

# *CURRENT STATUS*

Management of Genetic Marker and Genotyping Data

R. Valerio

---

---

# Overview

## DESKTOP APPLICATION

- New GEMS DLL functions
- New module in ICIS workbook for loading Genotyping data
- Script for Loading SSR data into ICIS from GCP template

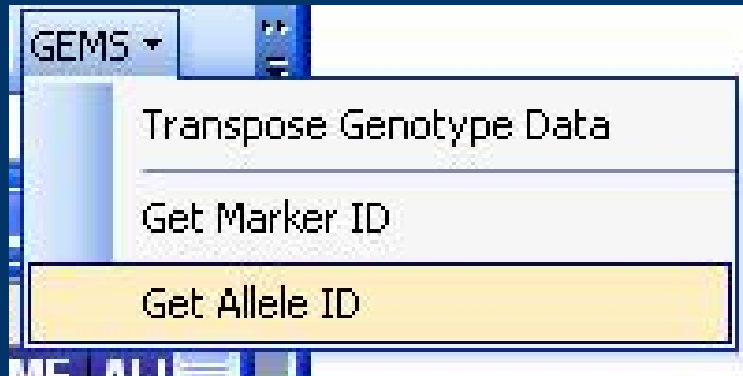
## WEB APPLICATION

- Script for Retrieving molecular data
  - Web interface for molecular data retrieval
- 
-

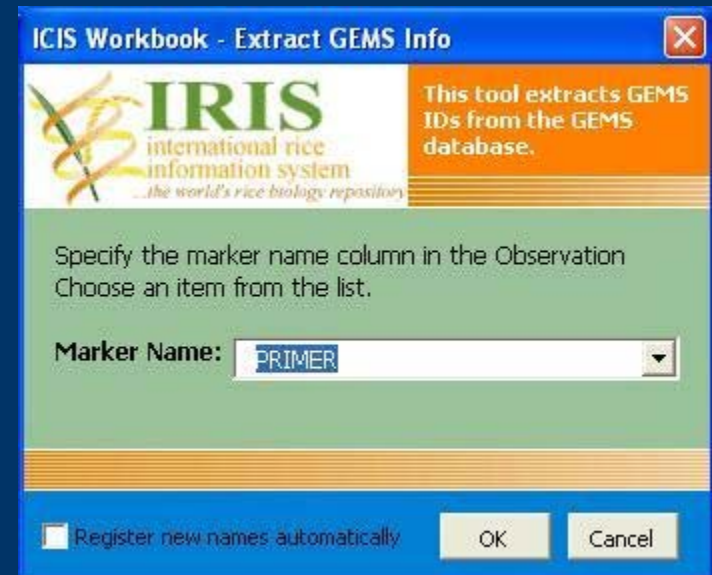
# New GEMS DLL Functions

- GEMS\_openDatabase(LPCSTR szIniFile);
  - GEMS\_autoCommit(BOOL autoCommit);
  - GEMS\_commitData(void);
  - GEMS\_closeDatabase(void);
  - GEMS\_getGemsMID(GEMS\_INFO \*data, int fOpt);
  - GEMS\_getGemsMVID(GEMS\_INFO \*data, int fOpt);
  - GEMS\_getMaxID(CHAR \*szTblName, CHAR \*szFieldName);
  - GEMS\_addGEMSName(GEMS\_NAME \*recName);
  - GEMS\_addMD(GEMS\_MARKER\_DETECTOR \*recTblMD );
  - GEMS\_getPdComp(PD\_COMP\_INFO \*pdCompInfo, LONG fopt);
  - GEMS\_addPdComp(GEMS\_PD\_COMP \*recTblPdComp );
  - GEMS\_getProtocolID(GEMS\_INFO \*recProcInfo , LONG fopt);
  - GEMS\_findPID(PROP\_INFO \*PropInfo) ;
  - GEMS\_findPropid(GEMS\_PROP \*recTblProp);
  - GEMS\_findMethod( GEMS\_METHOD \*recTblMethod);
  - GEMS\_addMethod(GEMS\_METHOD \*tblMethod) ;
  - GEMS\_findScale(GEMS\_SCALE \*tblScale);
  - GEMS\_addScale(GEMS\_SCALE \*tblScale) ;
  - GEMS\_findPdid(GEMS\_PD \*tblPd);
  - GEMS\_addLocus(GEMS\_LOCUS \*tblLocus);
  - GEMS\_getNameID(GEMS\_INFO \*data);
  - GEMS\_getAlleleID( GEMS\_INFO \*recNameInfo, LONG fopt) ;
- 
-

# ICIS Workbook Tool



- retrieve markerid and alleleid
- import genotyping data
- register new markers, alleles
- add new protocol



# Workbook Tools

- Description Sheet

Microsoft Excel - GEMS\_SSR\_Template

File Edit View Insert Format Tools Data Window Help Adobe PDF ICISworkbook

Study - Setup - Utilities - Add-ins - Help

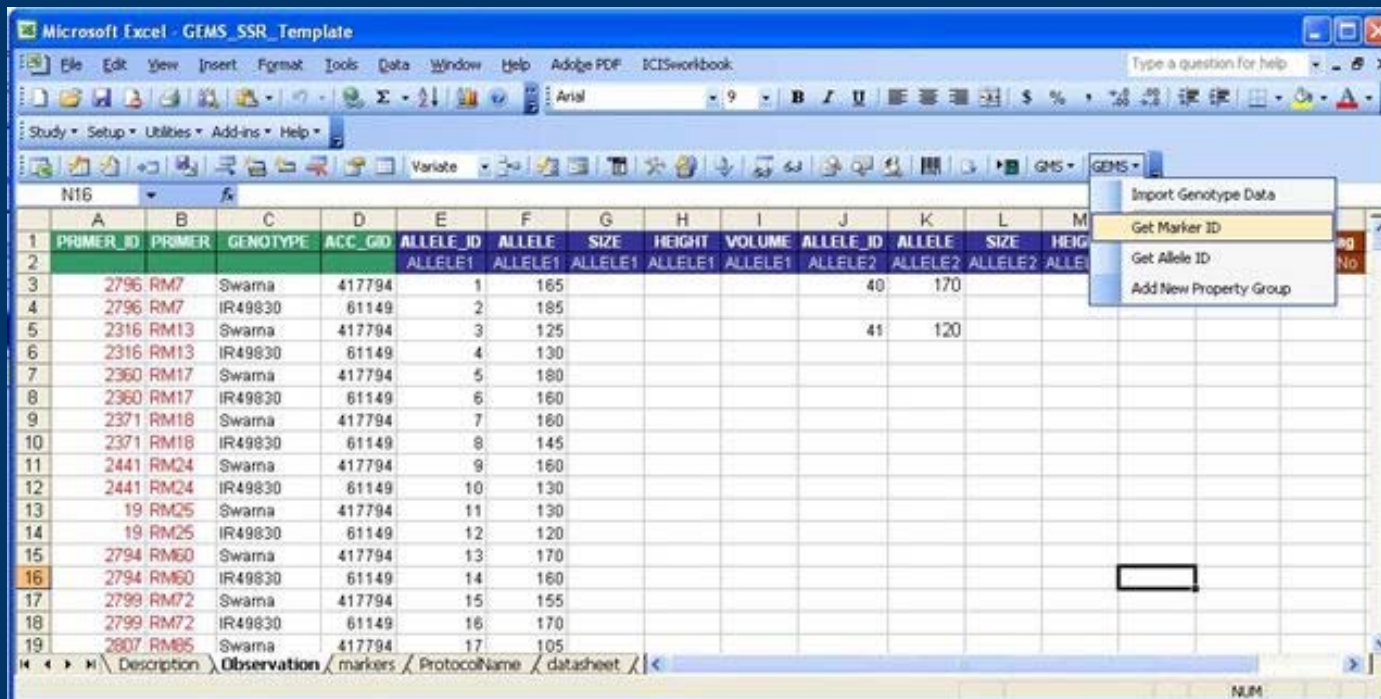
B26 Allele ID

1	STUDY	SSR_Study						
2	TITLE	GEMS SSR template						
3	PMKEY	0						
4	OBJECTIVE	Template for upload of SSR Data						
5	START DATE	20051001						
6	END DATE	20060301						
7								
8	CONDITION	DESCRIPTION	PROPERTY	SCALE	METHOD	DATA TYPE	VALUE	LABEL
9	INSTITUTE	Institute Name					Institute Name	STUDY
10	INVESTIGATOR1	Investigator Name	INVESTIGATOR	INVESTIGATOR NA	Not specified	C	Researcher Name	STUDY
11	SUPERVISOR1	Investigator Name	INVESTIGATOR	INVESTIGATOR NA	Not specified	C	Supervisor	STUDY
12	DNA_METHOD	Reference about DNA extraction	DNA EXTRACTION	Reference Name	Reference	C		STUDY
13								
14	FACTOR	DESCRIPTION	PROPERTY	SCALE	METHOD	DATA TYPE	VALUE	LABEL
15	PRIMER_ID	Primer Locus ID	LOCUS	ID	Not specified	N		PRIMER_ID
16	PRIMER	Primer Locus Name	LOCUS	NAME	Not specified	C		PRIMER_ID
17	GENOTYPE	Parental Line	CULTIVAR	VARIETY NAME	NOT SPECIFIED	C		GENOTYPE
18	DNA_GID		CULTIVAR	GID	NOT SPECIFIED	N		GENOTYPE
19	ACC_GID	Accession GID	CULTIVAR	GID	NOT SPECIFIED	N		GENOTYPE
20	ALLELE_NO		MOLECULAR_VARI	Number	Not specified	N		GENOTYPE
21								
22	CONSTANT	DESCRIPTION	PROPERTY	SCALE	METHOD	DATA TYPE	VALUE	
23								
24								
25	VARIATE	DESCRIPTION	PROPERTY	SCALE	METHOD	DATA TYPE	ALLELE_NO	
26	ALLELE_ID	Allele ID	MOLECULAR_VARI	ID	Not specified	N	1	
27	ALLELE	Allele Size	MOLECULAR_VARI	SIZE	Not specified	C	1	
28	SIZE						1	
29	HEIGHT						1	
30	VOLUME						1	
31	ALLELE_ID	Allele ID	MOLECULAR_VARI	ID	Not specified	N	2	
32	ALLELE	Allele Size	MOLECULAR_VARI	SIZE	Not specified	C	2	
33	SIZE						2	
34	HEIGHT						2	
35	VOLUME						2	

NUM

# Workbook Tools

- Observation Sheet



The screenshot displays a Microsoft Excel window titled "Microsoft Excel - GEMS\_SSR\_Template". The spreadsheet contains a table with columns for primer information, genotype, accession numbers, and allele data. A context menu is open over the "GEMS" tab, listing options: "Import Genotype Data", "Get Marker ID", "Get Allele ID", and "Add New Property Group".

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	PRIMER_ID	PRIMER	GENOTYPE	ACC_GID	ALLELE_ID	ALLELE	SIZE	HEIGHT	VOLUME	ALLELE_ID	ALLELE	SIZE	HEIG
2					ALLELE1	ALLELE1	ALLELE1	ALLELE1	ALLELE1	ALLELE2	ALLELE2	ALLELE2	ALLE
3	2796	RM7	Swarna	417794	1	165				40	170		
4	2796	RM7	IR49830	61149	2	185							
5	2316	RM13	Swarna	417794	3	125				41	120		
6	2316	RM13	IR49830	61149	4	130							
7	2360	RM17	Swarna	417794	5	180							
8	2360	RM17	IR49830	61149	6	160							
9	2371	RM18	Swarna	417794	7	160							
10	2371	RM18	IR49830	61149	8	145							
11	2441	RM24	Swarna	417794	9	160							
12	2441	RM24	IR49830	61149	10	130							
13	19	RM25	Swarna	417794	11	130							
14	19	RM25	IR49830	61149	12	120							
15	2794	RM60	Swarna	417794	13	170							
16	2794	RM60	IR49830	61149	14	160							
17	2799	RM72	Swarna	417794	15	155							
18	2799	RM72	IR49830	61149	16	170							
19	2807	RM85	Swarna	417794	17	105							

# Workbook Tools

- Marker Sheet

The screenshot shows a Microsoft Excel spreadsheet titled "GEMS\_SSR\_Template". The spreadsheet contains a table with the following columns: A (Marker), B (Protocol), C (Chrom), D (Position\_cM), E (Motif), F (Forward\_Primer), G (Reverse\_Primer), H (Min\_allele), I (Max\_allele), and J (Accession\_GenBar). The data rows list various markers (RM428 to RM231) with their corresponding protocols, chromosomal positions, motifs, primers, and allele information. The cell G17 is highlighted, containing the sequence "cacctcatcctgtaacgcc".

Marker	Protocol	Chrom	Position_cM	Motif	Forward_Primer	Reverse_Primer	Min_allele	Max_allele	Accession_GenBar
RM428	ProtocolName	1	19.3	(AG)15	aacagatggcctcctctcc	cgtctcctcactctctgtg	262	288	AQ271919
RM490	ProtocolName	1	51	(CT)13	atctgcacactgcaaacacc	agcaagcagtgctttcagag	93	107	AQ510019
RM583	ProtocolName	1	58.9	(CTT)20	agatccatcctctgtagag	ggcaactcctggtgtaacc	156	192	AF001278
RM24	ProtocolName	1	78.4	(GA)29	gaagtgatcctctgaacc	tacagtggcggcgaagtcg	152	198	AF343841
RM246	ProtocolName	1	115.2	(CT)20	gagctccatcagcattcag	ctgagtgctgctgcgact	97	118	AF344066
RM212	ProtocolName	1	148.7	(CT)24	caacttcagctactaccag	cacccttctctctctctatg	112	134	AF344033
RM485	ProtocolName	2	0	(TA)18	caacttcagctctctcc	caacttcctctctctctcc	291	361	AQ365678
RM110	ProtocolName	2	6.9	(GA)15	tccagccatccaccaccag	tccagccatccaccaccag	138	156	D22203
RM279	ProtocolName	2	17.3	(GA)16	ggggagagggatctct	ggctaggagttaccctcgcg	148	174	AF344105
RM174	ProtocolName	2	47.5	(AAG)(GA)10	agcagcgcacagcagtcggg	tccagctcagcgcacagcggg	207	222	D48756
RM324	ProtocolName	2	66	(CAT)21	ctgatccacacacttctgc	gattccacgtcaggatcttc	135	180	AF344149
RM561	ProtocolName	2	74.1	(GA)11	gagctgtttggactacggc	gagtagtcttctccaccacc	188	198	AQ913300
RM497	ProtocolName	2	150.8	(CAC)11	tctcttcactatgggtgg	gccagtgctaggagagttgg	133	143	AQ574392
RM250	ProtocolName	2	170.1	(CT)17	ggttcaaccacagctgaca	gatgaagccttccacagag	154	174	AF344070
RM482	ProtocolName	2	187.5	(AT)9	tctgaagcctcgtactcag	gtcaattgcagtgcccttcc	184	196	AQ365132
RM207	ProtocolName	2	191.2	(CT)25	ccattctgagagatctga	cacctcatcctgtaacgcc	71	139	AF344028
RM60	ProtocolName	3	0	(T)SAATCT(A	agtcacatgctcacttccg	atggctactgctgactacc	167	171	AF343863
RM231	ProtocolName	3	15.7	(CT)16	ccaaatttctctcactc	carctatactctcactc	169	191	AF344051

# Workbook Tools

- Protocol Sheet

Microsoft Excel - GEMS\_SSR\_Template

File Edit View Insert Format Tools Data Window Help Addge PDF ICDworkbook

Study Setup Utilities Add-ins Help

E19

PCR Recipe				
chemical	Volume	scale	Initial Concentration	scale
DNA	3	µL		ng/µl
H2O	3.75	µL		
Buffer	1	µL	10 X	
MgCl2		µL	10	mM
dNTPs	1	µL	1	mM
Taq	0.25	µL	4	U/µl
primer_Reverse	0.5		5	µM
primer_forward	0.5		5	µM
M13		mL		

PCR Condition			
Method	Property	Value	Scale
Initial denaturation	temperature	94	C
Initial denaturation	duration	5:00	minute
Initial denaturation	cycle	1	
Denaturation	temperature	94	C
Denaturation	duration	01:00	minute
Denaturation	cycle	35	
Annealing	temperature	55	
Annealing	duration	01:00	minute
Annealing	cycle	35	
Elongation	temperature	72	
Elongation	duration	01:30	minute
Elongation	cycle	35	
Final Elongation	temperature	72	C
Final Elongation	duration	7:00	minute
Final Elongation	cycle	1	
PCR cycle			

Electrophoresis			
Method	Property	Value	Scale
PCR reaction	Volume		2 µl
Buffer concentration	Volume		1 X
Method	Method Name	manual	

NUM



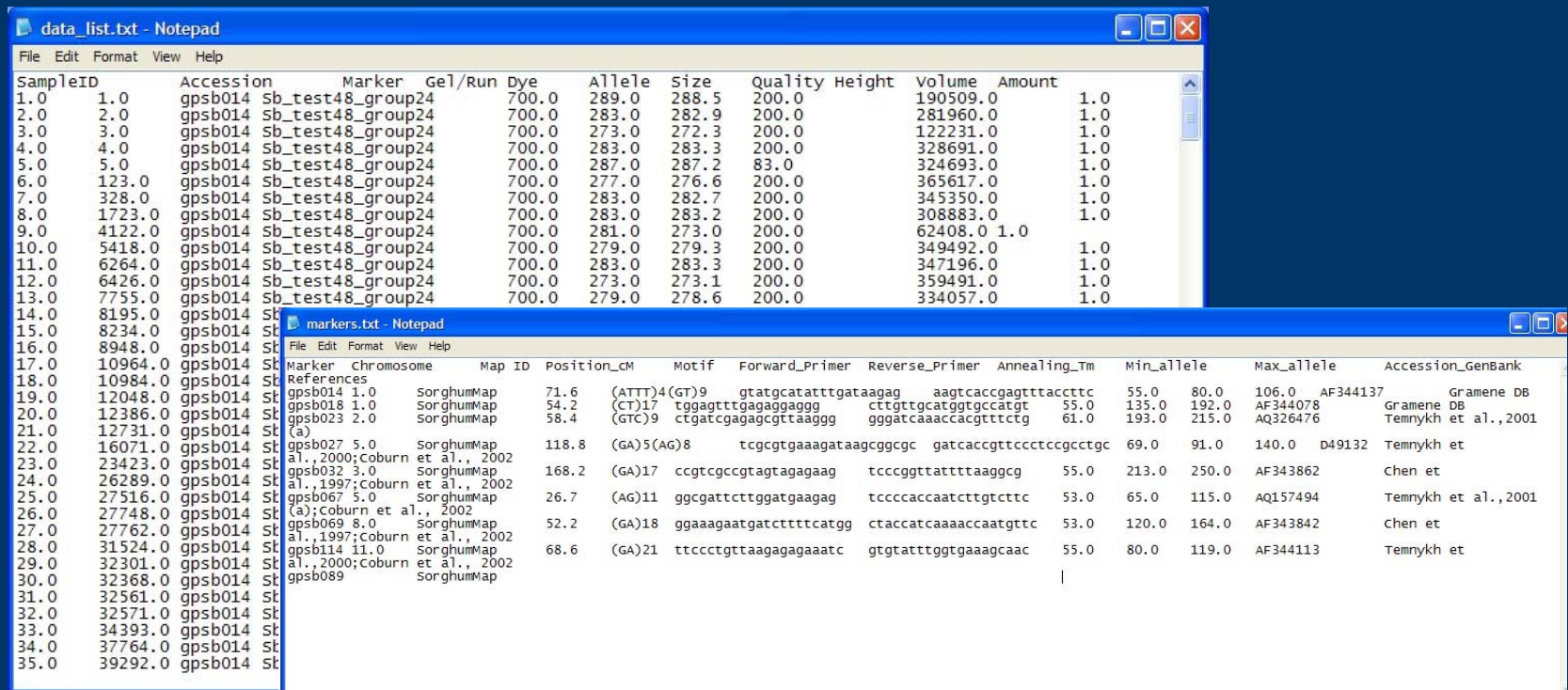
# *Script for Loading SSR data from GCP template*

- A response to a request for a script for loading GCP genotyping data from a GCP Template
- Written in postgresSQL taking advantage of its stored procedures feature
- Performs the main tasks of workbook for loading Genotyping Data
  - Retrieves marker IDs and Allele IDs from GEMS database
  - Adds new marker and allele names in the database



# Script for Loading SSR data from GCP template

## GCP template



The image shows two Notepad windows. The top window, titled 'data\_list.txt', contains a table of sample data. The bottom window, titled 'markers.txt', contains a table of marker information with columns for Marker, Chromosome, Map ID, Position\_CM, Motif, Forward\_Primer, Reverse\_Primer, Annealing\_Tm, Min\_allele, Max\_allele, and Accession\_GenBank.

SampleID	Accession	Marker	Gel/Run	Dye	Allele	Size	Quality	Height	Volume	Amount
1.0	1.0	gpsb014 Sb_test48_group24		700.0	289.0	288.5	200.0		190509.0	1.0
2.0	2.0	gpsb014 Sb_test48_group24		700.0	283.0	282.9	200.0		281960.0	1.0
3.0	3.0	gpsb014 Sb_test48_group24		700.0	273.0	272.3	200.0		122231.0	1.0
4.0	4.0	gpsb014 Sb_test48_group24		700.0	283.0	283.3	200.0		328691.0	1.0
5.0	5.0	gpsb014 Sb_test48_group24		700.0	287.0	287.2	83.0		324693.0	1.0
6.0	123.0	gpsb014 Sb_test48_group24		700.0	277.0	276.6	200.0		365617.0	1.0
7.0	328.0	gpsb014 Sb_test48_group24		700.0	283.0	282.7	200.0		345350.0	1.0
8.0	1723.0	gpsb014 Sb_test48_group24		700.0	283.0	283.2	200.0		308883.0	1.0
9.0	4122.0	gpsb014 Sb_test48_group24		700.0	281.0	273.0	200.0		62408.0	1.0
10.0	5418.0	gpsb014 Sb_test48_group24		700.0	279.0	279.3	200.0		349492.0	1.0
11.0	6264.0	gpsb014 Sb_test48_group24		700.0	283.0	283.3	200.0		347196.0	1.0
12.0	6426.0	gpsb014 Sb_test48_group24		700.0	273.0	273.1	200.0		359491.0	1.0
13.0	7755.0	gpsb014 Sb_test48_group24		700.0	279.0	278.6	200.0		334057.0	1.0
14.0	8195.0	gpsb014 Sb_test48_group24								
15.0	8234.0	gpsb014 Sb_test48_group24								
16.0	8948.0	gpsb014 Sb_test48_group24								
17.0	10964.0	gpsb014 Sb_test48_group24								
18.0	10984.0	gpsb014 Sb_test48_group24								
19.0	12048.0	gpsb014 Sb_test48_group24								
20.0	12386.0	gpsb014 Sb_test48_group24								
21.0	12731.0	gpsb014 Sb_test48_group24								
22.0	16071.0	gpsb014 Sb_test48_group24								
23.0	23423.0	gpsb014 Sb_test48_group24								
24.0	26289.0	gpsb014 Sb_test48_group24								
25.0	27516.0	gpsb014 Sb_test48_group24								
26.0	27748.0	gpsb014 Sb_test48_group24								
27.0	27762.0	gpsb014 Sb_test48_group24								
28.0	31524.0	gpsb014 Sb_test48_group24								
29.0	32301.0	gpsb014 Sb_test48_group24								
30.0	32368.0	gpsb014 Sb_test48_group24								
31.0	32561.0	gpsb014 Sb_test48_group24								
32.0	32571.0	gpsb014 Sb_test48_group24								
33.0	34393.0	gpsb014 Sb_test48_group24								
34.0	37764.0	gpsb014 Sb_test48_group24								
35.0	39292.0	gpsb014 Sb_test48_group24								

Marker	Chromosome	Map ID	Position_CM	Motif	Forward_Primer	Reverse_Primer	Annealing_Tm	Min_allele	Max_allele	Accession_GenBank
10964.0	SorghumMap		71.6	(ATT)4(GT)9	gtatgcataattgataagag	aagtcaccgagtttaccttc	55.0	80.0	106.0	AF344137 Gramene DB
10984.0	SorghumMap		54.2	(CT)17	tggagtttgagaggagg	cttgttgcattggtgccatgt	55.0	135.0	192.0	AF344078 Gramene DB
12048.0	SorghumMap		58.4	(GTC)9	ctgatcgagagcgtaaggg	gggatcaaacacgtttctg	61.0	193.0	215.0	AQ326476 Temnykh et al.,2001
12386.0	SorghumMap			(a)						
12731.0	SorghumMap		118.8	(GA)5(AG)8	tcgcgtgaaagataagcggcgc	gatcacctgtccctccgcctgc	69.0	91.0	140.0	D49132 Temnykh et
16071.0	SorghumMap			(a);Coburn et al., 2002						
23423.0	SorghumMap		168.2	(GA)17	ccgtcgccgtagtagagaag	tcccggttattttaagcgc	55.0	213.0	250.0	AF343862 Chen et
26289.0	SorghumMap			(a);Coburn et al., 2002						
27516.0	SorghumMap		26.7	(AG)11	ggcgattccttgatgaagag	tcccaccaaactctgtcttc	53.0	65.0	115.0	AQ157494 Temnykh et al.,2001
27748.0	SorghumMap			(a);Coburn et al., 2002						
27762.0	SorghumMap		52.2	(GA)18	ggaaagaatgatcttttcattg	ctaccatcaaaaaccaatgttc	53.0	120.0	164.0	AF343842 Chen et
31524.0	SorghumMap		68.6	(GA)21	ttccctgttaagagagaatc	gtgtatttggtagaagcaac	55.0	80.0	119.0	AF344113 Temnykh et
32301.0	SorghumMap			(a);Coburn et al., 2002						
32368.0	SorghumMap			(a);Coburn et al., 2002						
32561.0	SorghumMap			(a);Coburn et al., 2002						
32571.0	SorghumMap			(a);Coburn et al., 2002						
34393.0	SorghumMap			(a);Coburn et al., 2002						
37764.0	SorghumMap			(a);Coburn et al., 2002						
39292.0	SorghumMap			(a);Coburn et al., 2002						

# *Script for Loading SSR data from GCP template*

- Setup the variables
  - Create table in the database with the same structure as the marker\_list and data\_list tab-delimited text file
  - Check marker and allele names if existing in the database if not, adds new names to the database
  - Setup the study, marker and allele factors
  - Load study into DMS
- 
-

# *Script for Loading SSR data (GCP template)*

Advantages:

- faster loading time
- not limited by excel's row and column limitations

Disadvantage:

- database specific



# GEMS interface

curation of Marker  
and Allele Names

The screenshot displays the Gene Management System (GEMS) interface. The window title is "Gene Management System". At the top, there is a search bar and radio buttons for "Marker Name", "Allele Name", "Primer", and "Marker Type".

The main interface is divided into several sections:

- MARKER DETECTOR**: A list of marker detectors on the left. "OSR 3" is selected and highlighted.
- MARKER DETAIL FOR MARKER DETECTOR ID: 16**: A form showing details for the selected marker detector.
  - Marker Detector: OSR 3
  - Marker Type: SSR
  - Forward Primer: tgatacgtggtagctgacgc
  - Reverse Primer: taatcgcttccctaccctcg
  - User's ID: 0
  - Location: 0
  - Date: 20060926
  - Reference: 0
- MARKER POSITION**: A section for marker location.
  - Marker Name: OSR 3
  - Chromosome: 1
  - Position: 154.8
- MOLECULAR VARIANTS OF OSR 3**: A table with columns: mvid, markerid, molecular variant name, status, mvtype, weight. The table is currently empty.
- MARKER DETECTOR PROTOCOL INFORMATION**: A section with tabs for Protocol, PCR Recipe, PCR Condition, Gel Recipe, and electrophoresis. The "PCR Condition" tab is active, showing a table of protocol parameters for "1".

Method	Property	Value	Scale
Initial denaturation	temperature	94	C
Initial denaturation	duration	4:00	minute
Initial denaturation	cycle	1	-
Denaturation	temperature	94	C
Denaturation	duration	1:00	minute

# WEB APPLICATION

Objective : Create a web interface for retrieving genotype and phenotype data

- Server side programming/scripting in creating database warehouse
  - functions or procedures can be executed just like any query
  - less client/server transaction



# *WEB APPLICATION*

Some useful scripts in MySQL and PostgreSQL

- Get\_dataset (phenotype study)
- Get\_dataset (genotype study)
- Create\_array\_dataset (genotype table)
- Join genotype and phenotype studies by GID



# *WEB APPLICATION*

- AJAX application





# *Documentations*

- Technical Documentation  
[http://cropwiki.irri.org/icis/index.php/TDM\\_Gen\\_e\\_Management\\_System](http://cropwiki.irri.org/icis/index.php/TDM_Gen_e_Management_System)
  - [http://cropwiki.irri.org/icis/index.php>Loading\\_SR\\_Data\\_into\\_ICIS](http://cropwiki.irri.org/icis/index.php>Loading_SR_Data_into_ICIS)  
[Http://](http://)
  - [http://cropwiki.irri.org/icis/index.php>Loading\\_of\\_DArT\\_Data\\_into\\_ICIS](http://cropwiki.irri.org/icis/index.php>Loading_of_DArT_Data_into_ICIS)
- 
-