

Combining decentralised genotyping data for diversity analysis

GenDiversity

Gautier Sarah, Xavier Argout, Manuel Ruiz, Pierre Larmande (CIRAD)

Collaboration with other SP4 subtasks (IRRI, CIMMYT, Bioversity)

“Happy” users: Brigitte Courtois, Jean-François Rami, Claire Billot (CIRAD)

Project Rationale

- As an output of SP1 projects, very large datasets involving genotypes of thousands of accessions are now available.
- A query and analysis application combining genotyping data from diverse data sources is developed in support of diversity studies.
- We focus on SSR as the predominant genotyping method in the GCP. But the informatics development has to be generic enough to be used for any genotyping technology (including DArT and SNP).

Result

- GenDiversity: a new Web application for querying different genotyping studies
- A component of the GCP Platform
- Three main functionalities:
 - Querying decentralised genotyping data from different databases
 - Combining 2 or more different genotyping studies
 - Converting outputs to different file formats for diversity analysis

Querying decentralised genotyping data

Species: <ul style="list-style-type: none"><input type="checkbox"/> Check/Uncheck all<input type="checkbox"/> Barley<input type="checkbox"/> Chickpea<input type="checkbox"/> Coconut<input type="checkbox"/> Lentil<input type="checkbox"/> Rice<input type="checkbox"/> Sorghum<input type="checkbox"/> SugarCane	Database: <ul style="list-style-type: none"><input type="checkbox"/> Check/Uncheck all<input type="checkbox"/> GCP Central Repository<input type="checkbox"/> TropGene
---	---

Users can select:

- the species to be queried (Sorghum, Rice, etc....)
- the database(s) (GCP central repository, TropGene)

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_file663
- gcpcr_file662

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

List of available genotyping studies (by species and by databases)

If you select nothing, everything will appear in the result

Select your Germplasm(s)

GCP_48_P04_0001
GCP_48_P04_0002
GCP_48_P04_0003
GCP_48_P04_0004
GCP_48_P04_0005
GCP_48_P04_0006
GCP_48_P04_0007
GCP_48_P04_0008
GCP_48_P04_0009
GCP_48_P04_0010
GCP_48_P04_0011
GCP_48_P04_0012
GCP_48_P04_0013
GCP_48_P04_0015
GCP_48_P04_0016
GCP_48_P04_0017
GCP_48_P04_0018
GCP_48_P04_0019
GCP_48_P04_0020
GCP_48_P04_0021

Select your Marker(s)

Sb4-72
Sb5-206
Sb6-84
SbAG-B02
Xcup02
Xcup14
Xcup53
Xcup61
Xcup63
Xisep0310
Xtxp010
Xtxp012
Xtxp015
Xtxp021
Xtxp040
Xtxp057
Xtxp114
Xtxp136
Xtxp141
Xtxp145

Ploidy

2

Pre-checked Germplasms
for elimination when Missing Data

are above

%

In progress:

submit

reset

Filters based on Passport information: Collections, Cultivars, Pedigree, Collecting Locations, Descriptors, etc... and on Markers

For each selected studies: list of (DNA samples and/or Germplasm ID) and markers

Users can extract sub-samples from the genotyping dataset

Germplasm	CNZ03	CNZ03	CNZ40	CNZ40	CNZ42	CNZ42	CnCir 119	CnCir 119	CnCir 126	CnCir 126	CnCir 147	CnCir 147	CnCir 2	CnCir 2	CnCir 206	CnCir 206	CnCir 215	CnCir 215	Missing Data	Germplasm	<input type="checkbox"/>
1	95	91	152	152	158	158	202	202	179	184	211	211	222	222	99	131	126	126	0.0%	1	<input type="checkbox"/>
2	95	95	152	152	146	158	202	202	179	179	211	211	222	222	99	131	126	126	0.0%	2	<input type="checkbox"/>
3	95	95	152	152	146	166	202	202	179	184	211	211	222	224	97	99	126	126	0.0%	3	<input type="checkbox"/>
4	91	95	152	152	158	166	202	202	179	179	211	211	222	222	97	99	126	126	0.0%	4	<input type="checkbox"/>
5	91	95	146	152	146	166	202	202	179	184	211	211	222	224	99	131	126	126	0.0%	5	<input type="checkbox"/>
6	91	91	146	152	158	158	202	202	179	184	211	211	222	224	99	131	126	126	0.0%	6	<input type="checkbox"/>
7	91	95	146	152	146	166	202	202	179	184	211	211	222	224	99	131	126	126	0.0%	7	<input type="checkbox"/>
8	91	91	146	152	146	166	202	202	179	179	211	211	222	224	97	99	126	126	0.0%	8	<input type="checkbox"/>
9	95	95	152	152	146	166	202	202	179	184	211	211	222	222	97	131	126	126	0.0%	9	<input type="checkbox"/>
10	95	95	152	152	146	158	202	202	179	179	211	211	222	222	97	99	126	126	0.0%	10	<input type="checkbox"/>
11	91	91	152	152	166	166	202	202	184	184	211	211	222	222	97	131	126	126	0.0%	11	<input type="checkbox"/>
12	91	91	146	146	158	166	202	214	184	184	211	211	222	238	97	97	126	126	0.0%	12	<input type="checkbox"/>
13	91	91	150	152	158	158	214	214	184	184	211	211	222	222	97	97	-99	-99	11.1%	13	<input checked="" type="checkbox"/>
14	91	95	152	152	158	158	202	202	184	184	211	211	222	222	97	131	126	126	0.0%	14	<input type="checkbox"/>
15	91	91	146	152	158	166	202	214	184	184	211	211	222	222	97	131	126	126	0.0%	15	<input type="checkbox"/>
Missing Data	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	6.7%	6.7%		Missing Data	
Monomorphic	No		No		No		No		No		Monomorphic		No		No		Monomorphic			Monomorphic	
<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>				
	CNZ03		CNZ40		CNZ42		CnCir 119		CnCir 126		CnCir 147		CnCir 2		CnCir 206		CnCir 215				

Germplasm	CNZ03	CNZ03	CNZ40	CNZ40	CNZ42	CNZ42	CnCir 119	CnCir 119	CnCir 126	CnCir 126	CnCir 147	CnCir 147	CnCir 2	CnCir 2	CnCir 206	CnCir 206	CnCir 215	CnCir 215	Missing Data	Germplasm	<input type="checkbox"/>
1	95	91	152	152	158	158	202	202	179	184	211	211	222	222	99	131	126	126	0.0%	1	<input type="checkbox"/>
2	95	95	152	152	146	158	202	202	179	179	211	211	222	222	99	131	126	126	0.0%	2	<input type="checkbox"/>
3	95	95	152	152	146	166	202	202	179	184	211	211	222	224	97	99	126	126	0.0%	3	<input type="checkbox"/>
4	91	95	152	152	158	166	202	202	179	179	211	211	222	222	97	99	126	126	0.0%	4	<input type="checkbox"/>
5	91	95	146	152	146	166	202	202	179	184	211	211	222	224	99	131	126	126	0.0%	5	<input type="checkbox"/>
6	91	91	146	152	158	158	202	202	179	184	211	211	222	224	99	131	126	126	0.0%	6	<input type="checkbox"/>
7	91	95	146	152	146	166	202	202	179	184	211	211	222	224	99	131	126	126	0.0%	7	<input type="checkbox"/>
8	91	91	146	152	146	166	202	202	179	179	211	211	222	224	97	99	126	126	0.0%	8	<input type="checkbox"/>
9	95	95	152	152	146	166	202	202	179	184	211	211	222	222	97	131	126	126	0.0%	9	<input type="checkbox"/>
10	95	95	152	152	146	158	202	202	179	179	211	211	222	222	97	99	126	126	0.0%	10	<input type="checkbox"/>
11	91	91	152	152	166	166	202	202	184	184	211	211	222	222	97	131	126	126	0.0%	11	<input type="checkbox"/>
12	91	91	146	146	158	166	202	214	184	184	211	211	222	238	97	97	126	126	0.0%	12	<input type="checkbox"/>
13	91	91	150	152	158	158	214	214	184	184	211	211	222	222	97	97	-99	-99	11.1%	13	<input checked="" type="checkbox"/>
14	91	95	152	152	158	158	202	202	184	184	211	211	222	222	97	131	126	126	0.0%	14	<input type="checkbox"/>
15	91	91	146	152	158	166	202	214	184	184	211	211	222	222	97	131	126	126	0.0%	15	<input type="checkbox"/>
Missing Data	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	6.7%	6.7%		Missing Data	
Monomorphic	No		No		No		No		No		Monomorphic		No		No		Monomorphic			Monomorphic	
<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>		<input type="checkbox"/>				
	CNZ03		CNZ40		CNZ42		CnCir 119		CnCir 126		CnCir 147		CnCir 2		CnCir 206		CnCir 215				

Missing data

Monomorphic locus

Germplasm	CNZ03	CNZ03	CNZ40	CNZ40	CNZ42	CNZ42	CnCir 119	CnCir 119	CnCir 126	CnCir 126	CnCir 147	CnCir 147	CnCir 2	CnCir 2	CnCir 206	CnCir 206	CnCir 215	CnCir 215	Missing Data	Germplasm	
1	95	91	152	152	158	158	202	202	179	184	211	211	222	222	99	131	126	126	0.0%	1	<input type="checkbox"/>
2	95	95	152	152	146	158	202	202	179	179	211	211	222	222	99	131	126	126	0.0%	2	<input type="checkbox"/>
3	95	95	152	152	146	166	202	202	179	184	211	211	222	224	97	99	126	126	0.0%	3	<input type="checkbox"/>
4	91	95	152	152	158	166	202	202	179	179	211	211	222	222	97	99	126	126	0.0%	4	<input type="checkbox"/>
5	91	95	146	152	146	166	202	202	179	184	211	211	222	224	99	131	126	126	0.0%	5	<input type="checkbox"/>
6	91	91	146	152	158	158	202	202	179	184	211	211	222	224	99	131	126	126	0.0%	6	<input type="checkbox"/>
7	91	95	146	152	146	166	202	202	179	184	211	211	222	224	99	131	126	126	0.0%	7	<input type="checkbox"/>
8	91	91	146	152	146	166	202	202	179	179	211	211	222	224	97	99	126	126	0.0%	8	<input type="checkbox"/>
9	95	95	152	152	146	166	202	202	179	184	211	211	222	222	97	131	126	126	0.0%	9	<input type="checkbox"/>
10	95	95	152	152	146	158	202	202	179	179	211	211	222	222	97	99	126	126	0.0%	10	<input type="checkbox"/>
11	91	91	152	152	166	166	202	202	184	184	211	211	222	222	97	131	126	126	0.0%	11	<input type="checkbox"/>
12	91	91	146	146	158	166	202	214	184	184	211	211	222	238	97	97	126	126	0.0%	12	<input type="checkbox"/>
13	91	91	150	152	158	158	214	214	184	184	211	211	222	222	97	97	-99	-99	11.1%	13	<input checked="" type="checkbox"/>
14	91	95	152	152	158	158	202	202	184	184	211	211	222	222	97	131	126	126	0.0%	14	<input type="checkbox"/>
15	91	91	146	152	158	166	202	214	184	184	211	211	222	222	97	131	126	126	0.0%	15	<input type="checkbox"/>
Missing Data	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	6.7%	6.7%		Missing Data	
Monomorphic	No	No	No	No	No	No	No	No	No	Monomorphic	No	No	Monomorphic							Monomorphic	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
	CNZ03	CNZ40	CNZ42	CnCir 119	CnCir 126	CnCir 147	CnCir 2	CnCir 206	CnCir 215												

Users can exclude markers/individuals with a given level of missing data

Germplasm	CNZ03	CNZ03	CNZ40	CNZ40	CNZ42	CNZ42	CnCir 119	CnCir 119	CnCir 126	CnCir 126	CnCir 147	CnCir 147	CnCir 2	CnCir 2	CnCir 206	CnCir 206	CnCir 215	CnCir 215	Missing Data	Germplasm	
1	95	91	152	152	158	158	202	202	179	184	211	211	222	222	99	131	126	126	0.0%	1	<input type="checkbox"/>
2	95	95	152	152	146	158	202	202	179	179	211	211	222	222	99	131	126	126	0.0%	2	<input type="checkbox"/>
3	95	95	152	152	146	166	202	202	179	184	211	211	222	224	97	99	126	126	0.0%	3	<input type="checkbox"/>
4	91	95	152	152	158	166	202	202	179	179	211	211	222	222	97	99	126	126	0.0%	4	<input type="checkbox"/>
5	91	95	146	152	146	166	202	202	179	184	211	211	222	224	99	131	126	126	0.0%	5	<input type="checkbox"/>
6	91	91	146	152	158	158	202	202	179	184	211	211	222	224	99	131	126	126	0.0%	6	<input type="checkbox"/>
7	91	95	146	152	146	166	202	202	179	184	211	211	222	224	99	131	126	126	0.0%	7	<input type="checkbox"/>
8	91	91	146	152	146	166	202	202	179	179	211	211	222	224	97	99	126	126	0.0%	8	<input type="checkbox"/>
9	95	95	152	152	146	166	202	202	179	184	211	211	222	222	97	131	126	126	0.0%	9	<input type="checkbox"/>
10	95	95	152	152	146	158	202	202	179	179	211	211	222	222	97	99	126	126	0.0%	10	<input type="checkbox"/>
11	91	91	152	152	166	166	202	202	184	184	211	211	222	222	97	131	126	126	0.0%	11	<input type="checkbox"/>
12	91	91	146	146	158	166	202	214	184	184	211	211	222	238	97	97	126	126	0.0%	12	<input type="checkbox"/>
13	91	91	150	152	158	158	214	214	184	184	211	211	222	222	97	97	-99	-99	11.1%	13	<input checked="" type="checkbox"/>
14	91	95	152	152	158	158	202	202	184	184	211	211	222	222	97	131	126	126	0.0%	14	<input type="checkbox"/>
15	91	91	146	152	158	166	202	214	184	184	211	211	222	222	97	131	126	126	0.0%	15	<input type="checkbox"/>
Missing Data	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	6.7%	6.7%		Missing Data	
Monomorphic	No	No	No	No	No	No	No	No	No	Monomorphic	No	No	No	No	No	No	Monomorphic	Monomorphic		Monomorphic	

Total Missing Data : 4.0%

This query contains 1014 germplasms and 30 markers

delete checked germplasms and markers

reset

Save in DARwin format

All queries of this session

Change species and database

Change studies

New query

- Basic statistics including total level of missing data

- Export

In progress: allelic

frequencies

Combining 2+ different genotyping studies

- General Description
 - Users can select 2 or more studies and the application get the common markers and/or common germplasms (DNA samples) between selected studies.
 - The application checks the allele values for the common marker/germplasm combination between selected studies.
- Main options
 - Users can choose to combine using only common markers or using only common germplasms, or using common marker/germplasm combination.

Combining 2 or more different genotyping studies

Species: <ul style="list-style-type: none"><input type="checkbox"/> Check/Uncheck all<input type="checkbox"/> Barley<input type="checkbox"/> Chickpea<input type="checkbox"/> Coconut<input type="checkbox"/> Lentil<input type="checkbox"/> Rice<input type="checkbox"/> Sorghum<input type="checkbox"/> SugarCane	Database: <ul style="list-style-type: none"><input type="checkbox"/> Check/Uncheck all<input type="checkbox"/> GCP Central Repository<input type="checkbox"/> TropGene
<input type="button" value="Submit"/>	<input type="button" value="Reset"/>

- Combine All Result (This will create missing Data when markers and/or germplasm are not common)
- Combine Results using common Germplasm
- Combine Results using common Markers
- Combine Results using common Germplasm and Markers

Germplasm	CNZ03	CNZ03	MissingData	Germplasms	<input type="checkbox"/>
1	95 91 333 444	95 91 333 444	0.0%	1	<input type="checkbox"/>
Missing Data	0.0%	0.0%			
	Monomorphic				
<input type="checkbox"/>		<input type="checkbox"/>			
	CNZ03				

Total Missing Data : 0.0%

This query contains 1 germplasms and 1 markers

delete checked germplasms and markers

Save in DARwin format

All queries of this session

Change species and database

Change studies

New request

Inconsistencies between two studies are highlighted (orange, different allele values separated by pipe characters)

Order By :

- Date
- Ploidy
- Total Missing Data
- Number of Germplasms
- Number of Markers
- Number of Germplasms with Missing Data

Order now

Request 1

- Study: GCP_Sorghum_3400accessions_39SSR
- Created the 4/8/2007 at 12H 27min 33sec
- Ploidy: 2
- Total Missing Data: 3.1%
- Number of Markers: 14
- Number of Germplasms: 14
- Germplasms with missing data: 3



Request 2

- Study: gcpcr_file663
- Created the 4/8/2007 at 12H 28min 23sec
- Ploidy: 2
- Total Missing Data: 0.4%
- Number of Markers: 15
- Number of Germplasms: 15
- Germplasms with missing data: 1



show checked studies

Reset

Delete checked Requests

Save output datasets and corresponding queries.

Format Conversion

The application will export to different data format for easy input into the range of software packages presently used by geneticists.

- DARwin formats : VAR and DON files (<http://darwin.cirad.fr/>)
- *In progress*: Convert format, Structure format...

gpcr_file662num0 [VAR file](#) [DON file](#)

You can download Darwin software [here](#)

[All queries of this session](#) [Change species and database](#) [Change studies](#) [New query](#)

```
@DARwin 5.0 - ALLELIC - 2
6      12
N?     CNZ03  CNZ03  CNZ40  CNZ40  CNZ42  CNZ42  CnCir 119  CnCir 119  CnCir 126
1      89      89      134    138    162    164    208    210    179    184    223    231
2      89      89      148    148    162    162    208    216    184    184    211    225
3      89      89      134    148    162    166    212    216    184    184    223    223
4      89      89      130    148    162    162    214    216    184    184    211    227
5      89      89      130    146    162    162    214    214    184    184    223    229
6      89      93      130    138    162    162    208    214    184    184    211    229
```

```
@DARwin 5.0 - DON -
6      1
N?     GermplasmID
1      2486
2      2487
3      2488
4      2489
5      2490
6      2491
```

Product Delivery: GenDiversity tool

- Prototype available through the CIRAD intranet network
- A public internet access *in progress*:
<http://sat.cirad.fr/sat/gendiversity/servlet/ChoixInitiale>
- Code freely available on CropForge, part of the GCP Pantheon project : <http://cropforge.org/>
- User documentation (*in progress*)

Product Delivery: Data Sets Available

- Now
 - Genotyping data
 - GCP central repository data (excel files): currently 11 species available (uploaded)
 - TropGene : Musa, Coconut, Rice, Sorghum, Sugarcane
- *In progress*
 - Genotyping data
 - IMIS, IWIS, IRIS : Maize, Wheat, Rice
 - Germplasm data
 - GCP central repository data (excel files)
 - IMIS, IWIS, IRIS : Maize, Wheat, Rice