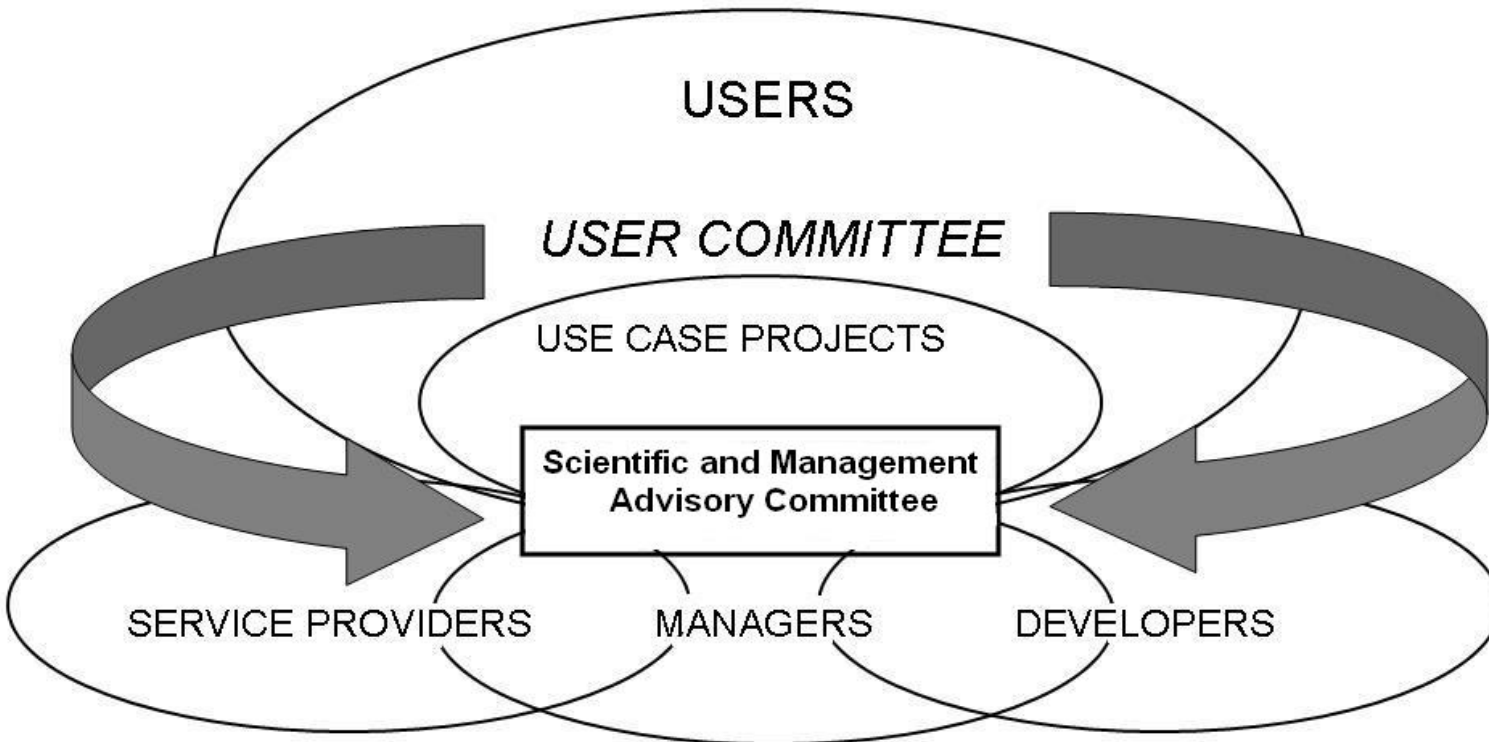
- Web technology from the GCP Informatics Platform to create the Portal,
- Domain model and semantic standards for data integration,
- Middleware, datasource and application programming interfaces, and web services technology,
- The Molecular Marker Information and Toolbox,
- GCP Phenotyping protocol and analysis projects,
- Genetic Resource Support Service (GRSS),
- Genotyping Support Service (GSS),
- SP4 Analysis Helpdesk, and
- IP and policy Helpdesk.



Quality Assurance

- Germplasm identity – fingerprinting and sample tracking
- Phenotyping –environment characterization, experimental design, protocols, data management and analysis
- Genotyping – reliable markers, standards and data management
- Analysis – efficient standardized analyses of genotypic and phenotypic data

Platform Management- Stakeholders

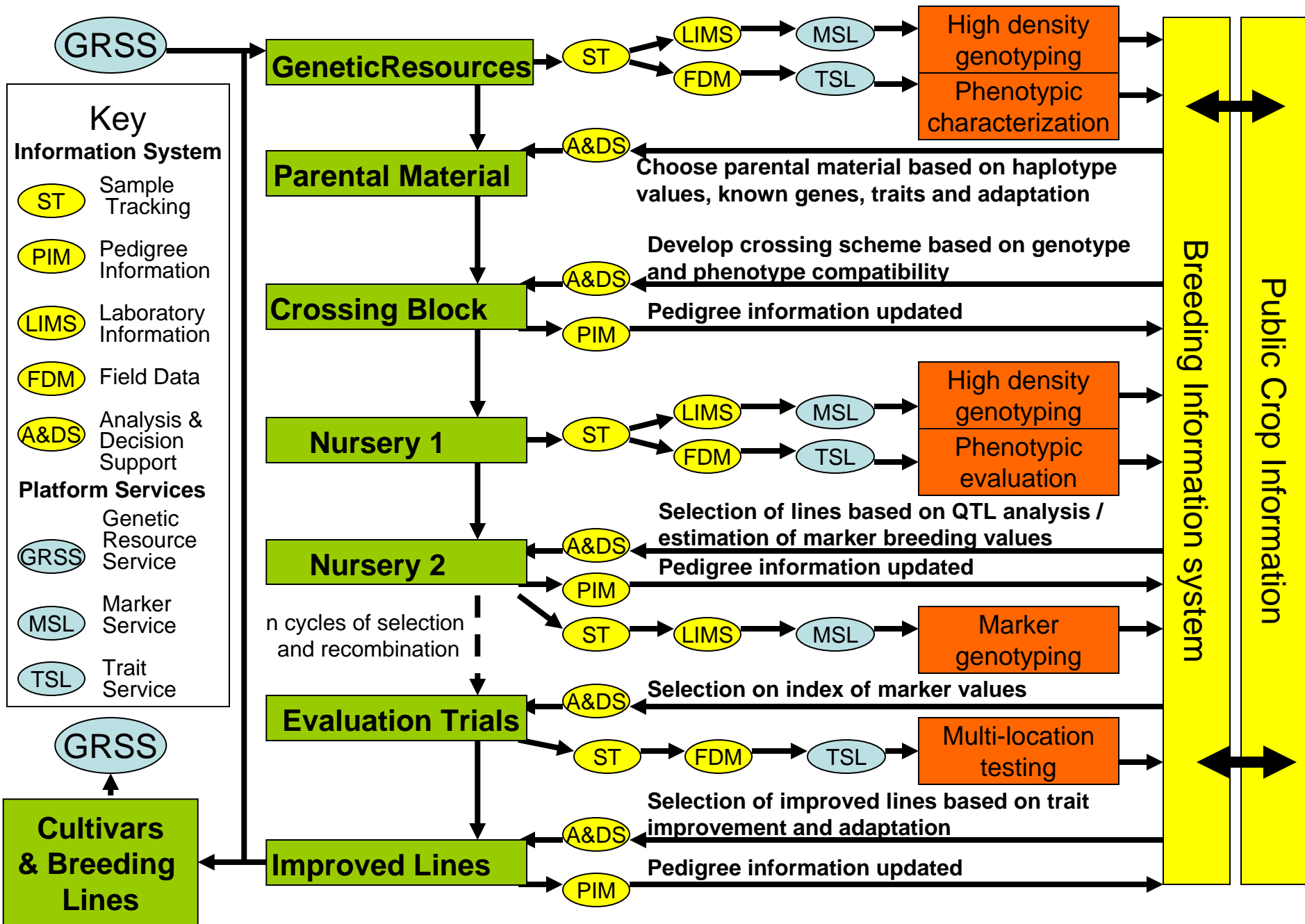


Platform Management - Structures



- The Scientific and Management Advisory Committee will:
 - oversee the evolution of the platform from a strategic perspective.
- Platform Manager to:
 - manage the Portal and Help Desk,
 - manage the Platform Services
 - coordinate regional activities through site visits
 - coordinate human resources development.
- A user committee will be constituted to
 - advise on the design of the platform elements,
 - to test different elements of the platform and
 - to provide feedback on the evolution of the portal.

Interaction of breeding workflow and platform elements



Objective 2.1 Make existing tools for data management and breeding logistics available to molecular breeding projects through the MBP.



- Identify, deploy and support tools facilitating management of germplasm lists, pedigrees, intellectual property and other passport data
- Identify, deploy and support tools for management of phenotypic characterisation and evaluation
- Identify, deploy and support tools for management of genotypic characterization

ICIS Applications: Standalone Tools



List Manager

List Selector: ARYT2002DS

Designation: 2000 F1 WET
 Group Name: ADDITIONAL
 Entry Code: ADDTNL_F1
 Source: AOYT2000DS
 GID: AOYT2000WS
 Tag: AOYT2001DS
 AOYT2001WS
 AOYT2002DS
 ARYT2000DS
 ARYT2000WS
 ARYT2001DS
 ARYT2001WS

Edit Window

List Characteristics

Name: ARYT2002DS Type: RYT Date: 2002-1-4 Levels for Group Name: 1

Title: REPLICATED YIELD TRIALS 2002 DRY SEASON

Entry Code Naming Convention: No. of entries: 450

Designation	Group Name	Entry Code	Source	GID	Entry ID	Tag
IR 71676-34-1-1	IR 65564-22-2-3PSBRC 2	B1001	B3814	437395	1	
IR 71676-90-2-2	IR 65564-22-2-3PSBRC 2	B1002	B3815	437440	2	
IR 71677-161-2-3	IR 65564-22-2-3MR 64	B1003	B3817	437687	3	
IR 71693-111-6-2-2	IR 65600-96-1-2-2MR 64	B1004	B3818	462599	4	
IR 71694-28-3-3	IR 65600-96-1-2-2PSBRC 2	B1005	B3819	438179	5	
IR 71700-247-1-1-2	IR 66159-164-5-3-5MR 64	B1006	B3821	462373	6	
IR 71703-587-1-3	IR 66736-118-1-2PSBRC 2	B1007	B3823	438944	7	
IR 71707-657-3-1	IR 66736-118-1-2PSBRC 2	B1008	B3826	439948	8	
IR 71780-1-1-3-2	IR 69161-99-2MR 59673-93-	B1009	B3751	462262	9	
IR 72161-311-5-4-2-2	IR 44962-161-2-4-4MR 6615	B1010	B9133	640311	10	
IR 72176-140-1-2-2-3	IR 68312-20-4-2-1-1MR 6615	B1011	B3839	640217	11	

SetGen

ICIS Workbook

Microsoft Excel - New DMS Input File Book2

CONDITION	DESCRIPTION	TRAIT	SCALE	METHOD	DATA TYPE	VALUE	LABEL
1	STUDY						
2	TITLE						
3	PMKEY						
4	OBJECTIVE						
5	START DATE						
6	END DATE						
7							
8							
9							
10							
11	FACTOR						
12	GID	GERMPLASM IDENTIFICATION	GID	Not Specified	N		GID
13	DESIGNATION	GERMPLASM IDENTIFICATION	Variety Name	Not Specified	C		DESIGNATION
14							
15							
16	CONSTANT						
17							
18							
19	VARIATE						
20	SOURCE	GERMPLASM IDENTIFICATION	Name of seed source	Seed Source	C		
21	DESTINATION						
22	SELTYPE						
23	NPLANTS						
24	STARTPLANT						
25	V5	WOOR	SES Score Vigor 5pt	At Seeding Stage	N		
26	FL	FLOWERING	No. of days	90% Flowering	N		
27	RL	LEAF BLAST	SES Score Blast Nursery (0-9)	Field Trial	N		
28	BD1	BACTERIAL BLIGHT	SES Score Lesion Area 5pt	Race 1 Field Trial	C		
29	BB2	BACTERIAL BLIGHT	SES Score Lesion Area 5pt	Race 2 Field Trial	C		
30	BPH1	BROWN PLANTHOPPER	SES Score Field 5pt	BPH Biotype 1	N		
31	BPH2	BROWN PLANTHOPPER	SES Score Field 5pt	BPH Biotype 2	N		
32	BPH3	BROWN PLANTHOPPER	SES Score Field 5pt	BPH Biotype 3	N		
33	OLH	GREEN LEAFHOPPER	SES Score Damage 6pt	Field Trial	N		
34	RTD	RICE TUNGRO VIRUS	SES Score Symptoms 5pt	Greenhouse Evaluation	N		
35	AMY	AMYLOSE CONTENT OF GRAIN	Percent	Standard laboratory pro	C		
36	GELTEMP	GELATINIZATION TEMPERATURE	Code H4L5	Cooking Test	C		
37	AROMA	AROMA	Description	Cooking Test at Maturit	C		
38	GRN	GRAIN ELONGATION			N		

Inventory Tracker

Inventory Management System - Query

List Selector: 2001WS_T, MR2001WS, 2002 DS_T, MR2002DS, 2002 WS_T, MR2002WS, PVS_T, PVS, PVL_T, PVL, VON_T, VON, IRON_T, RON, G01_T, G01, G02_T, G02

Inventory Filter: Location: [] Clear Locations

Inventory Units: []

Germplasm Scope: Exact GID, Management Neighborhood, Deducative Neighborhood

Tag	GID	Entry Code	Desig	Lot ID	Lot Location	Lot Units	Avail. Bal.	Actual Bal.	Remarks	
x	-32	320035	50001302	IR6C 141	15002	Green Box 4	gm (Middle)	11.3	0	From Entry Code No. 048
x	-32	320036	50001303	IR6C 141	15012	OCP Storage	gm (Middle)	27.7	0	
x	-32	299480	50001306	IR6C 144	15161	Green Box 4	gm (Middle)	123.8	0	From Entry Code No. 040
x	-32	321337	50001308	IR6C 157	15163	OCP Storage	gm (Middle)	89.7	0	
x	-32	9979	50001310	IR6C 181	15164	Green Box 4	gm (Middle)	109.4	0	From Entry Code No. 79
x	-32	9979	50001310	IR6C 181	15916	OCP Storage	gm (Middle)	105.9	0	
x	-32	321337	50001308	IR6C 157	15163	Green Box 4	gm (Middle)	404.5	0	From Entry Code No. 956
x	-32	321337	50001308	IR6C 157	15016	OCP Storage	gm (Middle)	806.7	0	
x	-32	9978	50001310	IR6C 181	15164	Green Box 4	gm (Middle)	308.1	0	From Entry Code No. 482
x	-32	9979	50001310	IR6C 181	15916	OCP Storage	gm (Middle)	340.4	0	



Objective 2.2 Develop a suite of analysis, prediction and simulation tools for MAB



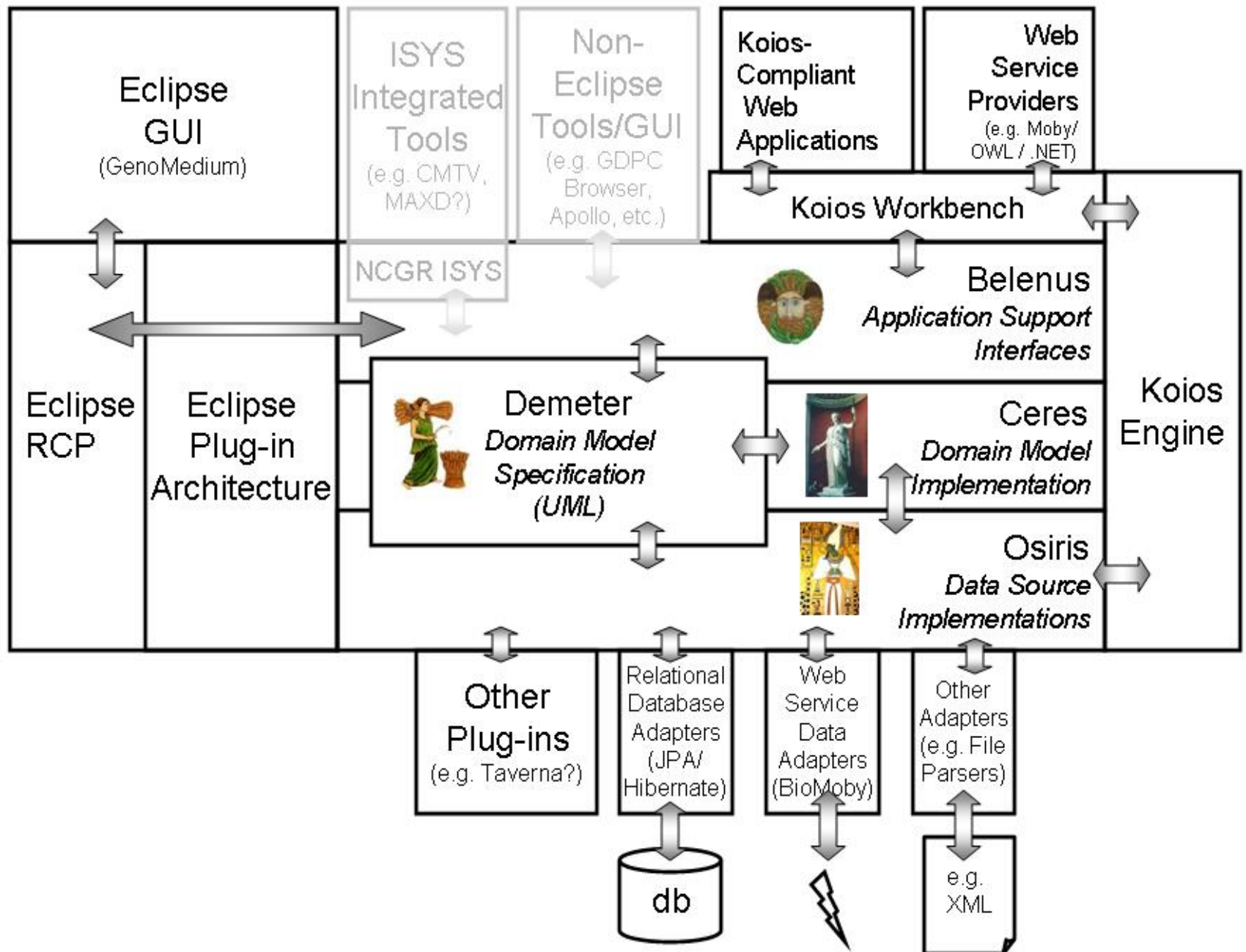
- Develop and deploy statistical and genetic analysis for molecular breeding
- Develop and deploy cross prediction and selection methodology for molecular breeding
- Develop and deploy simulation tools for complex G-E systems

Objective 2.3 Develop an information network, decision support applications and a workflow management system for molecular breeding



- Establish middleware infrastructure for networking databases and applications
- Develop and integrate visualisation and decision support applications
- Implement a configurable workflow system for molecular breeding

GCP Platform Architecture 2008 – Updated Details



Objective 3.1 Provide access to critical molecular breeding services.



- **Genetic Resource Support Service**
 - Make existing quality genetic resources accessible
 - Work with the GR community to identify useful genetic resources and make them available
 - Maintain and distribute genetic resources
- **Marker Services**
 - Provide information on markers and service providers
 - Facilitate international movement of DNA/samples
 - Negotiate service contracts/conditions with labs
- **Trait and Metabolite Services**
 - Facilitate access to trait evaluation services
 - Facilitate access to metabolite/quality screening

Objective 3.2 Provide assistance with a range of molecular breeding support services.

- Business Plan Development
- Information Management
- Data Curation
- Design and Analysis
- Phenotyping Sites and Screening Protocols
- Genotyping Support Services
- IP Helpdesk





Potential Risks

- The platform is technology push and not demand-driven
- The platform is not sustainable
- The platform is not adopted/used by breeders
- Access to the data generated through the platform is limited