



# Sorghum SSR reference kit

- [http://sat.cirad.fr/sat/sorghum\\_SSR\\_kit/](http://sat.cirad.fr/sat/sorghum_SSR_kit/)

Sorghum bicolor genetic diversity SS... Allelic reference

## A SSR Kit to study Sorghum genetic diversity


The sorghum SSR kit has been elaborated in the frame of a large genotyping project (3000 sorghum accessions x 48 SSR loci) supported by the [Generation Challenge Program](#), Sub-Programs 1 and 5 (grants dedicated to Daniel Fonceka and M'Baye N'Doye Sall, formerly working at the Ceraas, Senegal). It is composed of the allelic sizes of 3 controls for each of the 48 SSR locus.

**Composition of the controls:** includes the accession number (as provided by ICRISAT), in the genotyping order, and there belonging to the controls. The 3 controls are composed of 10 DNA samples mixed in 2 pools of 3 and 1 pool of 4 individuals. The 10 individuals were chosen from 48 Sorghum samples presenting a fair picture of the overall genetic diversity, in order to represent a large range of allelic diversity, both in term of allele number and allele sizes. Each control is amplified for each marker, and allelic sizes are used as control sizes.

**Characteristics of the 48 markers:** includes, for each locus, its name, the microsatellite motive, forward and reverse primer sequences, the number of alleles observed among the controls and their size range, and linkage group assignment (following Kim et al. (2005) nomenclature). The 48 loci are located regularly throughout the genome, so that there are roughly 5 markers per chromosome. Most of the markers were previously published markers (Brown et al., 1996; Taramino et al., 1997; Battramakki et al., 2000; Kong et al., 2000; Schloss et al., 2002). Additional markers were developed from microsatellites enriched libraries using the [SAT pipeline](#).

**Allelic patterns of control samples:** shows the Licor IR2 locus image, allelic content of each sample and of the controls. For each marker, the 10 individuals and the 3 mix controls were amplified and analyzed on Licor IR2 sequencer. A subset of the individual alleles detected for each marker were sequenced in order to determine their exact sizes. The size of the remaining alleles were determined relatively to the sequenced alleles based on stutter information.

All experiments related to this kit were performed on the [Languedoc Roussillon Genotyping Platform](#), hosted by the [CIRAD](#), under the supervision of Claire Billot and with the collaboration of Ronan Rivallan, Jean-François Rami and Monique Deu.

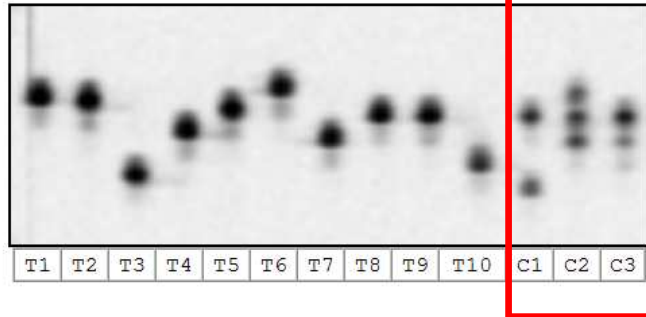


Terminé

Marker_Name	Motive	Forward_Primer	Reverse_Primer	Reference	T <sub>a</sub>	Size_Range	Allele_Nb	LG
gspb067	(GT)10	TAGTCCATACACCTTTCA	TCTCTCACACACATTCTTC	unpublished, Agropolis-Cirad-Genoplante	49	170-180	8	8
gspb069	(TC)12	CCCATAATACTGACCTTC	ACTTACTCCCTCTGTCCC	unpublished, Agropolis-Cirad-Genoplante	50	169-197	4	9
gspb069	(TG)9	ATCAGGTACAGCAGGTAGG	ATGCATCATGGCTGGT	unpublished, Agropolis-Cirad-Genoplante	50	105-173	4	1
gspb123	(CA)7+(GA)5	ATAGATGTTGACGAAGCA	GTGGTATGGGACTGGA	unpublished, Agropolis-Cirad-Genoplante	50	288-296	5	8
gspb148	(TC)3+(CA)5	CAACCACAAACCAAGAG	ATAGAAATGGGTGGAG	unpublished, Agropolis-Cirad-Genoplante	50	135-147	4	5
gspb151	(CT)12	ATACCAAGTTCCCTTACCT	GTTGGGGAGAGTTTT	unpublished, Agropolis-Cirad-Genoplante	50	106-128	7	4
ISEP0107	(TGG)4	GCCGTAAACAGAGAGGATGG	TTTCCGCTACCTCAAAAACC	unpublished, Iorisat	59	199-205	3	3
ISEP0310	(CCAAT)4	TGCCTTGCCCTGTTATCT	GGATC.GATGCC.TATC.TC.GTC	unpublished, Iorisat	60	164-219	3	2
mSbCIR223	(AC)6	CGTTCCAATGACTTTTTCTC	GCCAATGTGGGTGTATAAT	unpublished, Agropolis-Cirad-Genoplante	55	108-118	5	2
mSbCIR238	(AC)26	AGAAGAAAAGGGTAAGAGC	CGAGAAAACAATTACATGAACC	unpublished, Agropolis-Cirad-Genoplante	55	79-107	8	2
mSbCIR240	(TG)9	GTTCTTGCCCTACTGAAT	TCACCTGTAACCTGTCTTC	unpublished, Agropolis-Cirad-Genoplante	55	108-112	2	8
mSbCIR246	(CA)7.5	TTTTGTTGCACTTTTGAGC	GATGATAGCGACCACAATC	unpublished, Agropolis-Cirad-Genoplante	55	98-100	2	5
mSbCIR248	(GT)7.5	GTTGGTCACTGGTGATAAA	ACTCCATGTGCTGAATCT	unpublished, Agropolis-Cirad-Genoplante	55	89-101	4	10
mSbCIR262	(CATG)3.25	GCACCAAAATCAGCGTCT	CCATTACCCGGGATTAGT	unpublished, Agropolis-Cirad-Genoplante	57	208-220	4	7
mSbCIR276	(AC)9	CCCCAATCTAATATTTGGT	GAGGCTGAGATGCTCTGT	unpublished, Agropolis-Cirad-Genoplante	53	230-234	4	3
mSbCIR283	(CT)6 (GT)6.5	TCCTCTGAGCTGTAAAT	CAAGTCACTACCAATGGAC	unpublished, Agropolis-Cirad-Genoplante	54	113-139	6	7
mSbCIR286	(AC)9	GCTTCTATACCTCCCTCCAC	TTTTATGATAGGATGCTCTGC	unpublished, Agropolis-Cirad-Genoplante	55	112-134	4	1
mSbCIR300	(GT)9	TTGAGAGCGCGAGGTAA	AAAAGCCCAAGTCTCAGTGCTA	unpublished, Agropolis-Cirad-Genoplante	61	104-110	4	5
mSbCIR306	(GT)7	ATACTCTGTAAGCTGGCTCA	ATACTCTGTAAGCTGGCTCA	unpublished, Agropolis-Cirad-Genoplante	56	120-124	3	1
mSbCIR329	(AC)8.5	GCAGAACATCACTCAAGAA	GCAGAACATCACTCAAGAA	unpublished, Agropolis-Cirad-Genoplante	55	109-117	5	10
sb4-72	(AG)16	TGCCACCCTCTGAAAAAGGCTA	CTGAGGACTGCCCAATGTAGG	Brown et al., 1996	55	183-203	7	9
sb5-206	(AC)13/(AG)20	ATTCTATCATCTCATCTCTGTAGAA	AAAAACCAACCCGACCCAACCTC	Brown et al., 1996	55	106-146	10	6
sb6-84	(AG)14	CGCTCTGGGATGAATGA	TAACGGCACTAACAATGATT	Brown et al., 1996	55	183-217	7	2
SbAG-B02	(AG)36	CTCTGATATGTCTGTGTGCT	ATAGAGAGGATAGCTTATAGCTCA	Taramino et al., 1997	55	90-154	6	5
Xoup02	(GCA)6	GACGCAGCTTTGCTCTATC	GTCCAACCAACCCACGTATC	Schloss et al., 2002	54	192-204	4	6
Xoup11	(GCTA)4	TACCGCCATGTCATCATCAG	CGTATCGCAAGCTGTGTTTG	Schloss et al., 2002	54	165-172	2	3
Xoup14	(AG)10	TACATCAGCAGGGACAGG	CTGGAAAAGCCGAGCAGTATG	Schloss et al., 2002	54	211-225	4	3
Xoup53	(TTTA)5	GCAGGAGTATAGGCAGAGGC	CGACATGACAAAGCTCAAAACG	Schloss et al., 2002	54	186-198	3	1
Xoup61	(CAG)7	TTAGCATGTCCACCACAACC	AAAGCAACTGCTGTATCCC	Schloss et al., 2002	54	198-201	2	3
Xoup62	(GAA)6	CGAGAAGATCGAGAGAACC	TGAAGACGACGACGACAGAC	Schloss et al., 2002	54	190-193	2	1
Xoup63	(GGATGC)4	GTAAAGGGCAAGCAACAAG	GCCCTACAAAATCTSCAAGC	Schloss et al., 2002	54	133-145	3	2
Xbcp010	(CT)14	ATACTACCAAGGGGAGC	AGTACTAGCCACACGTCAAC	Kong et al., 2000	50	135-151	4	6
Xbcp012	(CT)22	AGATCTGGCGCAACG	AGTCAACCATCGATCATC	Kong et al., 2000	55	161-205	6	4
Xbcp015	(TC)16	CACAAACACTAGTGCCTTATC	CATAGACACCTAGGCCATC	Kong et al., 2000	55	199-223	6	10
Xbcp021	(AG)18	GAGCTGCCATGATTTGGTCG	ACCTCGTCCACCTTTGTTG	Kong et al., 2000	60	169-199	7	4
xbcp040	(GGA)7	CAGCAACTTGCACTTGTG	GGGAGCAATTTGGCACTAG	Kong et al., 2000	55	129-141	4	5
Xbcp057	(GT)21	GGAACTTTTGACGGTAGTGC	CGATCGTGTATGCCCAATC	Kong et al., 2000	55	223-257	8	9
Xbcp114	(AGG)8	CGTCTTACCCGGCTCCT	CATAATCCCACTCAACAATCC	Kong et al., 2000	50	211-217	3	3
Xbcp136	(GCA)5	GCGAATAGCATCTTACAACA	ACTGATCATTGGCAGGAC	Kong et al., 2000	55	240-243	2	10
Xbcp141	(GA)23	TGATGGCCCTAGCTTATCT	CAACAAAGCAACCTAAA	Kong et al., 2000	55	135-167	9	7
Xbcp145	(AG)22	GTTCTCTGCCAATFACT	CTTCCGCACATCCAC	Kong et al., 2000	55	208-244	8	9
Xbcp265	(GAA)19	GTCTACAGCGTGCAATAAAA	TTACCATGCTACCCCTAAAAGTGG	Kong et al., 2000	55	188-234	7	9
Xbcp273	(TTG)20	GTACCCATTAATGTTTGCAGTAG	CAGAGGAGGAGGAGGAAAGG	Kong et al., 2000	55	169-193	6	8
Xbcp278	(TTG)12	GGGTTTCACTGACCTACCGAATCTCT	ATGCCTCATCATGTTGCTTTGCTT	Kong et al., 2000	50	243-252	3	5
Xbcp295	(TC)19	AAATCATGCATCCATGTTCTCTC	CTCCCGTCAAGAGTACATTATAGCTTA	Kong et al., 2000	55	153-183	6	5

### mSbCIR329

Locus Image



Allelic content in the samples

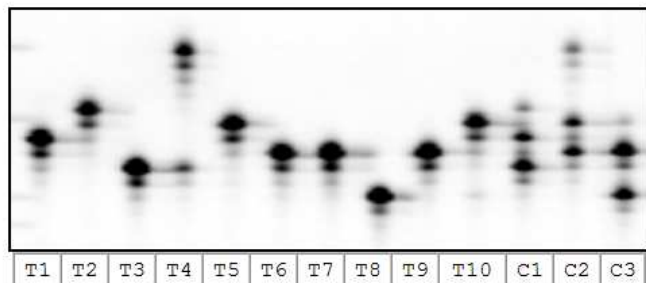
DNA_Sample	Control	Allele_Size
T1	C1	115
T2	C1	115
T3	C1	109
T4	C2	113
T5	C2	115
T6	C2	117
T7	C3	113
T8	C3	115
T9	C3	115
T10	C3	111

Allelic content in the controls

Control	Allele_Size
C1	109,115
C2	113,115,117
C3	111,113,115

### sb4-72

Locus Image



Allelic content in the samples

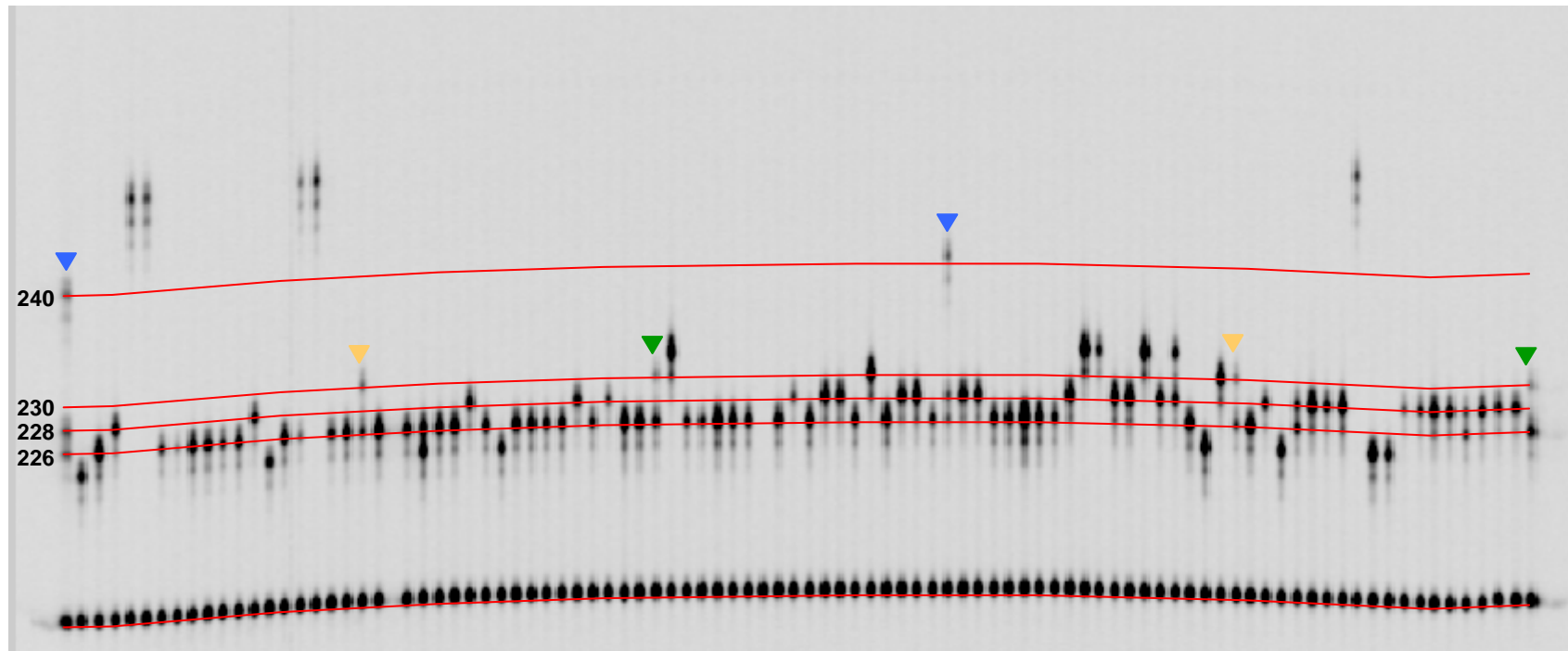
DNA_Sample	Control	Allele_Size
T1	C1	191
T2	C1	195
T3	C1	187
T4	C2	187
T4	C2	203
T5	C2	193
T6	C2	189
T7	C3	189
T8	C3	183
T9	C3	189
T10	C3	193

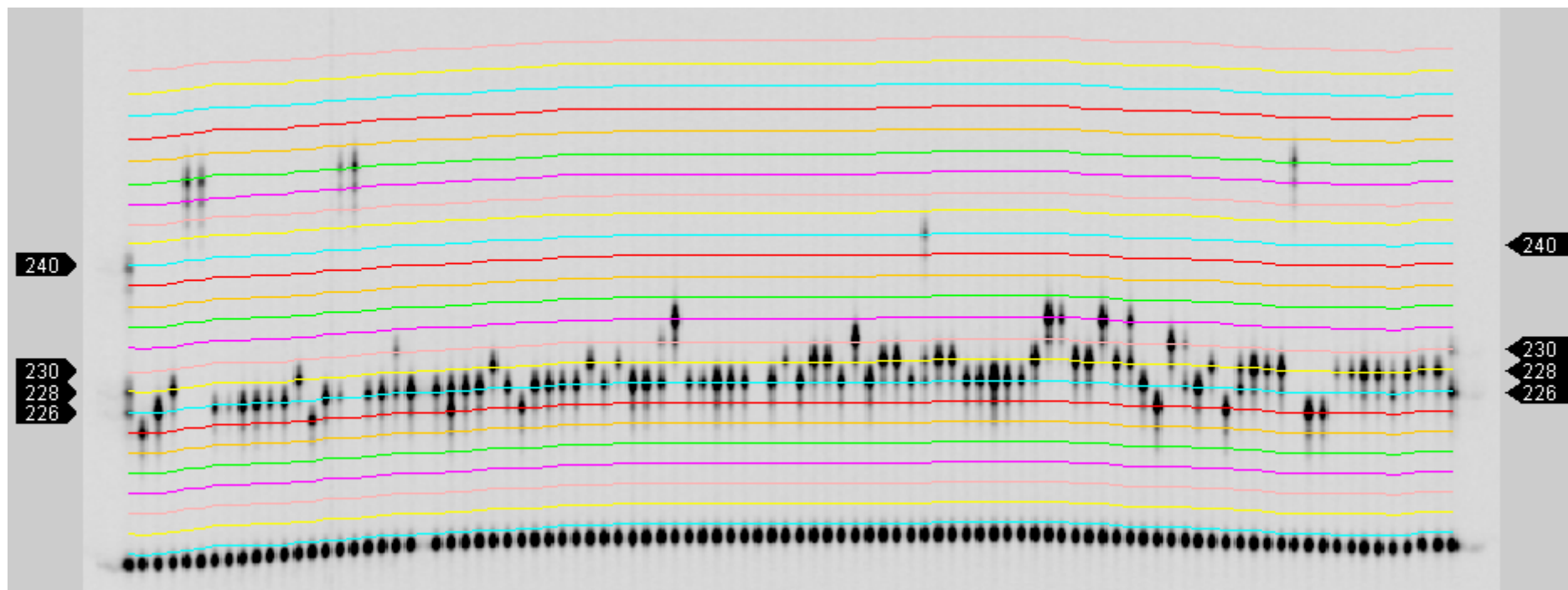
Allelic content in the controls

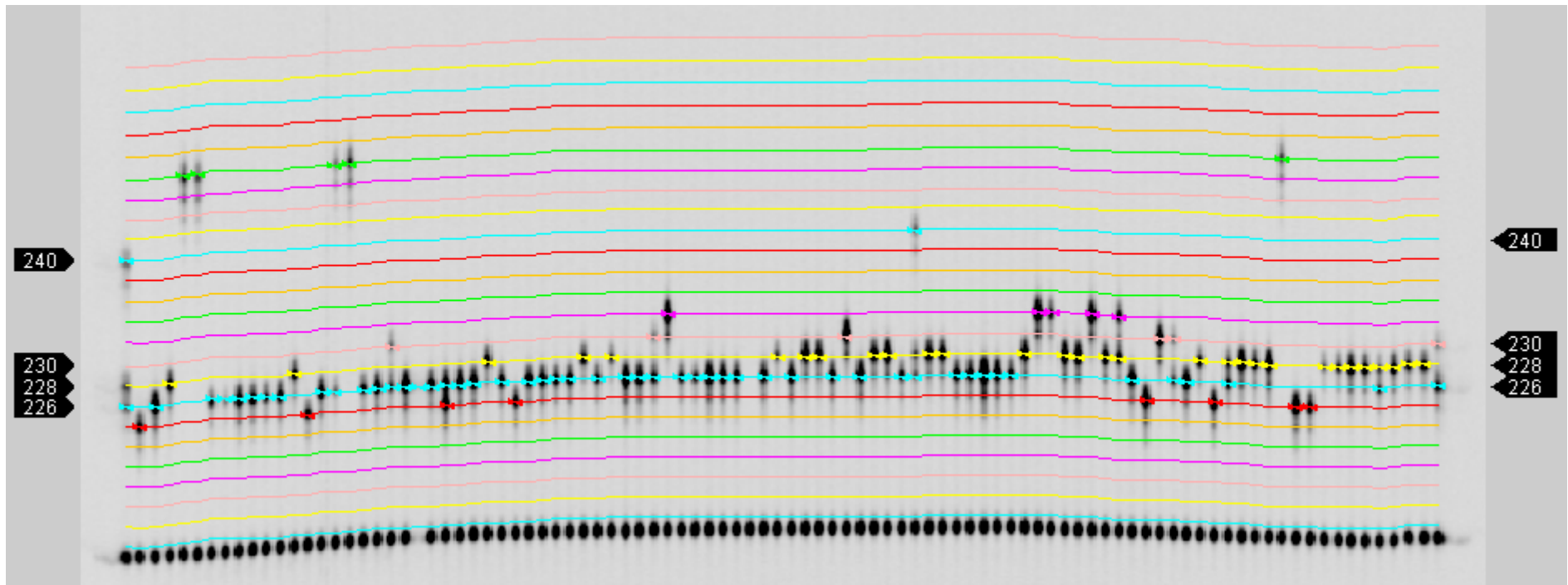
Control	Allele_Size
C1	187,191,195
C2	187,189,193,203
C3	183,189

Control		A	B	C
Xcup14	226	1	1	1
	228	1		
	230		1	1
	240	1		

Allelic control  
Softwares  
-SAGA software use  
-AlleloBin







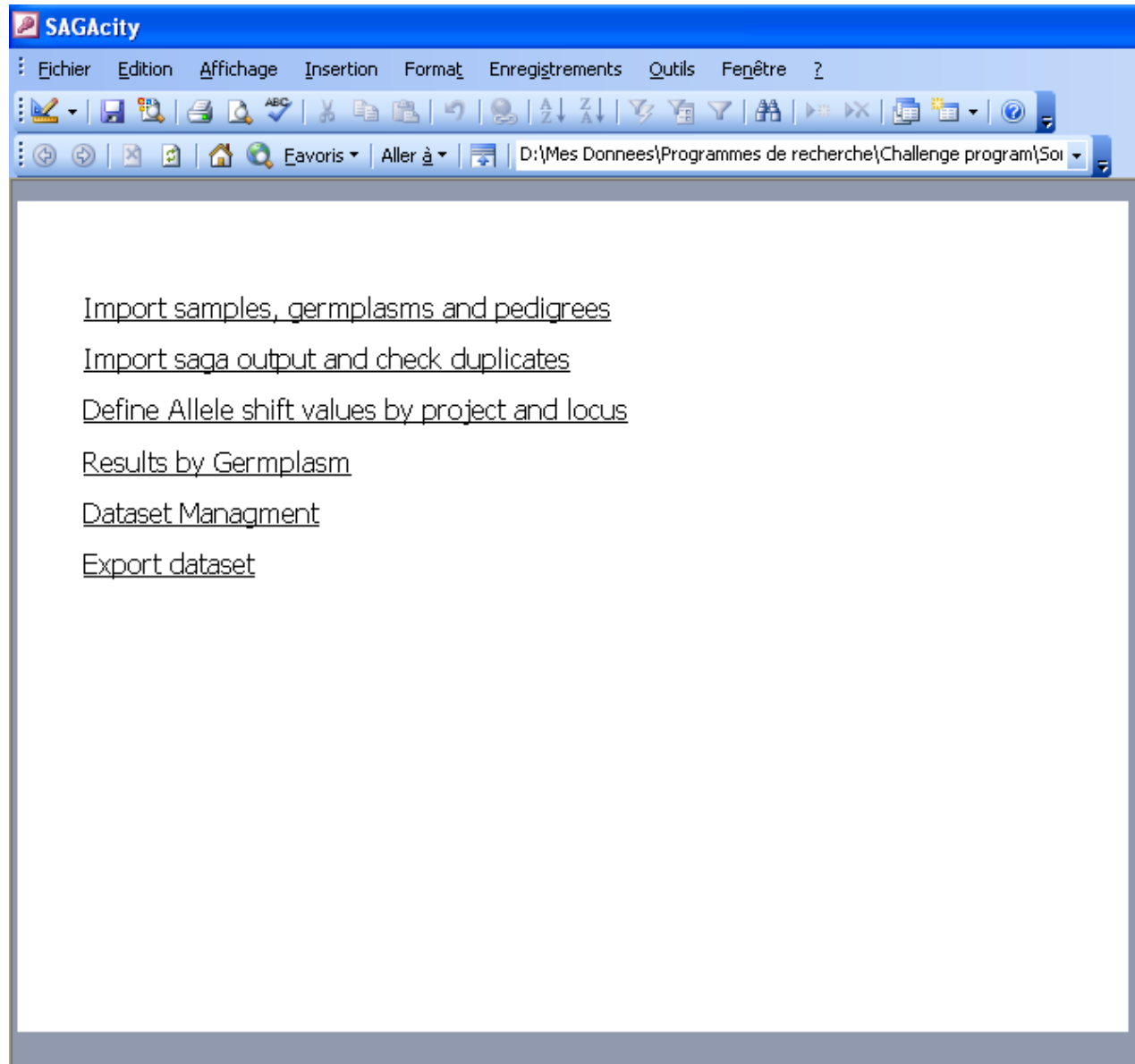
# Sorghum Dataset – Merging Data

- At Cirad:
  - No LIMS available
  - Construction of an access db ([SAGAcity](#)) which enables the storage of germplasm-sample correspondance, genotyping data, versionning and cleaning of the data
- At Icrisat:
  - LIMS
  - Extraction of a final dataset, including the control samples
- Between labs:
  - Performed in SagaCity, based on control bp sizes
- Submission to GCP of the dataset that is analysed
  - i.e. with less than missing data

# SAGAcity

- Access db, oriented towards SAGA outputs, but adaptable (used to import data obtained from Icrisat)
- Build by JF Rami (Cirad) to extract specific datasets
  - Dealing with germplasm-accession-DNA extraction
  - Enables the merging of datasets
  - Control the level of experiment repeats and user input on the quality of the data

# SagaCity – Main menu



# SAGAcity – Defining the samples

## Importing samples

- Sample\_ID: DNA sample
- Germplasm\_ID: seeds, etc...
  - The spirit of the db is to synthesise what has been obtained by germplasm
- Pedigree\_ID = accession ID

Sample\_import : Formulaire

Choose an existing project or type a new on [ ]

Copy and Paste the Sample identifiers to Import. All the fields are mandat

	SAMPLE_ID	GERMPLASM_ID	PEDI_ID
▶			

Enr : [ ] 1 [ ] sur 1

Import

# SAGAcity – Import data and check for duplicates

1- Choose datafile exported from SAGA

2- Import it into a project

3- Check for non-coherent duplicates  
*User control*

4- Validate the data

Step 1 :  
Import a SAGA output file.  
The file is a "complete" report  
It needs to be cleaned from the allelic frequencies information.  
The first line should look like:  
SampleID|Locus|Gel|Dye|Allele 1|Peak 1|Qual 1|Volume 1|Allele 2|Peak 2|Qual 2|Volume 2|Mendelian|Comment

Saga output file: D:\Mes documents\sagaout.txt

Step 2 : Add saga import to the results of a proje  
Choose a project or type a new one:  Add saga import

Step 3 : Look for non-coherent duplicates in all the results of the proj  
You can exclude some samples from the duplicate search by entering them in the list below. This is a "like" statement: any sample matching the specified words will be excluded

SampleID
control
*

Discard	Numero	SampleID	LocusID	Gel	Dye	Genotype
<input type="checkbox"/>	#Nom ?	#Nom ?	#Nom ?	#Nom ?	#Nom ?	#Nom ?

Look for duplicates

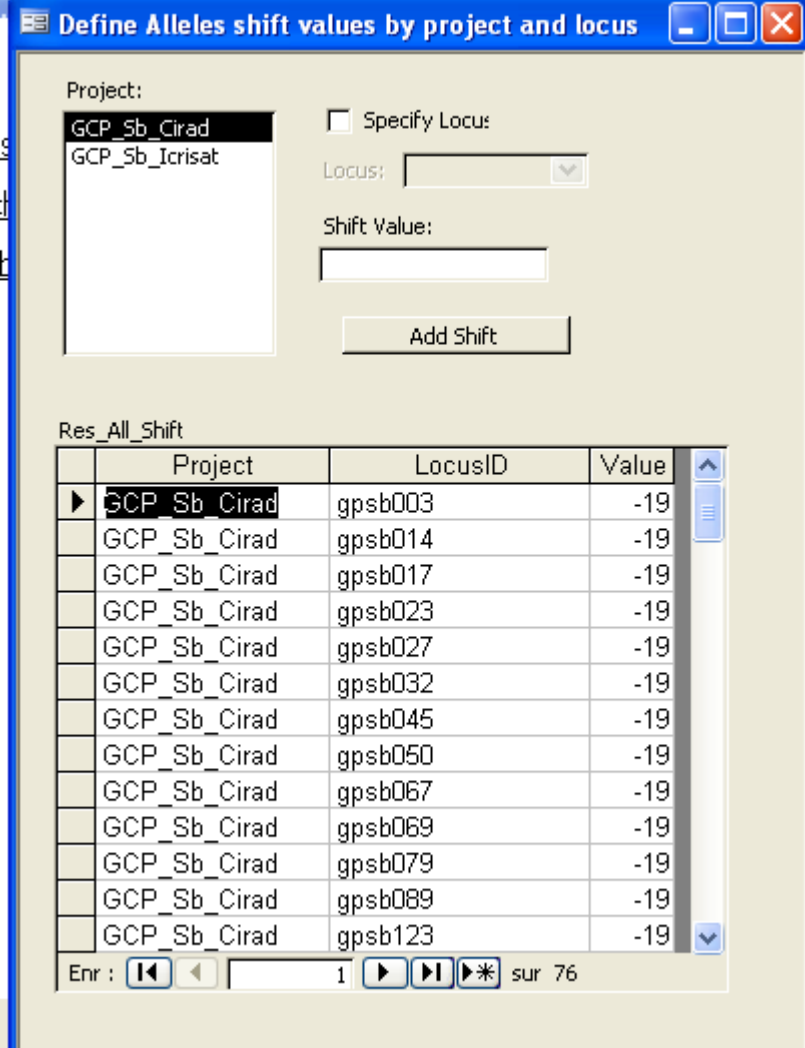
Enr : 1 sur 1

Validate

# SAGAcity – Merging data – Shifting values

According to control values,  
you can shift all data

- 19 bp for the M13tail
- shift between labs



Define Alleles shift values by project and locus

Project:  
GCP\_Sb\_Cirad  
GCP\_Sb\_Icrisat

Specify Locus  
Locus:   
Shift Value:   
Add Shift

Res\_All\_Shift

	Project	LocusID	Value
▶	GCP_Sb_Cirad	gpsb003	-19
	GCP_Sb_Cirad	gpsb014	-19
	GCP_Sb_Cirad	gpsb017	-19
	GCP_Sb_Cirad	gpsb023	-19
	GCP_Sb_Cirad	gpsb027	-19
	GCP_Sb_Cirad	gpsb032	-19
	GCP_Sb_Cirad	gpsb045	-19
	GCP_Sb_Cirad	gpsb050	-19
	GCP_Sb_Cirad	gpsb067	-19
	GCP_Sb_Cirad	gpsb069	-19
	GCP_Sb_Cirad	gpsb079	-19
	GCP_Sb_Cirad	gpsb089	-19
	GCP_Sb_Cirad	gpsb123	-19

Enr : 1 sur 76

# SAGAcy – Merging data – Shifting values

According to control values,  
you can shift the entire  
data

- 19 bp for the M13tail
- shift between labs

Project:

- GCP\_Sb\_Cirad
- GCP\_Sb\_Icrisat

Specify Locus

Locus:

Shift Value:

Add Shift

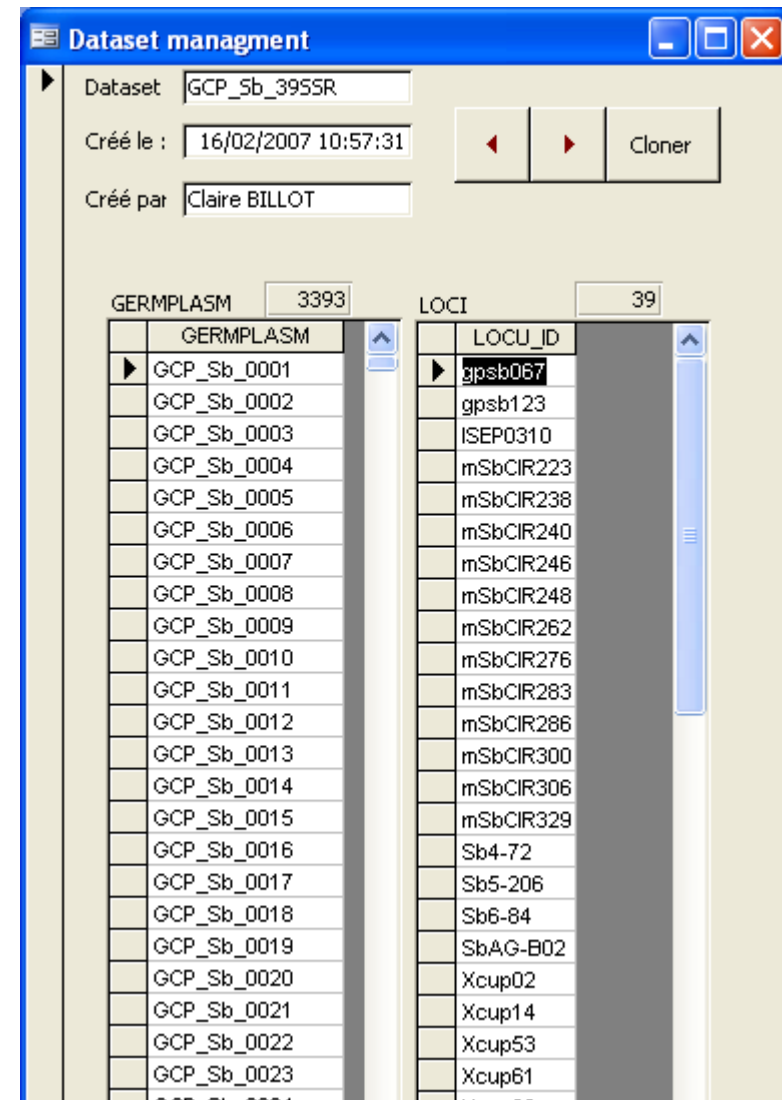
Res\_All\_Shift

	Project	LocusID	Value
▶	GCP_Sb_Icrisat	ISEP0107	-1
	GCP_Sb_Icrisat	ISEP0310	1
	GCP_Sb_Icrisat	mSbCIR223	3
	GCP_Sb_Icrisat	mSbCIR238	6
	GCP_Sb_Icrisat	mSbCIR240	2
	GCP_Sb_Icrisat	mSbCIR248	-1
	GCP_Sb_Icrisat	mSbCIR276	1
	GCP_Sb_Icrisat	mSbCIR283	0
	GCP_Sb_Icrisat	mSbCIR286	6
	GCP_Sb_Icrisat	mSbCIR306	0
	GCP_Sb_Icrisat	mSbCIR323	0
	GCP_Sb_Icrisat	Sb4-72	1
	GCP_Sb_Icrisat	Xcup11	2

Enr :       sur 24

# SAGAcity – Defining dataset

- Create a dataset
- Choose the germplasm list
- Choose the locus list



# SAGAcy – Exporting dataset

Export dataset

Datasets:

- GCP\_Sb\_3955R
- GCP\_Sb\_4155R
- GCP\_Sb\_Light\_4055I

Output:

@DARwin 5.0 - ALLELIC - 2  
3393□78  
N°□gpsb067□gpsb067□gpsb123□gpsb123□ISEP0310□ISEP0310□mSbCIR223□mSbCIR223□mSbCIR238  
□mSbCIR238□mSbCIR240□mSbCIR240□mSbCIR246□mSbCIR246□mSbCIR248□mSbCIR248□mSbCIR26  
2□mSbCIR262□mSbCIR276□mSbCIR276□mSbCIR283□mSbCIR283□mSbCIR286□mSbCIR286□mSbCIR3  
00□mSbCIR300□mSbCIR306□mSbCIR306□mSbCIR329□mSbCIR329□Sb4-72□Sb4-72□Sb5-206□Sb5-  
206□Sb6-84□Sb6-84□SbAG-B02□SbAG-  
B02□Xcup02□Xcup02□Xcup14□Xcup14□Xcup53□Xcup53□Xcup61□Xcup61□Xcup63□Xcup63□Xtxp010  
□Xtxp010□Xtxp012□Xtxp012□Xtxp015□Xtxp015□Xtxp021□Xtxp021□Xtxp040□Xtxp040□Xtxp057□Xt  
xp057□Xtxp114□Xtxp114□Xtxp136□Xtxp136□Xtxp141□Xtxp141□Xtxp145□Xtxp145□Xtxp265□Xtxp26  
5□Xtxp273□Xtxp273□Xtxp278□Xtxp278□Xtxp320□Xtxp320□Xtxp321□Xtxp321  
1□180□180□296□296□204□204□116□116□083□091□112□112□100□096□099□099□214□214  
□234□234□131□131□114□114□104□106□120□120□113□117□183□191□148□148□181□183□  
096□096□192□192□213□213□186□194□198□198□145□145□151□151□183□191□199□199□17  
5□175□138□138□241□241□214□214□240□246□151□151□214□214□207□213□190□190□243  
□249□263□263□198□198  
2□172□172□290□290□204□204□108□108□081□081□108□108□100□100□089□099□216□216  
□230□230□113□113□114□114□108□108□120□120□115□115□193□193□116□116□199□199□  
096□096□204□204□215□215□198□198□201□201□145□145□135□135□185□185□219□219□17  
5□175□138□138□241□241□214□214□243□243□135□135□242□242□198□198□190□190□243  
□243□287□287□204□204  
3□170□170□288□290□204□204□118□118□079□079□112□112□100□100□089□089□216□216  
□230□230□119□119□114□114□108□108□124□124□115□115□189□189□112□112□217□217□  
096□096□198□198□211□211□194□194□201□201□145□145□143□143□175□193□211□211□17  
5□175□138□138□241□241□211□211□240□240□167□167□244□244□234□234□190□190□249  
□249□278□278□218□218  
4□180□180□296□296□204□204□108□108□093□093□114□114□100□100□089□091□216□216  
□230□230□139□139□130□130□104□104□122□122□119□119□191□191□114□114□183□183□  
096□096□198□198□211□211□194□194□201□201□145□145□151□151□173□173□211□211□17  
5□175□138□138□251□251□214□214□243□243□151□151□214□214□210□210□190□190□249  
□249□287□287□198□198  
5□172□172□290□290□204□204□108□112□085□085□108□108□100□100□099□099□214□214  
□230□230□113□113□114□114□104□104□120□120□115□115□193□193□112□112□213□213□  
116□116□204□204□215□215□186□186□198□198□145□145□135□135□183□183□219□219□17  
5□175□141□141□241□241□211□211□243□243□151□151□240□240□204□204□166□166□249  
□249□275□275□212□212  
6□180□180□296□296□204□204□108□108□093□093□114□114□100□100□089□091□216□216  
□230□230□139□139□130□130□104□104□122□122□119□119□191□191□114□114□183□183□  
096□096□198□198□211□211□194□194□201□201□145□145□151□151□173□173□211□211□17  
5□175□138□138□251□251□214□214□243□243□151□151□214□214□210□210□190□190□249  
□249□287□287□198□198

Don File (Darwin):

@DARwin 5.0 - DON  
3393□3  
N°□LOT\_ID□PEDI\_ID  
1□GCP\_Sb\_0001□IS303  
2□GCP\_Sb\_0002□IS929  
3□GCP\_Sb\_0003□IS2156  
4□GCP\_Sb\_0004□IS2262  
5□GCP\_Sb\_0005□IS2263  
6□GCP\_Sb\_0006□IS2807  
7□GCP\_Sb\_0007□IS3421  
8□GCP\_Sb\_0008□IS3967  
9□GCP\_Sb\_0009□IS4821  
10□GCP\_Sb\_0010□IS6193  
11□GCP\_Sb\_0011□IS6745  
12□GCP\_Sb\_0012□IS7889  
13□GCP\_Sb\_0013□IS8685  
14□GCP\_Sb\_0014□IS9597  
15□GCP\_Sb\_0015□IS11119  
16□GCP\_Sb\_0016□IS12531  
17□GCP\_Sb\_0017□IS14331  
18□GCP\_Sb\_0018□IS14414  
19□GCP\_Sb\_0019□IS16186  
20□GCP\_Sb\_0020□IS19453  
21□GCP\_Sb\_0021□IS19455  
22□GCP\_Sb\_0022□IS20016  
23□GCP\_Sb\_0023□IS22282  
24□GCP\_Sb\_0024□IS22294  
25□GCP\_Sb\_0025□IS23178  
26□GCP\_Sb\_0026□IS23254  
27□GCP\_Sb\_0027□IS23644  
28□GCP\_Sb\_0028□IS28409  
29□GCP\_Sb\_0029□IS29233  
30□GCP\_Sb\_0030□IS29407  
31□GCP\_Sb\_0031□IS30405  
32□GCP\_Sb\_0032□IS30538  
33□GCP\_Sb\_0033□SSM29  
34□GCP\_Sb\_0034□SSM295

File format :

- PowerMarker
- Darwin
- CONVERT
- Liste

Export

Complete

Save As

Save and run CONVERT

Save As

That's it !!!