

Integrating GEMS (prototype interface) and DMS (workbook) for Managing Molecular Data

Sample Molecular Data

Study/Project
Information

	A	B	C	D	E	F	G	H
1	SSR genotyping details page							
2	GENERAL INFORMATION							
3	Name of researcher		Bert Collard					
4			Rhoda Manzano					
5	Supervisor(s)		Drs. Ken McNally, David Mackill					
6								
7	Aims of project (brief)		Parental screening for genotyping of 3 populations for drought QTLs					
8			SSRs were selected based on linkage to ESTs involved in response to drought					
9								
10	Dates (start/end of screening)		October 2005 - March 2006					
11								
12	GERMPLASM							
13	Genotype name		Code	Source				
14	IR64			Varoy				
15	Azucena							
16	Teqing							
17	Lemont							
18	Vandana							
19	Moroberekan							

Sample Molecular Data

- Genotype Data

Germplasm names

SSR marker database template Bert example

	A	B	C	D	E	F	G	H	I	J
1	Parental genotypes									%
2	Primer	Chr	Position	IR64	Azucena	Tequiq	Lemont	Vandana	Moroberekan	
3	RM1880	CH01	13.2	105	140, 145	150, 155	147, 152	md	md	6
4	RM5552	CH01	20.2	110	105, 160	110	105, 110	110	105, 160	6
5	RM3360	CH01	27.3	175	190	175	190	175	185	6
6	RM1220	CH01	37.4	198	175	196	178	198	175	6
7	RM8083	CH01	46.3	100, 120	135	100, 120	100, 160	md	135	6
8	RM580	CH01	50.8	195	225	195, 215	192, 210	195	208	6
9	RM449	CH01	73.1	130	125	138	150	128	122	6
10	RM5	CH01	98.5	135	140	138	135	133	146	6
11	RM3143	CH01	113	90	100	90	88	90	100	6
12	RM1183	CH01	127.3	135	150	135	140	150	135	6
13	RM486	CH01	142.5	105	100, 105	107	104	100, 105	100, 102	6
14	RM6407	CH01	176.3	150	145	158	145	160	168	6
15	RM6840	CH01	181.8	135	150	138	138	138	150	6
16	RM5897	CH02	32.8	155	154	170	166	170	172	6
17	RM322	CH02	36.3	145	150	135	133	135	150	6
18	RM145	CH02	36.8	180	195	182	176	184	195	6
19	RM475	CH02	92.5	205	225	222	205	205	225	6
20	RM1342	CH02	118	135	132	138	110	110	130	6
21	RM3421	CH02	126.4	150	145	152	155	140	148	6
22	RM425	CH02	138	135	145	138	137	137	145	6
23	RM207	CH02	156.3	125	115	130	120	103, 150	115	6
24	RM132	CH03	3.9	155	120	135	120	135	125	6
25	RM231	CH03	11.5	170	155	168	155	172	185	6
26	RM6849	CH03	15.2	125	112	105	112	112	125	6
27	RM7197	CH03	44.4	145	160	145	160	180	165	6
28	RM6832	CH03	88.95	170	185	195	188	172	md	6
29	RM1230	CH03	146.4	160	152	158	148, 154	156	md	6
30	RM514	CH03	158.2	300	270	300	260	268	269	6
31	RM3820	CH04	88	140	150	140	130	132	150	6
32	RM5709	CH04	109.9	180	205	192	md	md	200	6

Genotyping information \ SSR marker allele data

markers

Allele sizes

Sample Molecular Data

Genotyping
Information

SSR marker database template Bert example

	A	B	C	D	E	F
21	DNA extraction method	Zheng et al. 1995 ("IRRI method")	OR			
22		Wang et al. 1993 ("NaOH-Tris" method)				
23						
24	PCR					
25	PCR recipe	(Table - see below)				
26						
27	GEL ELECTROPHORESIS					
28	Amount of total PCR volume loaded on gel		6 - 8 ul			
29	Acrylamide Gel			6%		
30	Time and voltage for electrophoresis		1 hr 25 min			
31						
32						
33	PCR recipe					
34	PCR		Source/company			
35	H2O		-			
36	Buffer		Homemade			
37	MgCl2		-			
38	dNTPs					
39	Taq		Homemade			
40	Total volume =	10 ul				
41						
42						
43	Acrylamide Gel Recipe					
44	Ingredient		Vol			
45	Water		39.45			
46	TBE		2.5			
47	Acrylamide		7.5			
48	APS		500 ul			

Genotyping information | SSR marker allele data

Sample Molecular Data

Marker
Information

Edwards_50SSR_Panels_with_comments									
	A	E	F	G	H	I	J	K	W
1	Marker	Chromosome	Position cM	TIGRv3	Forward Primer (Labeled)	Reverse Primer	Annealing tm	PCR Cycles	References
2	RM495	1	2.8	213775	aatccaagg	caacgatga	55	30	Temnykh et al.,2001(a)
3	RM1	1	29.7	4633595	gcgaaaaca	gcgttggtg	55	30	Panaud et al.,1996
4	RM283	1	31.4	4883717	gtctacatga	cgccatgac	61	30	Temnykh et al.,2000;Coburn et al., 2002
5	RM259	1	54.2	7443424	tggagtttgac	cttgttgcata	55	30	Chen et al.,1997;Coburn et al., 2002
6	RM312	1	71.6	14890771	gtatgcatatt	aagtcaccg	55	30	Temnykh et al.,2000;Coburn et al., 2002
7	RM5	1	94.9	23952076	tgcaacttctc	gcacccgat	57	30	Panaud et al.,1996;Coburn et al., 2002
8	RM237	1	115.2	26795365	caaatcccg	tggaagagc	55	30	Chen et al.,1997
9	RM431	1	178.3	38874702	tcctgcgaac	agagcaaaa	55	30	Temnykh et al.,2001(a);Coburn et al., 200
10	RM154	2	4.8	1083895	accctctccg	ctcctctcc	61	30	Temnykh et al.,2000;Akagi et al.,1996
11	RM452	2	58.4	9564377	ctgatcgaga	gggatcaaa	61	30	Temnykh et al.,2001(a)
12	RM489	3	29.2	4316348	acttgagacc	tcacccatg	55	30	Temnykh et al.,2001(a);Coburn et al., 200
13	OSR13	3	53.1	7109988	catttgtcgt	agccacagc	53	40	Akagi,1996
14	RM338	3	108.4	13210600	cacaggagc	ggcaaaccc	55	40	Temnykh et al.,2000;Coburn et al., 2002
15	RM55	3	168.2	29002484	ccgtcgccg	tcccggttat	55	30	Chen et al.,1997;Coburn et al., 2002
16	RM514	3	216.4	35229618	agattgatctc	cacgagcat	55	30	Temnykh et al.,2001(a);Coburn et al., 200
17	RM307	4	0	12979675	gtactaccga	ctgctatgca	55	30	Temnykh et al.,2000;Coburn et al., 2002
18	RM124	4	150.1	34485157	atcgctgccc	catggatca	67	30	Temnykh et al.,2000;Coburn et al., 2002
19	RM507	5	0	71397	cttaagctcc	ctcaccctc	55	30	Temnykh et al.,2001(a)
20	RM413	5	26.7	2181391	ggcgattctt	tcccaccac	53	30	Temnykh et al.,2001(a);Coburn et al., 200
21	RM161	5	96.9	20714463	tgcagatgact	gtgtcatca	61	30	Temnykh et al.,2000;Coburn et al., 2002
22	RM178	5	118.8	24923084	tcgctgaaa	gatcaccgt	69	30	Temnykh et al.,2000;Coburn et al., 2002
23	RM334	5	141.8	28285978	gttcagtttc	gactttgatc	55	30	Temnykh et al.,2000;Coburn et al., 2002

How do we store it in ICIS?

What are to be stored?

- **Project/Study Information**
- **Germplasm (GIDs) used in the study/project**
- **Marker information (marker name, polymorphism detector used)**
- **Information on related loci, markers and mapping positions**
- **Genotype data (molecular variants)**

Our Answer:

DMS working together with GEMS

Storing Genotype Data in DMS

Description Sheet

SSR_MARKER_PBGB-gems									
	A	B	C	D	E	F	G		
1	STUDY	SSR_MARKER_PBGB							
2	TITLE	Parental screening for genotyping of 3 populations for drought QTLs							
3	PMKEY	0							
4	OBJECTIVE	Parental screening for genotyping of 3 populations for drought QTLs SSRs were selected based on linkage to ESTs involved in response t							
5	START DATE	20051001							
6	END DATE	20060301							
7									
8	CONDITION	DESCRIPTION	PROPERTY	SCALE	METHOD	DATA TYPE	VALUE		
9	INVESTIGATOR1	Investigator Name	INVESTIGATOR	INVESTIGATOR NA	Not specified	C	Bert Collard	STUD	
10	INVESTIGATOR2	Investigator Name	INVESTIGATOR	INVESTIGATOR NA	Not specified	C	Rhoda Manzano	STUD	
11	SUPERVISOR1	Investigator Name	INVESTIGATOR	INVESTIGATOR NA	Not specified	C	Dr. Ken McNally	STUD	
12	SUPERVISOR2	Investigator Name	INVESTIGATOR	INVESTIGATOR NA	Not specified	C	Dr. David Mackill	STUD	
13	DNA_METHOD	Reference about DNA extraction	DNA EXTRACTION	Reference Name	Reference	C	Zheng et al. 1995 (*)	STUD	
14									
15	FACTOR	DESCRIPTION	PROPERTY	SCALE	METHOD	DATA TYPE	VALUE		
16	MARKERID	MARKER ID	MARKER	ID	Not specified	N		MARK	
17	MID								
18	MARKER	MARKER NAME	MARKER	NAME	Not specified	C		MARK	
19	GENOTYPE	Parental Line	CULTIVAR	VARIETY NAME	NOT SPECIFIED	C		GENO	
20	GID		CULTIVAR	GID	NOT SPECIFIED	N		GENO	
21									
22	CONSTANT	DESCRIPTION	PROPERTY	SCALE	METHOD	DATA TYPE	VALUE		
23									
24									
25	VARIATE	DESCRIPTION	PROPERTY	SCALE	METHOD	DATA TYPE	Allele_No		
26	1st ALLELE_ID	Allele ID	MOLECULAR_VARI	ID	Not specified	N	1		
27	MVID1								
28	1st ALLELE	Allele Size	MOLECULAR_VARI	SIZE	Not specified	C	1		
29	2nd ALLELE_ID	Allele ID	MOLECULAR_VARI	ID	Not specified	N	2		
30	MVID2								
31	2nd ALLELE	Allele Size	MOLECULAR_VARI	SIZE	Not specified	C	2		

Storing Genotype Data in DMS

Observation Sheet

SSR_MARKER_PBGB-gems

	A	C	D	E	F	H	I	K	L	M	N
	MARKERID	MARKER	GENOTYPE	GID	1st ALLELE_ID	1st ALLELE	2nd ALLELE_ID	2nd ALLELE		Header	
2		RM1880	IR64			105					
3		RM1880	Azucena			140		145			
4		RM1880	Teginq			150		155			
5		RM1880	Lemont			147		152			
6		RM1880	Vandana			md					
7		RM1880	Moroberekan			md					
8		RM5552	IR64			110					
9		RM5552	Azucena			105		160			
10		RM5552	Teginq			110					
11		RM5552	Lemont			105		110			
12		RM5552	Vandana			110					
13		RM5552	Moroberekan			105		160			
14		RM3360	IR64			175					
15		RM3360	Azucena			190					
16		RM3360	Teginq			175					
17		RM3360	Lemont			190					
18		RM3360	Vandana			175					
19		RM3360	Moroberekan			185					
20		RM1220	IR64			198					
21		RM1220	Azucena			175					
22		RM1220	Teginq			196					
23		RM1220	Lemont			178					
24		RM1220	Vandana			198					
25		RM1220	Moroberekan			175					
26											
27											
28											
29											
30											
31											
32											

Description Observation Gt

Storing Genotype Data in DMS

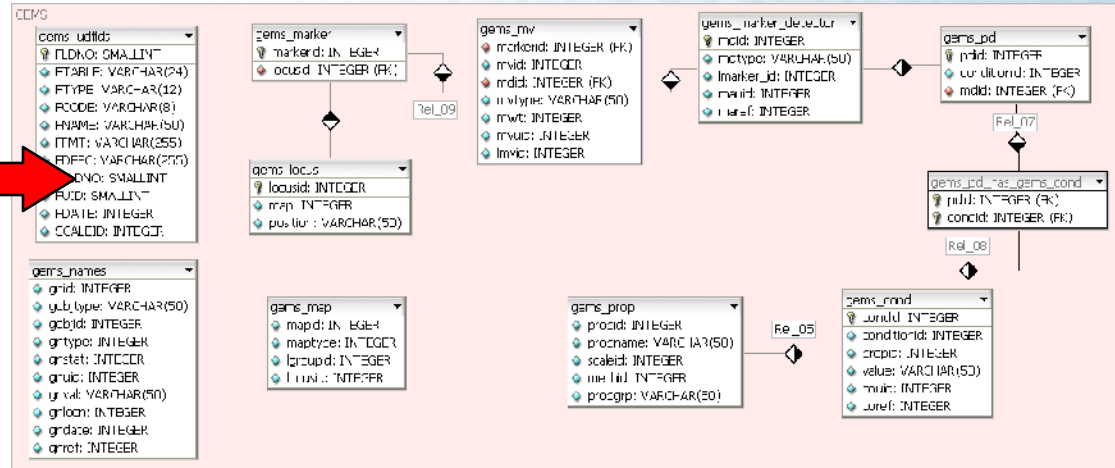
Observation Sheet

• Links to GEMS database

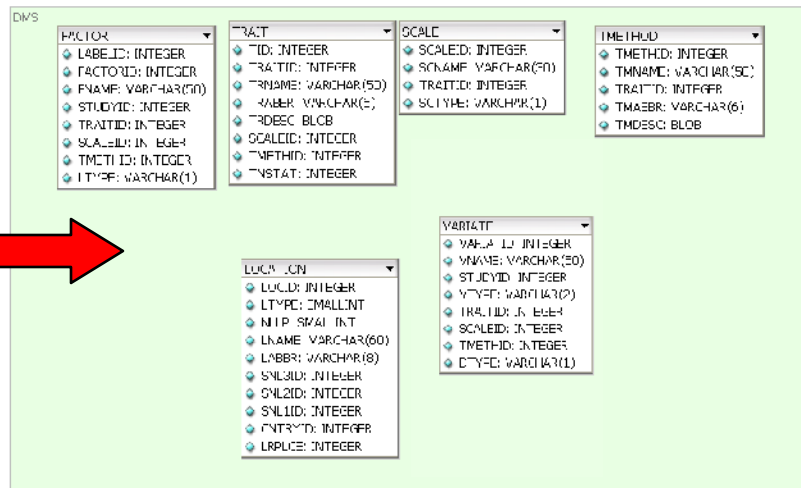
	A	C	D	E	H	I	K	L	M
2	MARKERID	MARKER	GENOTYPE	GID	1st ALLELE_ID	1st ALLELE	2nd ALLELE_ID	2nd ALLELE	Header
3	1	RM1880	IR64		1	105			
4	1	RM1880	Azucena		2	140	3	145	
5	1							155	
6	1							152	
7	1								
8	1								
9	2	RM5552	IR64		9	110			
10	2	RM5552	Azucena		10	105	11	160	
11	2	RM5552	Teging		9	110			
12	2	RM5552	Lemont		10	105	9	110	
13	2	RM5552	Vandana		9	110			
14	2	RM5552	Moroberekan		10	105	11	160	
15	3	RM3360	IR64		12	175			
16	3	RM3360	Azucena		13	190			
17	3	RM3360	Teging		12	175			
18	3	RM3360	Lemont		13	190			
19	3	RM3360	Vandana		12	175			
20	3	RM3360	Moroberekan		14	185			
21	4	RM1220	IR64		15	198			
22	4	RM1220	Azucena		16	175			
23	4	RM1220	Teging		17	196			
24	4	RM1220	Lemont		18	178			
25	4	RM1220	Vandana		15	198			
26	4	RM1220	Moroberekan		16	175			
27									
28									
29									
30									
31									
32									

GEMS and DMS Schema

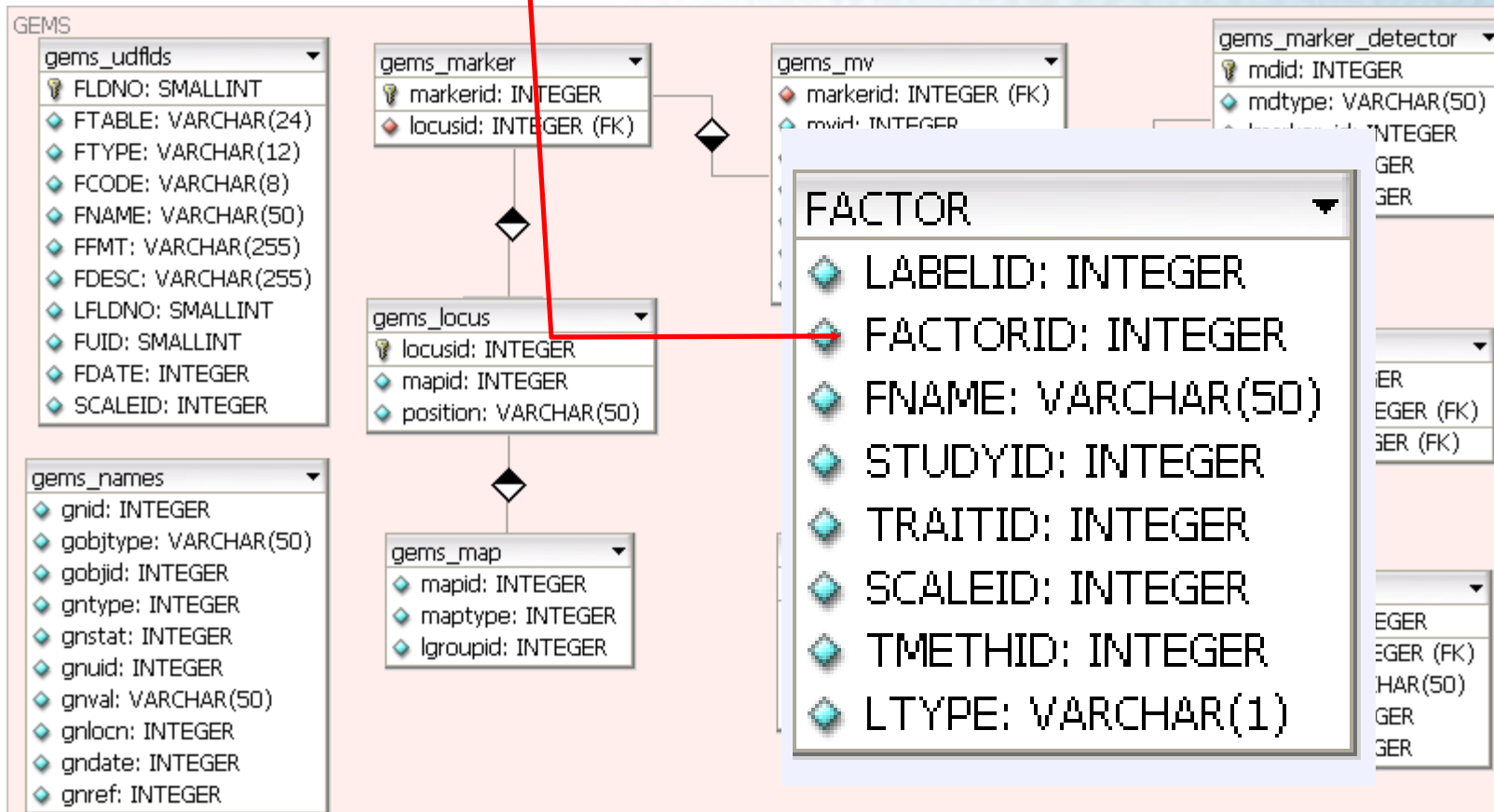
• **GEMS**



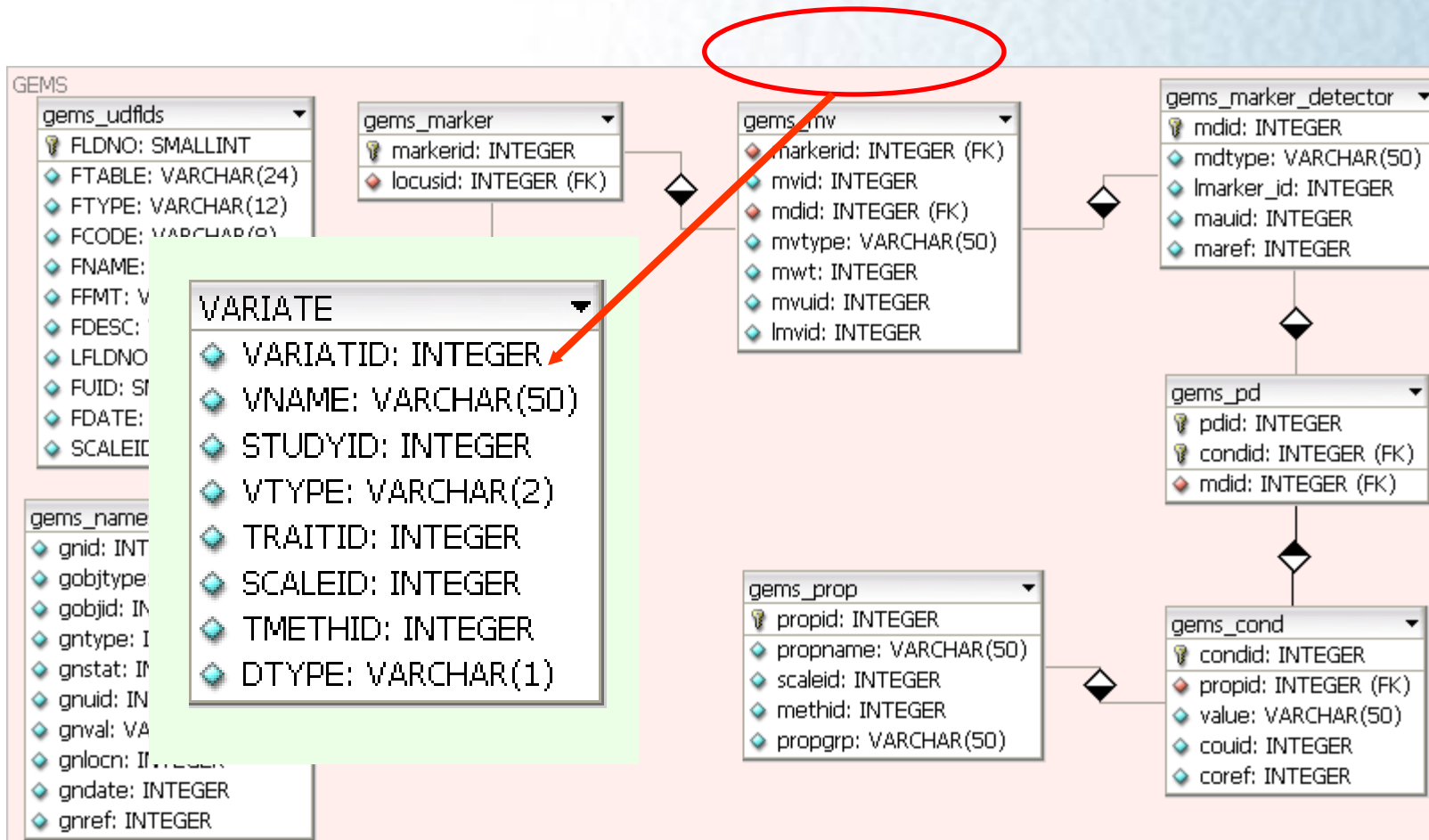
• **DMS**



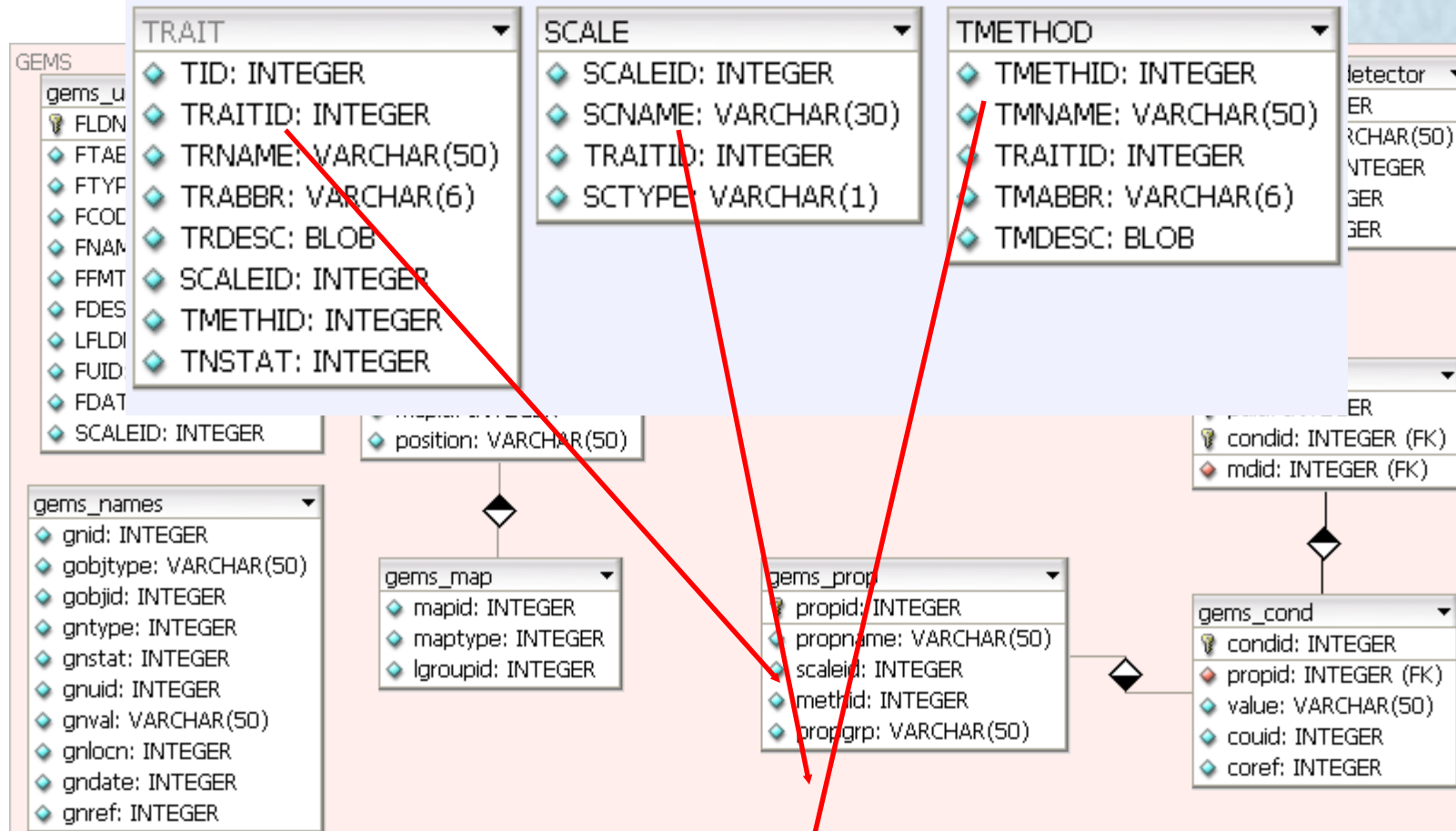
Links of GEMS and DMS Schema



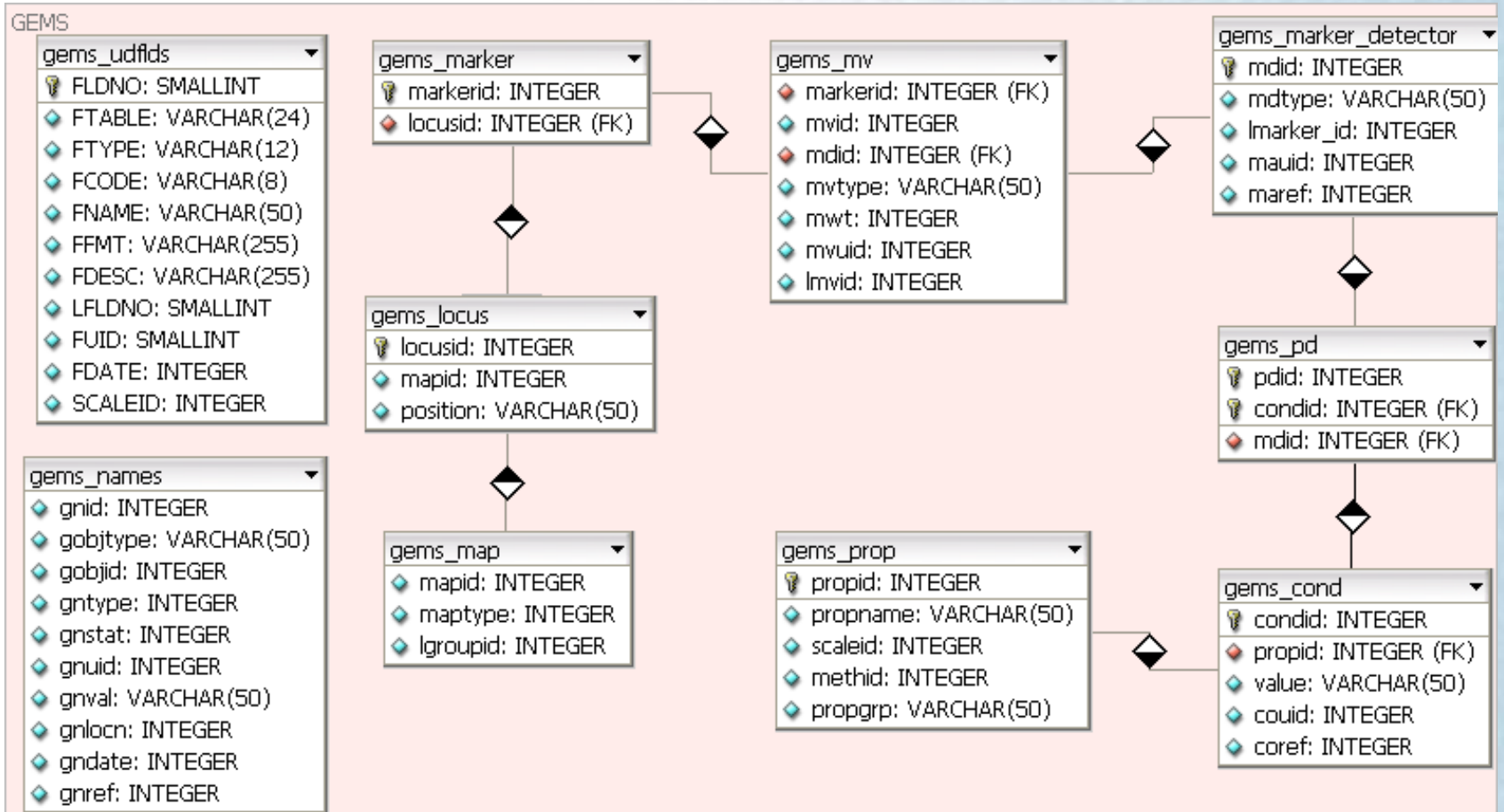
Links of GEMS and DMS Schema



Links GEMS and DMS Schema



GEMS Schema



GEMS schema features

- **Accommodates synonyms**

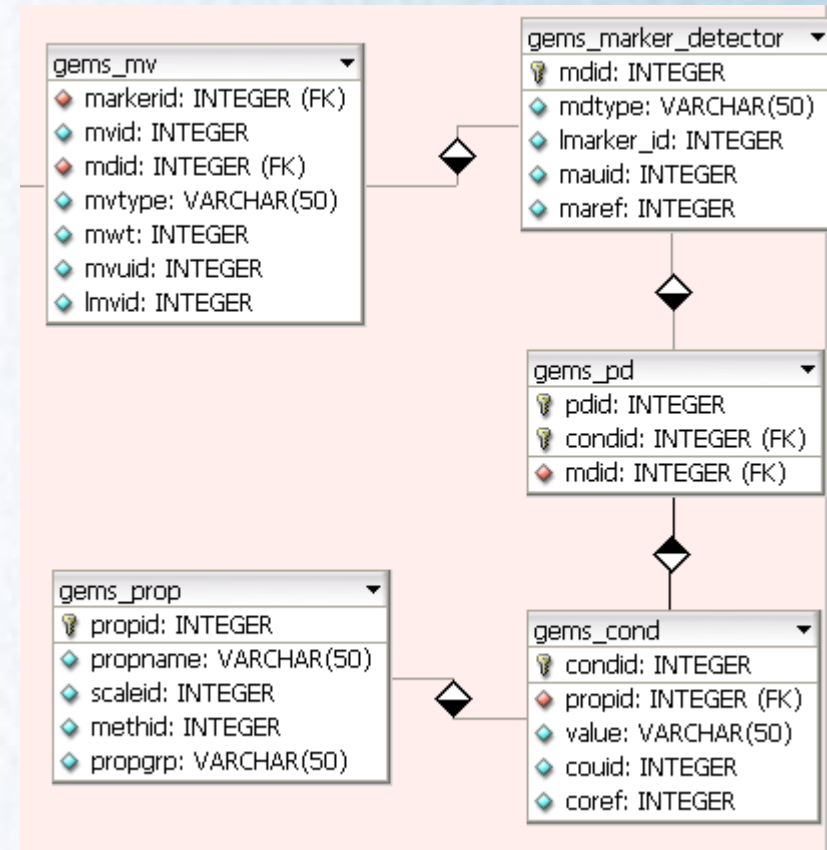
- Similar to the NAMES table of GMS
- All Names for Markerids and MVIDs are stored
- Gnstat is used to allow Synonyms for marker, alleles or for any object in the GEMS schema
- Gnlocn, Gndate, Gnref –refers to the location, date and references in naming the marker

```
gems_names
◆ gnid: INTEGER
◆ gobjtype: VARCHAR(50)
◆ gobjid: INTEGER
◆ gnstype: INTEGER
◆ gnstat: INTEGER
◆ gnuid: INTEGER
◆ gnval: VARCHAR(50)
◆ gnlocn: INTEGER
◆ gndate: INTEGER
◆ gnref: INTEGER
```

GEMS schema features

Flexibility with Marker Detector Descriptors

- Addition of marker detector descriptors is possible
- Adding a new descriptor mean adding a new row and not a column
 - Not limited by columns
 - Controlled Vocabulary



GEMS Prototype Interface

- Main User Interface

- Marker Name
- Marker Details
- Marker Synonyms
- Marker Locus
- Marker Alleles
- Marker Detector Conditions

The screenshot displays the GEMS interface with the following sections highlighted by red circles:

- Marker Name:** RM1880
- Marker Details:** Marker Name (RM1880), Marker Type (SSR), Date (20060328), Location (ARTHURTON (SA)), Reference (0), Synonyms (AY018555, MRG0880.RGP0880).
- Marker Locus:** Chromosome (1), Position (13.0).
- Marker Alleles:** Table with columns: mwid, molecular variant name, status, mvtype, weight.

mwid	molecular variant name	status	mvtype	weight
10001	4		99	1
8	RM1880_md	0		0
7	RM1880_152	0		152
6	RM1880_147			147
4	RM1880_150			150
3	RM1880_145			145
- Marker Detector Conditions:** Forward Primer, Reverse Primer, Primer Origin, DNA extraction Method, Visualization (Ethidium bromide).

GEMS Prototype Interface

Search facility

result

The screenshot shows the GEMS interface with a search bar containing 'rm5*' and a 'SEARCH' button. The search results list several markers, with RM5552 selected. The details for RM5552 are shown in a table format.

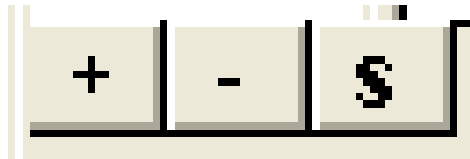
MARKER DETAILS FOR MARKER ID: 2

Marker Name	RM5552	locus	
Marker Type	SSR	Chromosome	Position
Date	20060328	1	20.2
Location	BUTE (SA)	0	
Reference	0		
Synonyms	AUT10244		

ALLELES FOR MARKER : RM5552

mvid	molecular variant name	status	mvtype	weight
11	RM5552_160			160
10	RM5552_105			105
9	RM5552_110			110

GEMS Prototype Interface



-  • Add
-  • Delete
-  • Update/Save

Gene Management System

SEARCH Marker Name Allele Name Primer Marker Type

Marker Name	Marker Type	Date	Location	Reference	Synonyms
RM1880					
RM5552					
RM3360					
RM1220					
RM8083					
RM580					
RM449					
RM5					
RM3143					
RM1183					
RM486					
▶ RM6407					
RM6840					
RM5897					
RM322					
RM145					
RM475					
RM1342					
RM3421					
RM425					
RM207					
RM132					
RM231					
RM6849					
RM7197					

MARKER DETAILS FOR MARKER ID: 12

Marker Name: RM6407 locus
 Marker Type: Isozyme Chromosome: Position:
 Date: 2
 Location:
 Reference:
 Synonyms:

ALLELES FOR MARKER : RM6407

mvid	molecular variant name
56	RM6407_168
55	RM6407_160
54	RM6407_158
53	RM6407_145
52	RM6407_150

Add New MV

Add New Molecular Variant

Molecular Variant ID: 10003
 MV Name: 56
 Status:
 MV Type:
 MarkerDetector ID:
 Molecular Weight:
 MV User's ID:
 MarkerID: 12
 Local MVID:
 Reference:

Initial Denaturation: Temp [] Duration []
 Amplification Conditions: Number of Cycles []
 Annealing: Temp [] Duration []
 Elongation: Temp [] Duration []
 Terminal Elongation: Temp [] Duration []

Record: [] [] [] [] [] []

GEMS schema features

An SSR Marker Detector with the following descriptors:

Primer

Forward primer

Reverse primer

PCR Reaction

H2O

MgCl₂

dNTPs

Electrophoresis Composition

H2O

Acrylamide

TBE

PCR condition

Annealing Temp

PCR cycle

Note:

- Property group

- Property name

GEMS schema features

The screenshot displays a software interface with a table of records on the left and a detailed view on the right. The table lists records with IDs: RM475, RM1342, RM3421, RM425, RM207, RM132, RM231, RM6849, and RM7197. The detailed view is titled 'MARKER DETECTOR INFORMATION' and has several tabs: 'Primers/Probe', 'PCR Condition', 'PCR Reaction', 'electrophoresis', and 'Enzyme Combinations'. The 'PCR Condition' tab is selected and circled in red, with an arrow pointing to the label 'Property group'. Below the tabs, there are several sections of input fields: 'Initial Denaturation' (Temp, Duration), 'Amplification Conditions' (Number of Cycles, Temp, Duration), 'Annealing' (Temp, Duration), 'Elongation' (Temp, Duration), and 'Terminal Elongation' (Temp, Duration). The labels for 'Initial Denaturation' and 'Amplification Conditions' are circled in red, with an arrow pointing to the label 'Property name'. The input fields for temperature and duration are also circled in red, with an arrow pointing to the label 'Property value'.

Property name

Property value

GEMS schema features

- An SSR Marker Detector with the following descriptors:

Primer

Forward primer

Reverse primer

PCR Reaction

H₂O

MgCl₂

dNTPs

Electrophoresis condition

Duration

Current

Voltage

Electrophoresis Composition

H₂O

Acrylamide

TBE

PCR condition

Annealing Temp

PCR cycle

Melting temperature

GEMS schema features

RM475
RM1342
RM3421
RM425
RM207
RM132
RM231
RM6849
RM7197

+ - S

Record:

+ - S

MARKER DETECTOR INFORMATION

Primers/Probe PCR Condition PCR Reaction electrophoresis Enzyme Combinations

Initial Denaturation: Temp Duration

Amplification Conditions: Number of Cycles

Temp Duration

Annealing Temp Duration

Elongation Temp Duration

Terminal Elongation Temp Duration

Development

- Devoted much of the effort in designing a flexible GEMS schema than developing the GUI
- Improvement on the speed of accessing GEMS data from workbook
- In the process of deciding on a more powerful Application Programming Language that will better accommodate the flexibility of the Schema

Thank You...

Flexibility of the GEMS Schema

gems_pd : Table

pdid	conditionid	mdid
1	1	1
1	2	1
1	3	1
1	5	1
1	8	1
1	9	1

Record: 6 of 6

gems_prop : Table

propid	propname	propgrp	method
1	forward primer	primer	
2	reverse primer	primer	
3	primer origin	primer	
4	Initial Denaturation Temperature	PCR Conditions	
5	Initial Denaturation Duration	PCR Conditions	
6	Amplification Temperature	PCR Conditions	
7	Amplification Duration	PCR Conditions	
8	Amplification Number of Cycles	PCR Conditions	
9	Annealing Temperature	PCR Conditions	
10	Annealing Duration	PCR Conditions	
11	Elongation Temperature	PCR Conditions	
12	Elongation Duration	PCR Conditions	
13	Terminal Elongation Temperature	PCR Conditions	
14	Terminal Elongation Duration	PCR Conditions	
15	Electrophoresis Voltage	Electrophoresis	
16	Electrophoresis Current	Electrophoresis	
17	electrophoresis time	Electrophoresis	
18	Water	Electrophoresis	
19	TBE	Electrophoresis	
20	Acrylamide	Electrophoresis	

Record: 16 of 34

gems_cond : Table

condid	propid	value	coid	coref
1	1	aatccaaggtgcagagatgg	0	0
2	2	caacgatgacgaacacaacc	0	0
3	9	55	0	0
4	8	30	0	0
5	28	8	0	0
6	29	6FAM	0	0
7	30	blue	0	0
8	31	1x	0	0
9	32	(CTG)7	0	0
10	33	148	0	0
11	34	160	0	0
12	1	gcgaaaacacaatgcaaaaa	0	0

Record: 23 of 55