Agropolis Resource Centre for Crop Conservation, Adaptation and Diversity

Scientific project (2009-2014)

Final report

December 2014

Project supported by Agropolis Fondation (No ARCAD 0900-001)

www.arcad-project.org

Contents

A. INTRODUCTION

1. Context

The overall aim of ARCAD (Agropolis Resource Centre for Crop Conservation, Adaptation and Diversity) is to set up an open multi-function platform devoted to the assessment and improved use of plant agrobiodiversity in Mediterranean and tropical regions. It is being jointly implemented by CIRAD, INRA, IRD and Montpellier SupAgro and is supported by Agropolis Fondation and the Languedoc-Roussillon Region.

The ARCAD concept was developed in 2007-2008, but it is rooted in a much older overall ambition of research and higher education institutions (mainly INRA, CIRAD, IRD and Montpellier SupAgro) to give Montpellier-Languedoc Roussillon Region world-wide visibility in the field of agrobiodiversity, plant genetics and genomics.

The Languedoc-Roussillon Region proposed in 2007 to support the development of a genetic resource conservation facility to the tune of EUR 5 million in order to give more visibility to the wealth of genetic resource collections maintained in Montpellier. While the largely publicized launch of Global Seed Vault in Svalbard was certainly inspiring in this move, it became quickly obvious that this initiative, in addition to its physical facility component, had also to mobilize the related scientific research potential.

The ARCAD programme therefore includes 3 components: Infrastructure, Equipment, and Scientific project.

The scientific programme of ARCAD was initially funded for 4 years from 2009 by Agropolis Fondation to the tune of EUR 3 million. The project duration has been extended by one year and ended in June 2014. In this report, we describe the activities carried out and the main achievements of that scientific programme.

Overall objectives and project organizational structure

ARCAD's scientific agenda focuses on the study of the history and patterns of crop domestication and adaptation as well as on the analysis of key parameters underpinning adaptation and diversity, at various time scales, through studies of evolutionary genomics, population genetics and social sciences. This thematic focus, as well as the geographical focus on Africa and the Mediterranean Basin, was strongly encouraged during the maturing phase of the programme by Agropolis Fondation's Scientific Council.

Most research activities conducted by ARCAD reflect a change of scale compared to those conducted by the teams involved before ARCAD was launched. The 3 major research sub-projects SP1 to SP3 focus on *Population comparative genomics*, *Adaptation to climate change*, and *Cereals in Africa*, respectively. These research activities are complemented by methodological and technological sub-projects. The SP4 *Bioinformatics* and SP5 *Pangenomic study of diversity* were designed to support and extend the diversity analysis capacity of the ARCAD programme. The SP6 *DNA bank* and SP7 *Cryopreservation aimed at increasing the potential in biological resources conservation and management.* Another major objective of the ARCAD programme is to set up a demand-oriented capacity-building platform, based on the educational facilities provided by different universities in Montpellier, along with the development of specific training modules (SP8 *Training*). See **Figure 1**. About 2/3 of the budget were allocated to SP1, SP2 and SP3 while SP4 Bioinformatics was also given a high priority (budget: EUR 400K).

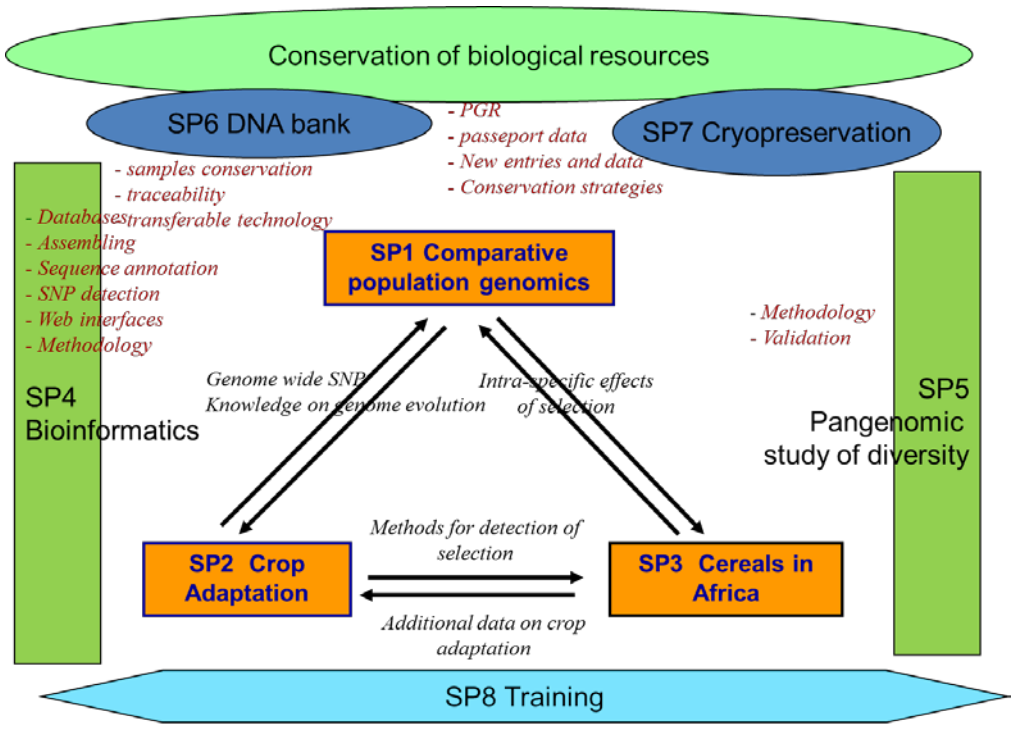


Figure 1. ARCAD organizational chart

About 80 permanent scientists were involved in the ARCAD scientific programme. They mostly belong to two research units, AGAP and DIADE and 3 research institutes (INRA,

CIRAD, and IRD). ARCAD also involved scientists belonging to other institutions in Montpellier (CNRS), as well as French and foreign partners (see **Annexe 1**).

The three main sub-projects were elaborated and have been carried out following different partnership and organisation models. While the SP1 aimed at designing and sharing a methodology across a set of teams as large as possible in order to develop a comparative study on the genomics of domestication, the SP2 brought together a few teams with a high expertise on the population genetics of adaptation. The SP3 mobilized a group of scientists in Montpellier with different backgrounds in crop diversity studies in order to consolidate a collective expertise on the in situ diversity of cereals in Africa. Because of its focus, it strongly involved South partners, especially in Guinea, Kenya and Morocco.

2. Governance and coordination mechanisms

The coordination team (JL Pham[1], project leader & JP Labouisse[2], project coordinator) has fulfilled several main functions (see **Annexe 2** for the record of activities):

- daily management of the project;
- internal and external communication;
- focal point of the ARCAD project, participation in various meetings and events;
- development of infrastructure and equipment funding requests;
- implementation of 'good practices' to access and exchange plant material.

Sub-projects were led by 1 to 4 leaders, depending on the size of the SP (**Table 1**). Each of the SPs had its own internal coordination mechanism.

A coordination committee met regularly since the inception of the project. It met 5 times a year on average and brings together the coordination team as well as the SP leaders. Management and scientific issues of collective interest were discussed.

For instance, when it was decided to strongly reshape the SP5 component[3], the coordination committee considered several options before choosing to focus on methodological development for the pangenomic study of diversity.

[1] Jean-Louis Pham is seconded to Agropolis Fondation from IRD (80% full time)

[2] Jean-Pierre Labouisse is seconded to Agropolis Fondation from CIRAD (50% full time)

[3] Initial SP5 project proposed to develop association genetics methods in allogamous Mediterranean and tropical plants i) to reconstruct haplotypes for more accurate estimation of LD, ii) to correct for structure and residual kinships in LD estimation, and iii) to infer population history from LD estimation (local or extended) .
Since the writing of ARCAD proposal, works have been performed by partners or other teams in this area. It has been shown by Rogers and Huff (2009) that the measurement of LD based on haplotypic data (r^2 haplotypic) is similar to LD based on genotypic data (r^2 genotypic). In addition, we have developed novel measures of LD that correct the bias due to population structure and relatedness (Mangin et al 2012, Heredity 108, 285-291). Finally, methods were being developed within SP1 in order to infer population history from LD estimation. These reasons led us to ask for modifying SP5 project .

Table 1. List of the co-leaders of sub-projects and work packages

Sub-Project (SP)	*Sub -project and work packageb (WP)*	*Surname, name*	*Affiliation*
Coordination	Programme leader	**PHAM Jean-Louis**	AGROPOLIS FONDATION
	Coordinator	**LABOUISSE Jean-Pierre**	AGROPOLIS FONDATION
SP1 - Comparative population genomics	SP1, WP2 & WP3	**DAVID Jacques**	UMR AGAP, SUPAGRO
	SP1, WP2 & WP4	**GLEMIN Sylvain**	UMR ISE-M, CNRS
	WP1	**SANTONI Sylvain**	UMR AGAP, INRA
	WP1	**RISTERUCCI Ange-Marie**	UMR AGAP, CIRAD
	WP1	**MORCILLO Fabienne**	UMR DIADE, IRD
	WP2	**GALTIER Nicolas**	UMR ISE-M, UM2
	WP5	**CHANTRET Nathalie**	UMR AGAP, INRA
	WP5	**DE KOCHKO Alexandre**	UMR DIADE, IRD
SP2- Adaptation to climate change	SP2, WP1 & WP3	**RONFORT Joelle**	UMR AGAP, INRA
	SP2, WP1	**VIGOUROUX Yves**	UMR-DIADE, IRD
	SP2, WP2 & WP3	**AHMADI Nour**	UPR AGAP, CIRAD
	SP2, WP2	**GAY Laurène**	UMR AGAP, INRA
SP3 - Cereals in Africa	SP3, WP4	**NOYER Jean-Louis**	UMR AGAP, CIRAD
	SP3, WP3	**GHESQUIERE Alain**	UMR DIADE, IRD
	SP3, WP2 & WP3	**ROUMET Pierre**	UMR AGAP, INRA
	SP3, WP2	**LECLERC Christian**	UMR AGAP, CIRAD
	WP2	**MULLER Marie-Hélène**	UMR AGAP, INRA
	WP1	**DEU Monique**	UMR AGAP, CIRAD
	WP1	**GOUESNARD Brigitte**	UMR AGAP,INRA
	WP4	**PHAM Jean-Louis**	AGROPOLIS FONDATION
	WP4	**BILLOT Claire**	UMR AGAP/CIRAD
SP4 - Bioinformatics	SP4	**RUIZ Manuel**	UMR AGAP, CIRAD
SP5 - Pangenomic study of diversity	SP5	**THIS Patrice**	UMR AGAP, INRA
	SP5	**SEGUIN Marc**	UMR AGAP, CIRAD
SP6 - DNA Bank	SP6, WP1 & WP2	**SANTONI Sylvain**	UMR AGAP, INRA
	SP6, WP1 & WP2	**RISTERUCCI Ange-Marie**	UMR AGAP, CIRAD
	SP6, WP3	**PROSPERI Jean-Marie**	UMR AGAP, INRA
	WP1 & WP2	**SABAU Xavier**	UMR AGAP, CIRAD
SP7 - Cryoconservation	SP7	**ENGELMANN Florent**	UMR DIADE, IRD
SP8- Training	SP8	**NOYER Jean-Louis**	UMR AGAP, CIRAD
	SP8	**PHAM Jean-Louis**	AGROPOLIS FONDATION

B. PRESENTATIONS AND ACHIEVEMENTS OF SUB-PROJECTS

SP1 - Comparative population genomics in wild and crop plants: a genome-wide phylogenic approach

1. Rationale

Domestication strongly impacted phenotypic and genomic evolution in crop species. Crop species typically exhibit lower genetic diversity than their wild ancestors, and may show dramatic phenotypic changes in their morphology, phenology and metabolism (Doebley et al. 2006). Understanding the domestication process is thus a key to crop breeding but also a unique opportunity to study rapid evolutionary processes on a short time scale. Differences among genomic patterns are still not adequately explained. For instance, in many crops, how many genes - and which - are involved in domestication and artificial selection is still unclear. Comparing the domestication process in a range of species, varying from ancient domesticated species (*Vitis, wheat, Sorghum*) to more recently cultivated species (*Coffea*, *cocoa*) should provide key information on the dynamics of adaptation and the correlated evolution of polymorphism patterns.

It is also especially important to compare molecular evolutionary patterns among species with contrasted life-history or ecological traits. Life-history or ecological traits may influence genome evolution through their effect on key population genetic parameters (effective size, recombination rates, and mutation rates). Genomic patterns may also vary among phylogenetically distant species because of specific molecular mechanisms such as recombination and repair mechanisms.

Knowing the molecular functions that are targeted by selection is also of interest to increase our understanding of adaptation. Thus, studying the evolution of gene families and relating it to expression data across lineages may help to identify which molecular functions play a key role in adaptation. Today, a comparative population genomic approach among many species is both essential and possible thanks to massive sequencing technologies.

2. Main objectives

First, we aimed to conduct comparative analyses of the effect of domestication on genome evolution in different crop species to:

- quantify the loss/recovery of diversity associated with domestication;
- test the hypothesis of a domestication cost due to an increased probability to fix weakly deleterious mutations;
- identify genes selectively involved in the domestication process;
- investigate variations in domestication patterns among crop species based on their life history, domestication depth, or phylogenetic position.

Second, we aimed to investigate the genomic selective patterns among angiosperm species and possible causes of variation using a comparative approach, taking into account the life-history traits and the genomic environment (e.g. GC-content) of the species to:

- quantify the whole genome level of selective constraints and the proportion of adaptive substitutions;
- investigate the causes of the phylogenetical variation of base composition (selection, GC-biased gene conversion, mutation);
- test the predictions of the effects of life-history traits (breeding systems, life span) on polymorphism within species and divergence among species.

Third, we aimed to investigate in more detail how genes functionally evolve in the different species to:

- identify general trends, in terms of gene or gene family content, along the angiosperm phylogeny and lineage-specific families, subfamilies or clades;
- compare selective constraints between genes from different functional categories, and genes belonging to subfamilies with different expansion dynamics;
- quantify selective constraints in gene families involved in specific metabolic networks.

This project aimed also to demonstrate that such investigation can be made on orphan species as well as on model species. A high-throughput RNAseq technology was retained for feasibility reasons, for a direct access to the coding part of the genome and to expression levels, and for possibilities of comparing the fate of some genes in different species. The Montpellier community, enlarged to Avignon, has acknowledged competences on tropical and Mediterranean crops and the management of their genetic resources. Using these resources, network and knowledge, the project proposed a joint action: 13 crops were finally recruited and funded by ARCAD : coffee, vine grape, diploid wheat (einkorn), yam, banana, palm tree, African rice (*O. glaberrima*), fonio, sorghum, cocoa, alfalfa, pearl millet and tomato. Two other species joined the project on their own funding for data production and benefited from the project's infrastructure: cotton and olive tree. As two outgroup species were used for phylogenetic purposes, 45 species were finally involved in the study.

Involving such a large group of scientists was a challenge for the use of standardized procedures, capacity building in bio-informatics and evolutionary analysis. The added value of this project is the comparative approach.

3. General approach

For each crop, genetic diversity was first investigated in 10 cultivated and 10 wild individuals and phylogenetic patterns were established using 2 outgroups species, whose genomes are somewhat divergent from the crop (**Figure SP1-1**).

Each out of the 15 crops was thus represented by a quadruplet (**Table SP1-1**). For each quadruplet, the domestication process was investigated by comparing gene polymorphism patterns between the wild and the domesticated species. Selective constraints, adaptive evolution, and GC-content evolution were investigated using classical frameworks, and both polymorphism and divergence data and recent tools developed by the ISEM partner. It is worth noting that model species that still have numerous genomic resources were also used in the comparison. Data were expected for several thousands of genes in each quadruplet. The SP1 project also aimed at using local resources and at developing up to date skills and capacity in NGS pre and post-processing.

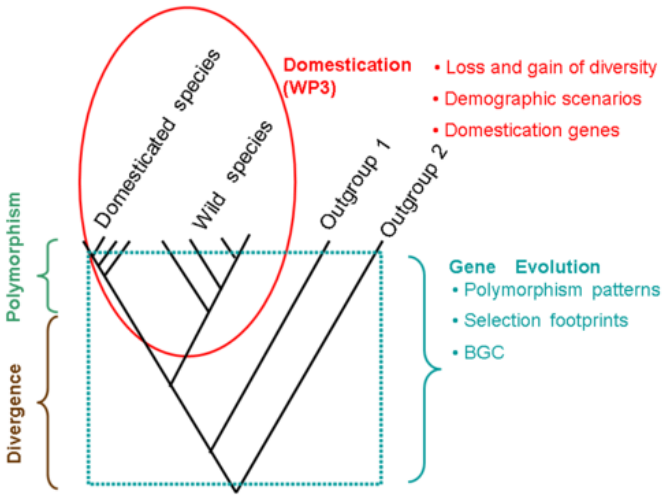


Figure SP1-1: Species sampling design

Table SP1-1. List of the crops studied in ARCAD SP1

Crop	Family	Domestication	Cultivated taxon	Life span	Mating system	Life form	outgroups
African rice	Poaceae	Old	*Oryza glaberrima*	Annual	selfing	herb	*O. sativa, O. meridionalis*
Banana	Musaceae	Old	*Musa acuminata*	perennial	outcrossing	herb	*M.balbisiana, M. becarii*
Cocoa	Malvaceae	Old	*Theobroma cacao*	perennial	mixed	tree	*T. speciosa, Herrania nitida*
Coffee	Rubiaceae	recent	*Coffea canephora*	perennial	outcrossing	tree	*Empogona ruandensis, Bertiera laxa*
Einkorn wheat	Poaceae	Old	*Triticum monococcum*	Annual	selfing	herb	*Eremopyrum bonaepartis, Taeniatherum caput-medusae*
Grapevine	Vitaceae	Old	*Vitis vinifera ssp. sativa*	perennial	outcrossing	vine	*V. romaneti, V. riparia*
Medicago	Fabaceae	Old	*Medicago sativa*	perennial	outcrossing	herb	*M. truncatula ,M. marina*
Oil palm	Arecaceae	?	*Elaeis guineensis*	perennial	outcrossing	tree	*Phoenix dactylifera, Mauritia flexuosa*
Pearl millet	Poaceae	Old	*Pennisetum glaucum*	Annual	outcrossing	herb	*P. polystachyion, P. alopecuroides*
Sorghum	Poaceae	Old	*Sorghum bicolor ssp bicolor*	Annual	selfing	herb	*S. brachypodium, Zea mays*
Tomato	Solanaceae	Old	*Solanum lycopersicum*	Annual	selfing	herb	*Capsicum annuum, Solanum melongena*
Yam	Dioscoreaceae	Old	*Dioscorea rotundata*	perennial	outcrossing	herb	*D. trifida, D. alata*
Fonio	Poaceae	Old ?	*Digitaria exilis*	Annual	selfing	herb	*D. longiflora, D. sanguinalis*
Cotton	Malvaceae	Old	*Gossypium hirsutum*	Annual	selfing	shrub	*G. barbadense, G. mustelinum, G. raymondii;G. herbaceum*
Olive tree	Oleaceae	Old	Olea europea	perennial	outcrossing	tree	*Phillyrea latifolia*

4. Activities carried out and main results

a) WP 1: Data acquisition

Collection and conservation of DNA samples

For some crops, collections were built from international genebanks (e.g., einkorn, tomato) and/or local collections (e.g. vine grape, African rice), while for others the project implied to launch specific collaborations and collections in the wild or in foreign countries (e.g., yam, coffee, banana). Samples were constituted by adding different organs, including leaves and flowering tissues. Standardization lead to produce almost all RNA samples, all cDNA libraries in the GE2POP AGAP laboratory (under the responsibility of S. Santoni) for their tagging and pooling before being sent to different sequencing platforms. Three hundred and thirty two individual cDNA libraries were produced.

The whole collection of the cDNA libraries has been transferred to the SP6 program (DNA Bank) and will constitute an valuable resource for the community.

Sequencing

At the beginning of the project the 454 Roche technology was used on the GENOTOUL platform but this technology was rapidly abandoned for the Solexa High-seq technology. For the latter, all data were obtained using a total of 35 lanes leading to a gross number of 13.3 billions reads of 75 to 100 bp. The main partner for the sequencing was the regional platform Montpellier Genomix monitored by the CNRS (http://www.mgx.cnrs.fr). All data are now available for the 15 species. These data were used for the SP1 project but were also made available for different scientist (not SP1 partners) for other purposes (SNP development, gene annotation, etc.).

Lab protocols for RNA preparation and Illumina libraries production and sequencing are given in **SP1-Annex1** (p.27).

b) WP 2: Sequences pre-treatment and database (with *Bioinformatics* sub-project SP4)

Data were then stored and analyzed by the SP4 research team in close interaction with a group of SP1 scientists and in collaboration with the *PopPhyl* ERC project of Nicolas Galtier (CNRS, ISEM). A specific SP1 methodological contribution was the development of a likelihood method for detecting paralogous genes, impairing the proper identification of intragenomic polymorphism. This method (called *Paraclean*) was implemented in R and C++, integrated in the genotype calling tool *Read2snp* developed by the *PopPhyl* group, and applied to a set of non-model animal species in the *PopPhyl* project (Gayral et al. 2013). A program (called *Homeosplitter*) was also set to disentangle the homeo alleles in allo polyploid species based on the expression bias with an application on durum wheat (Ranwez et al, 2013).

After the process, polymorphisms were called for all individuals and annotated as coding vs non coding, synonymous vs. non synonymous sites. Transcriptomes were assembled for all species (see SP4) and polymorphism data were investigated on all species (**Table SP1-2**).

Table SP1-2. Polymorphism data obtained on the focal species of the quadruplets. # wild and # cultiv. are the number of individuals finally kept for the analysis. Contigs are the number of independent reads assemblies obtained from the raw sequence (see WP4). SNP are the number of single nucleotide polymorphisms obtained in the wild and cultivated compartments of the different crops.

Espece	Group	# wild	Contig	SNP	# cultiv.	Contigs	SNP
Coffea canephora	Coffea	12	14581	148508	12	15722	49790
Dioscorea rotundata	Dioscorea	10	15981	139368	9	16708	80335
Eleais guineensis	Eleais	10	16975	54238	10	17220	51722
Musa acuminata	Musa	10	21422	286784	10	277146	274179
Olea europa	Olea	10	19549	209402	10	21168	168372
Oryza glaberrima	Oryza	9	12215	NA	7.000	12215	NA
Pennisetum glaucum	Pennisetum	10	15714	143297	10	17586	94058
Solanum lycopersicum	Solanum	10	18887	73778	10	16396	40189
Sorghum bicolor	Sorghum	10	11397	46298	10	13231	41266
Theobroma cacao	Theobroma	10	13901	83613	10	13899	76114
Triticum monococcum	Triticum	10	5847	11148	10	14066	20776
Vitis sylvestris	Vitis	12	14397	133697	12	16222	195603

The quantity of data obtained is satisfying for each species and beyond the expectation at the start of ARCAD. During the ARCAD program, sequencies technologies and offers evolved rapidly and the whole program benefited from it. On average, 99,300 SNP have been discovered in the cultivated individuals, with a large variation, from 20,776 for the diploid wheat (einkorn) to more than 274,000 for banana (*Musa acuminata*). This discrepancy could come of course from the relative diversity level of each species and also from a different sequencing period. The first to be sequenced (einkorn) had less coverage than the last sequenced (Coffea, Yam and Vitis)

The sequencing was also successful on the outgroups (data not shown). The ancestral state of the SNP is thus available for a large number of polymorphisms.

All data were publicly available through the ARCAD web site.

c) WP 3: Comparative population genomics of the domestication process

C.1 A pilot approach on African Rice, *O. glaberrima*.

The African rice was used as a pilot species to set up the data processing and the population genomic analyses (Nabholz et al. 2014).

A severe bottleneck

The African cultivated rice (*Oryza glaberrima*) was domesticated in West Africa 3000 years ago and is still cultivated in some areas. It is often used as a source of adaptation for *O. sativa* in African breeding programs. The study was carried out on more than 12,000 transcripts having data for 9 cultivated *O.glaberrima,* 10 wild current *O. barthii* as representative of the wild ancestor and one outgroup *O. meridionalis* individual. With a synonymous nucleotide diversity π_s = 0.0006 per site, *O. glaberrima* appears as the least genetically diverse crop grass ever documented. Using approximate Bayesian computation, we estimated that *O. glaberrima* experienced a very severe bottleneck during domestication since the loss of diversity is around 61%.

This demographic scenario almost fully accounts for the pattern of genetic diversity across *O. glaberrima* genome as we detected very few outliers regions where positive selection may have further impacted genetic diversity.

Cost of domestication

The strong demographic bottlenecks experienced by the initial crop populations are expected to promote the fixation of weakly deleterious mutations and to depreciate the fitness of the first domesticated plants. The large excess of derived nonsynonymous substitution that we detected suggests that the *O. glaberrima* population indeed suffered from the 'cost of domestication' (figure SP1-2).

As lowly expressed genes are expected to be less constrained, mutations on these genes should be less deleterious and more prone to fixation by genetic drift induced by the domestication process. We tested this hypothesis in the African rice by quantifying the excess of non synonymous mutations fixed in the crop compared to those fixed in the wild compartment according to the gene expression level. Those mutations are indeed expected to be mostly deleterious (or neutral) and usually removed by natural selection.

As predicted, we found a strong excess of fixation of deleterious mutations in the crop, especially for the lowly expressed genes (**Figure SP1-2**, Nabholz 2014).

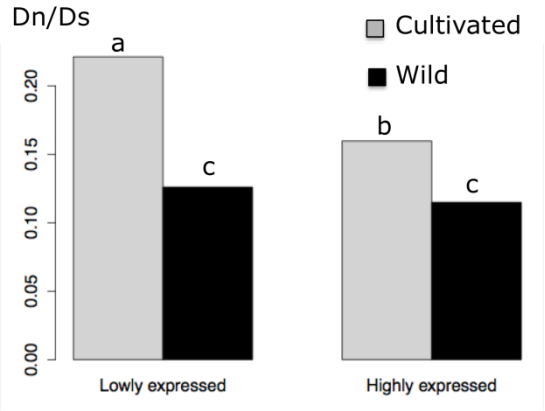


Figure SP1-2: Ratio of non-synonymous (Dn) over synonymous (Ds) fixed mutations in the cultivated and wild compartments of African Rice (*O. glaberrima)* for lowly and highly expressed genes. Different letters correspond to significantly different values.

This result suggests that useful alleles are available in the wild and shows that wild relatives could be used in breeding program not only for specific characters (e.g., resistances, rusticity) but also more globally to compensate the load accumulated during domestication.

Genomic influence

In addition, we used this genome-scale data set to demonstrate that (i) *O. barthii* genetic diversity is positively correlated with recombination rate and negatively with gene density, (ii) expression level is negatively correlated with evolutionary constraint, and (iii) one region on chromosome 5 (position 4–6 Mb) exhibits a clear signature of introgression with a yet unidentified *Oryza* species. This work represents the first genome-wide survey of the African rice genetic diversity and paves the way for further comparison between the African and the Asian rice, notably regarding the genetics underlying domestication traits.

Genomic regions under selection

To pick up genes under selection, we first built a simple demographic scenario using Approximate Bayesian Computations (ABC) assuming the genome evolved mostly neutrally. Genes for which diversity statistics significantly deviated from this scenario were candidate for adaptive evolution. ABC simulations (Nabholz et al. 2014) demonstrated that

the detection power of these outlier genes was dependent of bottleneck intensities. In African rice, this bottleneck (61%) was too strong and no clear outlier was detected.

Such investigations were also carried out on other species, Tomato, Sorghum and Vitis.

In tomato, the loss of diversity was less severe (around 33%). Chromosomic regions associated with a high loss of diversity were identified across the genome (fig. SP1-3) supported by Tajima's D (Sauvage & al. in prep). Around four percent of the studied genes showed footprints of positive or balancing selection. The analysis of the site frequency spectrum across the genome revealed at least two regions showing hard selective sweeps (chr 9 and chr 11 on **Figure SP1-3**).

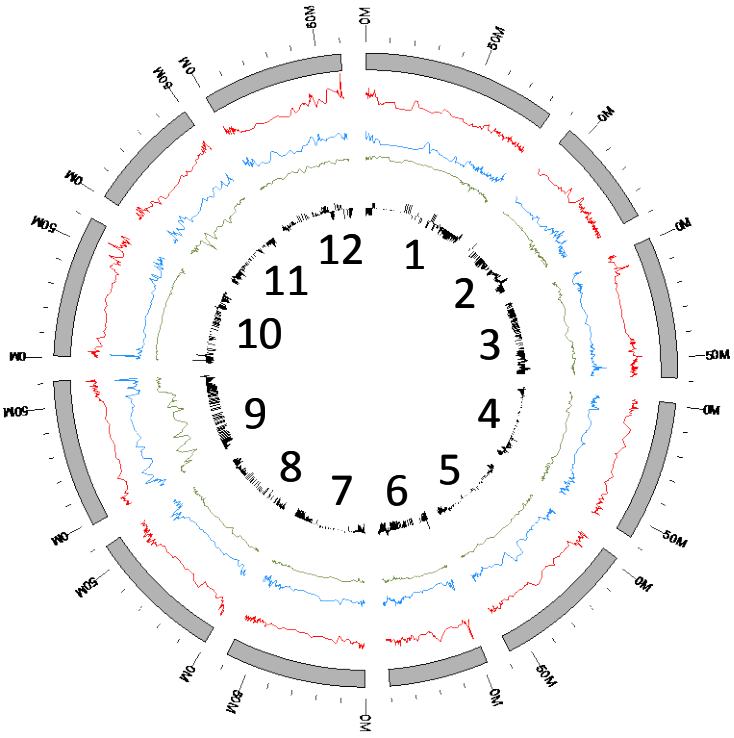


Figure SP1-3. Circular diagram depicting the genomic landscape of the twelve tomato chromosomes (Chr1 to Chr12 on a Mb scale). The nucleotide diversity variation (π) for the cultivated (red line) and wild accessions (blue line) as well as the patterns of reduction of nucleotide diversity (green line) and the values of the Tajima's D test (black line) for all polymorphic CDS are represented.

Identification and validation of candidate gene

In sorghum, the length of the vegetative cycle (i.e. duration between sowing and flowering) is one of the key drivers of plant adaptation to different climatic regions. Here, it was tested if genes already known to modify flowering time were associated with the transition from wild to cultivated genotypes. A domestication scenario was established and 15 outlier genes harbouring a significant higher loss of diversity were detected (**Figure SP1-4**). Two genes, Auxin Responsive Factors 3 and 4, were in the outliers. These two genes have been previously functionally identified for their potential role in the molecular mechanisms underlying flowering time. Using differentiation approach between the wild and the cultivated sorghum, another good candidate gene, whose allelic frequencies clearly opposes wild and cultivated has been identified. Mutant of this gene is known in *Arabidopsis thaliana* to heavily impact flower development (Pot & al, in prep).

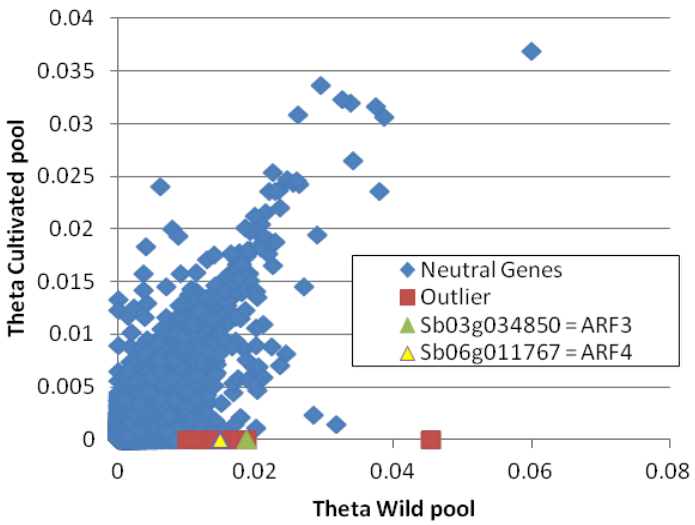


Figure SP1-4: Identification of genes harbouring an extensive loss of diversity within the cultivated pool in comparison to the wild pool of sorghum. Two genes corresponding to Auxin Responsive factors (ARF) are highlighted, they are involved on a homology based analysis in the molecular mechanisms underlying flowering time. *Theta* stands for Watterson's nucleotidic diversity parameter.

C2. Comparative genomics

The data set was finally assembled and consolidated in the spring of 2014.

This report provides the first conclusions, more data analysis are still needed.

Losses of genetic diversity

The raw comparison of nucleotide diversity between the wild and the cultivated forms is given in **Figure SP1-5**.

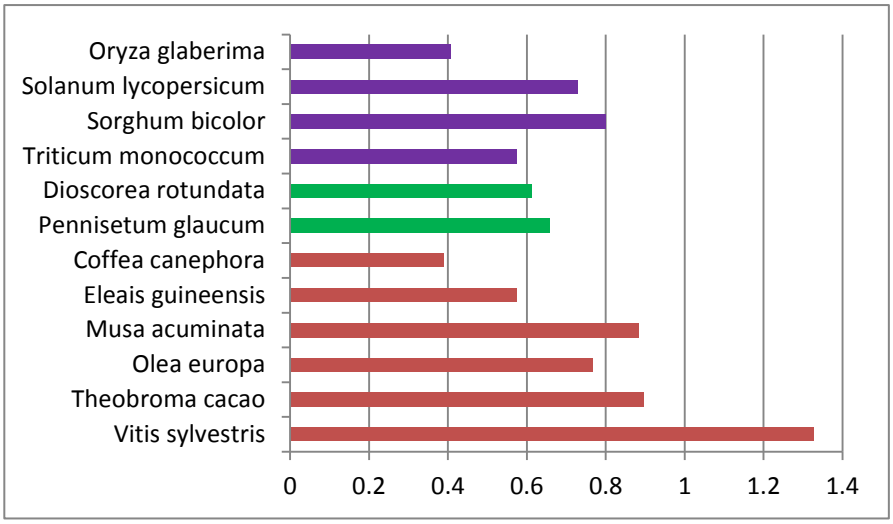


Figure SP1-5. Ratio of synonymous nucleotidic diversity between the cultivated and the wild compartments.

In red: perennial crops, in green annual and outcrossing crops, in blue: annual and selfing crops. Yam was considered here as an annual since its growth habit makes it more similar to an annual.

The situation is much contrasted between species and some results were not expected. Thanks to these data and preliminary structure analysis, we first examined if some of these situations could be explained by some caveats in the sample assemblies : for *Coffea canephora* the cultivated compartment (Conilon genotypes, which derived from a unique

sub-group of diversity) appeared extremely poor in diversity compared to the whole African wild compartment (2 main genetic groups and several sub-groups); *Eleais* cultivated accessions were apparently artificial hybrids between two different wild species. *Vitis* seems to have a more complex evolutionary pattern than previously expected. Recent litterature suggests that cultivated *Vitis* could have benefited from some introgressions from another species while the wild V*itis* has not. This could explain the fact that diversity of the cultivated compartment exceeds the wild one. As it was not possible to get flowering *Vitis* tissues from Eastern Europe accessions for RNA-seq, it is possible that our wild sample was not fully representative of its whole diversity, even if the discrepancy between the wild and the cultivated compartment is really weird.

For some other species, such as yam, the low level of knowledge about the direct ancestors of the cultivated forms may also have resulted in sampling in a too much restricted area in West Africa (mostly Benin) and our results suggest that a larger area has to be prospected to disclose the wild diversity of yam.

All these results demonstrate that ARCAD was able to point out that for many crops domestication is not fully understood and that in many cases the mere identification of the wild ancestor and its full geographical distribution remains ambiguous.

All teams will be able to use all these data to go further in their knowledge of their species.

As ARCAD planned a comparative study on some standardized scenarios, an original study was thus carried out on 3 annual species (einkorn; sorghum and *Pennisetum*) and 3 perenial species (cocoa, banana and olive tree) for which the samples fitted our requirements: good coverage of wild and cultivated distribution.

On these 6 species, we used an approach based on a rapid resolution of diffusion equations instead of Approximate Bayesian Computation we previously used for the pilot study in rice (**∂a∂i** method, Gutenkunst et al. 2009). First, outgroups were used to identify the likely ancestral state of each SNP. The spectrum of the frequencies of derived alleles was then computed for all the available SNP for the wild and cultivated groups. The joint distributions for Einkorn and Banana are presented in **Figure SP1-6**.

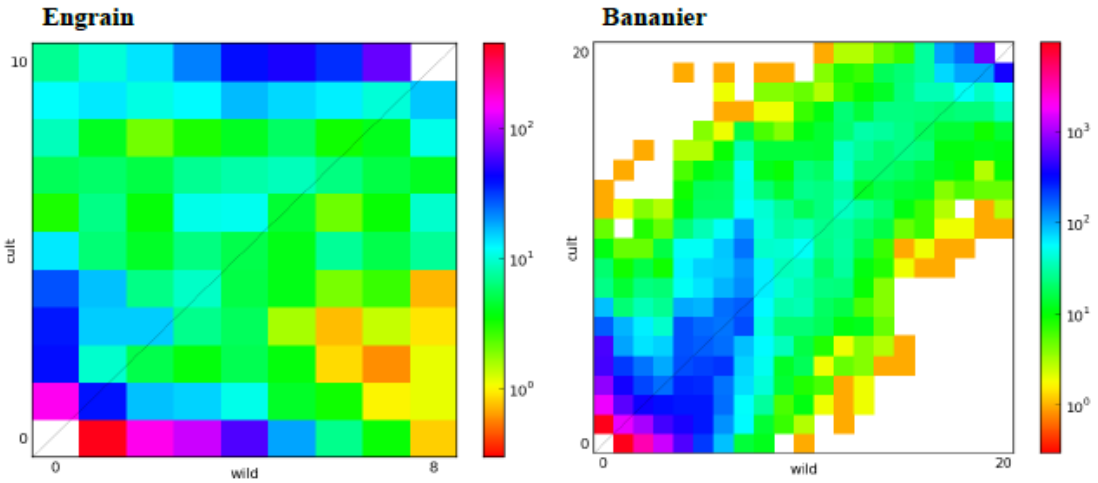


Figure SP1-6. Joint wild-cultivated distributions of derived allelic frequencies of Einkorn (left) and Banana (right). *The colour reflects the density of SNPs in every box. For example, blue colour means that a relatively high fraction of derived SNPs were found in the x and y coordinates indicating frequencies (frequencies are given according the number of copies found in the sample, e.g., from 0 to 8 in wild Einkorn (X axis).*

These joint distributions were used on einkorn to sort three different demographic scenarios and estimated the related parameters: bottleneck intensities, bottleneck durations and migration rates between wild and cultivated. All information is used at once (**Figure SP1-7**)

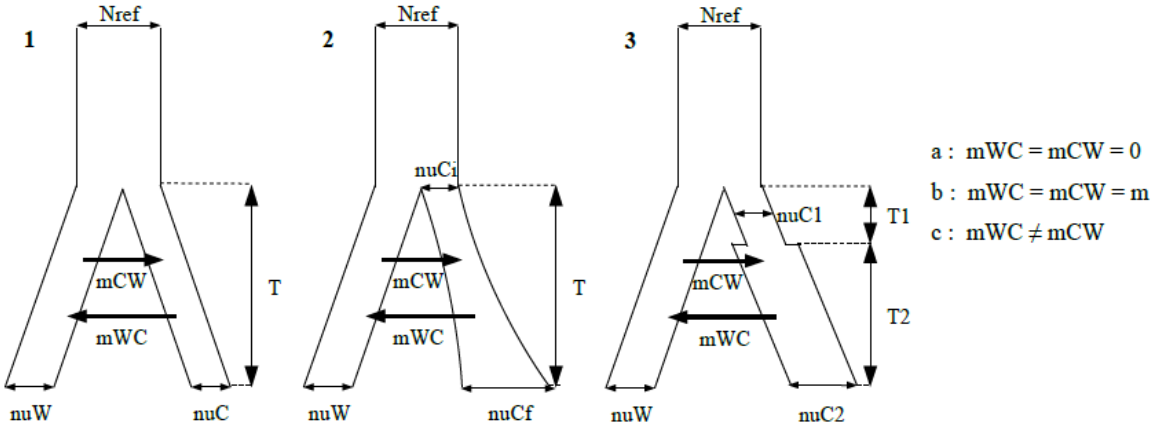


Figure SP1-7 Modeling domestication demographic scenarios foir einkorn
*Three models were tested: **1** simple split model with no migration, **2** split and expansion of the cultivated population, **3** split followed by a bottleneck. For all three models, three migration modalities were tested. **a** no migration, **b** symetrical migration & **c** asymetrical migration.*

In brief, results confirmed that annual plants (at least einkorn and sorghum) experienced strong bottlenecks and accumulated a severe domestication cost. All other species, except bananas also had a slight increase in their rate of fixation of non synonymous mutation in the cultivated compartment compared to the wild one (domestication cost) but this difference were not significant. Banana is puzzling. It suggests clearly that it should follow a specific feature which has now to be modelized correctly with a deeper analysis and discussion with banana specialists.

Another important feature was clear on einkorn. Modelling migration during and post domestication was clearly improving greatly the fit of the model to the data. This is an important contribution of the new **∂a∂i** method. It clearly suggests that cultivated forms should have regularly incorporated new alleles from the wild for a long period after domestication. It could also suggest that the wild form could have been shaped by recurrent gene flows from the cultivated compartment.

d) WP4: Life history traits and genome evolution

Beyond domestication, it is increasingly recognized that life history traits may shape genomic patterns of within species polymorphism and between species divergence. The large range of life history traits and the availability of wild species in the project allow exploring the link between species biology and genomic patterns.

We started this approach by comparing genomic patterns between species with contrasted mating systems. When compared to outcrossing, selfing is expected to reduce the level of polymorphism and the efficiency of selection. Moreover, selfing also affects molecular process such as GC-biased gene conversion (gBGC), a recombination-associated process that favours G and C over A and T bases during meiosis at heterozygous locus. This mechanism has been proposed to be a major determinant of GC-content in many species,

including plants and especially grasses as we showed in a large preliminary study (Serres-Giardi et al. 2012). Because selfing reduces heterozygosity it should also reduces the intensity of GC-biased gene conversion.

A comparison between wild pearl millet (outcrosser) and wild African rice (selfer) is in agreement with the theoretical prediction (**Table SP1-3**): polymorphism, selection efficiency (inferred by the π_N/π_S ratio), and GC-biased gene conversion intensity are lower in African rice than in pearl millet. Of course, for the sake of comparative power, more species need to be added in such comparisons. The species in ARCAD SP1 are well balanced for different life traits, especially for mating system.

Table SP1-3: Comparison of some genomic patterns between African rice and pearl millet.

	π_S (per kb)	π_N / π_S	GC-biased gene conversion*
Pearl millet (outcrosser)	7.13	0.17	> 0
African rice (selfer)	1.54	0.24	N.S.

*: the intensity of GC-biased gene conversion was estimated through the frequency spectrum of mutation using a method developed by ISEM partner (Muyle et al. 2011)

In flowering plants, GC content at third codon positions (GC3) strongly vary within and between genomes, some exhibiting peculiar patterns. What causes these patterns is still poorly known and difficult to disentangle. Among possible evolutionary forces, selection favoring synonymous codons which optimize translation (Selection for the codon usage SCU) and GC-biased gene conversion (gBGC), a recombination associated mechanism which favors the fixation of G and C alleles, are known to be active in some groups of plants, especially monocots. This point is strongly debated in the international scientific community.

We developed several nested population genetic models to test for the occurrence and estimate the intensity of GC-biased gene conversion and/or selection on codon usage.

To disentangle gBGC and SCU, we used the expression patterns provided by the ARCAD data. Following the assumptions that gBGC strength is expected to increase with GCcontent and SCU strength is expected to increase with gene expression, we separated our SNPs into 4 independent classes (based on gene expression and GC content) and estimated gBGC and SCU strength in each class (**Figure SP1-8**).

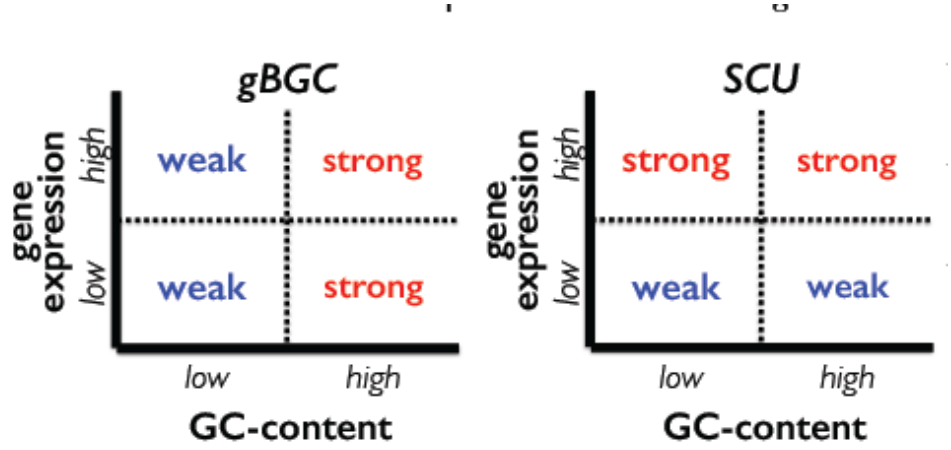


Figure SP1-8. Expected variation of strength of gBGC and SCU in different GC-content and gene expression.

We illustrated this method for banana (**Table SP1-4**). Here, gBGC strength increases with GC-content, but SCU strength does not increase with gene expression. So, in the banana genome, gBGC is active but not SCU.

Table SP1-4. gBGC and SCU strength estimations in banana. *p-value < 0.05.

gBGC			
gene expression	high	0.3199	0.3414*
	low	0.0292	0.7973*
		low	*high*
		GC-content	

SCU			
gene expression	high	0.1740	0.3912
	low	-0.1389	0.8718*
		low	*high*
		GC-content	

Applied to the other ARCAD species (**Table SP1-5**), the results suggest that:

- No evidence for selection on codon usage acting in any of the species in this study
- Signatures of gBGC in GC-rich Monocots (Grasses, palm tree, banana) but also in some GCpoor Eudicots
- According to theoretical predictions, lack of gBGC in selfing grass species (deficit in heterozygote positions

Table SP1-5: Estimation of SCU and gBGC strength in other species

Common name	Species name	Group	Mating type	GC3	SCU	gBGC
Robusta coffea	***Coffea canephora***	**Eudicot**	**Outcrossing**	**0.42**	**No**	**Yes**
Olive tree	*Olea europaea*	Eudicot	Outcrossing	0.42	No	No
Cocoa tree	***Theobroma cacao***	**Eudicot**	**Outcrossing**	**0.42**	**No**	**Yes**
Grape vine	*Vitis vinifera*	Eudicot	Outcrossing	0.44	No	No
Tomato plant	*Solanum lycopersicum*	Eudicot	Selfing	0.38	No	No
African rice	*Oryza glaberrima*	Monocot	Selfing	0.56	No	No
Einkorn wheat	*Triticum monococcum*	Monocot	Selfing	0.48	No	No
Yam	*Dioscorea rotundata*	Monocot	Outcrossing	0.46	No	No
African oil palm tree	***Eleais guineensis***	**Monocot**	**Outcrossing**	**0.49**	**No**	**Yes**
Wild banana	***Musa acuminata***	**Monocot**	**Outcrossing**	**0.52**	**No**	**Yes**
Pearl millet	***Pennisetum glaucum***	**Monocot**	**Outcrossing**	**0.53**	**No**	**Yes**

ARCAD results provided a unique window into neutral polymorphism and underline the importance of meiotic recombination and neutral forces like gBGC in shaping polymorphism and GC-content in land plants. Our results showed that GC-biased gene conversion is the major force acting on GC-content evolution in land plants, suggesting that this neutral process is widespread.

e) WP 5: Comparative functional genomics

Functional diversification is known to be partially promoted by gene duplication (Lynch, 2007). Duplicated gene copies may be retained and preserved on the long-term because of adaptive functional changes. The retention rate (*i.e.* the proportion of duplicated genes that are maintained in genomes) varies according to several factors and consequently extreme discrepancies in duplication/retention rates between taxa are observed in gene families phylogenies.

To study if adaptation is involved in the maintenance, and possible lineage specific expansion, of duplicated genes, seeking for positive (Darwinian) selection footprints in plant genome data is a possibility. The codon evolution models, *i.e.* analyzing non-synonymous on synonymous mutation rate (ω=dN/dS) sound appropriate to address this question at the genome scale of several angiosperm species whose genomes are completely sequenced. The idea is to test if positive selection can be observed more frequently in lineage specific duplicated genes than in non-duplicated genes.

Using *GreenPhylDB*, a database containing plant protein families and phylogenies for several completely sequenced plant genomes, ultraparalog (= only related by duplication; UP) and ortholog (=only related by speciation; OT) clusters were identified (http://www.greenphyl.org, Rouard et al. 2010) (**Figure SP1-9**). Five monocot and five dicot genomes were retained for their annotation quality: *Musa acuminata*, *Oryza sativa* subsp. *japonica*, *Brachypodium distachyon*, *Zea mays*, *Sorghum bicolor*, *Vitis vinifera*, *Arabidopsis thaliana*, *Populus trichocarpa*, *Glycine max*, and *Medicago truncatula*, most of them being in the ARCAD SP1 project.

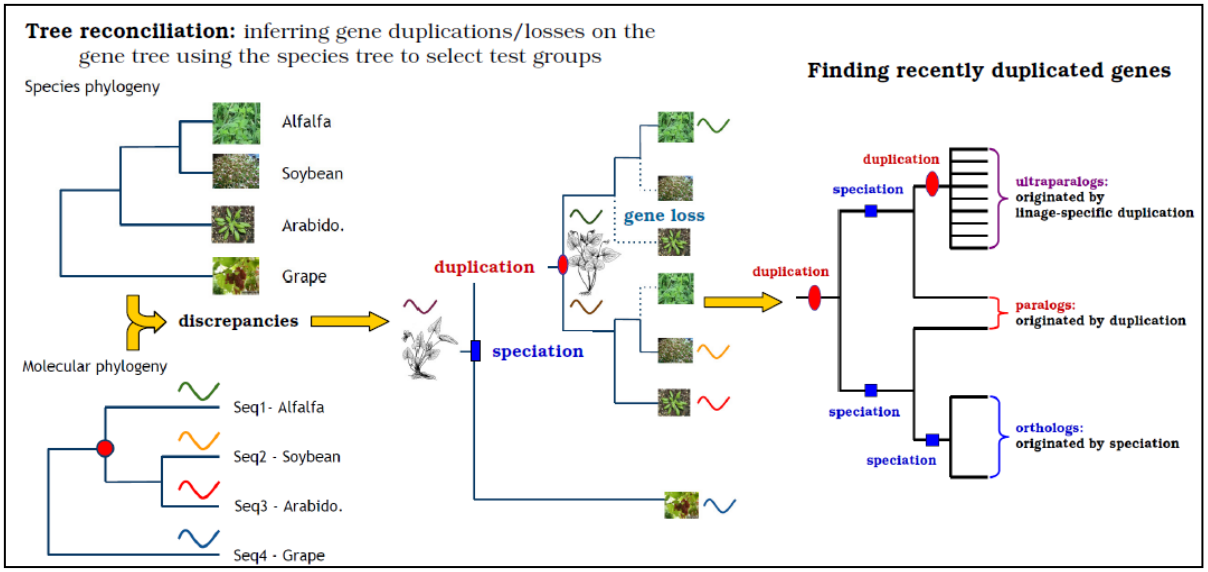


Figure SP1-9. Tree reconciliation and selection of ultraparalog (UP) and ortholog clusters (OT).

Selection was detected by computing ω either among sites or branches of UP and OT phylogenetic sub-trees. All these steps were performed by homemade scripts (mainly in java and python) using especially *Egglib* python package (DeMita & Siol 2012). On average, 6.37% of UP clusters show evidence for positive selection while only 0.41% of

the OT clusters (15 times less than UP) have a ω value indicating positive selection acting at codon level (**Table SP1-6**).

Table SP1-6: Number of ultraparalogous clusters under positive selection in different angiosperms with complete sequenced genome. OT: Ortholog clusters.

Species	Clusters used in Final Analysis	Clusters under selection (%)
M. acuminate	186	5 (2.7)
O. sativa	236	15 (6.4)
B. distachyon	98	3 (3.1)
Z. mays	359	13 (3.6)
S. bicolor	175	8 (4.6)
V. vinifera	208	10 (4.8)
A. thaliana	217	18 (8.2)
P. trichocarpa	264	39 (14.8)
G. max	245	8 (3.3)
M. truncatula	525	36 (6.9)
Total	**2,513**	**155 (6.17)**
OT	485	2 (0.41)

The analysis of ω among UP and OT phylogenetic trees branches showed that a significantly higher proportion of branches harbours a ω value above 1 in UP (11.64) than in OT (0.22) (**Table SP1-7**).

Table SP1-7: Results of branch analysis

	Orthologs (OT)	Paralogs (UP)
Nb branches analysed	4,154	23,976
Nb branches with ω>1 (%)	9 (0.22)	2,791 (11.64)
Mean ω	0.28 ±0.16	0.59 ±0.46

All together, these results published in BMC Plant Biology (2014), show that lineage specific duplicated genes are a much more important substrate for positive selection to act on than single-copy genes. This is – to our knowledge – the first genome-scale study to empirically demonstrate that duplicated genes fuel adaptation in angiosperms.

The developed pipe-line has been applied to two candidate gene families: the very large 'LRR-RLK' gene family (coll. A. Dievart & J.F. Dufayard, UMR AGAP) and the A1b albumin family which are entomotoxic molecules involved in insecticide reactions in legume species (coll. Y. Rahbé, INRA INSA-Lyon, UMR BFII).

It is also available for other species partners to evaluate the role of adaptation in the evolution of their favourite gene family.

5. Conclusions about scientific results

This collaborative project gathered 15 teams of plant scientists, from different research institutions (CIRAD, INRA, IRD), from different research units (UMR AGAP, UMR DIADE, UR GAFL), associated with a CNRS unit (UMR ISEM) specialized in molecular evolution. It produced a unique dataset to fulfil an ambitious challenge about crop evolution, domestication and to bring new knowledge about the evolution of plant genomes. This endeavour started by assembling original germplasm samples and mobilized up-to-date biotechnologies to produce numerous cDNA resources and sequences.

We produced much more data and for more species than initially planned. They are now available and could be accessed publicly by the international community. More than 13 billions sequences (the equivalent of more than 400 human genomes) have been produced, stored and processed. An important contribution of ARCAD SP4 sub-project and strong collaborations with the CNRS ERC *PopPhyl* initiative allow the development of efficient and powerful pipe-lines for shaping the raw data for evolutionary analysis purposes.

The task was challenging and it took time to get properly assembled data; sometimes on species that had no reference genomes. Many changes and pipe lines have been tested and developed (see SP4). This echoes the international effort and activities around RNAseq approach for *de novo* assembled genomes. All polymorphism data has been made available in 2014 to the teams. The analysis of this huge data set started with an unexpected delay and if some evolutionary trends are now really established, the publications of the scientific advance with high quality standards will still require time.

Post doc students hired in ARCAD all performed a very efficient job by pursuing methodological and innovative approaches. For two of them (B. Nabholz and Iris Fisher) they published their results. The results about SCU and gBGC are really innovative and should be soon published by Yves Clément, the last post doc hired.

In the master (M2) study of Alice Theisen (ISEM-AGAP), a very innovative comparative study of 6 species was carried out and revealed extremely important results about the comparison of annuals and perenials. The cost of domestication and migration are important biological features that will have to be investigated on more species.

All teams have also now data to develop their own research on their favourite crop (specific domestication studies, genetic structuration, diversity content, functional diversity) or tools for their specific purposes (SNP development, genome annotation, evolutionary studies, gene targeting).

Some (still few) species benefited from the first methodological works: African rice, sorghum and tomato mostly. Other teams also started to work on cleaned data (vitis, olive tree, pearl millet). These results were not mentioned in this document. The first investigations demonstrated the quality of the data and its richness. Previous known results on bottlenecks due to domestication, obtained on smaller number of genes were confirmed and new ones were produced: domestication cost in rice, detection of genomic regions and genes in sorghum and tomato. The comparative approach already produced significant results on the evolution of multiple copy gene family and there are premises on the impact of life history traits on genome evolution and the interplay with domestication. All these results are highly original and are assessed on a real statistical power thanks to a genome wide approach. Publications are in preparation.

Some work remains to be carried out to exploit properly these data sets, but capacities were built in the ARCAD community for such analysis.

6. Partnership

Building a scientific community

The partnership developed within the SP1 community was essential since it was a challenge to assemble so many people and skills. Access to the resources, standardisation of the biotech and bio-informatics procedures, development of methodological approaches sustained by specific recruitments of technicians, engineers and post doc scientists was a key for the success of SP1. The sample preparation and data production were ensured by a unique technological team. This permitted to scientists working on orphan species (*e.g.,* fonio) or with no specific facilities (*e.g.,* yam, olive tree) to access to up-to-date biotechnologies and to get their data rapidly and properly. With a similar efficiency, SP4 people (bioinformatics) in contact with the *PopPhyl* project developed standardized bio-informatics pipe lines in close interaction with a small group of SP1 researcher that will now be used by the community to establish *de novo* transcriptomes and the related polymorphism tables. It therefore had a strong impact on the handling of New Generation Sequencing by the Montpellier community.

SP1 organized a one-week seminar in 2011 in common with SP4 to disseminate tools and concepts. This was really appreciated and efficient to build capacities for the use of large amount of sequence data in the SP1 community.

A number of seminar, discussion and collaboration have been set up within the SP1 community and the associated SP4. Percolation of approaches, ideas and concepts is one of the great successes of SP1.

The mastering of NGS data production and analysis in ARCAD also had large repercussions on the proposal of new research projects (see beyond) and it is now acknowledged in the regional community; national and international collaborations were set with different teams to prepare samples and produce the data in collaboration with the MGX regional platform (Birc Aarhrus, CNRS Lille, INRA CBGP, ENS Lyon, Univ Ancona, CRA Foggia). Infrastructure developed in ARCAD SP1 is especially efficient in this regard.

7. Visibility, Attractiveness, national and international positioning

The data produced in SP1 allowed to the different plant teams to develop a number of collaborations at the national and international level.

- Tomato data was used within an international consortium to produce an SNP array (Sim et al, 2012). Identically sorghum teams took part in the project entitled « Sorghum polymorphism and divergence » funded by the JGI and coordinated by A.H. Paterson (Georgia University, USA).
- ARCAD RNA-seq data for grapevine have been presented at the third year Annual Meeting of the Cost Action FA1003: “East-West Collaboration for Grapevine Diversity Exploration and Mobilization of Adaptive Traits for Breeding” (Bucarest, 10-11 June 2013). The AGAP-DAAV team is leader of the work-package

"Genotyping and phenotyping methodologies". Presenting this rich dataset allowed the DAAV team to reinforce its leadership and to start new collaborations with the COST partners.

- INRA Lusignan used *M. sativa* data for international discussion about the settings of an alfalfa initiative. Data were also made available to different partners in coffee (Brasilian partners), banana (CIRAD Guadeloupe) and eggplant (outgroup species of tomato) for developing a linkage map in collaboration with CIRAD (UR Genetom , J. Dintinger) as well as developing a SNP genotyping array in pepper (outgroup species of Tomato).
- For durum wheat, the wheat group participated to the submission of a European project within the 7th KBBE program on durum wheat genetic diversity and domestication (project RECIPE, p.i. N. Stein IPK Gatersleben, Germany). This project was unfortunately not funded.
- Strong collaborations were also build in 2014 with the Pr. Roberto Papa (University Ancona, Italy) for the preparation of 144 cDNA banks covering wild and domesticated individuals of durum wheat, their sequencing and bio informatics treatments. Data analysis and publication of this data will be made in common with AGAP-GE2POP. ARCAD expertise was clearly the trigger of this collaboration.
- Eggplant data, an outgroup for Tomato, permitted to establish a collaboration with Anne Frary's lab at the University of Izmir, Turkey as part of a project of genetic mapping and the detection of QTLs for resistance to Ralstonia (ongoing PhD).
- For olive tree, a project has been funded by Agropolis Fondation (OliveMed 'Linking genes under domestication to phenotype traits in the Mediterranean olive tree: towards sustainable management by building a network of phenotyping platforms for association mapping studies'). This project involves Spain, Turkey and Morocco. ARCAD SP1 data were the back bone of the project. A post-doc scientist (18 months) has been recruited in 2014. Collaboration has been set up with the Laboratoire de Génétique et Évolution des Populations Végétales, CNRS-Univ Lilles to study sporophytic non compatibility in olive. An *ANR Blanc* has been submitted (PESSIOL: 4 UMR (AGAP, CEFE, UMR 8198 Lilles et UMR 5174 EDB Toulouse, et two Italian partners.

Leverage effect

SP1 clearly permitted the construction of new projects since it provided skills and methods encouraging the use of NGS in evolutionary biology and a enhance efficiency in the use of wild and exotic germplasm for breeding.

Projects:

ARCAD SP1 experience and data was beneficial for new projects : *ANR Blanc* on the evolution of mating system (ANR Trans: p.i. S. Glémin), CropDL a project funded by INRA within a meta program on genomic selection and an other innovative project EPO funded by INRA in durum wheat to test genotyping by sequencing (GBS) using RNA seq.

Based on the results of the grapevine RNA-seq data obtained within the ARCAD project, a multidisciplinary project involving archaeobotanists, geneticists, statisticians and the private sector has been proposed for funding at the Agropolis « Open Science » call for 2014, category "full projects". Title of the project: Historical genetics of grape domestication. SP1 clearly permitted the construction of such projects since it provided

skills and methods encouraging the use of NGS in evolutionary biology and a enhance efficiency in the use of wild and exotic germplasm for breeding.

Avignon-URGAFL was successful in a EU FP7-CIG Career Integration Grant - Bourse Marie-Curie (#PCIG10-GA-2011-304164- 4 ans 2012-2016, 100 000€) for the project: Molecular evolution of the Solanaceae: micro and macroevolutionary processes linked to domestication. SP1 clearly permitted the construction of such projects since it provided skills and methods encouraging the use of NGS in evolutionary biology and a enhance efficiency in the use of wild and exotic germplasm for breeding.

Human resources and post docs

V. Ranwez, a newly recruited professor of Evolutionary genomics and bio-informatics in Montpellier Supagro joined the ARCAD SP1 community. With N. Chantret (SP1 WP 5 leader) they obtained one year post-doc grant funded by Montpellier Supagro.

Iris Fischer (post-doc ARCAD from April 2012 to June 2014) recently obtained a DFG (Deutche Forschungsgemeinschaft) postdoctoral fellowship to carry out research in the GE2pop team for two years starting July 2014.

Jacques Dainat: one year post-doc grant funded by Montpellier Supagro (January to December 2013) on WP5

Permanent recruitment of ARCAD temporary collaborators

Obviously ARCAD had a positive effect for the recruitment in permanent position and success in post doc contracts of its temporary collaborators. The capacity in producing, shaping and treating the data is worthy to get permanent positions either in public institutions or private companies. The skills challenged in SP1 are highly valuable in private/public research.

8. Training

SP1 hosted 5 Master2 interns in its different labs who received a high level training in bio informatics, evolutionary and molecular genetics. They all got good marks after the defence of their master thesis.

A week of training was held from May 9th to 13th 2011 at Supagro, Montpellier. This training focused on the analysis of polymorphism data generated in SP1 by sequencing technologies. The whole process was presented, from production data analysis, methodological and theoretical bases. Training was a based on intensive practical approaches on computers. Twenty scientits, involved mostly in ARCAD attended the training. Training was realized mostly by SP1 leaders and the ARCAD post docs (B. Nabholz and S. De Mita)

It helped the SP1 partners to feel comfortable with simple population genomics investigation: nucleotide diversity computation and interpretation, simple domestication simulation and testing, identification of outlier locus, candidate for adaptation.

9. Forthcoming activities

ARCAD data SP1 are very rich and their mining is far from being completed. At least two important publications are in preparation (comparative domestication, impact of life history traits) and more could be prepared by the crop teams. To enhance dissemination of the results, participations to international congresses have to be planned.

Another activity will also be to think about the use of the unravelled diversity patterns for building new resources for broadening genetic basis of crops and then promote their resilience and adaptation. Diversiy loss and domestication costs for annuals claims for the need of base broadening and pre-breeding programs based on evolutionary perspective (domestication cost) and tools (sequencing and gene annotation, expression determinism). Theory and methodology are indeed eagerly needed by breeders and farmers communities and the knowledge produced by the SP1 will surely modify our perception of what diversity means at the genome level when measured on genes at the molecular scale.

10. Future research areas

- A challenge will be to increase the species sampling for monocots as a model for comparative studies
- Among the genomic traits that could be taken into account: Genome size and their rapid evolution, influence of transposons and ETs
- Diversity in gene expression and the link on functional effects
- As in human cohorts, using NGS in order to have an idea of the mutation rate in wild and crop genomes: what are the respective roles of standing variation and new mutations in adaptation (see SP2)?
- Cumulate more data on recombination impact and its relationships with other key parameters (effective size, selection intensity)
- In ARCAD, we focused on genes. Promoters and regulation domain polymorphism is still relatively unknown. Are these areas very active in the creation of genotypic variance?
- Promote pre-breeding material and associated methodologies for a better of genetic diversity. Resources produced by ARCAD SP1 permit to think about different strategies to use the disclosed and forgotten diversity in the wild: elaborating base broadened population possibly monitored with molecular information (effective size, mating system, LD evolution, detection of adaptation).
- Think about the conservation biology especially for in situ programs for wild progenitors and traditional landraces: collaboration with interaction avec SP2, SP3, SP5.
- Thanks to the progress in sequencing and the decrease of the costs, samples could be enlarged and enriched

References

Buckler et al. 2001, Molecular Diversity, Structure and Domestication of Grasses. Genetical Research (77) , 03 :213-218.

De Mita S, Siol M. 2012. EggLib: processing, analysis and simulation tools for population genetics and genomics. *BMC Genetics*. 13:27. doi:10.1186/1471-2156-13-27

Doebley J, Gaut BS, Smith BD. 2006. The molecular genetics of crop domestication. *Cell* 127:1309-1321.

Gayral P, Melo-Ferreira J, Glémin S, Bierne N et al 2013. Reference-free population genomics from next- generation transcriptome data and the vertebrate– invertebrate gap *PLoS Genetics* 9(4): e1003457.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. PLoS Genet. 5(10):

Lynch, M 2007. The origin of genome architecture. Sinauer, Sunderland

Muyle, A., L. Serres-Giardi, A. Ressayre, J.S. Escobar, S. Glémin**.** 2011 GC-biased gene conversion and selection affect GC-content in the Oryza genus. *Mol. Biol. Evol.* 28(9):2695-2706.

Rouard M et al. 2010., GreenPhylDB v2.0: comparative and functional genomics in plants. *Nucl. Acids Res.* first published online September 22, 2010 doi:10.1093/nar/gkq811

Serres-Giardi L, Belkhir K, David J, and Glémin S. 2012. Patterns and evolution of nucleotide landscapes in seed plants. *Plant Cell* 24:1379-1397.

Sim S-C, Durstewitz G, Plieske J, Wieseke R, et al . 2012. Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. PLoS ONE 7(7): e40563.

SP1- Annexe 1

Preparation of RNA samples

Samples were grinded in liquid nitrogen and total cellular RNA was extracted using a Spectrum Plant Total RNA kit (Sigma, Inc., USA) with a DNAse treatment. RNA concentration was first measured using a NanoDrop ND-1000 Spectrophotometer then with the Quant-iT™ RiboGreen® (invitrogen) protocol on a Tecan Genius spectrofluorimeter. RNA quality was assessed by running 1 µL of each RNA sample on RNA 6000 Pico chip on a Bioanalyzer 2100 (Agilent Technologies, Inc., USA). Samples with an RNA Integrity Number (RIN) value greater than eight were deemed acceptable according to the Illumina TruRNA-Seq protocol. For each genotype, 80% of RNA from the inflorescence and 20 % from the leave were mixed to obtain 2 µg of tissue bulked RNA.

Illumina library production

The Illumina TruRNA-Seq kit (Illumina Inc., USA) was used according to the manufacturer's protocol with the following modifications. In brief, poly-A containing mRNA molecules were purified from 2 ug total RNA using poly-T oligo attached magnetic beads. The purified mRNA was fragmented by addition of the fragmentation buffer and was heated at 94°C in a thermocycler for 4 min. The fragmentation time of 4 min was used to yield library fragments of 250-300 bp. First strand cDNA was synthesised using random primers to eliminate the general bias towards 3' end of the transcript. Second strand cDNA synthesis, end repair, A-tailing, and adapter ligation was done in accordance with the manufacturer supplied protocols. Purified cDNA templates were enriched by 15 cycles of PCR for 10 s at 98°C, 30 s at 65°C, and 30 s at 72°C using PE1.0 and PE2.0 primers and with Phusion DNA polymerase. Each indexed cDNA library was verified and quantified using a DNA 100 Chip on a Bioanalyzer 2100 then equally mixed by ten (from different species and different genotypes, possibly for different programs…). The final library was then quantified by real time PCR with the KAPA Library Quantification Kit for Illumina Sequencing Platforms (Kapa Biosystems Ltd, SA) adjusted to 10 nM in water and provided to the Montpellier Genomix platform for sequencing.

Illumina library clustering and sequencing conditions

Final mixed cDNA library was sequenced using the Illumina mRNA-Seq, paired-end protocol on a HiSeq2000 sequencer, for 2 x 100 cycles. Library was diluted to 2 nM with NaOH and 2.5 µL transferred into 497.5 µL HT1 to give a final concentration of 10 pM. 120 µL was then transferred into a 200 µL strip tube and placed on ice before loading onto the Cluster Station, mixed library, from 10 individual indexed libraries, being run on a single lane. Flow cells was clustered using Paired-End Cluster Generation Kit V4, following the Illumina PE_amplification_Linearization_Blocking_PrimerHyb_v7 recipe. Following the clustering procedure, the flow cell was loaded onto the Illumina HiSeq 2000 instrument following the manufacturer's instructions. The sequencing chemistry used was v4 (FC-104-4001, Illumina) using software SCS 2.6 and RTA 1.6 with the 2 x 100 cycles, paired-end, indexed protocol. Illumina base calling files were processed using the GERALD pipeline to produce paired sequence files containing reads for each sample in Illumina FASTQ format.

SP2 - Crop adaptation to climate changes: genetic and evolutionary processes involved in the phenological response

1. Rationale

There is mounting evidence that climate change affects biological and ecological processes and puts strong selective pressure on natural populations (Penuelas et al., 2002; Kremer et al., 2012). This raises two main questions regarding crop plants: (i) how will climate change affect the phenotypic and genetic diversity of crop species and their wild relatives, and (ii) what type of material should we produce to withstand the new climate regime. These questions are particularly important in developing countries where human populations mainly rely on traditional rain fed cropping systems. To accurately predict responses of crop plants and their wild relatives to future environmental changes, we need to learn more about the genetic architecture of adaptation but also about the adaptive trajectory of natural/artificial populations under heterogeneous and/or changing environments. For instance, the role of specific evolutionary factors such as mutation, migration and recombination in adaptation remains to be understood. Moreover, as several crop species reproduce through selfing, understanding the specific impact of the mating system on adaptation routes and genetic mechanisms also constitutes an important issue.

The rise of genomics paves the way to tackle these challenges. For an increasing number of plant species, new sequencing and high-throughput genotyping technologies allow the study of patterns of genetic variation at hundreds of loci. Such data enable to make inferences about population structure and demographic processes (genetic drift, migration, frequency of recombination) and provide a multilocus null distribution of variation across the genome that can be used to reliably detect footprints of selection in candidate genes. Applied to samples comparing cultivated forms and their wild progenitors, this genome scan approach has enabled tremendous progress in our knowledge about the genetic architecture of domestication, which can be considered as an example of adaptation to cultivation and human needs (see for example: Wright, 2005; Burke et al., 2005). Few applications of this approach to samples collected on climatic or environmental gradients were available at the beginning of this project (see however Hancock et al., 2008 and more recently, Hancock et al., 2011), asking for both empirical data and theoretical developments in this area.

2. Objectives

The aim of this project was to explore the genetic and evolutionary mechanisms involved in local adaptation to spatially heterogeneous and temporally variable climatic conditions. The three main questions we addressed were: (1) what is the genetic architecture of plant adaptation to climate variation? (2) what is the relative role of standing genetic variation *versus* new mutations in the generation of genotypes adapted to new climate conditions?,

and (3) what is the role of evolutionary mechanisms like migration and recombination in the evolution of a population/species under rapid climate change?

To address these questions, we proposed to develop high throughput genome scan analyses on two kinds of samples: (*i*) samples collected along climatic gradients (hereafter referred as *spatial contrasts*) and (*ii*) populations collected from the same site at separate times (*temporal contrasts*). Climatic gradient analyses were expected to allow the identification of sets of candidate genes underlying response to climate-mediated selection; while monitoring the temporal evolution of populations during the 20-30 last years should provide evidence that evolution has happened (or not) and give insights into the demographic and selective trajectories of the populations under climatic variation.

The project was subdivided in three work-packages: first to test and develop methods for the detection of selection on environmental gradients using genome wide polymorphism data (WP1); then to use these methods to pinpoint genes or genomic regions exhibiting a distinct diversity signature on both spatial (WP2) and temporal contrasts (WP3).

3. Material and methods

Theoretical developments (WP1)

Available methodologies for selection scan include F_{ST}-based tests. These tests are based on the idea that if local selection occurs at a given locus, differentiation (assessed by F_{ST}) will increase at this locus compared with what is theoretically expected at neutral loci. They have been incorporated in several statistical frameworks (*frequentist methods*, Beaumont and Nichols, 1996, Vitalis et al., 2001; *Bayesian methods*, Beaumont and Balding, 2004; *Markov Chain Monte Carlo methods*, Foll and Gaggiotti, 2008). These methods do not actually take advantage of associated bio-geographical data (climate, soil but see Foll and Gagiotti, 2006). Recent approaches to this problem have been proposed and use the covariance of genetic variation with environmental data (correlation-based methods, Joost et al., 2007; Poncet et al., 2010; Coop et al., 2010). The performance and the rate of false positives of such tests have been addressed in only a few studies and correlation-based methods have not been included in such comparisons. In this part of the project, we thus aimed to *(i)* compare available methods to detect selection considering the special case of genes involved in adaptation along environmental gradients, *(ii)* compare the efficiency of these different methods under different migration models, mating systems and sampling strategies, *(iii)* develop a frequentist-based method adapted to environmental structure and *(iv)* analyze available methods allowing to search for footprint of selection in temporal samples.

To perform *(i)* and *(ii)*, we used a simulation approach to generate datasets under explicit models (Epperson et al., 2010). Simulations were performed using QUANTINEMO (version 1.0.3, Neuenschwander et al., 2008), an individual-centered simulation software. We modeled 100 populations arranged in a 10 x 10 regular grid and simulated diploid and hermaphrodite individuals, with non-overlapping generations. We considered 6 combinations of demographic parameters: 2 rates of self-fertilization (S=0 and S=0.95) x 3 migration models (the island model, IM; the hierarchical island model, HM and the stepping stone model, SS). Both neutral and selected loci were considered. For the later,

we applied a gradient of selection along the environment. Five different sampling strategies were defined on the simulated datasets: from one individual sampled in all the simulated populations to 24 random alleles randomly sampled in only eight populations.

A project focusing on three model species….

The project was built on three species: rice (*Oryza* sp), pearl millet (*Pennisetum glaucum*) and the legume model species *Medicago truncatula*. Regarding adaptation processes, these three species exhibit several complementarities. First, two species – rice and *M. truncatula* - are predominantly autogamous while pearl millet mostly reproduces through outcrossing. Second, pearl millet and rice are cultivated species; both natural and anthropic evolutionary forces will thus affect evolution and adaptation in these two species. In contrast, *M. truncatula* is a spontaneous species around the Mediterranean basin; its evolution thus mostly depends on natural selection and evolutionary forces. Finally, these three species represent different climatic conditions: *M. truncatula* is a temperate species that is highly sensitive to temperatures especially during winter while rice and pearl millet are tropical species, both sensitive to photoperiod.

…*and a single adaptive trait, i.e. flowering time*

Although the above methodologies were expected to lead to sets of genes involved in different functions, we proposed to primarily target flowering time candidate genes. Flowering time has repeatedly been shown to be of paramount importance for the adaptation of sessile organisms such as plants, ensuring reproduction to occur in favorable conditions with respect to climate, herbivory and pathogen pressures or pollinators behavior (Remington and Purugganan, 2003; Roux et al., 2006). Flowering time is arguably the trait that conditioned the speed at which cultivated plants spread from their respective centers of domestication (Diamond, 2002), and explains a sizeable proportion of the variation in species geographical distribution around the globe. Finally, the genetic architecture of this trait has been extensively studied in plants (Kobayashi and Weigel, 2007; Izawa, 2007), so that a large set of flowering time candidate genes was available at the beginning of the project.

Spatial contrasts

Genome scans on samples collected along climatic gradients were designed to identify genes/alleles involved in flowering time variation associated to climate conditions. To this aim, we proposed to compare patterns of SNPs diversity between anonymous genes and candidate genes for flowering time variation and to look for genes or genomic regions showing ‘abnormal’ patterns of variation, suggestive of selection. As a second step and in order to validate the involvement of these genes in flowering time variation, we planned to investigate these candidate genes through phenotype/genotype associations using either QTL populations or GWAS panels.

For the three species, large geo-referenced collections were available, as well as flowering time genes/genomic data. For *Medicago truncatula*, the sample to be considered was a core-collection of 192 individuals representing the whole species distribution. For rice, the spatial analysis was carried on using sets of rice accessions collected in Madagascar where rice is cultivated over a wide range of climatic conditions mainly influenced by altitude (0-

1900m). For pearl millet (*Pennisetum glaucum*), we considered 424 varieties from West-Africa where a strong environmental gradient is observed from wet coastal country to the dry northern limit of agriculture. The first part of the project consisted to develop genome wide SNPs in both “neutral” (or “naïve”) genes and in flowering time genes. For *Medicago truncatula* and rice, data from re-sequencing analyses were already available to build appropriate Illumina beadchips. For *Pennisetum glaucum*, we sequenced a set of flowering as well as random genes to identify appropriate SNPs. To detect selection footprints related to climatic variation, the level of precision of climatic and environmental variables used to describe the studied gradients was an important issue. We thus decided to carefully analyze the publicly available climatic databases and to establish partnerships with local climatic research units (AGROCLIM research unit, INRA Avignon).

Temporal contrasts

Temporal contrasts were designed (*i*) to determine how climate changes affect population genetic diversity, (*ii*) to identify genes involved in the population response to current climatic changes and (*iii*) to characterize the evolutionary mechanisms involved in this response. To this aim, we used sets of populations that had evolved under changing climatic conditions during the 30 last years. As for spatial contrasts, we aimed to study SNP data between anonymous (supposedly “neutral”) and candidate genes for flowering time variation, but now using individuals/populations sampled at different time points in the same location (“snapshots samples”). Two levels of analysis were proposed: one favoring the number of locations instead of the number of individuals per population and one focusing on a small number of populations.

For pearl millet, landraces sampled in the same villages in Niger in 1976 (192 individuals) and in 2003 (420 individuals) were already available and a significant shift in flowering time toward earlier varieties had already been observed between the 2 sampling years. For *Medicago truncatula* and rice, new collections were needed. We chose to focus on the south of Spain, the south of France and the Corsica Island for *M. truncatula* and on Guinea for rice, since landraces had already been collected in the 1980’s in this country.

For each species, we used data obtained through genome scan SNP genotyping (*a)* to estimate temporal variation in allele frequencies between the two considered time points. *(b)* For the two predominantly self-fertilizing species (rice and *Medicago*), we combined data obtained for each SNP to define individual multilocus genotypes and to infer recent outcrossing and recombination events and to detect recent founder or migration events. (c) Although in our scenario we hypothesized an evolutionary response, selection could also favor genotypes exhibiting differential GxE interaction and give a more or less plastic phenotype. To assess the role of plasticity in the population response to climate variation, we thus studied the relative plasticity of ancient and modern genotypes in certain geographic sites.

4. Activities carried out and main results

Theoretical studies and developments

Test of available methodologies to look for footprints of selection

This work was implemented by Stéphane de Mita during his post-doctoral fellowship (see De Mita et al., 2013). We simulated populations distributed along a selective gradient and explored different migration models (island model, IS; hierarchical island model, HM; stepping stone model, SS), selective intensities, self-fertilization rates and sampling schemes. Patterns of diversity obtained through those simulations were analyzed using eight methods designed to detect footprints of selection: three genotype/environment correlation-based methods and five differentiation-based methods (**Table SP2-1**).

Table SP2-1: List of methods tested in this study

Method (Reference)	Technique	Underlying model	Envir. Variable	Control loci
LR (Joost et al., 2007)	GLM	Independence of the observations	+	-
GEE (Poncet et al., 2010)	GEE	Independence of the clusters	+	-
CWDRP (Coop et al., 2010)	MCMC	Island model	+	+
FLK (Bonhomme et al., 2010)	Forward simulations	Multiple diversgence model	-	+
BN (Beaumont & Nichols, 1996)	Coalescent simulations	Island model	-	+
EHF (Excoffier et al., 2009)	Coalescent simulations	Hierachical island model	-	+
VDB (Vitalis et al., 2001)	Coalescent simulations	Pairwise divergence model	-	+
FG (Foll and Gaggiotti 2008)	RJ-MCMC	Island model	-	-

Legend:
BN: Beaumont and Nichols; CWDRP: Coop, Witonsky, DiRienzo and Pritchard; EHF: Excoffier, Holer and Foll; FG: Foll and Gaggiotti; GLM: Generalized linear model; GEE: generalized estimating equations; LR : logistic regression ; MCMC: Monte Carlo Markov Chain ; RJ-MCMC: Reversible Jump Monte Carlo Markov Chain; VDB : Vitalis, Dawson and Boursot. LR, GEE and CWDRP are correlation-based methods; FLK, BN, EHF, VDB and FG are differentiation-based methods.

We showed that methods based on genotype-environment correlations were substantially more powerful to detect selection than differentiation-based methods but that they generally suffer from higher rates of false positives. This effect was exacerbated whenever allele frequencies were correlated either between or within populations. Our analysis also showed that the reproductive regime has important consequences for the detection of selection. Indeed, combined with the island model of migration, selfing reduced the power of the differentiation-based methods, except for BN, which showed a good rate of true

positives (93%). LR and GEE methods were also affected by selfing, showing a consistent increase in false positive rates under all migration models.

Taken together, this study allowed us to formulate general guidelines for selecting a method to detect footprint of selection. Important advices are: (1) results obtained with correlation-based methods should be interpreted with caution, especially in situations that lead to between-population correlations. In such cases, recently developed statistical tools accounting for underlying correlation structure are recommended; (2) it is better to sample a few individuals in many populations than many individuals in a few populations; (3) under predominant outcrossing and the island model of migration, all the methods we tested are efficient and comparable; (4) under predominant outcrossing and SS or HM models of migration, differentiation-based methods are better ; (5) under high selfing rates and the IM of migration, the best option is to use the LR method and to sample one individual per population; (6) under high selfing rates and SS or HM migration model, the best methods are LR with one individual per population, BN with a few populations and FG with many populations.

Development of a specific tool to estimate selection coefficient from temporal sample of outbreeding population

Assessing the value of selection (s) from temporal samples is very useful information to understand the strength of selection *in situ*. We developed a simple approximate Bayesian approach to this endeavor. This approach takes into account sampling effect and drift, to estimate the value of s from temporal samples. We assessed the effectiveness of the method for varying effective size, sample size and strength of selection. The method was applied to pearl millet varieties sampled across two years. This specific approach was developed by IS Ousseini (Master then PhD student).

Development of a specific tool to analyse the haplotypic structure of selfing populations

Self-fertilization is expected to have huge impacts on the population genetic structure. When recurrent among generations, it rapidly increases the rate of homozygosity and creates multilocus genotypes that increase in frequency through generations. Populations of predominantly selfing species are thus expected to contain several multilocus genotypes in varying frequencies. While several tools are available to estimate the *per* locus level of homozygosity from molecular data, there is to our knowledge no specific tool allowing to identify and analyse the different multilocus genotypes available in a population. Such a specific tool was built by O. Defrain during a three months training course (training staff: V. Ranwez, L. Gay, J. Ronfort UMR AGAP). This tool can be used to analyse large molecular datasets (> 10000 SNPs). It identifies the different multilocus genotypes present in the studied sample, estimates the relative frequency of each multilocus genotype in the population as well as the genetic distance between each pair of multilocus genotypes. The software also allows identifying specific multilocus genotypes resembling recombinant inbred lines as expected if some rare outcrossing events are occurring in the populations.

Spatial contrasts

The legume model Medicago truncatula

Spatial contrasts analyses in *Medicago truncatula* were built on a core-collection of 192 individuals (CC192) initially collected in 163 geographic locations around the Mediterranean basin. Climatic data were obtained using the public database WorldClim (http://www.worldclim.org/), which gives access to monthly climatic parameters (min and max temperatures, total precipitations, and bio-climatic variables) averaged over 50 years (between 1950 and 2000), with a resolution of approximately 1 km. As a pilot study, we first designed an Illumina Beadship X384 SNP array targeting both flowering and naïve genes and used it for a preliminary analysis of the CC192 nucleotide variability (Loridon et al., 2013). In November 2011, as part of a partnership with the *Medicago* Hapmap NSF program (Leader N. Young, Univ. of Minnesota), we obtained whole genome re-sequencing data for the CC192 (mean coverage = 5X). The dataset contained more than 10,000,000 genome-wide SNPs. To analyze this dataset, we benefited from an INRA funded post-doctoral position (C. Burgarella, April 2011-December 2012).
We first extracted a set of 20,188 inter-genic and supposedly anonymous SNP, which we used to infer the genetic structure underlying the CC192. Using the DAPC method (Jombart, 2009), we inferred a clear subdivision of the sample in two groups: a western group containing most of the accessions originating from Morocco and the Iberian Peninsula and an eastern group containing all the localities lying in the East side of a line drawn between the south of France and Algeria (**Figure SP2-1**).

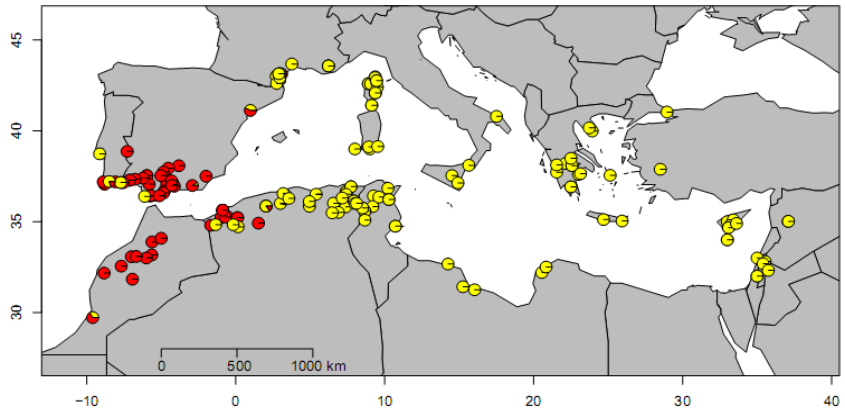


Figure SP2-1: Illustration of the stratification detected in the CC192 of *Medicago truncatula* using the DAPC method and a set of 20,188 genome-wide SNPs located in inter-genic regions.

We also extracted a set of 289 *Medicago truncatula* genes corresponding to different flowering pathways proteins. Applying filters to select single nucleotide polymorphisms (SNPs) showing a missing rate <15% and a minor allele frequency (MAF) <=5%, we obtained a set of 5229 SNP covering 225 candidate genes for flowering time. To detect covariations between climatic data and genetic variations, we used a mixed model approach as implemented in the general framework proposed for association studies. Using a threshold *P*-value = 10^{-4}, we found 61 SNPs located in 40 candidate genes for flowering time variation significantly associated with (one or several) environmental variable(s). As expected, most of the identified polymorphisms were located in genes involved in the plant response to external stimulus (i.e. photoperiod, light quality, temperature), while a minor number tagged genes involved in later stages of the flowering pathways (i.e. genes involved in the integration of signals from multiple pathways or in the development of the flower organs, see **Table SP2-2**).

Table SP2-2: Gene families associated with environment variables and flowering traits (p-value < 10-4) sorted by position in the regulatory network of flower induction.

Pathway*	Gene_family	Environment association		Flowering association		Common
		Genes	SNPs	Genes	SNPs	Genes
Photoperiod	FHY/FAR related proteins	5	8	6	10	
Photoperiod	COP	4	8	2	8	2
Photoperiod	Phytochrome	4	7	3	4	1
Photoperiod	Constans	2	2	1	1	
Photoperiod	Cryptochrome			1	1	
Photoperiod	Nuclear transcription factor Y subunit	1	1			
Photoperiod	FT/TFL1 family	1	1	2	2	
Photoperiod	XAP5 family			1	1	
Photoperiod	SENSITIVITY TO RED LIGHT REDUCED	1	1			
Vernalization	VRN/VIL family	3	6	1	7	1
Age	Squamosa promoter binding-like family	3	3	3	4	1
ambient temperature	Short Vegetative Phase			1	1	
Autonomous	Flowering Time Control Protein FCA			1	2	
Autonomous	Della	1	1	1	1	
Autonomous	Luminidependens	1	1	1	4	1
Gibberellin	Gibberellin	5	5			
flower development	Agamous	2	7			
flower development	Maternal Embryogenesis Control Protein	2	4			
flower development	Embryonic flower	1	1	2	8	
flower development	Apetala	2	2	2	2	2
flower development	WD-40 repeat family	1	1	1	1	
flower development	Curly Leaf	1	2			
		40	**61**	**29**	**57**	**8**

* Roux et al. 2006, Fornara et al. 2010.

Considering the same threshold value, association analyses performed with a set of 16 flowering traits resulted in 115 significant tests, tagging 57 SNPs located in 29 candidate genes (**Table SP2-2**). More than 60% of the genes associated with flowering time were located on chromosomes 5 and 7, all the genes detected on chromosome 7 being located in a window of 10 Mbp. In this region, the number of significant SNP was higher than by chance (23/57 SNP, Chi-squared test p-value ≤ 10-4). Interestingly, this region has already been reported as a QTL region for flowering time variation in *Medicago truncatula* as well as in a previous GWA study (Julier et al., 2007; Stantton-Geddes et al., 2013), supporting that identified genes may harbor or be close to functional variation. Multiple regression analyses showed that the 57 SNPs associated with flowering traits explained on average 55% of the variability in flowering time (mean adjusted-r^2 over flowering traits, excluding STD and Vernalization-related traits, range=[37%-67%]), while the 61 SNPs detected with environmental variables explained on average 40% of the flowering time variability (mean over flowering traits, excluding STD and Vernalization-related traits, range=[17.8%-55.4%]). Not unexpectedly, this latter value was significantly lower than the one obtained with SNPs discovered using flowering genes traits (Wilcoxon signed rank test p-value = 3.052e-05). Both values were however significantly higher than expected by chance (as shown by randomly sampling 10 000 sets of 57 and 61 SNPs within the 5206 available). This confirms the notion that environment analysis indeed allows to identify regions of the genome involved in the determination of flowering time. Individual effects of flowering time-associated SNPs were on average larger than climate-associated

SNPs (Figure 6, Mean effect= -13.47 and -43.57 for SNP_env and SNP_flo respectively; p_value (ANOVA) = 1.5E-3). Few genes (eight) and even fewer polymorphisms (three from as many genes) were identified in both association analyses (Table SP2-2). These genes are strong candidates for functional analyses.

Cultivated rice in Madagascar

The spatial analysis was carried on using several sets of rice accessions collected in Madagascar. A two-step approach was adopted consisting in (1) the use of previously developed resources (rice collection, genotypic data...) and (2) the development of new dedicated resources. Resources previously developed, in the framework of a Generation Challenge Program (GCP) project, consisted in a core collection of 150 accessions, representative of the rice genetic diversity and rice-growing ecosystems in Madagascar, genotyped with 1564 SNP.

Population structure analysis led to distinguish four subpopulations of uneven eco-geographical distribution. Two factors were determining in this distribution: the rice growing ecosystem, i.e. lowland-flooded (LL) versus upland-aerobic (UL), and the altitude. Three of these subpopulations corresponded to well-known genetic groups: *indica* with preferential distribution in LL ecosystem of low elevation (0-1000 m) areas, temperate *japonica* in LL of altitude above 1750 m, and tropical *japonica* in UL of 0-750 m. The fourth subpopulation, preferentially cultivated in the LL ecosystem of 1250-1750 m elevation, was specific to Madagascar, unknown anywhere else in the world. The stratification of the collection revealed through a sequencing analysis of 8 genes known for their major role in rice flowering time was very similar to the one observed with the 1564 neutral SNPs. Applying the methods and tools studied in WP1 on this dataset led to detect between 0 and 200 outlier loci, depending on the method. The correlation-based methods (LR, GEE and CWDRP) detected many more signatures of selection than differentiation-based methods (FLL, BN, VDB and FG), as expected when the population structure of the sample and the environmental variables are overlapping (De Mita et al. 2013). None of the candidate locus was declared outlier by all the genome-scan tests we implemented; moreover, no outlier loci co-localised, strictly, with any major known gene/QTL involved in rice flowering time and thermal sensitivity. Analysis of genes and gene families co-localising with outlier SNPs detected with at least three different methods, colocalised with a MADS-box transcription factors (involved in floral organ specification and other aspects of plant growth and development), Na+ and K+ transporters, and some defence genes, not with any of the 167 flowering time related genes known in rice.

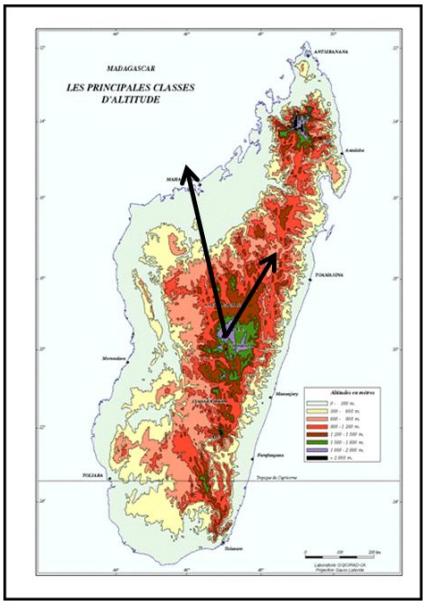


Figure SP2-2: Spatial contrast in Madagascar as illustrated by the 2 axes of collect during the 2012 campaign of collection

The above results confirmed the need for the development of dedicated resources as proposed at the beginning of the current project. We chose to focus on one rice growing ecosystem (lowland-flooded which represents 85% of the rice growing area in Madagascar) and two rice subpopulations (*indica* and the group specific to Madagascar) that exhibited the highest global

diversities and the smallest inter-group differentiation while covering the largest altitudinal interval (0-1850m). A new collecting campaign was undertaken in 2012 along the axis of Mahjunga (sea level) – Antsirabé (1850 m) regions and of Lac Alaotra (750m) – Antsirabé regions (**Figure SP2-2**).

Some 600 accessions were collected, 150 for each of the 4 following altitudinal intervals: 0-500m, 500-1000m, 1000-1500m and >1500m. The collected material was grown in greenhouse for for DNA extraction and seed increase. It was also grown in Madagascar for phenotyping for days to flowering. The panel was also genotyped using SSR markers and Genotyping-By-Sequencing (GBS) technology. Adaptation of the analysis pipeline is ongoing to optimise extraction of SNP and DArT markers in terms of both number and robustness. Once this step achieved, tools and knowledge developed in WP1 will be mobilised for detection of selection footprint. Climatic data (daily maximum and minimum temperatures and rainfall) for the 1961-1975 and 1995-2010 periods were acquired for 11 climatic stations distributed along the two axes of collection. Preliminary analysis of these data confirmed (1) the stability of mean temperatures between the two periods surveyed and (2) the existence of a gradient of about 15°C along the altitudinal axis, whatever the period of the year considered.

Pearl millet in West Africa

For spatial analysis, we assembled a sample of 424 individuals from West Africa. This sample was genotyped with a 384 SNP Illumina chip targeting both flowering and "naive" genes. We adapted our strategy and switched the initial aim to only use SNP genotyping chip. Following progress on the first sub-project of ARCAD ("Comparative population genomics"), we re-sequenced individually a set of 90 inbred lines using an RNAseq approach. In a parallel project, we assembled a reference transcriptom of 50,313 contigs, with a total of 36,479,993 nucleotides. Each inbred line was individually tagged and was sequenced. We obtained a final set of 112,924 variants constituted of 108,512 SNP and 4,412 Indel located on 20,240 different contigs. We performed an association mapping strategy using phenotypic data from 7 field trials. To reduce the risk of false positive in the association analysis, we used a mixed model with correction for similarities between individuals (simple EMMA model). After a FDR (False discovery rate) correction for multiple testing with a q-value threshold of 0.05, no variants were significant for flowering time, whereas 87 variants remained associated with the number of tiller, 169 with plant height and 234 with spike lenght. A publication of this result is ongoing.

Temporal contrasts

Pearl millet landraces evolution under climate change in Niger

We analyzed samples of pearl millet landraces collected in the same villages in 1976 and 2003 throughout the entire cultivated area of Niger (Vigouroux et al. 2011). We studied phenological and morphological differences in the 1976 and 2003 collections by comparing them over three cropping seasons in a common garden experiment (Figure SP2-3). We found no major changes in the main cultivated varieties or in their genetic diversity. However, we observed a significant shift in adaptive traits. Compared to the 1976 samples, samples collected in 2003 displayed a shorter life cycle (**Figure SP2-3**), and a reduction in plant and spike sizes. We also found that an early flowering allele at the *PHYC* locus increased in frequency between 1976 and 2003. We further validate the potential

direct role of *PHYC* on the observed genotype/phenotype association using pattern of selection around the *PHYC* locus, sequencing a BAC around *PHYC*, studying 3 F2 populations segregating for *PHYC* variants (Saïdou et al. 2014).
This increase exceeded the effect of drift and sampling, suggesting a direct effect of selection for earliness on this gene. We conclude that recurrent drought can lead to selection for earlier flowering in a major Sahelian crop (Mariac et al., 2011, Vigouroux et al., 2011). Surprisingly, these results suggest that diffusion of crop varieties is not the main driver of short term adaptation to climatic variation in this particular case.

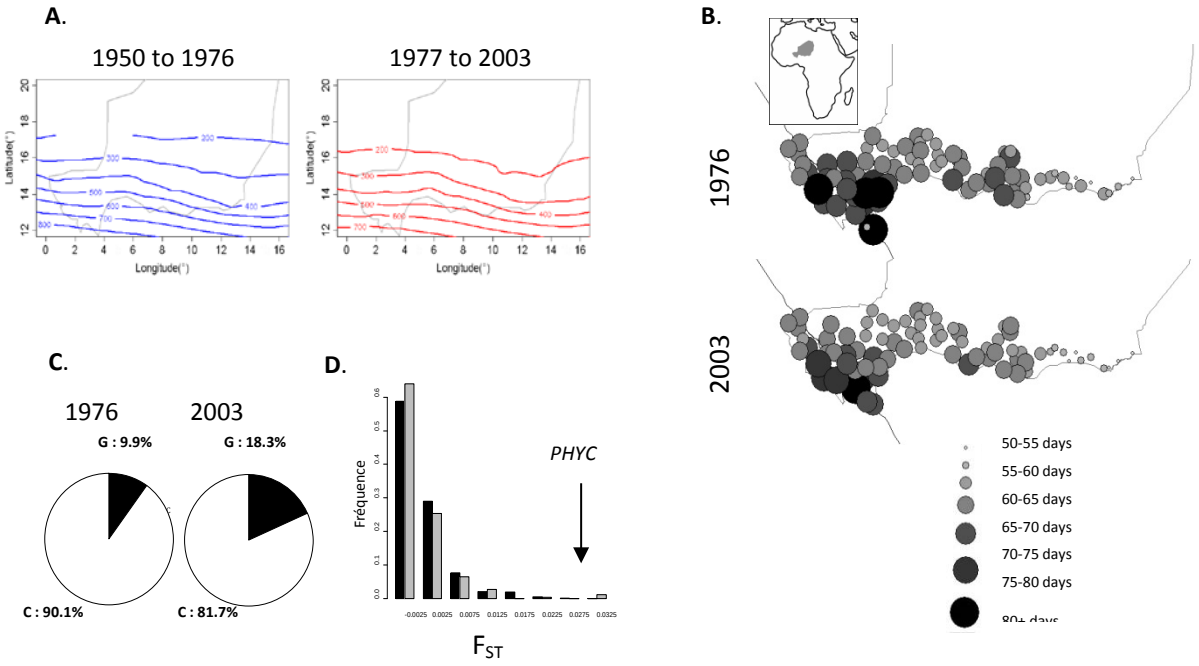


Figure SP2-3: Selection for earliness associated with climate variation in the Sahel. (A) The rainfall isohyets are plotted for the 1950–1976 periods (blue) and the 1977–2003 periods (red). The shift in isohyets 100 to 150 km further south between the two periods illustrates the average decrease in rainfall. (B) The average values of flowering time are plotted for each of the 79 villages sampled throughout Niger. Significant variations were observed between villages, i.e. varieties collected in 2003 flowered earlier ($p<0.001$). A significant increase (C) in early flowering alleles at the PHYC gene (SNP G) was observed between 1976 and 2003. The difference (F_{ST}) between the two samples in PHYC alleles exceeded the expected effect of drift or sampling (D) based on empirical (dark) or model-based (gray) F_{ST} distributions. These results suggest that selection led to an increase in early flowering alleles at the PHYC locus.

Temporal evolution of rice varieties in Guinea

For rice, the temporal study took place in Guinea where a large number of rice accessions were collected in 1979 and 1982, hereafter referred to as T1. Going back to the collection campaign notebooks, we digitalised passport data for some 780 accessions and geo-localized the places/villages of collecting. On the basis of this information we decided to focus the new collection campaign on a South –North transect covering 3 climatic zones (**Figure SP2-4**) and hosting the two main rice cultivation ecosystems, rainfed lowland (RLL) and rainfed upland (RUL).

In 2011 (hereafter referred to as T2), some 60 villages were visited twice to collect RLL and RUL varieties of different durations (number of days to flowering). Some 770 accessions were collected, while the number of accessions collected in T1 was 495 for the target villages. Seeds collected in The T1 and T2 panels were genotyped first with 12 SSR markers then through GBS method that produced some 28,000 SNP markers and 25,000 DArT markers. A second GBS genotyping experiment, using a different restriction enzyme is ongoing for *O. glaberrima* and *O. sativa japonica* accessions. The T1 and T2 panels collected in RUL ecosystem were phenotyped for in RUL conditions in Guinea, in 2012, for duration (day to flowering) and, photosensitivity in Guinea. Climate data were acquired for the 1961-2010 period, for 9 climatic stations distributed within the area covered by the T2 collection. Analysis of changes for the following sets of climate indicators is ongoing : (1) Tmax, Tmin and the difference between Tmax and Tmin; (2)Total rainfall, duration of the rainy season, date of the beguiling and the end of the rainy season, duration and frequency of short drought episodes during the rainy season; (3) Climatic risks (temperatures during the flowering time, and water balance), using eco-physiologic models (SarrH, for instance) to simulate rice development and water balance.

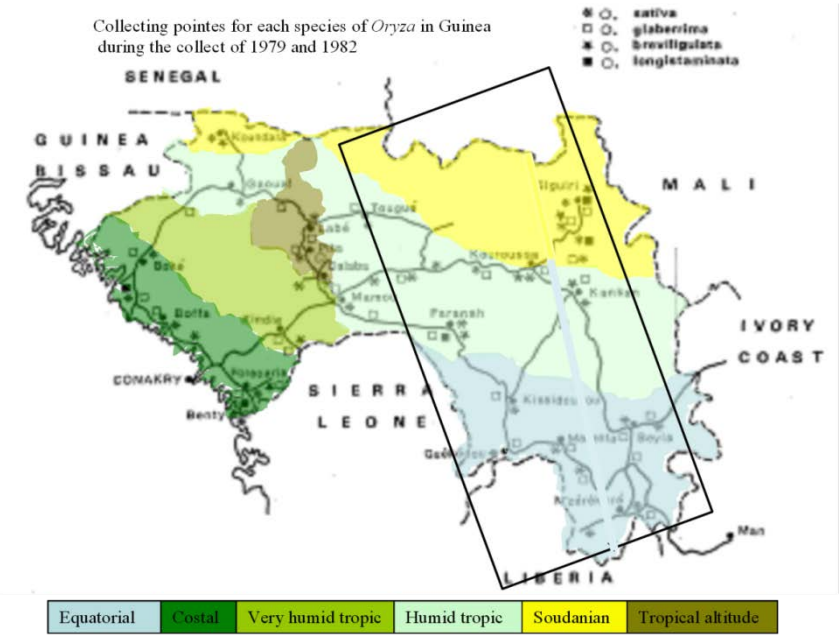


Figure SP2-4: Temporal contrast in Guinea. Symbols refer to the collection points in 1979-1982; the 2011 collection area is surrounded by a rectangle.

Preliminary analysis of SSR data revealed *(i)* a significant increase in frequency of RLL accessions to the detriment of RUL accessions, respectively 54% and 46% in T1 and 67% and 33% in T2; *(ii)* a significant decrease of the frequency of *O. glaberrima* at the benefit of *O. sativa* varieties, especially in the RLL ecosystem (respectively 23% and 77% in T1, 16% and 84% in T2); (*iii*) significant differences in the distribution of RLL versus RUL accessions between the north (Sudanean) and the south (forest) area of the T2 collection; (*iv*) no significant change in the mean number of alleles per locus between T1 and T2 and (*v*) a rather low number of accessions (< 10%) exhibiting the same multilocus genotypes in T1 and T2.

Analysis of phenotypic data revealed significant differences between the two panels for the number of days to flowering (DF): DF1979 = 107 and DF2011 = 104 **(Figure SP2-5**). The contrast was particularly high for *O. glaberrima* accessions (DF1979 = 101 and DF2011 = 94) and accessions collected in upland ecosystem ((DF1979 = 108 and DF2011 = 103).

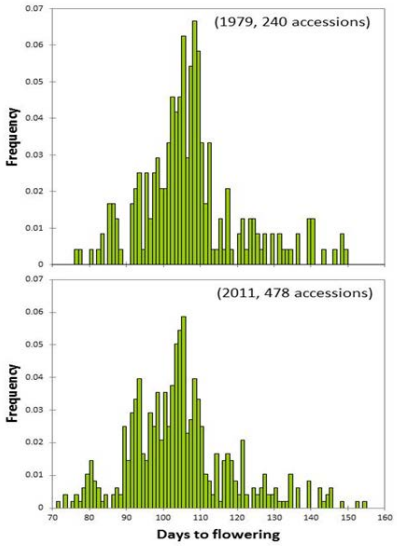


Figure SP2-5: Distribution of the number of day to flowering in the two rice panels established for analysis of temporal contrast

Screening of GBS data for quality score, missing data (<5%) and MAF (>5%) reduced drastically the actual number markers to about 6000 SNP and as many DArT. Global population differentiation over the period of 1979-2011 was very low for the three compartments of diversity considered: pairwise population FST = 0.014, 0.015 and 0.017, for *O. glaberrima, O. sativa indica* and *O. sativa japonica*, respectively. In the three cases, among population variation represented less than 2% of the total molecular variance. Preliminary search for loci under selection from hierarchical F-statistics, as described by Excoffier et al. (2009), detected very few such loci and the one detected did not co-localized with genes of known function. New methods are being explored.

Medicago truncatula and climate change in the Mediterranean Basin

At the beginning of the Arcad project, *M. truncatula* accessions collected between 1985 and 1990 in 110 sites located in Corsica, the south of France and Spain were already available in our laboratory. To determine if phenological changes have occurred in these three regions during the 30 last years, we first organized new campaigns of collection in these regions. Between 2005 and 2009, we revisited the initial 110 sites but populations of *Medicago truncatula* were found in only one third of these sites. New sites were thus defined and seeds were collected, giving access to a set of 72 localities for the 2005-2009 period (**Table SP2-3**).

Table SP2-3: Number of sites used to compare the flowering phenology of *Medicago truncatula* over a 30 year-period in three regions of the Mediterranean Basin.

	Spain	**Corsica**	**South of France**	**Total**
1985-1990	38	36	36	110
2005-2009	25	29	18	72
Common sites	20	26	10	56

We extracted a single seed from each of these populations. In 2010, these seeds were grown under greenhouse conditions to produce families for further analyses and to remove potential maternal effects. During the summer 2011, these families were grown under homogeneous greenhouse conditions for phenological analyses. Results from this study are reported in **Figure SP2-6.**

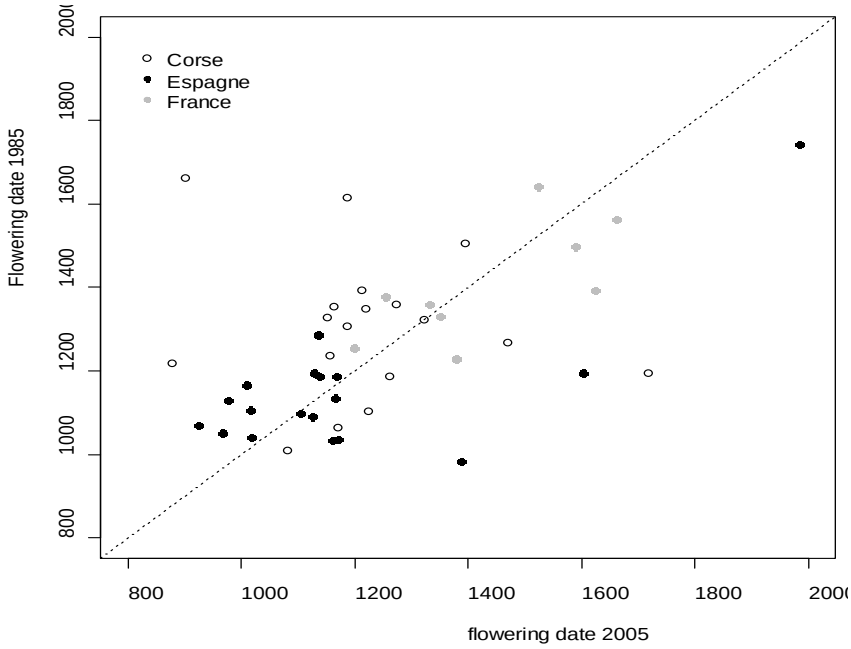


Figure SP2-6: Graphical representation of the flowering date of plants representing the same geographical locations but collected in 1985-1987 and then in 2005-2009. A significant increase in precocity was observed only in the Corsica sample (n= 26, Rank test p=0.049).

We detected no clear pattern of phenological change between plants collected in 1985 and 2005-2009, except in Corsica were plants collected in 2005-2009 flowered significantly earlier compared to plants collected in 1985. To go further into our understanding of the evolution of these populations, 8 geographical locations in which plants were collected both in 1985 and in 2005-2009 were chosen for further molecular and phenotypic analyses. Seeds were regenerated under greenhouse condition in order to obtain large samples (n=100) representing each year of collection in each of these 8 sites. In order to understand the evolutionary history of these populations during the last 25-30 years period, each plant was genotyped using a set of 20 SSR loci. At this time, only one population (F20.089) has been studied extensively for both molecular and phenotypic evolution. This population is located in the north of Corsica. In spring 2012, both old and modern individuals (collected respectively in 1987 and in 2009) were grown under homogeneous greenhouse conditions and characterized for their flowering phenology. This experiment showed that the modern genotypes flower significantly earlier than old plants (**Figure SP2-7**) as expected if adaptation for warmer temperatures had occurred. This study has been part of the Master study of J. Dhinaut (01/2014-06/2014).

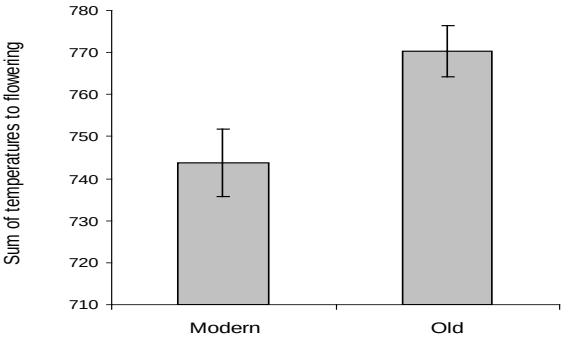


Figure SP2-7: Changes in flowering time between 1985 (old) and 2009 (modern). Flowering time is presented as a sum of temperatures and error bars stand for standard errors.

To test if this change in flowering phenology is adaptive, the following step will consist in a "temporal transplant" experiment introducing both modern and old genotypes into the current environment of the F20.089 population site. The "*in situ*" flowering phenology and the seed production of these plants will be studied as part of the project "SelfAdapt" (INRA funding MP ACCAF, Leaders: L. Gay UMR AGAP and M. Navascues UMR CBGP).

5. Foreseen valorisation of the ongoing analyses

Several publications are projected based on the results of the project: one on the estimation of the intensity of selection and its application to temporal samples, one on association mapping in pearl millet, one on spatial contrasts in *Medicago truncatula*, ... Most of those manuscripts are already in preparation.

6. Arcad added-value

Collaboration between the different ARCAD sub-projects helps adoption of techniques developed in other project to our study. Through training sessions, ARCAD allows to boost the use of next generation sequencing data in the community. Data analysis pipelines and methodological experimental protocols were developed and helped secure funding from other sources (French ANR BioAdapt, Metaprogram INRA ACCAF, INRA post-doctoral fundings). The ARCAD SP2 research agenda involved several master students and post doctoral researchers. Two post doctoral researchers involved in SP2 have obtained a secure permanent position.

7. Visibility, attractiveness, national and international positioning

The project helped secure new funding and to develop a research agenda in different directions.

One funded project focuses on the study of wild cultivated relative using methodology partly developed during the ARCAD project. This project includes: detection of selection approach along environmental gradient, association mapping, study of gene differential expression and effect of epigenetic changes in adaptation. The project is funded (514,000€) by ANR for 4 years "AdaptInWild: Identifying adaptive variation in the wild progenitors of two cereal crops, maize and pearl millet (2012-2016)".

Another related project uses a wild pearl millet whole African distribution to study phylogeography and adaptation (ANR, RPOC: MilDiv 162000€ 2012-2014). In *Medicago truncatula*, the temporal contrast analysis enabled us to develop a new project "SelfAdapt" (funded by the INRA Metaprogram "Adaptation to Climatic Changes, ACCAF", 50000€, 2013-2015) in collaboration with researchers developing theories in population genomics at the CBGP in Montpellier. This project aims at developing specific population genomics tools to look for footprints of selection in self-fertilizing species and at studying the population dynamics under climatic changes. Several teams involved in the Arcad project (SP1 S. Glémin, SP2 L. Gay and J. Ronfort, SP3 MH Muller) are also involved in the 'SEAD' project submitted to the BIOADAPT ANR program. This project aims at understanding the

specific demographic and evolutionary consequences of selfing on adaptation in plant species.

8. Future research areas

For the ARCAD SP2, the identification of genes associated with phenotypic variation could be a significant asset to study population *in natura*. The study of the dynamics of populations, estimation of selection acting on a year to year basis is now feasible if important alleles are identified. One of the research areas on the ARCAD SP2 project was to focus more thoroughly on the dynamics of the populations using both neutral and selected locus. During the course of the project, ARCAD SP2 leads on to new questions and challenges regarding adaptation: what is the role of epistasis and of cis-regulations in adaptation? What are the specific effects of self-fertilization on the adaptive trajectory of selfing populations?

References cited:

Beaumont MA, Balding DJ 2004. Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, 13, 969-980.

Beaumont MA, Nichols RA 1996. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 263, 1619-1626.

Burke, JM, Knapp, SJ and LH. Rieseberg 2005. Genetic Consequences of Selection During the Evolution of Cultivated Sunflower. *Genetics*, 171: 1933–1940

Clotault J, Thuillet AC, Buiron M, De Mita S, Couderc M, Haussmann BIG, Mariac C, Vigouroux Y. 1012. Evolutionary history of pearl millet (*Pennisetum glaucum* [L.] R. Br.) and selection on flowering genes since its domestication. *Mol Biol Evol*. In press.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK 2010. Using environmental correlations to identify loci underlying local adaptation. *Genetics*, 185, 1411-1423.

De Mita, S., Thuillet, A.C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J. and Y. Vigouroux. 2013. Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations, Molecular Ecology, *Mol. Ecol.* **22**:1383-1399.

Diamond, J 2002. Evolution, consequences and future of plant and animal domestication. *Nature*, 418: 700-707.

Epperson BK, McRae BH, Scribner K *et al.* 2010. Utility of computer simulations in landscape genetics. *Molecular Ecology*, 19, 3549-3564.

Foll M, Gaggiotti O 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, 180, 977-993.

Foll M, Gaggiotti O 2006. Identifying the environmental factors that determine the genetic structure of populations. *Genetics*, 174 : 875-891

Hancock AM, Faure N, Horton MW, Jarymowycz LB, Gianluca Sperone F, Toomajian C, Roux F, Bergelson J 2011. Adaptation to Climate Across the *Arabidopsis thaliana* Genome S*cience,* 334: 83-86

Hancock AM, Witonsky DB, Gordon AS, Eshel G, Pritchard JK, et al. 2008. Adaptations to climate in candidate genes for common metabolic disorders. *PLoS Genet* 4: e32.

Izawa T 2007. Adaptation of flowering-time by natural and artificial selection in Arabidopsis and rice. *J Exp Bot* 58:3091–3097.

Jombart T, Pontier D Dufour A-B 2009. Genetic markers in the playground of multivariate analysis. *Heredity*, 102:330-341.

Joost S, Bonin A, Bruford MW *et al.* 2007. A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology*, 16, 3955-3969.

Kobayashi Y, Weigel D. 2007. Move on up, it's time for change—mobile signals controlling photoperiodic-dependent flowering. *Genes and Development* 21, 2371–2384.

Kremer A, Ronce O, Robledo-Arnuncio JJ, Guillaume F, Bohrer G, Nathan R, Bridle JR, Gomulkiewicz R, Klein EK, Ritland K, Kuparinen A, Gerber S, Schueler S 2012 Long-distance gene flow and adaptation of forest trees to rapid climate change. Ecology Letters, 15: 378–392

Loridon, K., Burgarella, C., Chantret, N., Martins, F., Gouzy, J., Prospéri, J-M. and J. Ronfort. 2013 Single Nucleotide Polymorphism discovery and diversity in the model Legume *Medicago truncatula*. Molecular Ecology Notes, 13:84-95.

Mariac C, Jehin L, Saïdou AA, Thuillet AC, Couderc M, Sire S, Jugdé H, Adam H, Bezançon G, Pham JL and Vigouroux Y. 2011. Genetic basis of pearl millet population adaptation along an environmental gradient investigated by a combination of genome scan and association mapping. Mol Ecol. 20:81-91.

Neuenschwander S, Hospital F, Guillaume F, Goudet J 2008. quantiNEMO: an individual-based program to simulate quantitative traits with explicit genetic architecture in a dynamic metapopulation. *Bioinformatics*, 24, 1552-1553.

Penuelas, J. Filella, I. and Comas, P. 2002. Changed plant and animal life cycles from 1952 to 2000 in the Mediterranean region. *Global Change Biology*, 8:531-544.

Poncet BN, Herrmann D, Gugerli F *et al.* (2010) Tracking genes of ecological relevance using a genome scan in two independent regional population samples of *Arabis alpina*. *Molecular Ecology*, 19, 2896-2907.

Remington DL., Purugganan, MD 2003. Candidate genes, quantitative trait loci, and functional trait evolution in plants. *Int. J. Plant Sci*. 164:S7-S20.

Roux F, Touzet P, Cuguen J, Le Corre V 2006. How to be early flowering: an evolutionary perspective. *Trends in Plant Science*,11: 375-381.

Saïdou AA, Clotault J, Couderc M, Mariac C, Devos KM, Thuillet AC, Amoukou IA, Vigouroux Y. 2014. Association mapping, patterns of linkage disequilibrium and selection in the vicinity of the PHYTOCHROME C gene in pearl millet. Theor Appl Genet. 127:19-32.

Vigouroux Y, Barnaud A, Scarcelli N, Thuillet AC. 2011. Biodiversity, evolution and adaptation of cultivated crops. CR Biologies 334 : 450-457.

Vigouroux Y, Mariac C, Pham JL, Gérard B, Kapran I, Sagnard F, Deu M, Chantereau J, Ali A, Ndjeunga J, Luong V, Thuillet AC, Saïdou AA, Bezançon G. 2011. Selection for earlier flowering crop associated to climatic variations in the Sahel. PlosOne 6(5): e19563.

Vitalis R, Dawson K, Boursot P 2001. Interpretation of variation across marker loci as evidence of selection. *Genetics*, 158, 1811-1823.

Wright S, Vroh Bi I, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD and BS Gaut. 2005 The effects of artificial selection on the maize genome. *Science*, 308:1310-1314

SP3 – Cereals in Africa: from advanced to under-utilized crops

1. Introduction

Among all crops studied by Agropolis research teams, African cereals are likely those on which the largest expertise in diversity analysis has been developed, whether through pluridisciplinary approaches at the farm level in Southern agroecosystems or through the development of genomics-oriented projects on reference materials. The ARCAD project provided the opportunity to, and could take advantage of, pulling together these various research and training experiences and skills in order to organize them into a structured set of expertise and to apply this set to the analysis of the diversity and adaptive potential of underutilized African cereals.

Our objectives were: i) to develop a comprehensive understanding of wild and cultivated cereal diversity at a continental scale, ii) to investigate potential diversity hot-spots for cultivated and wild populations with the view to refining *in situ* conservation approaches, iii) to identify the main evolutionary factors responsible for the observed structures of diversity, with a special emphasis on the domestication processes, and to propose new indicators of diversity, and iv) to obtain preliminary results on the diversity of underutilized cereals in a limited number of study sites and develop a research initiative on these crops based upon a multi-stakeholder research network.

Using a transversal approach, agricultural systems were analyzed to understand the dynamics of crop diversity. Surveys were conducted intensively in 4 countries (Burkina Faso, Guinea, Morocco, and Kenya), focusing on crop diversity while analyzing the relative cultural and agronomic importance of species, as well as the temporal dynamics of agro-ecosystems. This approach allowed us to better understand the biological and socio-economic factors that affect the distribution of crop diversity.

The first section (A) highlights some activities and results at different scales: the temporal dynamics of *Oryza glaberrima* at a regional scale, the distribution at national scale of maize in Burkina Faso, the altitudinal distribution of durum wheat in Morocco, and, at the local scale using ethnic groups contact zone, the social distribution of sorghum in Kenya.

In the second section (B), a transversal approach (comparing sites and crops) summarizes the key results on crop diversity distribution and its temporal dynamics from 1970 to 2001.

In addition to the specific innovative and the transversal approach, complementary studies were

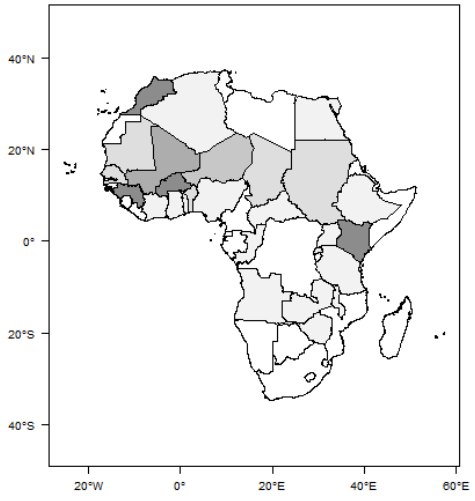


Figure SP3-1. The map displays the countries of the four intensive study sites (in black) and of origin of complementary genetic data (in grey) used for the integration with worldwide and regional data. Intensity of grey is proportional to the number of accessions compared across coutries for maize, pearl millet,

implemented in order to compare for each species the crop genetic diversity, which was analyzed in Arcad SP3 project, with the continental diversity (**Figure SP3-1**). These preliminary results are summarized in the third section (C).

2. Activities carried out and main results

Section A: Crop studies

This series of studies addresses the question of the temporal and spatial structure of crop diversity at various geographical scales and for a set of cereals that differ in their history (indigeneous/non-indigeneous, domestication process, uses) and life history traits (mating system).

a) Temporal dynamics of *Oryza glaberrima* varietal diversity in West Africa

Going back to the notebooks of the rice collecting campaigns in Burkina Faso, Botswana, Cameroun, Chad, Cote d'Ivoire, Guinea, Guinea Bissau, Mali, Nigeria, Senegal, Tanzania, and Zambia during the 70's, passport data for some 860 *O. glaberrima* accessions collected were digitized and partially geo-localised. Based on above mentioned update of information on 70's collects, three countries/sites (Burkina Faso, Guinea and Senegal) were selected for the analysis of dynamics of *O. glaberrima* varietal diversity. This analysis included (1) characterisation of social, farming and cropping system dynamics and hydro-climatic evolutions of the last 30 years, at the regional level, and (2) detailed characterisation of rice varietal dynamics at the village and farm level, and current status of *O. glaberrima* varieties.

Main results

Survey results from the Cascades region of Burkina Faso indicated drastic change in the status of *O. glaberrima* varieties, from the dominant position in the 1960s (almost only rice species cultivated) to an extremely marginal position nowadays. Factors determining this decline are deeply linked to the agrarian dynamics at the regional level characterised by introduction of cash crops and the intensification of rice cropping systems. But socio-cultural factors also play a non-negligible role: cultivation of *O. glaberrima* is associated with poverty, and old-fashioned behaviours.

Survey results from Maritime Guinea indicated contrasted dynamics. While *O. glaberrima* maintains its dominant position in the slash-and-burn rice cropping systems facing land scarcity and reduce duration of fallow (it is the species which yields well even when the bush is young), it loses importance to the benefit of *O. sativa* in lowland ecosystems and in places not facing land scarcity problem. In general *O. glaberrima* is perceived as the «economical rice for large (and somehow poor) families» whereas *O. sativa* is «the sweet and sugared rice that people prefer».

b) Maize diversity in Burkina Faso : the impact of modern varieties

In 2011, in Burkina Faso, 102 farmers of West of Burkina Faso in 43 villages were investigated by a survey on their farm, crops, maize varieties (**Figure SP3-2)**. During the survey, maize landraces were sampled. 113 of the maize landraces collected and 12 modern maize varieties were analyzed for 19 nSSR and seven chloroplastic SSR (cpSSR hereafter) markers in Montpellier. Nuclear SSR markers were identical to those used for a collection of American and European landraces analyzed by UMR-GV team, so that genetic origin of maize landraces of Burkina Faso could be studied. To complete the study of the origin of African maize, 133 African landraces of IITA, 144 Americain and 135 European landraces were also analyzed with seven cpSSR markers. Landraces and modern varieties collected in Burkina Faso were characterized with 24 phenotypic traits (morphology of plants and ears, earliness, yield components) in two trials in Burkina Faso (one well irrigated trial, one water deficit trial) and in three sub trials according to the maturity of the accessions (extra early, early, intermediate). Genetic values of accessions were estimated by a mixed model with the software Asreml, and the drought tolerance of accessions was evaluted..

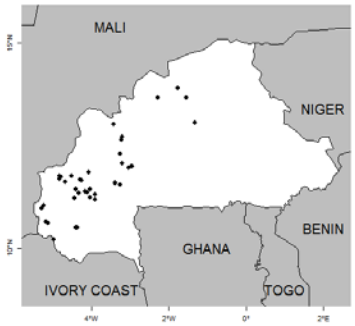


Figure SP3-2. Location of villages surveyed for maize in Burkina Faso

The 30 maize landraces sampled in Mali were analyzed with the same molecular markers. Due to a delay in seed transfer and difficulties in communicating with our partner in Mali, the Malian landraces will be phenotyped in 2013.

Main results

Around half of surveyed farmers conduct a semi-intensive agriculture (with fertilizers and animal for plowing), 30% a traditional agriculture (with manual or animal plowing, few fertilizers), and others conduct a mixed farming between semi-intensive and traditional. Farmers cultivate in average 1.88 varieties of maize. Around 26% of maize cultivars are released varieties. 2/3 of farmers tried modern varieties at least once.

Molecular data were treated in a matrix of presence/absence of alleles for the calculation of allelic richness and a matrix of allelic frequency for the calculation of Roger genetic distance. Mean allelic richness for landraces (2.86) was close to the allelic richness of modern varieties (2.85). The 12 modern varieties are composite varieties with a broad genetic base. Allelic richness per locus was slightly higher in landraces from the South of Burkina Faso than for the North where fewer modern varieties were released (**Table SP3-1**). We performed a PCA on allelic frequencies (**Figure SP3-3**). Modern varieties were close to the landraces on the first plan. Sanou (1996) compared landraces collected in 1994 from west of Burkina Faso to the diversity of the variety SR22 with isozymic markers. On the first plan of the PCA performed on allelic frequencies, SR22 was globally well differentiated from the landraces. In our study conducted in 2011, the differentiation was no longer observed. We then hypothesize that the landraces were partly introgressed by modern varieties.

Table SP3-1. Genetics characteristics of Burkina's maize accessions per surveyed zone (Richall : allelic richness, Hw : within population heterozygosity, He : Heterozygosity, Gst)

Zone	Richall	Hw	He	Gst
Total	6.29	0.44	0.58	0.21
Center	4.30	0.46	0.49	0.23
North	3.65	0.44	0.57	0.06
South	4.52	0.46	0.57	0.19
Varieties	4.47	0.46	0.56	0.17

The first plan of the PCA showed a weak structure of the set of accessions. The village of origin and the type of kernel do not structure the diversity. The only structure found on nuclear markers was the group made of landraces of the North zone (near Sahel) which were early for flowering time. We used Darwin program to plot the genetic distance between the populations. We also observed that early landraces tended to cluster together.

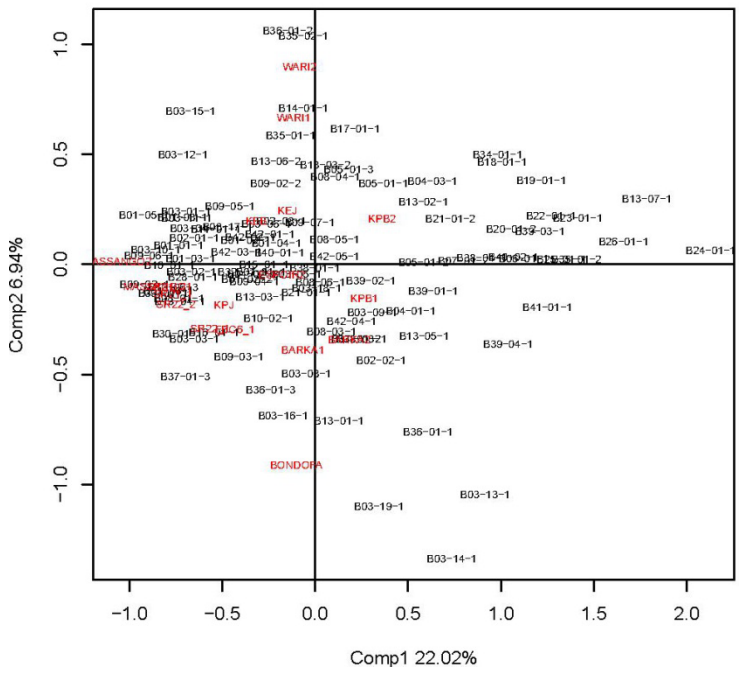


Figure SP3-3. Distribution of landraces (in black) and varieties (in red) of Burkina Faso on the first plan of PCA based on allelic frequencies.

We analyzed the structure of Burkina accessions with the software DAPC and Structure. With DAPC, one cluster, on the 3 found clusters, grouped early accessions. With Structure, the admixture rate was very high and the landraces from the North zone were grouped together.

The weak structure found on nuclear markers was also found on cytoplasmic markers as it is showed in **Figure SP3-4**. The cytotype G was the more frequent. There was no correlation between nuclear genetic distance and cytoplasmic genetic distance (Mantel test).

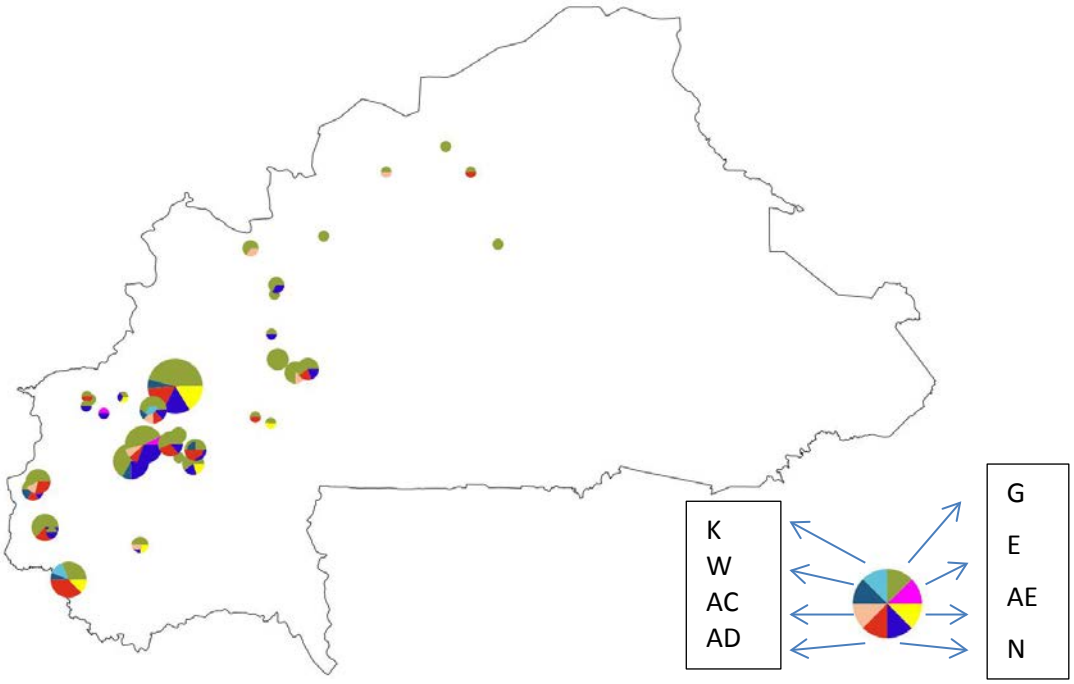


Figure SP3-4. The geographical distribution of cytotypes in Burkina Faso populations.

Phenotypic trials showed genetic variability among accessions for most of traits. Flowering date and maturity contributed most to the first PCA axis. From the two trials (well watered and water stress), we evaluated the drought tolerance of landraces vs modern varieties. Water stress has been severe with 59-71% loss performance on yield. For each sub trial, we calculated an index of selection on drought susceptibility value of grain yield and yield components whose weight is a function of the heritability of the trait and its correlation with yield. In extra early sub trial, the modern variety Wari was the best followed by two landraces. In the early sub trial, the modern varieties had an average ranking; few landraces were tolerant to water stress. In the intermediate sub trial, seven landraces had a selection index upper than modern varieties. Landraces which are tolerant to drought stress can be a source of variability for a breeding program.

We also compared landraces and modern varieties on biochemical values estimated by near-infrared spectroscopy (NIRS). Again, the modern varieties were not separated from the landraces on the first plan of PCA. Few modern varieties were characterized by the morphology or the constitution of kernels. Three landraces had a protein content upper than 13.5%.

Nuclear and cytoplasmic DNA analyzes showed a weak structure between maize landraces. Only landraces from the North part of the country clustered together on nSSR allelic frequency and morphological analyzes. Comparison between landraces and modern varieties suggested gene flow between the two types. The most anciently released modern varieties, such as SR21 and SR22, were likely creolized. Despite the influence of modern varieties, allelic richness of landraces was still high. Genetic erosion will become a risk if farmers start to grow, near their local populations, a limited number of modern varieties with a narrow genetic base. It is time to collect the landraces in order to preserve their diversity as guaranty of genetic progress.

c) Durum wheat diversity in Morocco

In Morocco, two regions were chosen, based on previous knowledge on the occurrence of traditional varieties of durum wheat: i) the Pre-Rif region, with rainfed agriculture and relative homogeneity of environmental conditions, and ii) a valley and satellite areas in High Atlas, with irrigated agriculture and a marked gradient of altitudes (from 1300 to 2300 meter absl). Each region was subdivided into 4 or 6 zones, according to geographical proximity (**Figure SP3-5**).

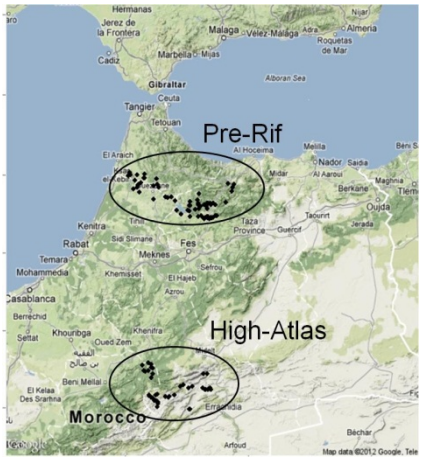


Figure SP3-5 : Location of the villages surveyed in Morocco

The interviews (n=264) focused on 3 main topics: (i) characteristics of the farmers and their farm (socio-cultural and agricultural features), (ii) durum wheat cultivated varieties (name, description), (iii) management practices, seed origin and destination, and selection. Results of the interviews were described using descriptive techniques to perform simple two-way and multi-way tables (Multiple Correspondence Analysis)

A subsample of the cultivated populations was constituted (63 populations in the Pre-Rif, 101 populations in High Atlas): for each population, 30 ears where collected at random within the field at harvest time. Each of them was described for morphological traits. After threshing, grain protein content was measured using NIRS analysis. In 2011, a field experiment was conducted in Rabat. Seeds from each sampled ear were sown in lines and the plants were described for three plant traits (height, heading date and spike shape).

Based on the spike shape and, when necessary, flow cytometry, each line was assigned to bread wheat (*Triticum aestivum*) or to a durum wheat subspecies (*Triticum turgidum* ssp. *dicoccum*, ssp. *carthlicum* or ssp. *durum*). Based on the plant height, each line was assigned to dwarf or non-dwarf class. One seed per population, excluding individuals scored as bread wheat, was genotyped with 14 nuclear SSR markers (nSSR hereafter).

In parallel a collection (conserved in the lab) encompassing the diversity of durum wheat subspecies (495 individuals) was analysed with the same nSSR markers. Genetic data were analysed with GENETIX, DAPC and DARwin, in order to compute classical population genetic indices of diversity and differentiation, analyse the structure of diversity within Morocco, with respect to the geography and the varietal names, and compare the Moroccan collection to the international collection.

Main results

- Varietal diversity

Durum wheat varietal diversity is presented in **Table SP3-2**. 19 different varieties were cited by farmers in the North among 251 censed populations, and 14 in the South, among 122 populations.

Table SP3-2 : Statistics describing the prevalence of durum wheat cultivation and its varietal diversity

Zone	Mean farm area devoted to durum wheat (ha)	Average farm richness	Zone richness	Average farm evenness	Zone evenness	Average divergence (between farms within zones)	Percent of modern varieties among the censed populations (%)	Percent of area devoted to modern varieties among the total surveyed area (%)	Percent of farms without modern varieties
ZN1	1,715	1,35	8	0,145	0,61	0,761	20,3	17,4	72,5
ZN2	0,745	1,30	5	0,125	0,59	0,789	38,4	57	50
ZN3	1,722	1,77	13	0,284	0,80	0,647	36,8	51,2	38,4
ZN4	1,225	1,06	2	0,031	0,40	0,922	0	0	100
North Pre Rif	1,611	1,54		0,206			29,9	36,2	55,8
ZS1	1,007	1,13	5	0,061	0,67	0,910	11,8	5,3	86,7
ZS2	0,520	1,27	8	0,132	0,76	0,825	14,3	8,7	81,8
ZS3	0,794	1,04	5	0,019	0,70	0,974	0	0	100
ZS4	0,750	1,00	4	0,000	0,68	1,000	0	0	100
ZS5	0,783	1,83	6	0,361	0,77	0,533	45,4	42,6	33,3
ZS6	7,891	1,30	2	0,138	0,38	0,631	0	0	100
South High Atlas	2,394	1,21		0,096			9	1,6	90,1

Farmers clearly distinguished modern varieties from the traditional ones based on the descriptors related both to plant cycle and grain quality traits.

We identified some individuals or whole populations as bread wheat, whereas farmers considered and named them as durum wheat. Spike morphologies indicative of diverse durum wheat subspecies were also identified, including ancestral forms (ssp. dicoccum). This diversity was observed within and among populations. There was variation among regions (some types were more frequent in one zone/region compared to others), and among varieties (some are durum or dicoccum types) **(Figure SP3-6)**.

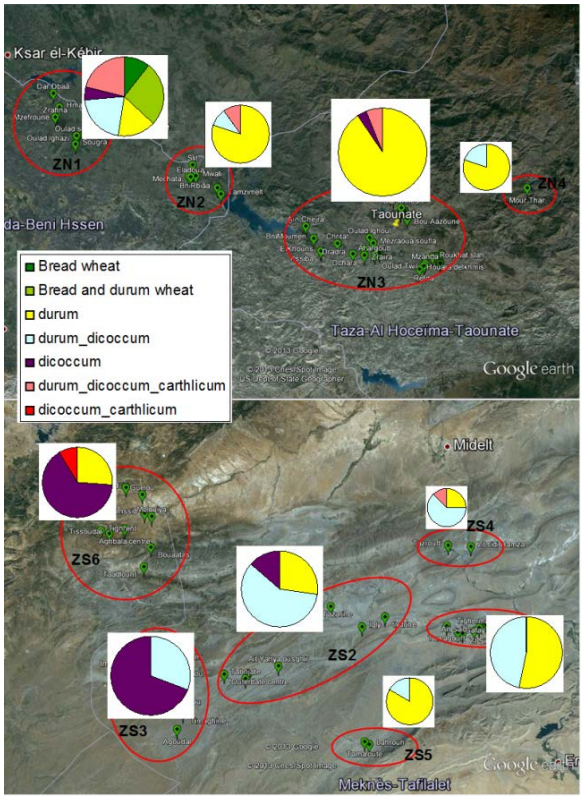


Figure SP3-6: Proportion of different types of wheat populations in the surveyed zones. Each population was assigned to a type according to its taxonomic composition

The genetic diversity of the Moroccan sample was compared to the diversity of a worldwide collection including the different domesticated subspecies of Triticum turgidum. The Moroccan genetic diversity was limited and quantitatively comparable to that found in the subset of the collection made of modern varieties. Moreover, using a cluster analysis (DAPC), the Moroccan sample clustered with a group of accessions including mainly free-threshing forms of T. turgidum, without apparent proximity with the ancestral form ssp. dicoccum. As a whole, the genetic diversity did not reflect the variation detected on spike morphology.

- Genetic diversity and differentiation

The genetic differentiation (*F*st) between the two regions was high (0.19). The fine analysis of multilocus genotypes showed that variety names were not strictly related to the genotypes. A given name most often included different genotypes. Reciprocally, a given genotype could correspond to different names. Some groups of varieties were grouped together, suggesting that different denominations actually corresponded to a single varietal group. In High Atlas, such a group was constituted with varieties of different names (Aberyoun and Ifermough) belonging to two zones known to regularly exchange seeds (ZS3 and ZS6). (**Figure SP3-7**).

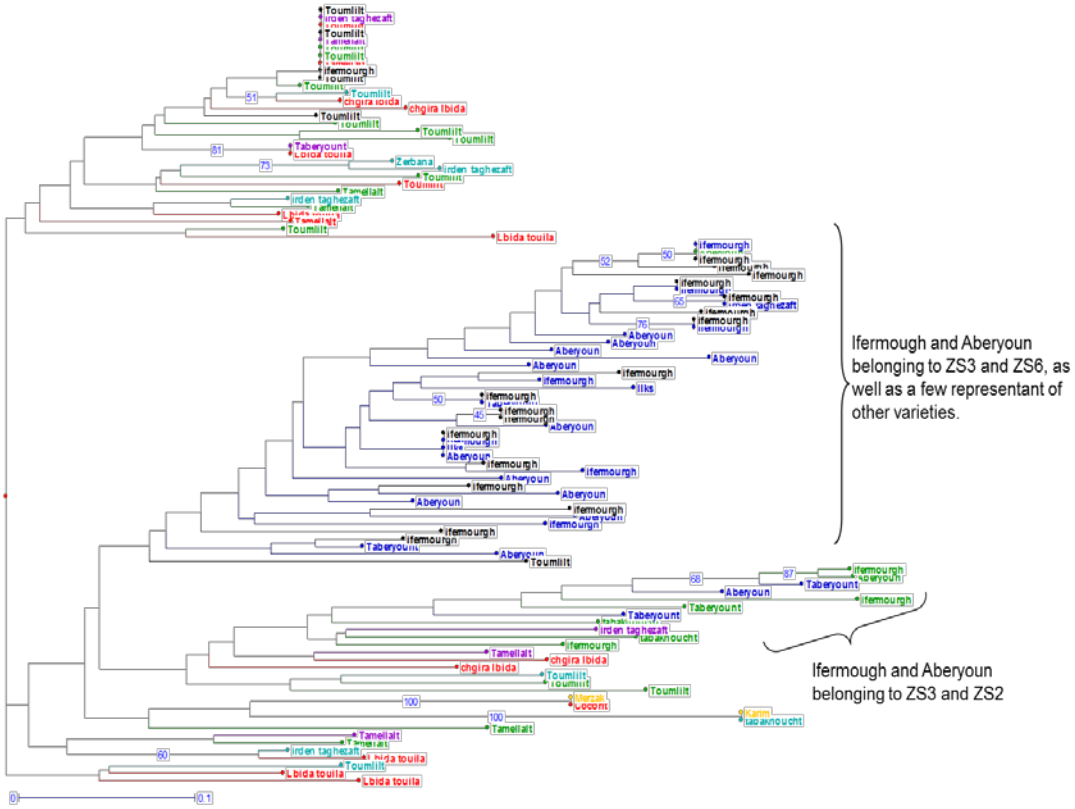


Figure SP3-7: Neighbor joining tree based on dissimilarities between multilocus genotypes (DARwin software). Each node is labelled by the variety names. The colours correspond to the study site subzones: red: ZS1, green: ZS2, blue: ZS3, light blue: ZS4, violet: ZS5, black: ZS6. Orange corresponds to two common modern varieties used as reference genotypes

In this region, two genetically and morphologically well-differentiated groups of varieties corresponding respectively to *dicoccum* and *durum* spike shape types were distributed among the 6 zones. One group (*dicoccum* flat spike) was mainly present in the highest area (ZS3) whereas the second group (identifiable through its squared *durum* spike) was largely represented in lower areas (ZS1, ZS2, ZS4 and ZS5). These 2 two types coexisted in the last area (ZS6) which is an important area for seed production. This suggests that farmers' practices maintain these two groups well-differentiated. In this region, the adoption of modern varieties remains rare: only 10% of farmers grew three modern varieties which covered only .5% of the durum wheat area.

The situation differs in the Pre Rif region where the adoption of modern varieties was much more important than in the High Atlas: the five modern varieties recorded covered 36% of the area allocated to durum wheat and were used by 44% of farmers. This importance of modern varieties could be related to the farmers' need to renew their seeds after bad climatic years, coupled to the difficulty to get seeds of traditional varieties. These modern varieties were cultivated in the same agricultural practices than the traditional varieties and the presence of identical, or similar genotypes shared between modern and traditional varieties suggests that they are "creolized".

d) Uneven distribution of crop species and sorghum varieties among neighborhood groups

In Kenya, an ecological anthropology approach was adopted to investigate the relation between crop diversity patterns and the social organization of *Tharaka* farmers. The *Tharaka* are organized in neighborhood-groups (*ntora*), in clans and in age-sets (Middleton 1953). We quantified the influence of these three major social institutions on both crop species and sorghum landraces diversity patterns, measured through the richness and composition of cropping systems. The study site was approximately ten square kilometers. It was selected for its edaphic and topographic uniformity, with a constant altitude of 700m above sea level (± 50 m). Individual interviews were conducted in 95 households randomly sampled (**Figure SP3-8**).

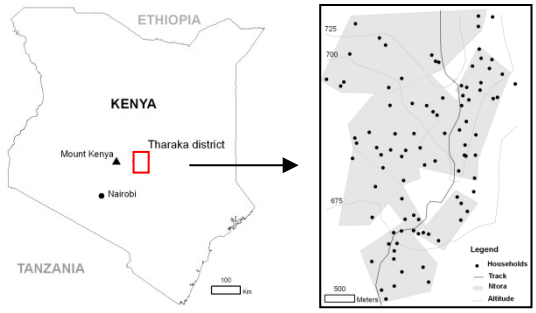


Figure SP3-8. Location of farmers surveyed in Tharaka district, Kenya.

The interviews focused on: (i) the social information concerning the household heads, (ii) the list of the crop species planted in the October 2010 cropping season, (iii) the inventory of sorghum landraces and the main source of each seed lot for the same season. We tested whether a relationship existed between the number of crop species cultivated per household (i.e. species richness) and the *ntora*, the clans, and the age-sets of farmers using a Poisson loglinear model followed by a Tukey's Honestly significant difference test. Household's crop species and landraces composition, which is the assemblage of crops that are cultivated together in the same household, were compared between *ntora*, clans or age categories using a Between-Class correspondence analysis (Chessel et al. 2004). This multivariate ordination analysis tested whether crop composition was more similar within groups than between. The significance of these differences was assessed using a Monte-Carlo test with 9999 simulations.

Main results

No significant relation was found between the household species richness and the clan or age class of the household head, whereas a significant relation was found with the *ntora (*neighborhood-groups). Indeed, species richness significantly differed between adjacent

ntora (Tukey HSD test p-value < 0.01).

There were significant differences for both species and sorghum landraces compositions between *ntora (*Monte-Carlo test p-value: species = 0.0171; landraces = 0.0010). This factor explained 9% of the total variance of species composition and 11% of the total variation in sorghum landraces composition (**Figure SP3-9**).

No significant differences were observed between clans and age categories neither for species (respectively for each factor: p-value = 0.0977 and 0.6238) nor for sorghum landraces (p-value = 0.7254 and 0.3915). This study confirmed that social organization of farmers shapes organization of crop diversity (Leclerc and Coppens d'Eeckenbrugge, 2012)

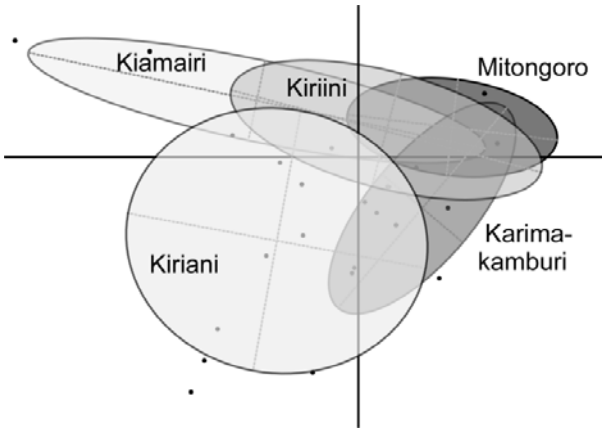


Figure SP3-9. Graphic display of the Between Class analysis for sorghum infraspecific composition between *ntora*: projection of farms' portfolios similarity on axis 1 and 2 of the Between-Class analysis with ellipse and gravity centers of each *ntora.* The first and second projection axes represented respectively 53% and 23% of the between-*ntora* variation

e) Uneven distribution sorghum varieties and their genetic diversity among ethnic groups

To assess the effect of social boundaries on the spatial distribution of sorghum diversity, a study was conducted in the contact zone between the Chuka, Mbeere and Tharaka ethnolinguistic groups in the Mount Kenya region **(Figure SP3-10**). Sorghum varieties were inventoried and samples collected in 130 households. In all, 297 individual plants derived from seeds collected under sixteen variety names were characterized using a set of 18 SSR molecular markers The genetic diversity of sorghum populations sampled in each ethnic group was assessed using several indexes. The genetic structure was investigated using both a Bayesian assignment method and distance-based clustering. The distribution of the varieties and the main genetic clusters across ethnolinguistic groups was described using a non-parametric MANOVA and pairwise Fisher tests.

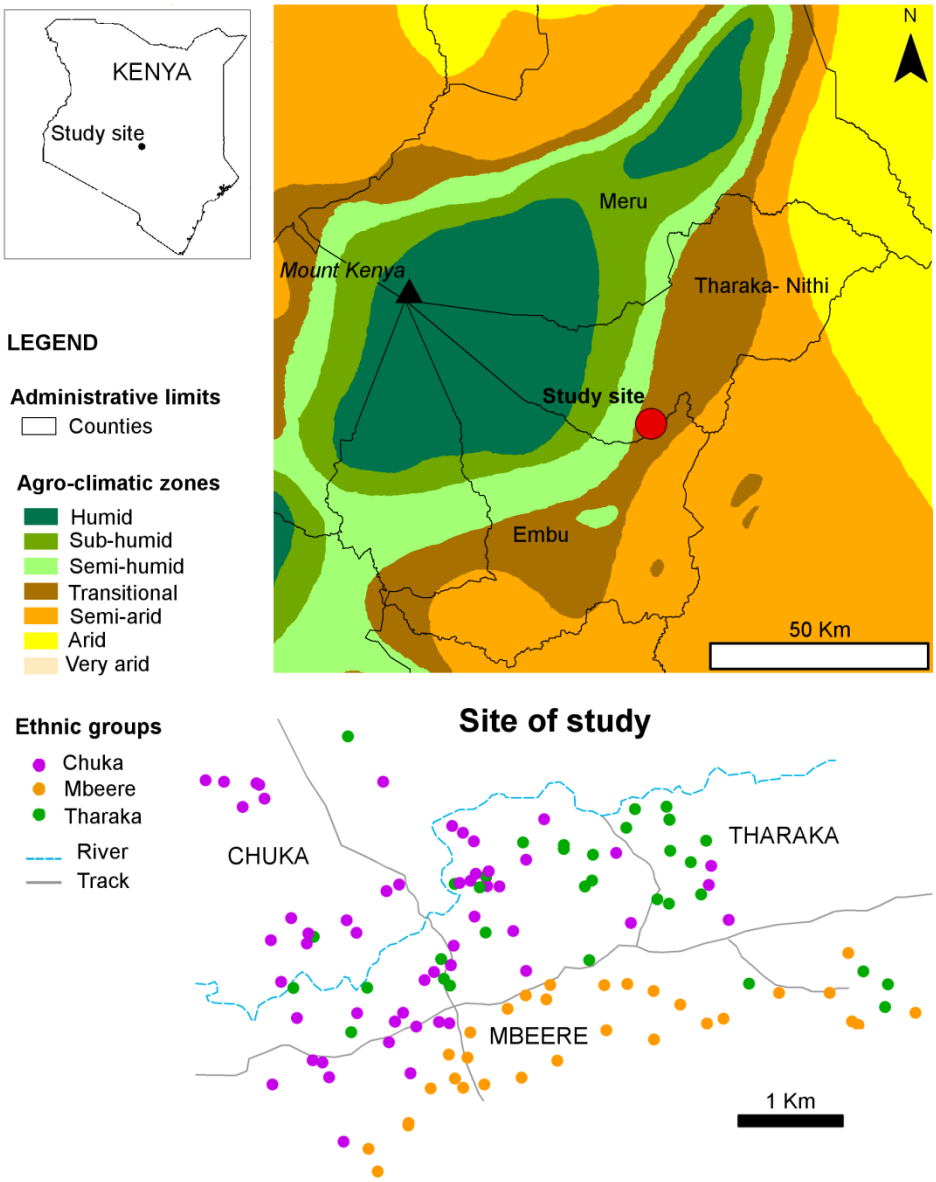


Figure SP3-10. Study site location. Map of the eastern side of Mount Kenya and location of the farms where sorghum samples were collected (colors correspond to the ethnic identity of the male household-head).

Main results

14 different varieties were respectively inventoried in the Chuka and Tharaka groups, and 10 in the Mbeere group. The non-parametric perMANOVA showed that sorghum variety assemblages differed significantly between ethnic groups, even though the ethnic partition explained a limited part of variability (pseudo-F2,121 = 4.971; p-value = 0.0002; R2 = 0.076). Pairwise Fisher exact tests confirmed that the frequency of three out of the five most frequent landraces differed significantly among ethnic groups, while the frequency of improved varieties (Gadam, Kaguru and Serendo) did not differ significantly among ethnic groups (**Figure SP3-11**).

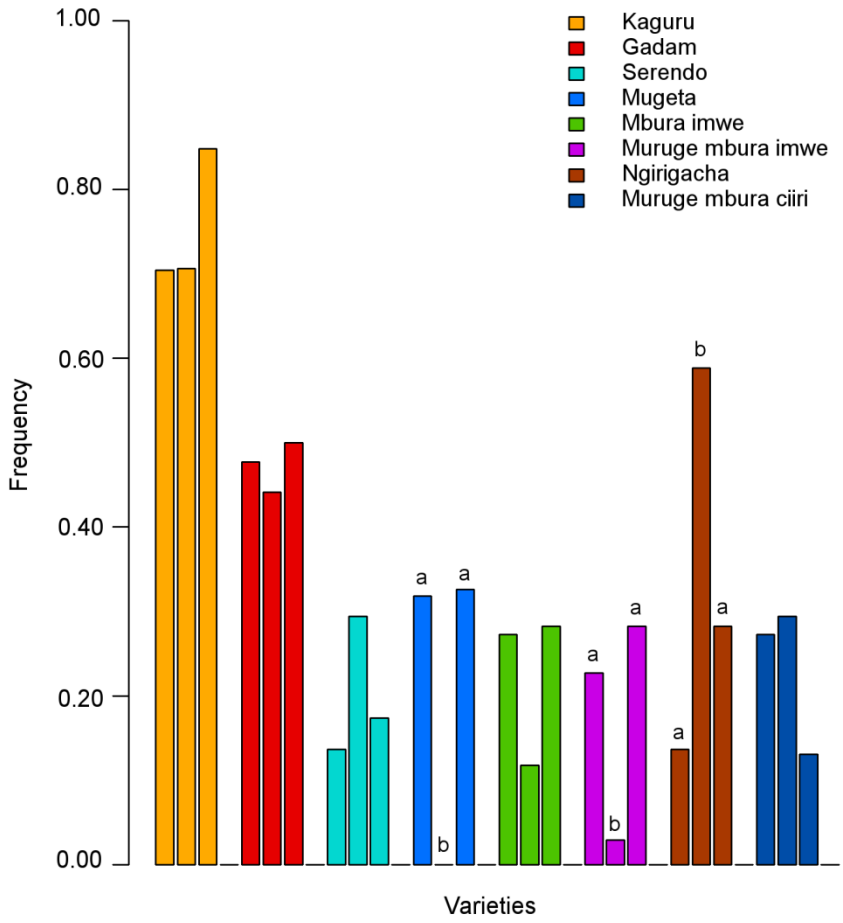


Figure SP3-11. Frequency of the eight major varieties in each ethnic group. The vertical axis displays the percentage of farms where each variety was cultivated. Ethnic groups are present in the following order for each variety: Chuka, Mbeere, Tharaka. The letters (a, b) on top of the bars indicate the statistical significance of differences (Fisher test) at a 5% level after correction for multiple testing (FDR).

The unbiased gene diversity estimates (He) of Chuka and Tharaka sorghum populations were significantly higher than that of the Mbeere (Wilcoxon test: p-value = 0.01). Similar results were found for the unbiased allelic richness. FIS was very high in the three groups, yet it was significantly lower in the Chuka population as compared to those of both the Tharaka and the Mbeere (Wilcoxon test: p-value = 0.05 for both pairwise comparisons), in relation with the higher heterozygosity found within the Chuka sorghum population (0.033) compared to the other two populations (0.022 for the Mbeere and 0.023 for the Tharaka). An exact G-test of genetic differentiation of sorghum across ethnic groups was significant (p-value = 0.0205). The Pairwise G-tests showed that genetic differentiation was highly significant between the sorghum populations of the three groups, being highest between the Chuka and both the Tharaka (p-value = 0.0001) and Mbeere (p-value = 0.0002) populations and lowest between the Tharaka and Mbeere populations (p-value = 0.0083). The FST values between the sorghum populations of the three ethnic groups were low: 0.027 between the Chuka and Mbeere sorghum populations and 0.019 between the Chuka and Tharaka populations, both significant; and non-significant between the Mbeere and Tharaka populations (FST = 0.010).

The genetic diversity of sorghum in the area of study, as assessed with molecular markers, is organized in four major groups (**Figure SP3-12).** These groups reflected the influence of improved variety dissemination and a differentiation in terms of cycle duration and phenology. Long-cycle landraces formed a genetically distinct cluster which was more frequently encountered in the Chuka sorghum population than in the Tharaka sorghum population. The improved Kaguru variety showed more admixture with the local landraces in the Chuka sorghum population than in the Mbeere and Tharaka ones. As a result of the unbalanced frequency of the different genetic clusters across ethnic groups, the genetic differentiation of their sorghum populations was significant.

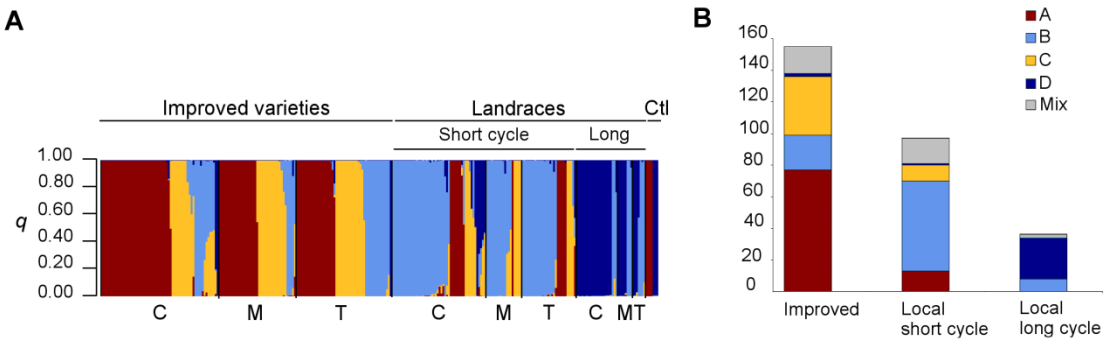


Figure SP3-12. Genetic structure of the sorghum cultivated on the area of study. *(A) Cluster assignment of 297 sorghum individuals estimated using STRUCTURE for K = 4. The genome of each individual is represented by a vertical line, which is partitioned into K colored segments that represent the admixture coefficient (q), i.e the estimated proportion of membership of its genome in each of the K clusters (Red: cluster A, light blue: cluster B, yellow: cluster C, dark blue: cluster D). Thick black lines separate the individuals identified by farmers as improved varieties, short-cycle landraces or long-cycle landraces, and control individuals (Ctrl), as labeled above the figure. Thin black lines separate individuals sampled in the different ethnic groups (Chuka: C, Mbeere: M, Tharaka: T, as labeled below the figure. The figure shown is based on the highest probability run at K = 4. (B) Number of individuals classified according to their origin and cycle length (farmers' information) assigned to each MMb genetic cluster. The vertical axis indicates the number of individuals assigned to each cluster. Individuals were assigned to a cluster when their estimated admixture coefficient (q) for this cluster was equal to or over 0.8. Admixed individuals are represented*

Section B: Transversal approach. Combining social and biological sciences

A common protocol was implemented in Burkina Faso, Guinea, Kenya and Morocco, in order to characterize trends in the crop and varietal dynamics of cropping systems reflecting various agroecological and economic conditions. For a total of 1412 farmers, 932 men and 480 women, were interviewed in Burkina Faso (100), Guinea (840), Kenya (208) and Morocco (264). Across the countries, 208 villages were visited. Farmers belong to 56 different ethnic groups.

The list of all species cultivated by farmers along their life was recorded by using the independent interview technique, whereby each farmer was interviewed individually and not in a group setting. Thus the responses given by an individual farmer were not influenced by those given by a different farmer. The area cultivated by each farmer was estimated from farmers' statements, and analyzed as ordered qualitative variables. The cropping system dynamics was assessed considering the crop assemblage of each local farming system from 1970 to 2010.

Main preliminary results

The number of crops cultivated per farmer is 4.8 in Morocco, 5.2 in Kenya, 6.6 in Burkina Faso and 6.9 in Guinea. Each study site has its specific crops. Fonio is mainly cultivated in Guinea and durum wheat in Morocco, while groundnut is cultivated by less than 20% of farmers (**Figure SP3-13**). Sorghum is usual in all countries, except Morocco.

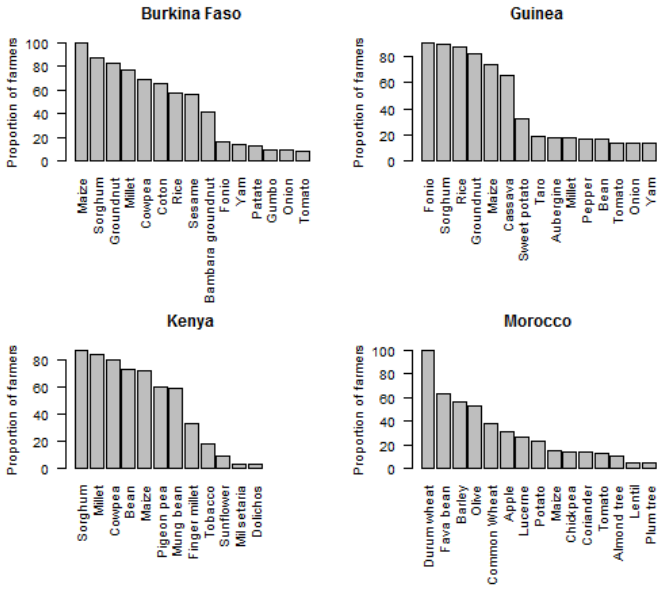


Figure SP3-13: Crop species popularity. Y axis: proportion of farmers (%) per cultivated crop in the 4 surveyed countries

The mean number of crops per village varies from 6 to 28 in Guinea (**Figure SP3-14**).

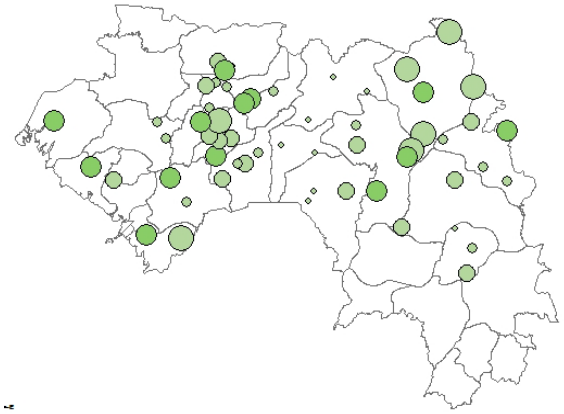


Figure SP3-14: Number of crops cultivated per village in Guinea. Circles represent the number of species per village (minimum 6, maximum 28)

Considering all surveyed sites, farmers have cultivated 6.8 crops along their life on average. The average number of crops still cultivated today is 6.4 crops per farmer. So, on average 0.5 crops per farmer were abandoned from 1970 to 2011. The number of abandoned species increases from 2000, but this result must be weighted by the number of farmers to be interpretable (**Figure SP3-15**).

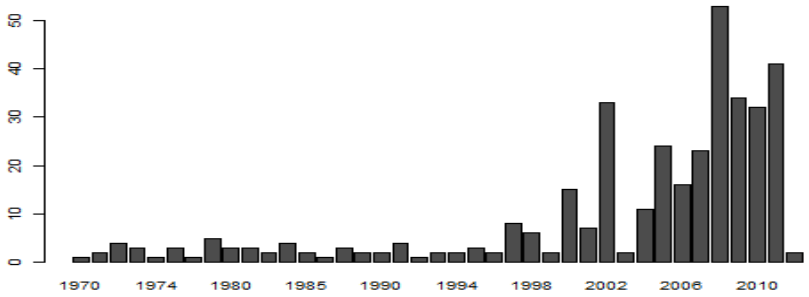


Figure SP3-15. Frequency of crop abandons in from 1970 to 2011 (all sites of the 4 countries included)

Cultivated crops can be distinguished by their status: local vs introduced, food vs cash crops, minor vs major crop. Unanimously, farmers consider that fonio and yam are 'local' crops, whereas rice is considered as an introduced crop. Together, 75% of crop are perceived as 'local' crops, inherited or acquired from a family member.

The number of cash crops was only assessed for 100 farmers in Burkina Faso. The mean number of cash crops is 5.93 (sd= 3.68). Groundnut, rice, sorghum, maize and cotton are the most cited cash crops. Most of those cash crops are also used as food crop contributing to food security. The difference might be in variety uses.

Crop can also be classified as minor or major crop. Cropping system surveys will allow us to assess the importance of neglected and underutilized crops (NUS) in traditional low input and rain-fed farming systems. Fonio for instance is considered as a NUS. However in Guinea it is one of the most important crops for food security.

Section C: Integration with worldwide and regional data

Specific and transversal studies provide the opportunity to pool together these various research and training experiences and skills in order to organize them into a structured set of expertise and to apply this set to the analysis of the diversity and adaptive potential of underutilized African cereals.

In the Arcad SP3 project, research activities have been conducted on major crops, notably, pearl millet and sorghum, in order to refine our methodology and get a global picture of domestication and adaptation processes, using them as models. Maize, durum wheat, pearl millet, and sorghum were studied, but we present only the two latter ones.

Integration with worldwide and regional data aims to compare genetic diversity within and between countries, using international collection as reference set. For each species, a common set of SSR makers was used for both ARCAD SP3 accessions and international accessions.

a) Major species used as models

Pearl Millet: genetic diversity and domestication process in Central and West Africa

The regional variation of genetic diversity in pearl millet is poorly known. We used a large dataset of 48 populations to document this variation (**Figure SP3-16**). The 38 studied wild populations clustered into three major groups from East to West: an East group (Chad, Sudan), a Central group (Niger, Mali) and finally a Western group (Senegal, Mauritania). For wild populations, diversity was also shaped by a North/South gradient. Diversity hotspots were found in the southern limit of wild population distribution. As expected, lower genetic diversity was observed in the cultivated varieties as compared to the wild populations. However, an unexpected result was that wild population diversity was positively correlated with the level of introgression from the cultivated compartment (results not shown). So cultivated to wild population gene flow led to diversity hot-spot in the wild populations. While this introgression is largely variable in space, it suggests that the cultivated crop shaped back diversity found in their closest relatives. The North/South gradient of diversity is partly associated with the isolation from the cultivated population. The discovery that wild diversity is enhanced by cultivated gene flow calls into question the widely held assumption that cultivated gene flow is only detrimental to wild populations. Our result also indicate to be careful when doing inference about domestication history based on current wild populations, if we do not deal with current and past gene flow.

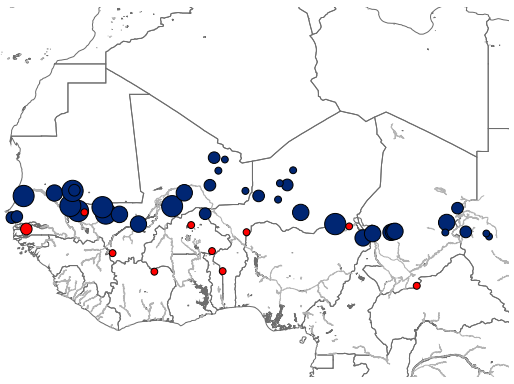


Figure SP3-16. Allelic richness variation of cultivated and wild pearl millet populations from Africa. Dot size is proportional to the allelic richness. Blue and red circles represent wild and cultivated values of diversity parameters, respectively.

Sorghum: genetic diversity and domestication process in West Africa

The previous sorghum collection assembled by Sagnard et al. (2011) in Mali in 2004 and 2005 (60 visited villages – 420 cultivated varieties and 83 wild-weedy forms) was completed in 2010 and 2011. Four regions were surveyed. Two were used to fill gaps in the existing collections (Gao region in Mali where specific varieties were grown, “sorghos de décrue et de mare”, and Nioro region, near to Mauritania where durra sorghums are abundant). Two other regions were selected to consider zones that are characterized by a large varietal diversity (Yanfolila region in South Mali and “pays Dogon” in North Mali). Collective interviews were conducted in 10 villages per zone, and all cultivated and named varieties were identified in each village. Then, each variety was provided by one farmer in one village. Around 300 samples (276 cultivated varieties and 26 wild-weedy forms) were genotyped with 28 nSSR markers, which were previously used to characterize genetic diversity of the Niger collection (484 varieties collected in 79 villages, Deu et al. 2008), the Burkina Faso collection (124 varieties collected in 10 villages; Barro et al. 2010), and the Malian collection. This set of markers was applied on the collections which have been assembled in Guinea at the end of 2012 and in 2013 (286 varieties collected in 36 villages as well as 10 wild-weedy forms) and would be applied to the collections which have been assembled in Senegal at the end of 2013 and in 2014 (165 cultivated varieties and 18 wild-weedy forms collected in 28 villages). The collections assembled in Guinea and Senegal were more exhaustive than those assembled in Mali, Niger and Burkina Faso, as we retained all the varieties grown by the 14 interviewed farmers in each village. This new sampling method was applied in order to access to the varietal and genetic diversity present in a village and to finely analyse the spatial distribution of the diversity, in relation to eco-geographical, climatic and ethnic factors.

The whole set of data allowed a synthesis of the sorghum varietal and genetic diversity in five countries of Western Africa. Genetic data were analysed with GENETIX, STRUCTURE and DARwin, in order to compute classical population genetic indices of diversity and differentiation, analyse the structure of diversity within Guinea, with respect to the geography and the ethical groups that grown the sorghum varieties, and compare the Guinea collection to the Western Africa collection.

The Malian new collection was also phenotyped in 2012 at the experimental station of Sotuba (Mali) for 14 traits (morphology, phenology) using sorghum descriptors (IBPGR), comparing 2 sowing dates.

To refine the sorghum domestication scenario, specially the hypothesis of a possible center of domestication for the “guinea margaritiferum” varieties in Western Africa, we analyzed a subset of the Malian and the Guinea collections assembled before Arcad project (around 100 accessions including both cultivated and wild and weedy forms) along with a worldwide subsample of cultivated and wild varieties maintained in gene banks with 12 cpSSR markers.

The goal of this study was to get a comprehensive understanding of wild and cultivated sorghum diversity in two contrasting regions of Africa (Western and Eastern). To achieve these objectives, additional collections in Kenya as well as genotyping of recent collections with the subset of 28 nSSR markers were envisaged. Despite a strong partnership developed in Kenya during a previous project, we were not able to realize these tasks (loss

of a part of the Kenyan collection, and the leaving of our main partner for the USA). We could yet overcome this problem by using genetic data for the genebank accessions from Eastern Africa produced by the Challenge Program Generation (Billot et al. 2013).

Main results

Genetic diversity and structure of the sorghum collection of Guinea

The numbers of farmers growing sorghum is uneven in Guinea: most Malinke farmers from the eastern part of the country cultivated sorghum contrarely to farmers from the western part belonging to the Susu people whose rice is the major cereal.

The cultivated accessions were clustered by STRUCTURE (as well as Neighbour joining tree based on dissimilarity values calculated between all accessions) into two major groups which are highly differentiated: one is mainly composed by accessions belonging to the "guinea" race (with larger grains) meanwhile the other one is composed by the "guinea margaritiferum" types (with smaller grains compared to the other "guinea"). A few accessions (around 2%) showed admixtures between these two clusters. This strong differentiation was expected and confirms the high divergence between "guinea margaritiferum" types and all other sorghums types detected in previous studies conducted at different scales (world-wide or country scale). Sorghums attributed to the "guinea margariferum" cluster are largely dominant (around 71% of the collected accessions were attributed to this cluster) but show a reduced diversity (He=0.25) compared to the "guinea"cluster (He= 0.42). However, they are grown in the whole country (from west to east) meanwhile the "guinea cluster" is restricted to the villages located in the north-eastern part of the country, which is mainly inhabited by the Malinke people (**Figure 3-17**). Our results indicate that the Malinke farmers maintain a larger diversity (growing both the two types of sorghum) than the Susu (in the west of the country), who only grow guinea margaritiferum types and highlight the Susu's preference for these types of sorghum, whose grains are usually cooked and eaten as rice in Western Africa.

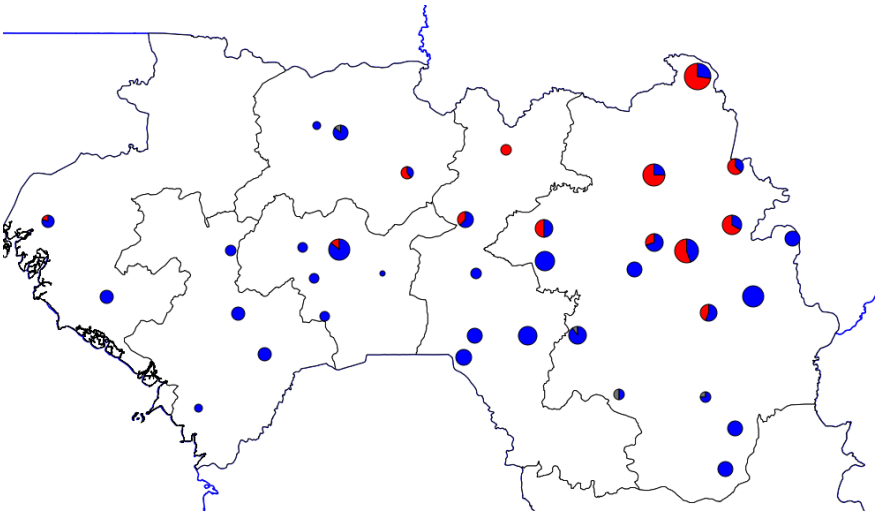


Figure 3-17. Projection of the two clusters obtained with STRUCTURE software.

Only accessions with more than 90% ancestry were attributed to a cluster. Red and blue indicate accessions attributed to clusters 1("guinea" cluster) and 2 ("guinea margaritiferum" cluster), respectively. Unattributed accessions are shown in grey. The size of the pies is proportional to the number of accessions collected per village.

A finer analysis was then performed within in each one of the two clusters. Three sub-clusters were detected within each cluster. Within the the cluster 2 of “guinea margariferum”, a sub-cluster (represented in dark blue in the **Figure 3-18**) is restricted to the north–eastern part of the country (inhabited by the Malinke people and characterized by lower annual rainfall) meanwhile the other one (in light blue) has a larger distribution going from the western (inhabited by the Susu people) to the eastern part of the country. Again, the Malinke people maintain a larger diversity within the “guinea margariterum” types compared to the other ethnic groups (especially the Susu). However, the Malinke people have a large distribution in the east of the country and only those located in the southern part (which is characterized by a high annual rainfall) shared a part of the “guinea margaritiferum” type with the Susu people. The distribution of genetic diversity within the “guinea margariferum” types appeared to be associated with both environmental and ethnical factors at the country scale. At a finer scale, the Malinké people cultivate two types of “guinea margaritiferum”, which were suspected to be associated with different climatic patterns. Complementary analyses will aim to analyse the effect of the different factors on the structure of the genetic diversity within the genetic group of “guinea margaritiferum”.

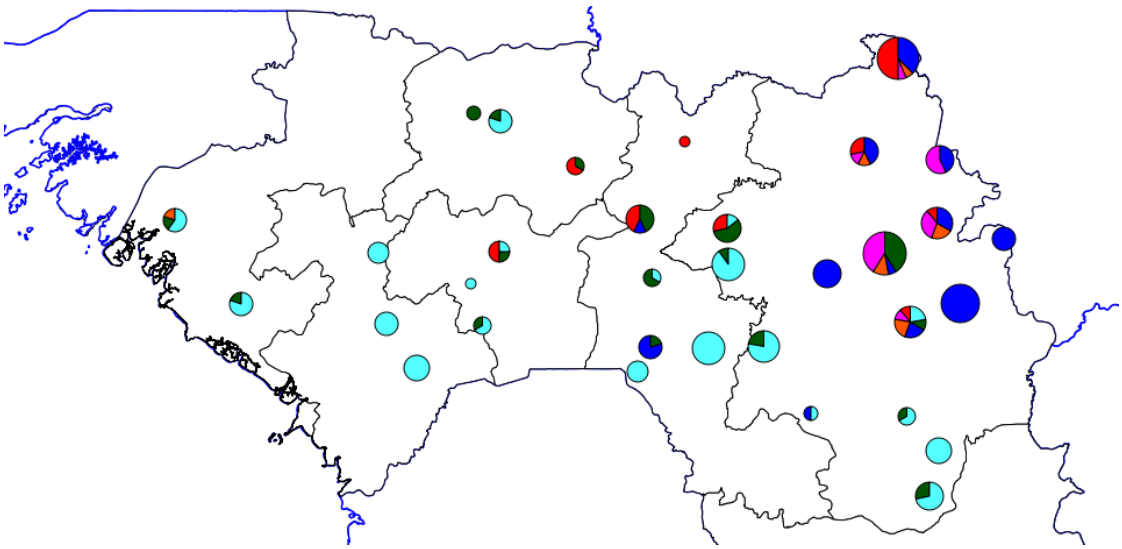


Figure 3-18. Projection of the sub-clusters obtained with STRUCTURE software when analysis was conducted separately within in each cluster defined at the upper level. Only accessions with more than 90% ancestry were attributed to a sub-cluster and represented in the map. Red, pink and orange indicate accessions attributed to sub-clusters 1a, 1b and 1c,respectively.Dark blue, green and light blue indicate accessions attributed to sub-clusters 2d, 2e and 2f, respectively. The size of the pies is proportional to the number of accessions attributed to a sub-cluster per village.

Genetic diversity and structure of the sorghum collection assembled in Western Africa

A preliminary analysis combining all the cultivated accessions collected in Burkina Faso, Guinea, Mali and Niger (1605 accessions) and genotyped with 28 SSRs indicated a contrasted pattern of genetic diversity among countries. Sorghums from Burkina Faso and Guinea are the least diverse in terms of both unbiased gene diversity and allelic richness compared to Mali and Niger in which greater racial sorghum diversity was observed. Our analyses confirm that sorghum genetic diversity seems to be associated with the botanical diversity.

The Neighbour joining tree (**Figure 3-19**) indicated that sorghum from Burkina Faso clustered in one major group composed by accessions classified as “guinea” types (not margaritiferum), based on their panicle and spikelet morphology. Only a few accessions from Burkina Faso clustered in the group of “guinea margaritiferum”. Accessions from Niger and Mali were dispersed all over the tree as expected. In these two countries, the different botanical types were found: “guinea” types, “guinea margariferum” types, “durra”, “caudatum” as well as intermediates between the botanical races. The “guinea margaritiferum” collected in Mali seemed to represent only a part of the diversity detected in Guinea for this group of sorghum. The Malian “guinea margaritiferum” types were closely associated with one cluster detected in Guinea: the cluster of “guinea margariferum” specific of the Malinke people. Our results probably indicate a common origin for these types. However, caution is need when interpreting these preliminary analyses: subsampling in the collection assembled in Guinea is necessary for a better comparison with the neighbouring countries whose collections were not so exhaustive and less “redundant” than in Guinea.

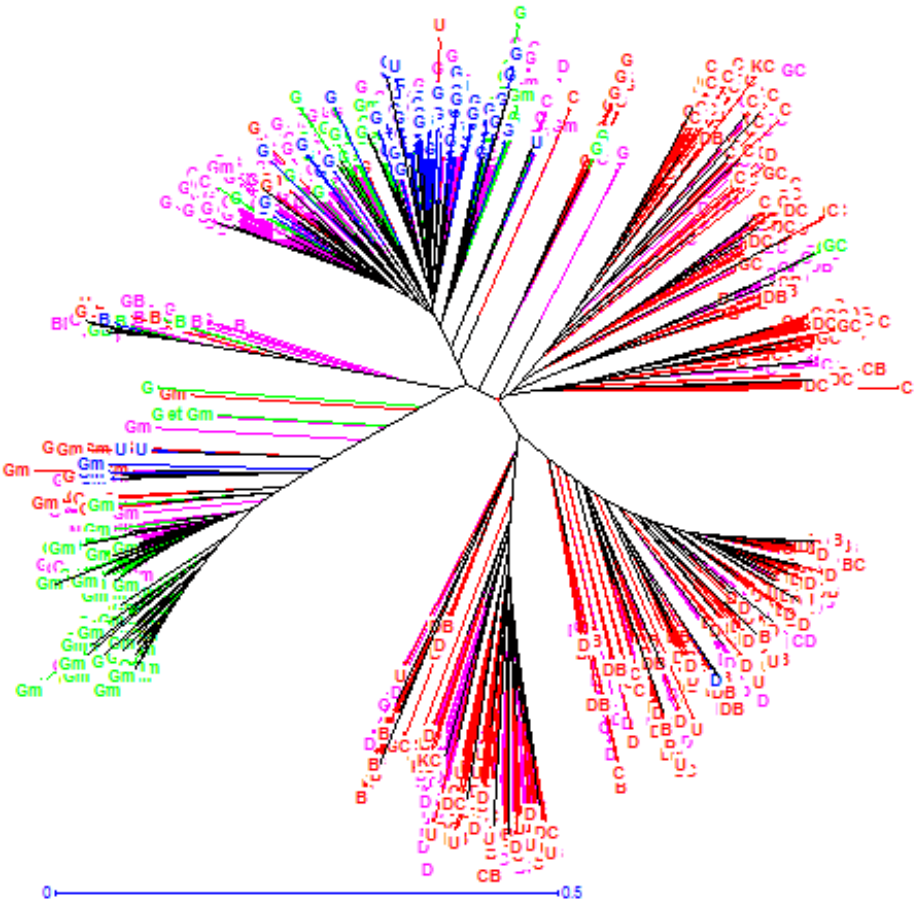


Figure 3-19: Unrooted neighbourjoining tree based on allelic data from 28 SSR loci among 1605 sorghum varieties collected in the four countries. Each country is represented by a color: accessions from Burkina Faso are represented in blue, Guinea: in green, Mali in pink and Niger in red.
Sorghum varieties are identified by their race (B: bicolor, C: caudatum, D: durra ,G: guinea , Gm: Guinea margaritiferum, XY: intermediate races between race X and Y, U: unclassified accessions).

Rice: transcriptomic and genomic sequencing to develop new resources

Based on the 2500 *O. glaberrima* accessions (IRRI, AfricaRice, IRD and Cirad), a reference set of 300 *O. glaberrima* accessions representative of the geographical diversity was jointly selected by AfricaRice and IRD. It was completed by a set of 100 accessions extracted from ORSTOM collection to represent true-wild forms of *O. barthii*. Genomic data from the NSF project OMAP (R. Wing, Tucson, Arizona) were also used in this project (CG14 accession). 384 specific SNPs were extracted from those resources as a proof-of-concept for a specific, highly efficient and fast design of SNP set for African rice diversity. The collection was genotyped in using an Illumina VeraCode technology. 235 SNP were polymorphic and allowed the correct separation in an inter-specific (*O. sativa* vs *O. glaberrima* vs *O. barthii*) as well as in an intra-specific way (**Figure SP3-20**).

First of all, the data were sufficiently precise to re-classify mixtures, off-types plants, and presumed introgressed accessions in their correct genetic category. The *O. barthii* populations are structured on a geographical origin basis (Chad Lake, Niger River, West Africa and East/Austral Africa). The main diversity observed in this survey was lying in the Lake Chad area.

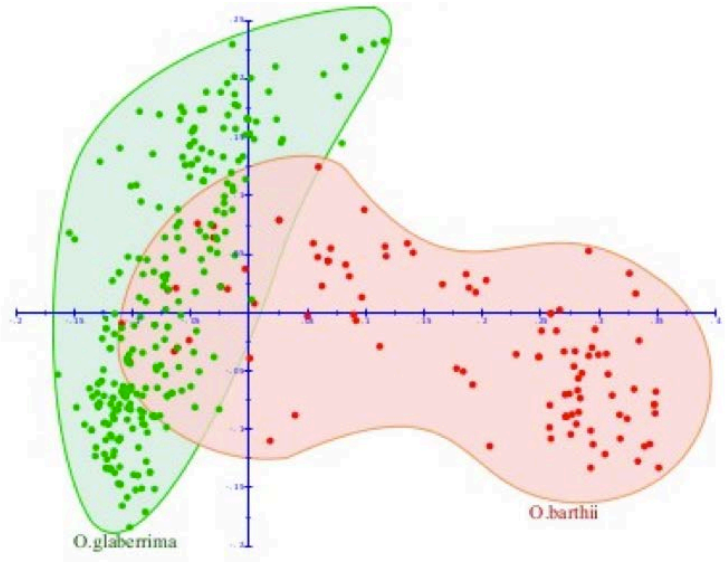


Fig SP3-20. Factorial analysis of African rice diversity (406 accessions – 194 SNP)

The *O. glaberrima* accessions showed a much lower diversity with no clear geographically structured distribution of the genetic diversity. It may come from the not yet sufficient resolution of the genetic diversity or/and from the movement of varieties after domestication, which could partially blur up the original structure. Focus was done on SNPs and other convenient RBIP markers in the S1 locus region, the main sterility locus separating the two rice cultivated species. It confirmed the absence of natural introgressions detectable at this locus and very probably the strict maintenance of the genetic isolation between the two species since the *O. sativa* introduction in West Africa.

Maize: origin of Burkinabe landraces based on nuclear SSR

From the allelic frequencies of American, European, African landraces (database of UMR GV Le Moulon), we studied the origin of Burkinabe landraces. We performed a Structure analysis: we took into account the 7 clusters (Andean, Caribbean, Corn Belt Dent, European, Italian, Mexican, Northern Flint) identified by Camus el al. (2006) on an American and European collection and we introduced Burkinabe landraces as supplementary accessions. The admixture rate of Burkinabe landraces was high. Among the 124 Burkina landraces, only 6 landraces had an assignment coefficient upper than 0.80. A few landraces, among which the populations of the North part of Burkina Faso, were grouped with the Italian cluster. This fact may be a footprint of the American origin of the Italian populations. The other landraces derived from Mexican and Caribbean clusters and to a lesser extent from Andean and Corn Belt Dent clusters. This result is to be discussed in the light of the bibliography. Porteres (1955) suggested two ways of introduction ways (Mediterranean, Egypt and the Nile; the Guinean gulf). Recently, Mir et al. (2013) used the structuration of the American populations to determine the possible origins of African landraces. Two main clusters contributed to West African landraces: middle South-America, Northern South-America. Westengen et al. (2012) reported two main origins for Western African landraces: Coastal Brazil and lowland South America. Because the Nile origin does not seem to be retained in the recent publications, we will reanalyze our data by comparison with the American continent only

Maize: contribution of cytoplasmic analysis to the structuration of American, European and African maize landraces

Twenty six cytotypes were found on the 588 samples among which 13 were present in more than 10 accessions; cytotypes G and AD were the most frequent. A cytotype (AA) was shared between Europe and the Northern part of North-America, which confirmed the introduction of North America flint maize into Europe. We found also an original cytotype (F) in Europe showing the originality of European maize. On the other side, we found no original cytotype in Africa (except cytotype AE for only 16 accessions). In **Figure SP3-21**, the African countries were not distinguished on the first plan of the factorial correspondence analysis. The 2 main cytotypes found in Africa were common in America and in Europe (G and AD). Moreover, we did not find a East-West structuration of the cytotypes as found for the nuclear molecular markers. Our study on cytoplasmic markers confirms the multiple ways of introduction of maize in Africa and the low structuration of the diversity.

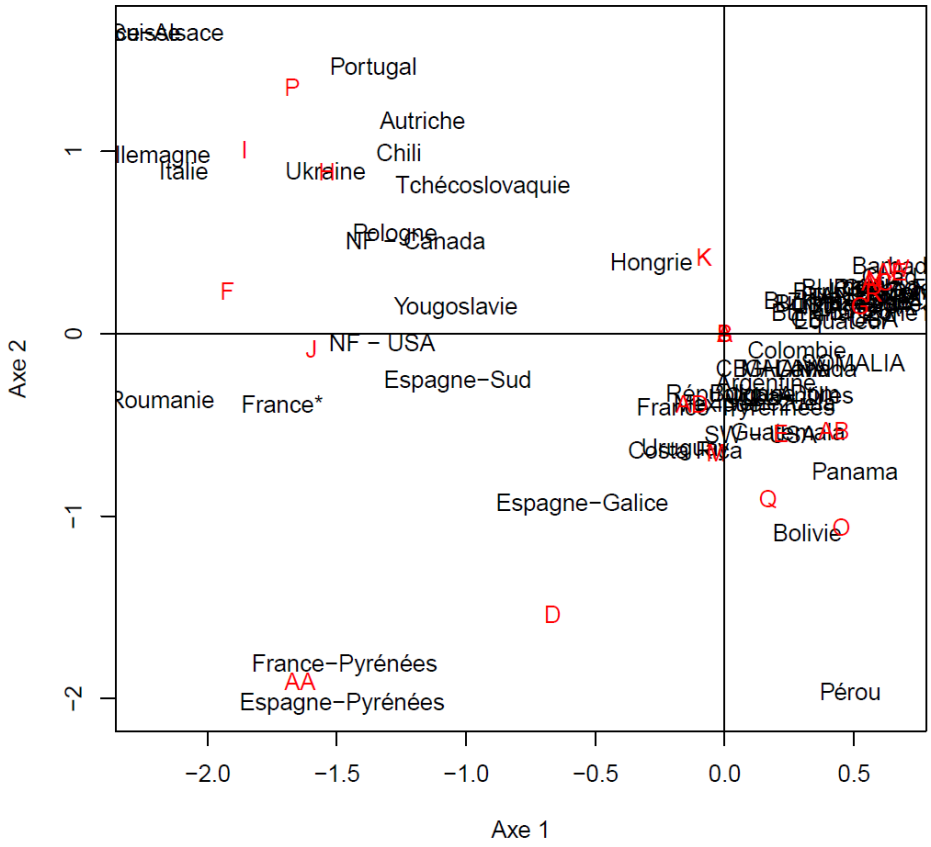


Figure SP3-21. Factorial correspondence analysis of cDNA data of global maize population provenance with name of cytotype as additional data

b) Transferring methodologies to underutilized crops to assess crop genetic diversity and adaptation

Underutilized crops do not benefit from the same research effort as major crops even if they are locally essential for food security. Fonio remains largely under-studied compared to other African cereals such as sorghum, millet and rice (Barnaud et al. 2013). So far, studies of genetic diversity focused mainly on white fonio, *Digitaria exilis* (Hilu et al. 1997; Adoukonou-Sagbadja et al. 2007) neglecting other cultivated fonio species (i.e. *D. iburua*) or wild relatives. Evolutionary and breeding disciplines are now bound to molecular techniques (e.g. phylogeny, phylogeography, landscape genetics, selection assisted by markers, association mapping, molecular breeding, etc.). These studies require genomic resources, such as molecular markers, genetic maps and sequences, which were not available for fonio. During the frame of Arcad SP3 project, 38 nSSR markers (Barnaud et al 2012), chloroplast sequences (Scarcelli et al. 2011) and 12 cpSSR markers (unpublished data) were released and new genetic resources are developed in the frame of both SP1 (expressed genome sequence) and SP5 (Genotyping by Sequencing)

We genotyped the largest collection available for fonio, including 641 accessions from 6 countries (Clement and Leblanc, 1984) with 21 nSSR (Barnaud et al. 2012) and 12 cpSSR (unpublished data). Using the same SSR set, we also conducted genetic diversity studies at the country scale in Guinea, Senegal and Niger, including reproductive system assessment. Finally we documented fonio domestication process through phylogenetic and

phylogeographic approaches of African Digitarias (ChloroDiv project).

Results show that reproductive system is most probably autogamous and that genetic diversity observed in Guinea is large compared to other countries (**Figure SP3-22**), confirming the common idea that Guinea could be a/the white fonio center or origin. Further analyses are ongoing (Structure, spatial analysis, time-scale analysis).

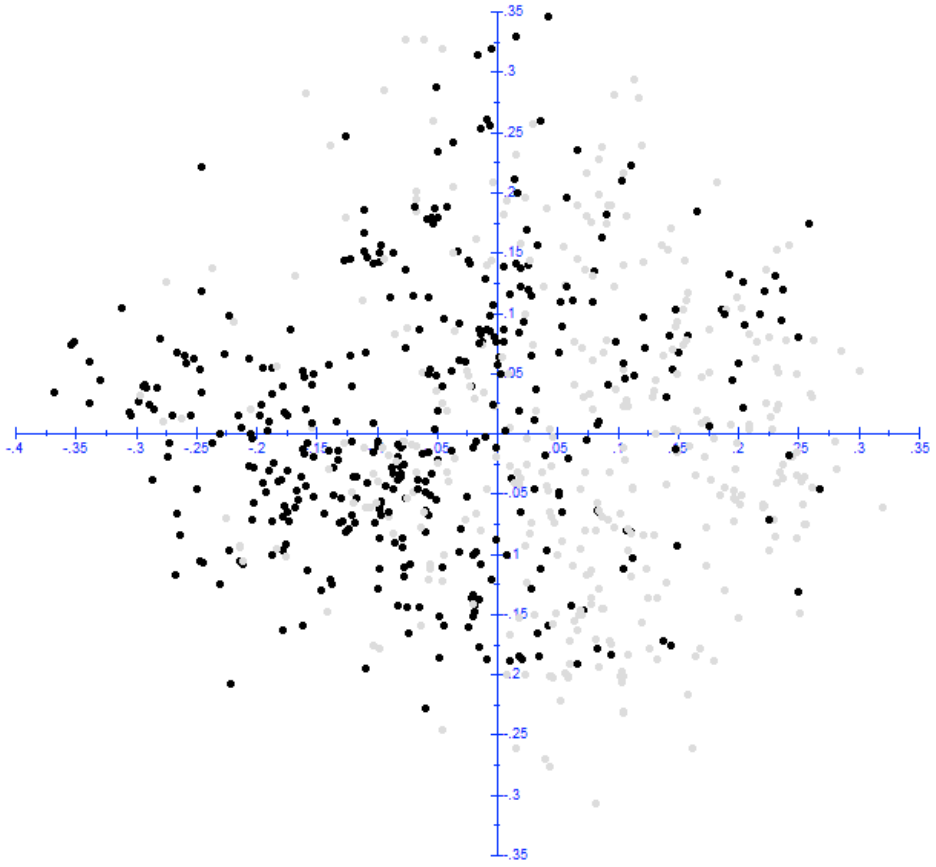


Figure SP3-22: Factorial analysis of fonio genetic diversity (21 SSR markers). Grey dots represent Guinean accessions.

3. Arcad added-value, visibility, attractiveness, national and international positioning

- Over the last 20 years, the study of crop diversity on-farm has become a lively field of research, promoted by the importance given to in situ-conservation on-farm of genetic resources in international arenas. Very often however, crop genetic diversity was approached through a proxy which is variety name diversity. ARCAD SP3 permitted to bring together teams focusing on cereal genetics only and teams interested in multidisciplinary approach to the dynamics of crop diversity in agroecosystems, and to reinforce the expertise and research potential of the Agropolis community in this area. It also extended the scope of study from tropical cereals to Mediterranean cereals.
- The SP3 teams were strongly involved in the development of the training module on

agrobiodiversity analysis (*Ecole thématique internationale Agrobiodiversité : des hommes et des Plantes. Outils et Méthodes d'analyse*) which has gained a high visibility in French-speaking African countries and beyond (>150 applications).

- The work done to prepare this training module as well as progress made in the multidisciplinary analysis of crop diversity will be the basis to prepare a methodological toolkit for the study of underutilized crops.
- ARCAD SP3 is having a strong impact in developing efficient partnerships and reinforcing the capacities in crop diversity studies of national partners, whether in Morocco, Guinea or Niger. Also, a research network on fonio is progressively being set up in West Africa.

4. Conclusion

- ARCAD SP3 can be described as a research project on the adoption of innovation by farming communities and its impact on local diversity. Non-indigeneous crops such as Asian rice are innovations that impacted the fate of local crops, e.g. the African rice *Oryza glaberrima*. So-called "modern" varieties are innovations of which the release may affect the genetic landscape. More research needs to be done to study the way farming communities mobilize existing and introduced crop diversity to address their needs, answer environmental and economic pressures, maintain (or not) the resilience of their cropping systems.

- One of the questions addressed by the SP3 is how human activities and social structures affect crop diversity. The methodological progress made by the SP1 and SP2 will permit to study the genetic signature of these factors.

- Finally, the development of the ARCAD component on biological resources conservation will call for research on "smart integrated strategies" for crop diversity conservation, based in particular upon a thorough knowledge on the status and dynamics of crop diversity on-farm. This includes research on the way to manage information on crop diversity, by institutions as well as by farming communities.

References cited

Adoukonou-Sagbadja, H., Wagner C., Dansi A., Ahlemeyer J., Daïnou O., Akpagana K., Ordon F., and Friedt W. 2007. Genetic diversity and population differentiation of traditional fonio millet (Digitaria spp.) landraces from different agro-ecological zones of West Africa. *Theoretical and Applied Genetics* 115: 917-931..

Barnaud, A., Vignes, H., Risterucci, A.-M., Noyer, J.-L., Blay, C., Buiron, M., Vigourou,x Y., and Billot C. 2012. Development of nuclear microsatellite markers for the fonio, Digitaria exilis (poaceae), an understudied west african cereal. *American Journal of Botany* 99: e105-7.

Barnaud A, Vigouroux Y, Barry B, Beavogui F, Camara M, Billot C, Noyer J-L, Pham J-L. & Bakasso Y. 2013. From advanced to underutilized crops: making fonio benefit from research on major cereals in Africa. Acta Horticulturae 979: 421-430.

Barro-Kondombo, C., F. Sagnard, J. Chantereau, M. Deu, K. Vom Brocke, P. Durand, E. Gozé, and J. D. Zongo. 2010. Genetic Structure among Sorghum Landraces as Revealed by Morphological Variation and Microsatellite Markers in Three Agroclimatic Regions of Burkina Faso. *Theoretical and Applied Genetics* 120:1511-1523.

Billot, C., P. Ramu, S. Bouchet, J. Chantereau, M. Deu, L. Gardes, J.-L. Noyer, et al. 2013. Massive Sorghum Collection Genotyped with Ssr Markers to Enhance Use of Global Genetic Resources. *Plos One* e59714.

Camus-Kulandaivelu L., Veyrieras J.B., Madur D., Combes V., Fourmann M., Barraud S., Pierre Dubreuil p., Gouesnard B., Manicacci D. and Charcosset A. 2006 Maize Adaptation to Temperate Climate: Relationship Between Population Structure and polymorphism in the Dwarf8 Gene. Genetics 172: 2449–2463.

Chessel, D., A. B. Dufour, and J. Thioulouse. 2004. The Ade4 Package - I : One-Table Methods. *R News* 4:5-10.

Clément, J., and J. M. Leblanc. 1984. Collecte IBPGR-ORSTOM de 1977 au Togo. In Prospection des Digitaria exilis (Fonio) en Afrique de l'Ouest, 3–7. ORSTOM, Marseille, France.

Deu, M., F. Sagnard, J. Chantereau, C. Calatayud, D. Hérault, C. Mariac, J.-L. Pham, et al. 2008. Niger-Wide Assessment of in Situ Sorghum Genetic Diversity with Microsatellite Markers. Theoretical and Applied Genetics 103:903-913.

Hilu, K., M'Ribu K., Liang H., and Mandelbaum C. 1997. Fonio millets: ethnobotany, genetic diversity and evolution. S. Afr. J. Bot. 63: 185-190.

Labeyrie, V., B. Rono, and C. Leclerc. 2013, accepted. How Social Organization Shapes Crop Diversity: An Ecological Anthropology Approach among Tharaka Farmers in Kenya. Agriculture and Human Values. Accepted for publication.

Leclerc, C., and G. Coppens D'eeckenbrugge. 2012. Social Organization of Crop Genetic Diversity the G X E X S Interaction Model. Diversity 4:1-32.

Middleton, J., and G. Kershaw. 1953. Central Tribes of the North-Eastern Bantu: The Kikuyu Including Embu, Meru, Mere, Chuka, Mwimbi, Tharaka, and the Kamba of Kenya. Etnographic Survey of Africa, East Central Africa, Part V. London: Hazell, Watson & Viney.

Mir C., Zerjal T., et al., 2013. Out of America: tracing the genetic footprints of the global diffusion of maize. Theor Appl Genet, 126 : 2671-2682.

Portères, R., 1955. L'introduction de maïs en Afrique. Journal d'agriculture tropicale et de Botanique Appliqué 221–231.

Sagnard, F., M. Deu, D. Dembélé, R. Leblois, L. Toure, M. Diakité, C. Calatalyud, et al. 2011. Genetic Diversity, Structure, Gene Flow and Evolutionary Relationships within the Sorghum Bicolor Wild–Weedy–Crop Complex in a Western African Region. Theoretical and Applied Genetics 123:1231-1246.

Sanou J., 1996. Analyse de la variabilité génétique de cultivars locaux de maïs de la zone de savane Ouest africaine en vue de sa gestion et de son utilisation (Thesis for Phd degree).Scarcelli N., Barnaud A., Eiserhardt W., Treier U.A., Seveno M., et al. 2011. A Set of 100 Chloroplast DNA Primer Pairs to Study Population Genetics and Phylogeny in Monocotyledons. PLoS ONE 6: e19954.

Scarcelli N., Barnaud A., Eiserhardt W., Treier U.A., Seveno M., et al. 2011. A Set of 100 Chloroplast DNA Primer Pairs to Study Population Genetics and Phylogeny in Monocotyledons. *PLoS ONE* 6: e19954.

Westengen O.T., Berg P. R., Kent M. P., Brysting A. K., 2012. Spatial Structure and Climatic Adaptation in African Maize Revealed by Surveying SNP Diversity in Relation to Global Breeding and Landrace Panels. PLoS ONE 7(10): e47832. doi:10.1371/journal.pone.0047832.

SP4 - Bioinformatics

1. Rationale

The ARCAD programme tackles the challenge of managing and analyzing huge datasets created in a relatively short time by using new sequencing and genotyping technologies. The Bioinformatics project aimed to provide an integrated bioinformatics support to take up the corresponding challenges in terms of computing power, integration of multi-scale data, and new algorithms and methods of analysis.

2. Objectives

We aimed to develop appropriate bioinformatics tools for the management and harnessing of the next generation sequencing data and to make them available first to all partners of the ARCAD project and subsequently to an extended community. This requires the best new technologies to explore (i) complex high throughput sequence analysis ranging from assemblers, multialignment searches to SNP detection and phylogenetic analyses, (ii) several data models and databases to store heterogeneous data that will enable interoperability between systems, and (iii) several Web interfaces to launch analyses, to synthesize, visualize, query and edit the results.

Our objectives were:

(i) to set up a collaborative development environment to avoid redundancy and to facilitate future bioinformatics developments across organizations,

(ii) to provide training in bioinformatics and support for bioinformatics projects hosted on the ARCAD platform,

(iii) to collaborate (share software, workshop, mailing lists, and good practices) with other national as well as international bioinformatics platforms

(iv) to ensure quality control in bioinformatics research though a scientific user committee, documentation, data traceability and reliability, CECILL licences, indicator measurement.

The project made use of computer facilities built around a 240 processors computer cluster, a low latency Infiniband network, and a 65 To storage capacity. The amount and nature of sequencing data brought many bioinformatics issues in terms of algorithms for NGS analyses and data storage.

3. Activities carried out and main results

a) We developed a **unique Web portal South Green** (http://southgreen.cirad.fr/) to give access to tools and databases for managing genetic and genomic resources of

tropical and Mediterranean plants, analyzing transcriptomes, predicting orthologs by phylogenomics, determining SSR and SNP, analyzing genetic diversity data, and performing structural, functional and comparative annotations. The South Green web portal contains currently 20 information systems and tools and targets about 30 plants. These tools are available on-line and are used massively (50,000 queries per month).

b) We developed new workflows for NGS (Next Generation Sequencing) sequence high-throughput analyses with different steps: cleaning, assembly, mapping, SNP detection, annotations, and phylogenetic analyses. We developed a package gathering the scripts for the analysis of high-throughput sequencing data from the ARCAD project. These scripts were mainly used in support of SP1 and SP5. The package is available at https://github.com/SouthGreenPlatform/arcad-hts

c) In collaboration with the SP1 component, we evaluated different methods for de novo short-read assemblies using data from two transcriptomes of crops with reference genomes: grape and sorghum. Then, we chose the best methods and parameters to release 29 new transcriptomes of plants including the key species and outgroups sequenced in the SP1. We produced functional annotations and SNPs detection for the different crop individuals. The data are available at http://arcad-bioinformatics.southgreen.fr/

d) We implemented a **Galaxy instance** (http://gohelle.cirad.fr/galaxy), a workflow manager which permits to run several bioinformatic analyses using a simple Web interface (**Figure SP4-1**). The South Green Galaxy instance is opened to anyone, but anonymous users are limited to 10 Mo data (*Maillol, V., et al., 2012, Role of Galaxy in a bioinformatic plant breeding platform, 2012 Galaxy Community Conference*). This instance contains a large collection of exclusive tools of the platform. The access to workflows developed for the ARCAD SP1 project is currently restricted to users with specific login.

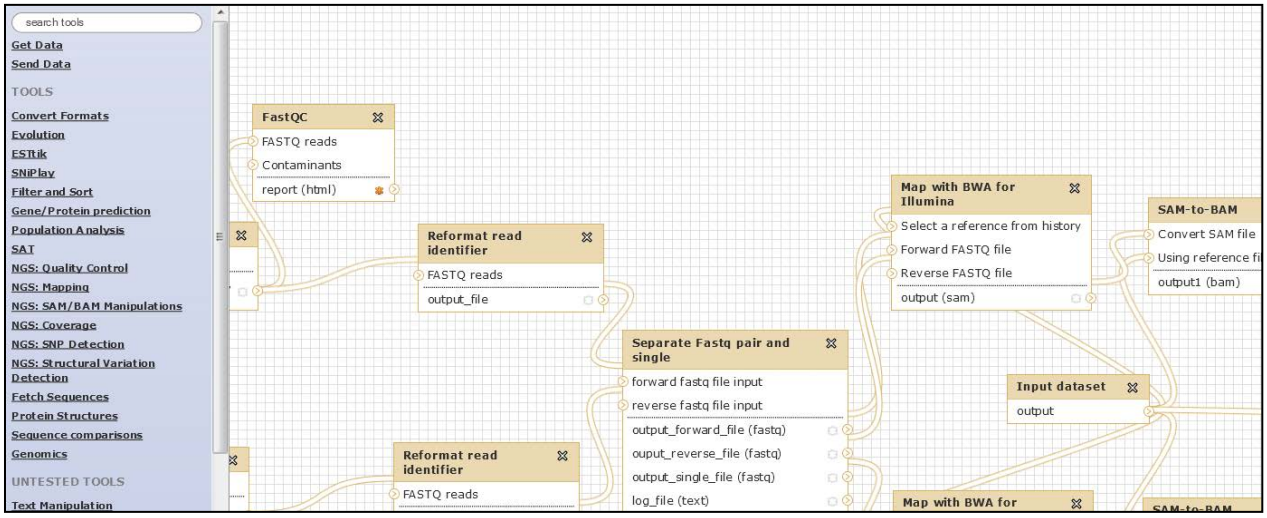


Figure SP4-1 – Graphical visualization of a Galaxy ARCAD workflow

e) In March 2013, we obtained the **AFNOR certification** (ISO 9001:2008) for the following activity: provision of bioinformatics software and equipment for the agronomy industry.

f) We developed a **Web-based tool** that provides the means to quickly build search interfaces over existing databases, without the need of any programming effort. It is particularly suited for scientific data that can conveniently be displayed in tables. This application was used for the information system TropGeneDB which will integrate part of the ARCAD data (*Hamelin C., et al., 2013. TropGeneDB, the multi-tropical crop information system updated and extended. Nucleic acids research*).

g) **GreenPhylDB is a Web resource** designed for comparative and functional genomics in plants. The database contains a catalogue of gene families based on complete genomes, covering a broad taxonomy of green plants. The version 3 of was released (http://www.greenphyl.org).

3. Partnership

South Green is a collaborative initiative with bioinformaticians from CIRAD, INRA, IRD and Bioversity.

4. Training

We organized 2 training sessions/year with an average of 20 attendants for each (http://southgreen.cirad.fr/?q=content/trainings).

5. Structuring effects and synergies created

ARCAD SP4 Bioinformatics is structuring a community on NGS bioinformatics tools to address biodiversity studies thanks to regular meetings, every 2-3 weeks, between biologists and bioinformatics specialists and training sessions organized for ARCAD community. Training workshops are now open to students and scientists of the South posted in diverse research teams of Montpellier, whether or not they are involved in ARCAD projects.

6. Visibility, attractive, national and international positioning

We are partner of the network of national platforms in bioinformatics, ReNaBi-IFB (French Institute of Bioinformatics), and we are currently developing a partnership with platforms of CGIAR centers (CIAT, IRRI, etc.).
We are key members of the Computational Biology Institute (http://www.ibc-montpellier.fr/) which is a multidisciplinary project aiming at the development of innovative methods and software to analyze, integrate and contextualize large-scale biological data.

With ARCAD partners (rice geneticists) we are involved on the funded France Génomique IRIGIN project (International Sequencing Initiative of Major Genetic Stocks and Resources for Rice Knowledge). IRIGIN project will provide deep resequencing data for African rice genomes and GBS for a large panel of populations of interest (NAMs, Genomic Selection, CSSL, etc.).

7. Conclusion

ARCAD contributed significantly to the development of the South Green platform, http://www.southgreen.fr/, which is a local network of scientists gathering Bioinformatics skills. Based on interactions with the strong Agropolis community in the field of agriculture, food and biodiversity, various bioinformatics applications and resources dedicated to genomics of tropical and Mediterranean plants have been developed and published.

Exchange and collaborative developments were also fostered through regular hands-on sessions on synergistic themes such as Galaxy, genome annotation or next generation genotyping.

The South Green web portal contains currently 20 information systems and tools and targets about 30 plants. Data produced by ARCAD projects are specifically available through the ARCAD BioInformatics Portal, http://arcad-bioinformatics.southgreen.fr/.

In order to be able to manage large-scale biological data like Genotyping By Sequencing data, we needed to explore new scalable computational solutions. Efficient storage of the huge resulting data sets for management, sharing and routine exploration purposes remain challenging. To address this issue, we explored new systems bases on non-relational or NoSQL databases. This work is still ongoing but a working prototype should soon be available.

SP5 - Pangenomic study of diversity through the reduction of genome complexity

1. Rationale

Due to the development of new sequencing methodologies (or NGS for New Generation Sequencing) and the increase of whole genome sequences, it is becoming more and more tempting to study the genetic diversity at whole genome scale and revealing allelic variability directly by DNA sequence comparison.

Pangenomic studies, especially for species with large genomes, may however require reduction of genome complexity in order to become useful and economically acceptable.

Various strategies for reducing genomic complexity in order to study only a targeted part of the genome can be implemented depending on the species (and the complexity of its genome) and the research question. One of them consists of sequencing only the transcriptomic part, *i.e.* expressed genes, based on mRNA purification and sequencing of cDNA. Such a strategy was chosen in the *Comparative population genomics* sub-project (SP1). NGS application to whole transcriptome sequencing is also called RNA-seq and different protocols have been developed for allelic diversity analysis.

Other strategies are applied on genomic DNA, but aim to concentrate sequencing effort on the non repeated part of the nuclear genome, by partial elimination of cytoplasmic DNA and highly repeated nuclear sequences. One of the methods used for that purpose is based on DNA restriction using methylation sensitive restriction enzymes.
Application to genotyping implies to set up fast, low cost and robust procedures, allowing reliable analysis of large sets of accessions.
Although, all these strategies and methods produce *de facto* genotype information by DNA sequencing, the term “Genotyping By Sequencing” and its acronym “GBS” are currently used, by the scientific community, in the restricted case of complexity reduction using restriction enzymes (Elshire *et al.* 2011).
Next Generation Genotyping (NGG) has been proposed as the generic term including all the genotyping approaches using NGS (M.C. Le Paslier pers.comm.).

Since many species analyzed within ARCAD project are orphan or neglected plant species, most of the time they lack the genomic information available on more studied plants, which may make the use of GBS more difficult.

2. Objectives

The SP5 sub-project *Pangenomic study of diversity* was an exploratory project, giving the opportunity to evaluate the potential of NGS for the analysis of pangenomic diversity even in the case of highly heterozygous, unsequenced and/or orphan species.

The research program was organized in 3 work packages (WP):

WP1: GbS tests and protocols setup:
For a set of species with diversified biological caracteristics (mating system, ploidy, genome size, heterozygosity) and different level of genetic and genomic knowledge, the objective was to test the faisability of SNPs detection and genotyping by GbS approach, on a reduced sample of accessions.

WP2: Segregation analysis by genetic mapping of SNP/GbS markers on heterozygous perennial species
The objective was to estimate by genetic mapping experiments (genetic determinism and segregation assessment) the number of reliable and informative SNP markers correctly genotyped by the GbS method.

WP3: Application of GbS to the genetic diversity characterization of an orphan crop
We initialy planed to analyse genetic diversity by GbS in fonio (*Digitaria exilis,* Poaceae), an orphan annual species without sequence or mapping information, and which has been studied within the *Cereals in Africa* sub-project (SP3). We planed to apply GBS on 96 fonio accessions in the frame of the current ARCAD project. Unfortunately, due to germinating problems in Senegal, plant material was not available at the time. Currently, genomic DNA samples of the 96 accessions has been extracted and purified (C. Billot pers. comm.) and are now available for GbS experiments.

The program of SP5 sub-project began later than the other ARCAD Sub Projects, since we have modified the objectives from the initial (2009) SP on linkage disequilibrium. The first meeting has been organized in November 2012 in order to launch the project and to decide on the species to be analyzed. In the “ARCAD progress report 2013” a first list of species was given, based on previous expressions of interest.

Due to unavaibility of biological material (leaf samples of accessions or DNA samples of sufficient quality) or of the launching of complementary research projects in the meantime, the list of species effectively studied in the WP1 was reduced and is given in **Table SP5-1**. This allowed us to increase the sequencing effort on the remaining species.

Four of the species studied in SP5 were also included in the SP1 sub-project (Cotton, *Coffea canephora*, Olive tree and Grapevine). It will allow comparing the data and results produced by the SP1 sub-project to those obtained using the genomic reduction method. These results will be useful to the entire ARCAD community.

Table SP5-1. List of the crops studied in ARCAD/SP5 and corresponding scientist

SP5/WP	Species	Corresponding scientist
WP1	Coffee (*Coffea arabica, C. canepora & C. arabusta*)	Thierry Leroy (UMR-AGAP)
WP1	Cotton (*Gossypium hirsutum & G. herbaceum*)	Jean-Marc Lacape (UMR-AGAP)
WP1	Coconut (*Cocos nucifera*)	Luc Baudouin (UMR-AGAP)
WP1	Breadfruit (*Artocarpus altilis*)	Jean-Pierre Labouisse (UMR-AGAP)
WP1	*Citrus* spp.	Patrick Ollitrault (UMR-AGAP)
WP2	Grapevine *(Vitis vinifera*)	Agnès Doligez (UMR-AGAP)
WP2	Olive tree (*Olea europaea*)	Bouchaïd Khadari (UMR-AGAP)
WP2	Rubber tree *(Hevea brasiliensis & Hevea benthamiana*)	Marc Seguin (UMR-AGAP)

3. Activities carried out and main results

3.1 WP1: Tests and setup of GbS laboratory protocols

We tested genomic complexity reduction methods on 7 species or species complexes retained in the WP1 (Citrus, Coconut, three 3 Coffee species, Cotton and Breadfruit, **Table SP5-1**).

Eight to twenty four accessions for each of the species have been analyzed. The choice of the accessions for each species was made by the corresponding scientist (**Table SP5-1**), and was based on the knowledge on genetic diversity of the species or species complexes. The genomic experiments were performed by the team of the AMM laboratory of the UMR-AGAP (Audrey Weber, Muriel Latreille and Sylvain Santoni – INRA, UMR-AGAP).

Extraction of total DNA of high quality is a critical step for GbS. It ensures quality and reliability between accessions of the DNA restriction, i.e. an efficiency of the reduction of DNA complexity similar between the accessions.

For the evaluation of the complexity reduction process, severall factors and factor combinations were tested:

1) Type of restriction enzyme: 7 enzymes, with restriction site of 4, 5 or 6 bases methylation sensitive or not (methylation sensitive restriction enzymes cut more frequently hypomethylated DNA which correspond in the genome of regions reach in expressed genes and low copy DNA).
2) Combinations of restriction enzymes (double restriction)
3) Sizing or not of the restricted total DNA
4) 2 different DNA sequencing (NGS) methods: Illumina High-Seq or MiSeq

Selecting DNA restricted fragments of smaller size (sizing), *i.e.* frequently cut, leads to isolate a genomic DNA fraction enriched in non repetitive DNA, with a lower content in highly repeated genomic and cytoplasmic DNA. It allows reducing DNA sequencing efforts by focusing on non-highly repeated DNA.

Sizing of restricted DNA can be controlled using specific equipment (Blue Pippin, Sage Science) allowing accurate size selection of DNA fragments (for instance DNA fragments with size comprised between 400 and 500 base pairs (pb)). But this procedure is optional as, in the DNA library construction process, the PCR amplification and the DNA fragment purification steps lead *de facto* to the selection of small DNA fragments (enriched in fragments over 200 bp and smaller than 500 bp).

Based on the tests of severall combinations of enzyme, sizing and NGS methods, for each of the studied species 1 to 3 combinations were specifically choosen for GbS application on all the accessions (**Table SP5-2**; see also **Table SP5-3**).

DNA sequencing is completed. We received the sequence data in July 2014 and the bio-informatic data treatments are in progress, with the help of the bio-informatic tools (scripts) developed in SP5 (see section "3.3 Bio-informatic development and analyses" below). As a first result of data analysis, number of DNA fragment sequences (reads) available for SNPs detection and genotyping is given per species in **Table SP5-2**)

It is however necessary to complete sequence data analysis and SNPs detection to conclude on the comparative efficiency of the laboratory conditions and combinations tested on the different species.

Table SP5-2. GbS development: summary of DNA sequencing performed on the species analysed in the frame of ARCAD/SP5

Samples		GbS			
Species	Number of accessions	Restriction enzyme	Sizing (bp)	Sequencing	Nb of pairs of reads
Citrus spp.	12	ApeK1	400-500	Hiseq 2 x 100	3 000 000
Citrus spp.	12	Pst1/Mse1	>200	Hiseq 2 x 100	17 000 000
Coffea canephora	8	ApeK1	400-500	Hiseq 2 x 100	8 000 000
Coffea arabica	8	ApeK1	400-500	Hiseq 2 x 100	10 000 000
Coffea arabusta	8	ApeK1	400-500	Hiseq 2 x 100	14 000 000
Cotton	24	ApeK1	400-500	Hiseq 2 x 100	32 000 000
Coconut	12	ApeK1	400-500	Hiseq 2 x 100	6 000 000
Breadfruit	8	Pst1/Mse1	>200	Miseq 2 x 250	2 774 562
Breadfruit	8	ApeK1	400-500	Miseq 2 x 250	85 724
Breadfruit	8	ApeK1	>200	Miseq 2 x 250	252 933

3.2 WP2: Test on heterozygous perennial species

Due to little background information on the use of this method in highly heterozygous species, we developped GbS markers on segregating populations in three allogamous perennial species (**Table SP5-1**), including 2 studied in the SP1 ARCAD subproject:

1. Grapevine: small (480 Mb/1C), fully sequenced genome (Jaillon et al. 2007)
2. Olive tree: medium size (950 Mb/1C), incompletely sequenced genome (Munoz-Mérida et al. 2013)
3. Rubber tree: large (2100 Mb/1C), partly sequenced genome (Rahman et al. 2013, BMC Genomics, 14:75)

The segregation of the SNPs in the mapping population and in particular the absence of biais in the segregation was then a good test of the different software & parameters to use for the identification of the SNPs.

Preliminary work for the methodological choice

For the 3 species analyzed within this WP, development work was also necessary. Preliminary works performed on the analysis of intravarietal diversity in grape (clonal variability) as well as *in silico* analysis of restriction sizes from the sequenced genotype (Pinot noir) were adapted to the analysis of the mapping populations. For this species, we tested 4 enzymes, leading to 10 to 50 Million reads, ApeK1, being the most interesting (Tableau SP5-3). This enzyme was thus selected.

For olive tree, preliminary experiments on 8 samples, using the DNA based protocol produced a small number of reads (<10 000; **Table SP5-3**). In addition, it was very difficult to assemble the genomic data for olive tree without a reference genome. We thus decided to test a low cost RNAseq based protocol for the analysis of 96 individuals and to assemble the expressed genome as performed in SP1. For Rubber tree given the other results, we decided to use ApeK1 as well.

For each of these 3 species, after the development phase, it was thus decided to perform the different libraries using ApeK1 restriction enzyme either with sizing (Grape, olive) or without (rubber tree). The first grape library was sequenced in house (Myseq apparatus) the other libraries were sequenced by Mgx platform on a Hiseq apparatus. The different experiments lead 20 to 164 million reads (table SP5-3) 2 x 250bp with the Miseq and 2 x 100bp or 1x 100bp with the Hiseq.

The genomic experiments were performed by the team of the AMM laboratory of the UMR-AGAP (Audrey Weber, Muriel Latreille and Sylvain Santoni – INRA, UMR-AGAP).

Table SP5-3. GbS development: summary of DNA sequencing performed on the species analysed in the frame of ARCAD/SP5

Samples		GbS			
Species	Number of accessions	Restriction enzyme	Sizing (bp)	Sequencing	Nb of pairs of reads
Grapevine	60	ApeK1	400-500	Miseq 2 x 250	20 000 000
Grapevine	120	ApeK1	300-400	Hiseq 2 x 100	96 000 000
Grapevine	24	ApeK1	>200	Hiseq 2 x 100	50 000 000
Grapevine	24	EcoT22i	>200	Hiseq 2 x 100	40 000 000
Grapevine	24	Pst1	>200	Hiseq 2 x 100	10 000 000
Grapevine	24	Hpa2	>200	Hiseq 2 x 100	22 000 000
Olive tree	96	ApeK1	400-500	Hiseq 2 x 100	11 000 000
Olive tree	8	EcoT22i	>200	Hiseq 2 x 100	10 000
Olive tree	8	Pst1	>200	Hiseq 2 x 100	3 000
Olive tree	8	Hinf1	>200	Hiseq 2 x 100	2 000
Olive tree	8	MluC1	>200	Hiseq 2 x 100	5 000
Rubber tree	275	ApeK1	-	Hiseq x 100	164 000 000*

* unlike all other crops in the table, for rubber tree the total number of reads is given (and not of pairs of reads) as for this species DNA sequencing was performed in single reads of DNA library with no pair ends.

Development of the genetic maps

Grape

The work on this species enabled to test several thresholds for the different parameters: minimum coverage for the markers versus the percentage of missing data, use of multiples SNPs in the reads in order to develop multi-allelic markers.

In total 2168/1630 markers were recovered and were mapped (**Table SP5-4, figure SP5-1**). Out of the 1630, 948 markers corresponded to single SNPs, while 682 markers corresponded to multiple linked SNPs (up to 14).

The number of markers with a threshold of 8X was much higher and enabled the development of a map with a lower number of large gaps (> 10 cM) but without more skews in the segregation (**Figure SP5-2**). With this threshold, the number of missing data was also lower (**Table SP5-5**), it should thus be preferred.

Tableau SP5-4. Grape. Comparison of the number of markers (according to their segregation types) at a coverage threshold of 8X versus 10x for at least 46 individual out of 64 from the first analysis with the Miseq.

Ségrégation	8X	10X
Aaxab	537	404
Abxaa	538	413
a0xab	0	0
abxa0	11	11
Abxab	680	494
Abxac	286	228
Abxcd	116	80
Total	**2168**	**1630**

Tableau SP5-5. Grape. Comparison of the number missing data when considering a threshold of 8X or of 10X.

	Number of SNPs present on each individual under the threshold of	
# indiv. without MD	10 X	8 X
44	7039	
45	6309	10463
46	5721	9533
50	3513	6357
55	1346	3196
59	425	1188

Finally as expected, the sequencing using the Hiseq apparatus provided more markers (2428) than using the Miseq apparatus (1630)

Olive tree

For olive tree, using the 81 individuals, 7856 SNPs were obtained. After reconstruction, they were transformed into 3499 markers, 2835 of which were mapped (1754 individual and 1081 reconstructed).

Tableau SP5-6. Olive tree. Number of markers according to their segregation types of the olive map

Segregation	Markers
aaxab	524
abxaa	577
abxab	1124
abxac	487
abxcd	123
Total	**2835**

Rubber tree

For rubber tree, we worked on a segregating population issued from an interspecific F1 cross. The 2 parents are highly heterozygous accessions. The progeny is conserved in the Cirad collection in French Guiana. DNA samples of sufficient quality for GBS was available for 271 F1 individuals and for their 2 parents, PB260 (*Hevea brasiliensis*) and FX3899 (*H. brasiliensis x H. benthamiana* hybrid) and the 2 maternal grandparents (*H. brasiliensis*). Severall restriction enzymes were tested. Based on the restriction pattern the ApeK1 was selected. For rubber tree, we applied a GbS protocol similar to the standard protocol of Elshire *et al.* (2011) for the construction of DNA libraries of the 275 genotypes, without pair-ends. Only one condition was used for the DNA libraries construction of all the individuals sent to DNA sequencing: ApeKI enzyme, no controlled sizing and HiSeq sequencing (**Table SP5-3)**. We did not use controlled sizing, but the analysis of the distribution of the fragment size in the *Hevea* libraries showed that 80 % of the fragments had a size comprised between 200 bp and 500 bp.

The laboratory experiments were performed by Ronan Rivallan and Pierre Mournet at the AGAP Genotyping Plateform (Montpellier, France). The HiSeq DNA sequencing was made at the Genotoul Plateform (Toulouse, France).

Preliminary bio-informatic analysis of the sequences (164 10^6 reads of good quality for the 275 genotypes) led to the identification of 121,000 SNPs and application of the ARCAD_SP5_GBS pipeline allowed identification of 12,000 potential SNP markers. Genetic mapping of these markers has now be performed, using a pseudo-test cross strategy, in order to determine the final number of informative SNP markers.

3.3. Bio-informatic development and analyses

The bio-informatician engineer, Mrs Hajar CHOUIKI, was hired from June 2013 to June 2014. Her work was to improve the bioinformatics pipeline implementing scripts facilitating the analysis of GbS data. She worked in connection with the ARCAD SP4 sub-project, and with ID team of UMR-AGAP, under the supervision of Gautier Sarah (INRA, UMR-AGAP). This ARCAD_SP5_GBS pipeline is publicly available on Git Hub: https://github.com/SouthGreenPlatform/arcad-hts/tree/master/sp5_gbs. She then managed the analysis of the raw data produced in the frame of the SP5/WP2 on grape, olive tree and rubber tree.

The ARCAD_SP5_GBS pipeline gave original tools such as scripts allowing comparison of two sets of SNPs independently detected on the same accession. It allows rapid comparison of SNPs detection results obtained with different softwares or pipelines such as STACKS, TASSEL and ARCAD_SP4 pipeline. Another original development was made allowing transformation of bi-allelic single SNP markers in more informative multi-allelic markers, using information multiple adjacent SNPs.

4. Perspectives

SP5 was a 2-year, small budget project, which started at the end of 2012 and part of the activities are still on-going. One of the objectives is to build up a network of bioinformaticians and scientists on the analysis of NGS.

For the different WPs the perspectives are the following.

WP1

- Complete the analysis of the sequences and the SNP detection
- Comparative analysis of the best GBS protocols

WP2

- Develop the rubber tree map

WP3

- Finalize the molecular work and the data analysis

5. Conclusions

- For Grape: sufficient coverage of the sequences in order to obtain about 10,000 SNPs for the map => 1500-3000 markers were mapped. This was a particular work with low diversity but served as protocol evaluation in order to guarantee quality of GBS analysis for diversity work.
- For olive tree: RNAseq protocol was good and enabled the development of a large number of mapped markers.
- Globally, we have acquired expertise on several species, controled the different protocols that will be available for the community through future ARCAD infrastructure.

References

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* 6:e19379

Jaillon O, Aury JM, Noel B, et al., French-Italian Public Consortium for Grapevine Genome C. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449:463-467. doi:10.1038/nature06148

Munoz-Merida A, Gonzalez-Plaza JJ, Canada A, Blanco AM, Garcia-Lopez Mdel C, Rodriguez JM, Pedrola L, Sicardo MD, Hernandez ML, De la Rosa R, Belaj A, Gil-Borja M, Luque F, Martinez-Rivas JM, Pisano DG, Trelles O, Valpuesta V, Beuzon CR. 2013. De novo assembly and functional annotation of the olive (*Olea europaea*) transcriptome. *DNA research*: 20:93-108. doi: 10.1093/dnares/dss036

SP6 –Facilities for plants DNA manipulation: purification, quality control, storage and management.

Towards a DNA bank

1. Rationale

The Regional Genotyping Platform of Montpellier (GPTR Génotypage, http://www.gptr-lr-genotypage.com/index.html) supports any program of molecular genetics applied to plants, both for programmes aiming at characterizing diversity and for genetic improvement projects. It offers significant opportunities for genotyping (SSR, AFLP, SNP EcoTilling, DArT molecular markers) and medium/high speed sequencing (Sanger sequences and low throughput NGS sequences). It also allows for the construction, operation, and management of genomic resources, providing technologies for developing various DNA banks (BAC, cDNA, SSR, SSH, DArT, etc.). Finally, it allows for technological monitoring and for the care and training of different types of staff (students, technicians, researchers) from Montpellier's scientific communities, both national and international. Lastly, it brings its expertise to the development and use of similar technologies on other sites.

The current equipment of GPTR Génotypage includes:

Genotyping and sequencing
1 Personal Genome sequencer NGS Illumina MiSeq
1 x 24-capillary Applied 3500 XL fragment analyser
3 x 16-capillary ABI 3130xl sequencers (2 GPTR, 1 IFV)
6 x Li-Cor® DNA fragment analysers (automatic sequencers)
A large number of Eppendorf and MJ Research PCR blocks (14 with 384 wells, 18 with 96 wells)
1 Lightcycler 480 Roche quantitative PCR device
DArT marker genotyping platform: 1 TECAN LS300 laser reader, 1 Microgrid spotter

Robotics
Colony Picker GENETIX QPIX II XT (picking, replicating, re-arraying, macroarrays)
Beckmann Biomek NX robot: 8 and 96 head (preparation of PCR, rearrangements).
Two Hamilton MicroLabStar 8-channel pipetting robots.

Activity of the platform is expected to expand. In this context, ARCAD Sub-Project 6 (SP6) aims at consolidating the facilities for handling plant DNA. It is meant to be an open, flexible, and efficient facility. Through close relations with the management of germplasm collections and Biological Resources Centres, these DNA facilities will help organize a DNA bank for Mediterranean and tropical crops.

2. Objectives

The main objective of ARCAD SP66 was to build up significant capacity for DNA extraction, conservation and distribution in order to support all ARCAD activities dealing with DNA samples.

This project aimed to:

a) add the equipment required for high quality DNA sample collection, extraction, quality checking and archival storage;
b) design appropriate procedures for the creation and setting up of a DNA bank for Mediterranean and tropical crops.

A methodological component was developed in a related project funded by the GIS IBiSA (Groupe d'Intérêt Scientifique *Infrastructures Biologie, Santé et Agronomie*). In order to fit with the recommendations of the IPGRI/Bioversity International report on the international plant DNA banks, we conducted a methodological study for the creation of a plant DNA bank associated with the Biological Resources Centers of some Mediterranean and tropical plants located in Montpellier.

3. Material and methods

The establishment of this DNA bank required 4 complementary actions.

a) Extraction / purification of Nucleic acids.
Compilation and critical analysis of protocols for plant DNA extraction
Definition of an objective test for validation of the quality of DNA extracts
Choice of the best performing protocols, writing procedures
Similar approach to the messenger RNA

b) Secure storage of DNA
Comparing conditions of storage of DNA solutions : for average duration at -18°C and for a very long term, in a metal capsule under controlled atmosphere and at room temperature (Imagène process)

c) Management of DNAs in the DNA bank.
Integrated management of DNA samples with CRB databases

d) Transfer of DNA / relationships with partners
Reception of DNA from foreign partners, brought into conformity with the standards of the DNA bank. Proposal for a specific MTA for DNA.

The costs of all studies on technical points were supported by the IBiSA project. ARCAD SP6 supported the installation of new laboratory equipment.

4. Activities carried out and main results

a) Extraction / purification of nucleic acids.

Compilation and critical analysis of protocols for plant DNA extraction

Conventional techniques of extraction plant DNA extraction can be divided into two main groups, each associated with an original protocol:

Method 1: SDS / AcK according to Dellaporta et al, 1983: the cell lysis buffer contains SDS (sodium dodecyl sulfate), a strong detergent and denaturing of protein structures. Denatured compounds are precipitated by chaotropic medium with high concentration of potassium acetate. Nucleic acids, which remain in the clarified cell lysate, are then precipitated in the presence of alcohol.

Method 2: CTAB / Chloroform, according to Murray and Thompson, 1980: the cell lysis buffer contains Cetyltrimethylamomonium bromide (CTAB), detergent and denaturing of proteins. CTAB can also, under certain conditions, precipitate the polysaccharides. Lipophilic and / or denatured and precipitated compounds are removed by the action of chloroform, organic solvent. Nucleic acids (water-soluble) remain in the aqueous phase. Then they may be precipitated by either the CTAB itself or with alcohols.
There are many variants based on those two biochemical principles (for review: Varma et al, 2007). In both cases, additional steps of removing lipophilic compounds (lipids and proteins) can be achieved by the use of organic solvents, phenol, phenol/chloroform, octanol. Enzymes (proteases, Rnases) can be used if necessary. Finally, in order to obtain optimal purity of DNA, specific techniques of capture on a solid support (ion exchange polymers, silica) are now compatible with the two types of extraction procedure.

The two types of method are different mainly because of the detergent present in the lysis buffer. SBS, more powerful, enables, in some cases not to use a step of removing lipophilic compounds using organic solvents. CTAB (and related compounds), well suited for the removal of polysaccharides requires the use of chloroform to get rid of lipids and proteins. The use of such organic solvent imposes serious constraints on the procedure to ensure the safety of the laboratory staff.

References:

Dellaporta et al. (1983). A Plant DNA Minipreparation : Version II. Plant Mol. Biol. Rep. 11 19-21
Murray MG et Thompson WF (1980): rapid isolation of high molecular weight plant DNA. Nucleic Acid Research, 8 (19): 4321-4325.
Varma et al, (2007), Plant genomic DNA isolation: An art or a science, Biotech J, 2(3):386-392.

Definition of an objective test for validation of the quality of DNA extracts

To obtain objective data on DNA quality and in order to compare several situations, we elaborated a robust Quality Control procedure to estimate several parameters:

Tests for quantity**:** DNAs were assayed by several methods, UV spectroscopy and fluorescence spectroscopy.

UV absorption curves were published for each DNA sample. The shape of the curve and the ratios of OD 260/280 provide indications of the presence of any pollutant compounds.
We validated techniques of spectrofluorometric assay in the presence of Benzimidazole Hoechst H33258 or Picogreen. Protocols written using the principles of Quality Assurance in Research (AQR) are now available.

Tests of DNA degradation**:** assessment of the DNA size was estimated by agarose gel electrophoresis or using bio-analyzer (Agilent).

Functional tests**:**
For amplification: multiplex PCR using specific microsatellite markers for each species.
For hybridization: quantitative PCR with TaqMan probe. We used a standardized protocol og qPCR Internal Control, based on those used for DNA assays of transgenic plants on plant products
For enzymatic digestion: tests with one or two AFLP selective primer systems.

In all cases, protocols written using the principles of Quality Assurance in Research (AQR) are now available.

Choice of the best performing protocols, writing procedures

Plant geneticists of Montpellier, working on Mediterranean and tropical plants, used many DNA extraction protocols adapted to various situations. We chose to focus on the development of optimized protocols that uses the principles of method 1, SDS/AcK without using organic solvent and purifying DNA by capture on silica.

Technical organization, installation of new laboratory equipment (supporting by ARCAD SP6)

Most genotyping analyses are carried out using molecular markers manipulated by PCR (Polymerase Chain reaction) The amount of DNA needed for each analysis is reduced (about 10 to 100 nanogram) and is extracted from a small quantity of biological material (50 to 200 mg of leaf blade, for example). Protocols implemented and the equipment used proved to be suitable for handling small amounts of material on hundreds or thousands of samples.

Samples treatment
After harvesting, samples of plants are freeze-dried in a large capacity freeze dryer (i.e. freeze dryer Pilote Cryotec, acquired thanks to ARCAD SP6 budget) designed for handling simultaneously 2000 samples stored in 21 DeepWell96 plates.
The freeze-dried samples are crushed by series of 192 using laboratory shredders type Rechts MM301 or for greater quantity or for non-standard tube, with a high capacity shredder (i.e. shredder GenoGrinder 2010, ARCAD SP6 budget).

Optimized line of DNA extraction
We developed a series of protocols for purification of genomic DNA based on principle of Method 1 in a 96 samples format, adaptable to a great diversity of plant samples and able to produce DNA of very high quality.
- Incubation of crushed samples in an extraction buffer on an Incubator shaker.

- Precipitation of proteins and lipids by potassium acetate
- Centrifugation and recovery of a clarified lysate enriched with nucleic acids
- Fixing of DNA on silica fibers in the presence of chaotropic salts (guanidium chloride).
- Washing and elution of DNA by centrifugation

Such protocols and associated reagents are offered as commercial kits (Qiagen DNA Easy). However, we master all the steps and we chose to use solutions developed in the laboratory. We can make adjustments (composition of buffer, volumes, incubation time, and incorporation of additional stages of purification) and we use a wide range of silica fibers of different features, in 96 well plates, and with controlling costs. We are anxious to ensure security for samples and comfort to the experimenter. For this, we use of a multi-pipetor of solutions or samples (i.e. Pipetor 96 20-200 µl Rainin Liquidator, ARCAD SP6 budget) in the protocol, usable at any stage.

It would then be possible for a person to extract in a day, high purity DNA from 400 samples previously crushed. This organization is very suitable for all plants worked. It is designed to introduce multiple variations in the protocols and to analyze important series.

We have developed a set of protocols for plant DNA extraction written according to the standards of Quality Assurance in Research. The protocols are organized into a decision support table (not included in this report) that allows us to choose the most suitable one. The table takes into account various parameters. Many variations can be made on the basis of a combination of basic protocol modules. Specific protocol that includes all relevant information is then edited for each individual case (see an example of protocol text in ARCAD Publications section). To date, 18 specific DNA extraction protocols have been developed and validated.

Similar approach to the messenger RNA
An accurate assessment of the various total plant RNA extraction protocols is underway. The quality of RNA extracts and conditions of use (convenience, danger, etc.) are taken into account simultaneously.

b) Secure storage of DNA.

The DNA extracts should be stored to ensure maximum safety. We have conducted a comparison of different conditions of storage of DNA: for average duration at -18°C and for a long term at room temperature.

- For regular use of DNA or for storage of short period (<5 years), we use specific polypropylene tubes (0.6 to 2.2 ml) and screw caps (Abgene or Matrix) that avoid any risk of evaporation and contamination, ordered in dedicated racks of 48 or 96 places. They are stored in secure freezers (-20°C) (ARCAD SP6 budget). Mechanisms of degradation of frozen DNA are already described. They only occur during multiple episodes of freezing/de-freezing that we will reduce and manage.

- For very precious samples and for long term storage (> 5 years), we plan to use the system of protection of DNA under controlled atmosphere in airtight metal capsule (Imagène, Process for long term preservation of DNA at room temperature, www.imagene.fr) compatible with a storage at room temperature. It is the only system

which can guarantee the integrity of the DNA molecule stored for a minimum period of 100 years

Publication : Clermont D., Santoni S., Saker S., Gomard M., Gardais E. and Bizet C. (2014). Assessment of DNA Encapsulation, a New Room-Temperature DNA Storage Method. Biopreservation and Biobanking : 12 (3), 170-183.

c) Management of DNAs in the DNA bank

The chain of custody between samples of plants and DNA samples, using bare-coded labels and appropriate readers is particularly careful.
The tubes or storage capsules are barcoded (2D DataMatrix).

The building of a specific database for the DNA samples, linked to the database of plant genetic resources is underway.

A first DNA bank containing 40,000 DNA samples from different plants (wheat, corn, olive, fig, apricot and grape) is being set up.

d) Transfer of DNA / relationships with partners

Reception of DNA from foreign partners, brought into conformity with the standards of the DNA bank.
Re-purification protocols by capture on silica of DNAs from different origins and purified using different techniques are validated.

Proposal for a specific MTA (Material Transfer Agreement) for DNA
Most MTA applying to seeds can be generalized to transfer DNA or any unmodified derivatives when the material is not subject to any industrial or intellectual property (exchange through the International Treaty on Plant Genetic Resources for Food and Agriculture, for example). For all other cases, a specific MTA should be used. A reflection concerning the regulatory conditions to transfert nucleic acids of plants is under progress at national scale.

5. Partnership

We developed an effective partnership with several DNAbanks:

- Institut Pasteur, Collections de l'Institut Pasteur, Département de Microbiologie, Paris, France
- Banque d'ADN et de cellules, Généthon, Evry, France

6. Conclusions and perspectives

- The ARCAD SP6 program is an intermediate program that provided us with the means to make the link between the IBISA Methodological DNAbank program (2010-2011), focused on the development of protocols and the ARCAD FEDER DNAbank program (2013-2015) focused on the launch of a major DNAbank for mediterranean and tropical plants.

- We now have all the protocols for purification, control and secure storage of DNA samples. We have implemented most useful laboratory equipment. We take care about traceability of the operations. We will soon have a specific database for the management of DNA samples.

- We control every step of the purification process of DNA and can adapt it to different species, different plant organs and for different uses. We are ready to use this DNA resource to support the future studies of genetic and genomic, increasingly precise, programmed by the plant geneticists.

SP7 - Cryopreservation

1. Rationale

Many of the world's major food plants produce orthodox seeds which are tolerant to extensive desiccation and can be stored dry at low temperatures for extended time spans. In contrast to orthodox seeds, the conservation in the form of seeds of a considerable number of species, predominantly tropical or subtropical, is impossible because the seeds they produce are recalcitrant or because they are propagated vegetatively. Cryopreservation (liquid nitrogen, -196°C) is currently the only option for safe and cost-effective long-term storage of genetic resources of such problem species.

Dramatic progress has been made during the last 20 years in plant cryopreservation, thanks to the development of efficient and broadly applicable vitrification-based techniques. However, if tolerance to liquid nitrogen exposure has been demonstrated for a large number of plant species, the large scale, routine applications of cryopreservation in genebanks is still restricted to a limited number of examples. This is particularly true for tropical plant species, which do not possess the dehydration and cold adaptation mechanisms which are found in temperate species.

Priority areas for research in the area of cryopreservation include the development of new (vitrification-based) techniques and their application to additional species, including vegetatively propagated and recalcitrant species, with a strong focus on tropical species; fundamental studies aiming at understanding mechanisms related to resistance to dehydration and liquid nitrogen exposure; evaluation of the potential use of cryopreservation to eliminate viruses (a process termed cryotherapy); and studies on the management of cryopreserved collections and on the integration of cryopreservation in global conservation strategies.

2. Objectives

The three main objectives of SP7 were the following:

- Development of a cryopreservation protocol for yam *in vitro* shoot tips
- Pilot testing of the cryopreservation protocols developed
- Application of the protocols developed to other plant species

3. Materials and Methods

Plant materials

In vitro cultures of various plant species have been used during the implementation of SP7, including: *in vitro* plantlets of yam, sugarcane, *Lithodora rosmarinifolia*, *Limonium serotinum*, *Prunus*, *Rubus*, *Clinopodium odorum*, grapevine, vanilla, hairy root cultures of *Rubia akane* and proembryogenic masses (PEMs) of date palm.

Techniques

The most recent cryopreservation techniques available, including encapsulation-dehydration, droplet-vitrification, D- and V-cryoplate have been used, depending on the species studied.

With yam and *R. akane*, we have also employed qualitative and quantitative histological techniques, image analysis tools and real time microscopy.

4. Activities carried out and main results

a) Yam

The droplet-vitrification cryopreservation protocol developed in collaboration with IITA (Ibadan, Nigeria) has been successfully tested on a total of 42 accessions of African and Asian yam, in IRD Montpellier and in Ibadan. The average recovery percentage after liquid nitrogen (LN) exposure was 29%. The cryopreservation protocol has been tested on several accessions of American yam (*Dioscorea trifida*). Results have shown that this species was very sensitive to LN exposure, producing low survival and random regeneration. The work currently focuses on the optimisation of the shoot tip recovery medium, by comparing various growth regulators (brassinosteroids, topolin, zeatin, etc.). Epibrassinolid is very effective in reducing oxidation phenomena, which are detrimental to shoot tip growth.

In collaboration with CIRAD Montpellier, a detailed histological study has been performed on yam shoot tips during their cryopreservation using the encapsulation-dehydration technique. We have shown for the first time, thanks to the combination of qualitative and quantitative analyses of shoot tip histological sections, that it was the sucrose dehydration step which was responsible for the main cellular plasmolysis, and not the subsequent physical dehydration step.

We performed a preliminary study on the impact of the successive steps of a cryopreservation protocol (droplet-vitrification) on gene expression. The genes selected for this study included six *Dioscorea* genes, highly homologous to TOUCH3, ANAC042, TIL, ABCB4, NDB2, ENO1 *Arabidopsis* genes, as verified by sequence alignment using *Dioscorea* ESTs (aminoacid sequence identity % > 74%; E value < 4,00E-45) whose expression level was modified at least two-fold during cryopreservation in Arabidopsis, as shown by Volk et al. (2011). We also studied nine genes potentially involved in detoxification of reactive oxygen species, including homologs to FSD2 (superoxide dismutase FeSOD), CSD1 (superoxide dismutase Cu/Zn SOD), MSD1 (superoxide dismutase MnSOD), APX2 and APX4 (ascorbate peroxidase), Cat1 (catalase) and GPX2, GPX6 and GPX7 (gluthatione peroxidase). Gene expression was studied using real-time RT-PCR after the following steps of the cryopreservation protocol: dissection; dissection + 2 days recovery; loading treatment; loading treatment + 2 days recovery; PVS2 treatment; PVS2 treatment + 2 days recovery; liquid nitrogen exposure; liquid nitrogen exposure + 2 days recovery. Except for APX2 and APX4, whose expression level remained unchanged throughout the cryopreservation process, the expression of all genes studied varied depending on the step of the cryopreservation protocol. On average, dissection was

the step, which induced the largest changes in expression level for the majority of the genes studied, followed by the loading treatment.

b) Sugarcane

In collaboration with CIRAD Guadeloupe and Montpellier, we have shown for the first time that apices of *in vitro* plantlets could be cryopreserved using the droplet-vitrification technique. However, regrowth was lower, with the two clones tested, than that observed with the original encapsulation-dehydration technique.

c) Prunus and Rubus

In collaboration with INRA Bordeaux and Montpellier, we have shown that it was possible to cryopreserve shoot tips of *in vitro* plantlets of sour cherry (*Prunus cerasus*) using the vitrification technique. Cold treatment of *in vitro* mother-plants could be efficiently replaced with a preculture of shoot tips on sucrose-enriched medium, thereby significantly simplifying and shortening the protocol.

Similarly, in collaboration with the Fruit Research Institute of Cacak (Serbia), we have obtained growth recovery and multiplication of apices of *P. cerasifera* and *Rubus fruticosus* using droplet-vitrification.

d) Vanilla

Preliminary experiments performed in collaboration with CIRAD La Réunion have allowed obtaining limited survival after cryopreservation of vanilla shoot tips using droplet-vitrification.

In the framework of an Open Science project funded by Agropolis Fondation, performed in collaboration with Universidad Veracruzana (Mexico), CEPROCOR (Argentina) and CIRAD La Réunion, we have successfully experimented the V-cryoplate technique for cryopreservation of vanilla shoot tips. The regrowth percentage obtained was still low (around 10%) but we have improved the recovery medium.

e) Grape

In the framework of our collaboration with INRA Montpellier and Zagreb University (Croatia), we have shown for the first time that it was possible to use the droplet-vitrification technique for cryopreserving shoot tips of grape *in vitro* plantlets. We have underlined the critical importance of the physiological state of the buds employed for cryopreservation, by comparing the recovery of buds sampled at different levels on stems of in vitro plantlets, and that of buds sampled on microcuttings cultivated for various periods on media added with different growth regulators. We have also studied the conditions for *in vitro* introduction of a series of Croatian grape cultivars.

In collaboration with Cirad Montpellier, we have studied the possibility of using cryopreservation for elimination of grape viruses (a process called cryotherapy), using virus immunolocalisation techniques. The immunolocalisation protocol has been adapted to our material and experiments have been performed to compare shoot tips of virus-free and virus-infected cultivars, before and after liquid nitrogen exposure. Interestingky, we

have noted that viruses can be eliminated even without liquid nitrogen exposure, by treatment with the highly concentrated vitrification solution employed. Further experiments will be necessary to confirm this observation.

f) Endangered species

We have shown for the first time that apices of *in vitro* plantlets of two endangered Mediterranean species (*Lithodora rosmarinifolia* and *Limonium serotinum*), provided by the University of Palermo (Italy), could withstand cryopreservation using droplet-vitrification, with modifications of the treatment conditions, compared with the original technique.

In collaboration with CEPROCOR (Argentina) research has been performed for the development of a cryopreservation protocols for the endangered Argentinian species *Clinopodium odorum*. High recovery has been achieved using the V cryo-plate technique and limited recovery with the D cryo-plate.

g) Rubia akane

In collaboration with RDA (Korea) and Cirad Montpellier, we have established a cryopreservation protocol (droplet-vitrification) for hairy roots of this species, and performed a qualitative and quantitative histological study of the impact of the successive steps of the protocol on their structural integrity. We have shown that plasmolysis was mostly obtained during the sucrose preculture. However, when comparing sections performed at two different levels, we have shown that in the median part, plasmolysis was observed already during the sucrose pretreatment step, while in apical segments, plasmolysis was observed only after the loading step. These differences of reactivity can be related to differences in histological structure and in survival, the apical segments displaying higher survival compared to median sections.

h) Date palm

Two cryopreservation protocols (droplet-vitrification and D-cryoplate) have been compared with date palm proembryogenic masses (PEMs) and the response of four date palm varieties to LN exposure is studied. When studying the effect of sucrose pretreatment on PEM survival, we have shown that sucrose pretreatment allows obtaining survival after LN exposure without treatment with a vitrification solution.

5. Partnership

The project has allowed strengthening existing partnerships with French and foreign research institutes and establishing new partnerships with research institutes both within and outside Europe.

6. Training

Individual training periods in cryopreservation techniques have been implemented for nine researchers and technicians, for a cumulative total of 24 training months.

7. Structuring effects, synergies created by the SP

The implementation of SP7 has had a very positive impact on cryopreservation research and development activities at the national level by allowing closer collaborations and interactions with INRA and CIRAD research teams and CRBs. Another important point is the inputs made into the future ARCAD conservation facility in Montpellier, which will include an *in vitro* conservation and cryopreservation laboratory.

8. Visibility, attractiveness and national and international positioning

The implementation of SP7 allowed increasing the activity level in cryopreservation research performed in IRD Montpellier. This has resulted in increased visibility through communications in scientific conferences and publication of scientific articles and in increased attractiveness, attested by the requests for collaboration from foreign researchers. Finally, this increased visibility and attractiveness have been instrumental in further strengthening the national and international position of the IRD cryopreservation research team.

9. Conclusion

The project has allowed to establish/optimize cryopreservation protocols for a range of different plant species from Mediterranean and tropical origin. In one case (yam), the protocol established has been tested at the pilot scale in the genebank context. In other cases (e.g. sugarcane, grape, *Prunus*, *Rubus*), the protocols established are ready for large scale testing. Finally, we have increased our understanding of the mechanisms involved in cryopreservation protocols through the use of analytical techniques such as histology. In conclusion, the main future research areas to be addressed are the following:

- Development of cryopreservation protocols for additional plant species, including both vegetatively propagated and recalcitrant seed species, with a strong focus on tropical and Mediterranean plant species.
- Improving our understanding of biological and physical mechanisms involved in cryopreservation processes, using different analytical tools.
- Testing the applicability of cryotherapy and comparing its efficiency with other existing virus eradication techniques.
- Large scale application of cryopreservation in the genebank context.
- Integration of cryopreservation in global conservation strategies for priority plant species.

CONCLUSIONS

1. Delivering results and creating synergies

From the very beginning, we managed and developed ARCAD with the vision that we were building a long-term initiative, of which the 4-year scientific project was only the 'germination stage'. The objective was therefore to establish the bases for a steady development, and to make ARCAD visible and recognized, in the regional, national and international landscapes.

ARCAD concept was built on the evidence that the Agropolis scientific community had a strong expertise in the field of crop genetic resources. One of the objectives was to make this community more visible, innovative and efficient by bringing it together in a common 'flagship project'. Under the unifying theme of 'domestication', the initial choice was however to acknowledge the diversity of research conducted by this community by aggregating activities dealing from molecular evolution to anthropology.

The first level to create synergy was the sub-project level. The main 3 sub-projects were managed by 2 to 4 co-leaders, which allowed the sharing of duties and, therefore, lightened the workload on scientists who are often in-charge of other duties. Sub-project co-leaders played a key-role to ensure interactions between scientists belonging to different teams or institutions. Even though it was not a strict rule, we paid attention to the gender-balance of the leading teams. We also encouraged the emergence of young research leaders.

At the global project level, the yearly *Journées ARCAD* brought together most, if not all, of the scientists involved in ARCAD activities. The exchange of information between teams was essential to initiate new activities (first sequencing of fonio, for example). We believe that these events have strongly contributed to a growing sense of ownership of the project by the scientists.

Numerous examples of interactions created between Agropolis teams are given in the former parts of this report. SP1 Comparative Population genomics was particularly successful in bringing together a large number of teams and in sharing methodologies and concepts. ARCAD SP4 *Bioinformatics* structured a community on NGS bioinformatics tools to address biodiversity studies thanks to regular meetings, every 2-3 weeks, between biologists and bioinformatics specialists and training sessions organized for ARCAD community. Training workshops were open to students and scientists of the South posted in diverse research teams of Montpellier, whether or not they were involved in ARCAD projects.

Another example of strong interactions between the ARCAD sub-projects was the research on a minor crop like fonio. This crop benefited from the study of its genetic diversity in Africa and interaction between genetic diversity and human society (carried out under SP3), the building of reference transcriptom and generation of thousand SNP markers (SP1, SP4), and the development of new in-depth approaches (GBS) for studying its diversity (SP5). The study of adaptation to climate variation of another crop like pearl millet (SP2) benefited from methodologies developed in other work packages of SP1 and SP5.

Training

The organization of training activities proved indeed to be a remarkable way to promote interactions between teams, between disciplines and also with international partners (see **Annexe 3**). In particular, this was instrumental in fostering interactions between teams involved in the SP3 *Cereals in Africa*. Thus, the second international course "Agrobiodiversité : des hommes et des plantes. Outils et Méthodes d'analyse" was organized in Rabat, Morocco, from May 3-14, 2010 with the support of the Institut Agronomique et Vétérinaire Hassan II. This course (given in French) provided trainees with a multidisciplinary approach to plant agrobiodiversity research. This 2-week course was organized in two main parts: population genetics on the first week, social sciences on the second week. It was delivered to 25 trainees coming from 12 countries, mobilized a dozen of trainers mainly from ARCAD SP3 but also from Bioversity International. Training materials were produced for the course and improved afterwards. They will be used for the next sessions of the course and might serve as the basis for writing a manual on agrobiodiversity analysis.

Annexe 4 provides the list of the staff recruited within the frame of ARCAD, whether funded or not by ARCAD. In total, 41 staff (16 female and 25 male) were recruited under ARCAD budget. As required by Agropolis Fondation rules, the PhD student (1) and the 5 Post-doctoral fellows were posted in a foreign country for more than 1 year prior to their recruitment. Of the 20 Msc students, 7 are foreign citizens from Angola, Brazil, Benin, England, Germany, Mali and Senegal. Three of the 5 post-doctoral fellows got a permanent position after their contract with ARCAD and sometimes before the end of the contract. This is a good indicator of their quality as scientists but also of the value of the ARCAD research.

Communication

The development of the web site (www.arcad-project.org) contributed to the overall visibility of the project, both internally and externally. It had a satisfactory record of visits (3800 per year on average).

The co-organization of the national conference on Genetic Resources (2011) with the FRB (French Foundation for Research on Biodiversity) was one the key-events that we used to bring ARCAD to the front of the French stage.

Numerous talks were given in international meetings to present ARCAD. In addition to the interest in ARCAD research activities that was often noticed, it was clear that having created a main, if not single, entry point to Agropolis conservation and research activities on genetic resources is very much appreciated by international partners. For example, this permitted to organize for the first time in 2010 a workshop between the Agropolis community and Bioversity International to discuss opportunities for scientific partnership. Likewise, the on-going discussions about a possible involvement of Agropolis scientists into the Platform for Agrobiodiversity Research were certainly initiated because ARCAD is now able to play a role of focal point in Montpellier.

We also developed communication activities towards the broad audience (see **Annexe 2** for details). Besides fulfilling an essential citizen communication duty, this contributed to make ARCAD more known in the local landscape.

Publications and new projects proposals

The updated list of publications produced by ARCAD community is provided in **Annexe 5**.

Annexe 6 presents the list of projects proposals related to ARCAD that were funded or are being evaluated. These proposals are important achievements as they demonstrate that ARCAD has played its role of “seed project”, not only in creating new research opportunities but also in attracting new partners.

Other products

Annexe 7 presents the list of the main data (of sequencing, genotyping and phenotyping) generated during ARCAD and **Annexe 8** the list of different tools set up (analysis tools, software, methods) as well as the websites created.

Management of biological resources

Developing a shared vision of the management of biological resources within the ARCAD project was one of the first approaches used to promote the project internal identity, given the diversity of practices among teams and institutions. Access to biological resources was essential to fulfil the objectives of ARCAD scientific project. A large proportion of those resources were already available in several biological resource centres located in Montpellier or in French overseas territories, and managed by the three main research institutes (INRA, CIRAD, IRD). Other sources of biological resources are international centers members of CGIAR consortium, regional research institutes, national agricultural research institutes, universities, botanical gardens, and private companies. New collecting activities in African and European countries were also required for wild material and landraces.
Increase ARCAD community awareness about the rules governing the access to biological resources as defined by international conventions (CBD and ITPGRFA) was a key issue when the project started. In 2010, under the supervision of ARCAD coordination, an MSc student in Biotechnologies and Law studied the current practices of ARCAD researchers in that regard and made recommendations for improvement. Information about the good practices was largely disseminated, particularly during the *Journées* ARCAD. As far as possible we made sure to comply with those rules and MTAs (Material Transfer Agreements) were signed with the different partners.

The list and status of the biological material used in ARCAD is in **Annexe 9.**

International partnerships

Annexe 1 presents the list of partnerships developed in the frame of ARCAD. Southern partners are essential to ARCAD, particularly in the implementation of the SP3 *Cereals in Africa*. We encountered difficulties not with partners themselves but because of the

political events in West Africa. Activities in Guinea were delayed by one year, but since then, partnership with IRAG has been extremely fruitful. On the other hand, field activities in Niger and Mali have been cancelled for security reasons. Through ARCAD, we were able to develop an efficient partnership on durum wheat diversity with Morocco, including the joint supervision by IAV Hassan II and Montpellier SupAgro of two Moroccan PhD students.

Numerous contacts were taken with CGIAR research centers and Research programmes (CRPs) as well as the Consortium on agrobiodiversity issues. Promising discussions took place, but no concrete partnerships had been developed until Bioversity International was closely associated to the writing of the proposal on information systems for agrobiodiversity management submitted to ANR Agrobiosphere 2013.

2. Perspectives

The challenge ahead of ARCAD is now to prepare the next phase of its development:

- Define its role and area of activities, in connection with the development of its physical facilities, which will need thorough discussions among the ARCAD consortium members and also the joint research units AGAP and DIADE;
- Define its scientific project, based on perspectives derived from the 2009-2014 project and partnership opportunities.

ARCAD physical facilities: current status

The objective of developing new facilities to host ARCAD activities has been part of the ARCAD programme since its very beginning, taking advantage of the commitment of the Languedoc-Roussillon Region to support the development of new facilities to the tune of EUR 5 millions.

The interinstitutional working group shaped in 2007 the overall design of the ARCAD facilities. They should include:

- A biological resources conservation component for :
  - the conservation of refrigerated orthodox seeds (to host seed collections maintained by CIRAD, INRA and IRD);
  - the cryopreservation of the genetic resources of recalcitrant plants;
  - the high-throughput, high-quality treatment of DNA (extraction, purification), its storage and distribution.
- A DNA analysis component
- Genotyping and sequencing facilities
- Office space to host scientists, partners and students.

This overall scheme, i.e. the development of facilities that integrate conservation, analysis and research facilities, is still the one which is shared by the members of the ARCAD consortium (INRA, CIRAD, IRD and Montpellier SupAgro).

In 2009, under the supervision or the Languedoc-Roussillon, plans were developed with an architect to design a project reshaping the Agropolis Museum building into ARCAD

facilities. The cost of 3500m2-project was estimated to EUR 14 millions, which was significantly higher than the available budget (EUR 10.5 million by that time).

Complementary funding was sought through two proposals submitted to the calls "Infrastructures nationales en Biologie-Santé" of the "Investissements d'Avenir" in 2010 (ARCAD+ proposal) and 2011 (CAReGen[4] proposal). These proposals were not funded.

In October 2012, the members of the ARCAD consortium decided to launch the building programme with the available funds (EUR 7.1 million). The location of the future building has been identified on the campus of La Valette.

In order to anticipate the availability of the future building (expected by early 2017), an equipment request proposal was submitted to a FEDER funding (Fonds européens pour le développement régional). The proposal was accepted to the tune of EUR 3.6 millions. The grant supports the development of the following platforms through the acquisition of new equipments and temporary staff positions: Seed phenotyping, Cryopreservation, DNA Bank, Genotyping, and Information system.

Which ambition for ARCAD?

Despite the CAReGen proposal was not funded in 2011, its writing contributed to further conceptualize the role that ARCAD could play within the French genetic resource conservation system.

It was thus written in the CAReGen proposal :

> *"The public sector actors and their coordination bodies join the [CAReGen] initiative (CIRAD, INRA, IRD, Montpellier SupAgro, FRB, Agropolis Fondation) together with representatives of the seed profession (GEVES) and conservation associations (FCBN), an international agricultural research center (Bioversity International) and a young private company (ADNid). All existing plant BRCs in mainland France and overseas will be involved, putting together critical mass, expertise, coordination and scope for impact, concretizing a high collective ambition. Montpellier will be used as the focal centre, building on the large local research and education community on PGR, the successful collective endeavour of ARCAD (Agropolis Resource Center for Crop Conservation, Adaptation and Diversity), the visibility enhanced by the recent arrival of the CGIAR (Consultative Group for International Agricultural Research) headquarters, at the core of international stakes for agriculture-based economic development and biodiversity preservation and valorization.*
>
> *Building on the current BRCs, the project will consolidate the infrastructure with complementary investments and address methodological and technological developments for seed conservation and management, tissue cryopreservation, DNA extraction, conservation and distribution, medium-throughput genotyping, information storage and exchange, training on PGR conservation and analysis.*
>
> *CAReGen will provide a one-gate entry to the users and expand currently existing services to:*
>
> - *the identification and documentation of information-rich germplasm subsamples for subsequent high throughput genotyping*

[4] Infrastructure nationale de **C**onservation et d'**A**nalyses des **Re**ssources **Gén**étiques Végétales

- *the analysis of global patterns with the view to setting collecting priorities*
- *a broader range of species and purposes*
- *the extension to (semi)elite materials for a fine integration with breeding processes and priorities*
- *updates and advises on intellectual property and access and benefit sharing rules and practices*

The uniqueness of the infrastructure will rest on: its mission to serve, at the appropriate level, all sorts of stakeholders; its broad geographic and climatic coverage; its broad crop and species (including wild species) coverage; its fine integration within a vibrant research and education community; the development of pilot tools, such as robotics-based conservation or software for phylogeographic analyses; the integration of multiple functionalities leading to a global offer for services from a one-gate entry."

Even though the national ambition of ARCAD has to be set down to a lower level, the support available for its physical facilities make possible to envision that ARCAD will be able in a near future to meet significant objectives:

- develop an upgraded and coordinated conservation of the genetic resource collections maintained in Montpellier
- set up a national platform for cryopreservation
- set up a national platform for DNA banking
- set up upgraded facilities for DNA analysis (Pre NGS lab and SNP lab)
- develop, in partnership with the INRA URGI in Versailles, a web portal and information system on the collections maintained by the Biological Resource Centres

ARCAD will therefore fully deserve its name by being able to offer biological and technological resources to the users. It will have to carry on its effort to deliver methodological, knowledge and training resources, as it is being done through the 2009-2014 scientific project. This will make more salient the need i) to develop research aiming at defining more efficient and smart conservation strategies and ii) to include a policy component into the future ARCAD activities, both to better understand how regulations, policies and collective arrangements affect the mobilization and management of crop genetic diversity and to contribute to the international debates on genetic resources.

ANNEXES

Annexe 1. Partners and nature of partnership

ARCAD sub-project	Name of the partner	Country	Formalized agreement signed	Research	Capacity building	Biological resources exchange	Design/submission of new projects
SP2	University of Minnesota (Medicago Hapmap, NSF, Prof. N. Young	USA	X	x	x		
SP3	KARI (Kenyan Agricultural Research Institute)	Kenya	X	x	x	x	?
SP3	IER (Institut d'Economie Rurale)	Mali	X	x	x	x	
SP3	IRAG (Institut de Recherche Agronomique de Guinée)	Guinée	X	x	x	x	x
SP3	LMI LAPSE (Laboratoire mixte international Adaptation des Plantes et des microorganismes associés aux Stress Environnementaux)	Sénégal		x	x		x
SP3	UCAD (Université Cheikh Anta Diop)	Sénégal			x		x
SP3	ISRA (Institut Sénégalais de Recherches Agricoles)	Sénégal				x	x
SP3	AfricaRice	Sénégal		x		X	
SP3	CERAAS	Sénégal		x			x
SP3	Université de Thiès	Sénégal		x			x
SP3	Université Abdou Mamouni	Niger			x		
SP3	INERA (Institut de l'Environnement et Recherches Agricoles)	Burkina Faso	X	x	x	x	
SP3	Institut Agronomique et vétérinaire Hassan II	Morocco	X	x	x	x	x
SP7	IITA (International Institute of Tropical Agriculture)	Kenya		x	x	x	
SP7	Fruit Tree research Institute Čačak	Serbia		x	x		
SP7	Universidad Veracruzana	Mexico	X	x			

ARCAD sub-project	Name of the partner	Country	Formalized agreement signed	Research	Capacity building	Biological resources exchange	Design/submission of new projects
SP7	CEPROCOR	Argentina	X	x	x		
SP7	Zagreb University	Croatia	X	x	x		
SP7	RDA (Rural Development Administration)	South Korea	X	x	x		
All	UMR AGAP	France/Antilles	X	x	x	x	x
All	UMR DIADE	France/La Réunion	X	x	x	x	x
SP1	UMR GALF	France		x		x	
SP1	UMR RPB	France		x			
SP1	UMR ISEM	France		x			x
SP1, SP2	Génopole Toulouse	France	X		x		
SP1, SP3	UMR GV (Génétique Végétale Inra Moulon)	France	X	x	x	x	
SP2	LIPM Toulouse	France		x	x		
SP7	UR UREF (INRA Villeneuve d'Ornon)	France		x		x	
SP7	UR ASTRO (INRA)	Guadeloupe		x		x	
SP7	UMR PVBMT (CIRAD, Univ. Réunion)	La Réunion		x		x	

Annexe 2 – Coordination activities

WP1- Scientific animation and evaluation

Animation

- Réunions du comité de coordination (coordination et responsables de SP et WP) à raison d'une réunion tous les 2 mois en moyenne.
- Journées ARCAD 17-18 novembre 2010 - 85 participants (scientifiques associés au projet et représentants des institutions membres du consortium ARCAD)
- Journées ARCAD 23-24 novembre 2011, comprenant : une matinée ouverte à la communauté scientifique avec une conférence invitée (Jérôme Salse, INRA Clermont-Fd) et des exposés sur ARCAD, et de 3 demi-journées réservées aux scientifiques associés au projet et aux représentants des institutions membres du consortium ARCAD. Succès de la matinée ouverte (plus de 120 présents), bonne participation d'ensemble au reste des journées.
  Présentations Journées ouvertes 23 novembre 2011
  Présentations Journées ARCAD 2011
- Pas de journées ARCAD en 2012, remplacées par un séminaire interne de 2 jours les 11-12 février 2013. Le séminaire a rassemblé les responsables de sous-projets et work-packages, ainsi que 6 personnes- ressources, soit au total 33 participants. Organisé pour préparer l'évaluation d'ARCAD prévue au deuxième trimestre 2013, le séminaire a permis de partager les informations sur le projet immobilier, de faire un bilan des activités à ce jour et de discuter des perspectives de recherche.
  Présentations Séminaire ARCAD 11 février 2013
- Les Journées de clôture du projet scientifique ARCAD se sont tenues les 27 et 28 octobre 2014. La première journée était consacrée à la présentation du Centre de ressources ARCAD avec des exposés sur les collections de ressources génétiques végétales et les plateaux techniques en cours de développement (conservation, traitement de l'ADN, génotypage, bioinformatique,…). Le 28 octobre était consacré à un séminaire scientifique « Histoires de plantes cultivées : domestication, adaptation, diversité » animé par les chercheurs du consortium scientifique ARCAD et par des conférenciers invités. Présentations Journées du 27 et 28 octobre 2014

Expertise

Participation of the project leader to :

- Groupe de travail et Comité d'évaluation 2011 ANR BioAdapt
- Comité d'évaluation 2011 ANR Biodiversa
- Comité d'évaluation ANR BioAdapt 2012
- Atelier organisé par le Consortium du CGIAR « « Towards a CGIAR Consortium Strategy on Agro-Biodiversity Research », Montpellier 24-26/07/12
- Membre du CS de l'Unité DIASCOPE

Project management

- Réunions sur le dossier du bâtiment ARCAD avec architecte programmiste, scientifiques et représentants des institutions du Consortium - 4 réunions en 2009
- Montage du projet ARCAD+ soumis à l'appel à propositions 2011 « Investissements d'avenir - Infrastructures Nationales Biologie Santé »
- Développement de partenariat avec la FRB : Réunion FRB Ecoscope 29 mars 2011
- Développement de partenariat avec Bioversity International : Réunion sur la composante agrobiodiversity du CRP 1.1, Rome, 2-3 décembre 2011
- Montage du projet CAReGen « Infrastructure nationale de Conservation et d'Anlayse des Ressources Génétiques des Plantes », soumis à l'appel à propositions 2012 « Investissements d'avenir - Infrastructures Nationales Biologie Santé » Atelier de montage organisé à Agropolis les 18-19/07/12
- Soumission d'un projet à l'AAP FRB-Cesab :2011 « Towards a global monitoring system of crop agrobiodiversity"
- Développement de partenariat avec la Platform for Agrobiodiversity Research : réunions avec T Hodgkin, coordinateur de la PAR en juillet 2012 et février 2013 à Montpellier et avec le Steering committee de la PAR à Rome en février 2013
- Participation à des réunions relatives aux CGIAR Research Programmes Dryland areas ; Water, Land, Ecosystems ; Roots, Tubers and Bananas; GRiSP
- Soumission du projet ARCAD au FEDER 2010-2015, décembre 2012 (3.6M€). Projet accepté (janvier 2013-juin 2015)
- Mise en œuvre du projet FEDER en liaison avec l'INRA (institution porteuse) : recrutements du personnel sur contrat, coordination des différentes composantes
- Réunions sur le projet immobilier ARCAD, 4 réunions en 2012, 6 réunions du comité technique (COTEC) depuis mars 2013
- Mission (2014) sur la statégie et gouvernance d'ARCAD Centre de Ressources

Evaluation

- Information au SC8
- Rapport annuel lors du SC9
- Rapport annuel + proposition des ToRs de l'évaluation ARCAD au SC10
- Finalisation des ToRs de l'évaluation ARCAD
- Tenue de l'évaluation externe du projet ARCAD les 5, 6 et 7 juin 2013
  Lien vers le rapport d'évaluation

WP2 Communication

Conferences/seminars oral communications.

- Atelier Consortium CIBA. Montpellier, octobre 2009
- Conference TDWG (Biodiversity Information Standards), Montpellier, novembre 2009
    - *Pham J.L., Labouisse J.P., 2009.* ***ARCAD- Agropolis Resource Center for Crop Conservation, Adaptation and Diversity: a new open multi-function platform devoted to agrobiodiversity. Objectives and challenges*** *[Abstract] In: Weitzman A. L. (Ed.). Proceedings of TDWG 2009 Annual Conference, Montpellier, France, 9-13 november 2009. URL: http://www.tdwg.org/proceedings/article/view/544*

- Yam agrobiodiversity workshop. Montpellier, décembre 2009
- Atelier Agropolis-Bioversity International. Montpellier, mars 2010
- Foundation Week – Investing in the future: A roundtable discussion on the role of foundations in advancing research in agriculture, food and biodiversity. Bruxelles-Belgique, juin 2010
- Ecole thématique internationale, Agrobiodiversité. Rabat-Maroc, mai 2010
- Atelier international « De la connaissance à la valorisation du fonio ». Niamey, Niger, décembre 2010
- Journée Ressources Génétiques Végétales de la FRB. Paris, décembre 2010
- Journée Marché et mise en patrimoine de la biodiversité , MNHN, 7 décembre 2011
- Journées des Centres de Ressources Biologiques INRA, Ploudaniel mars 2012
- CS de l'unité DIASCOPE, mai 2012
- 2 communications à l'atelier "Toward a CGIAR Consortium Strategy on Agro-Biodiversity Research", Montpellier , juillet 2012
- Comité de direction DIADE mai 2012 et janvier 2013
- Conférence Common Knowledge Commons, Louvain la Neuve, Louafi et al. 13 septembre 2012
    - *Louafi S, Arnaud E, Barthelemy D, Noyer J-L, Pham J-L. 2012.* ***Value, norms and practices in plant biodiversity-based research and innovation commons.*** *First thematic conference on the knowledge commons. Governing pooled knowledge resources: Building institutions for sustainable scientific, cultural and genetic resource commons. 12-14th September 2012, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.* Résumé ; Présentation PPT
- Conférence Genomics of Plant Genetic Resources, Corée du Sud 16-19 April 2013, Jeju, South Korea.
    - *Glaszmann J-C, Pham J-L, Labouisse J-P. 2013.* ***ARCAD, a structuring node in the French plant genetic resource conservation system.*** *3rd International Symposium on Genomics of Plant Genetic Resources, 16-19 April 2013, Jeju, South Korea.*
- Atelier Ressources Génétiques du GIS- Biotechnologies Vertes (mai 2014)

Co-organization of seminars and workshops

- Atelier de réflexion et débat sur la mise en œuvre au niveau français du TIRRPA et du protocole de Nagoya (CDB) –Montpellier- 7 juin 2011
- Atelier de doctorants travaillant sur la domestication des plantes, avec implication de chercheurs Arcad, Agropolis International, 14 avril 2011
- Colloque FRB : Les Ressources Génétiques face aux nouveaux enjeux environnementaux, économiques et sociétaux, Montpellier- 20-22 septembre 2011
- Atelier Yam Crop Wild Relatives, Montpellier 2-4 juillet 2013

Presentations to visitors and external invitations

- Bioversity International. Rome, janvier 2009
- ICARDA-ICRISAT-CIAT-AARINENA. Montpellier, mars 2009
- Département BIOS-CIRAD. Montpellier, mars 2009
- Département ES-CIRAD. Montpellier, 2009
- Genetrop-IRD. Montpellier, 2009
- SYNGENTA. Montpellier, mars 2010
- C Fauquet, ILTAB (USA). Montpellier, mars 2010

- JL Jannink, Univ Cornell (USA). Montpellier, mars 2010
- INRA, Atelier sur les CRB, Bordeaux, juin 2010
- J. Kroymann, CNRS Orsay. Montpellier, août 2010.
- Fondation Cariplo. Montpellier, septembre 2010
- Global Crop Diversity Trust. Rome-Italie, novembre 2010
- Atelier Agrobiodiversité GFAR-TIRPAA,
- Réunion projet Sud Expert Plantes – Bioversity, Montpellier, 21 mars 2011
- Alexandra. Jorge, AWARD, Banque de gènes ILRI – Ethiopie, 8 avril 2011
- Délégation Séminaire AWARD 8-9 juin 2011
- Soirée Agropolis de la Conférence G20 Recherche pour le Développement, 12-13 septembre 2011
- Atelier Agropolis-Laboratoires Pierre Fabre, 14 novembre 2011
- Délégation PROCISUR, 08 février 2012
- Délégation ICARDA, 25 avril 2012
- Visite du Ministre aux Affaires Européennes, 17 décembre 2012
- Paula Bramel, Global Crop Diversity Trust, mai 2014

Participation to events without formal communication

- Colloque Crops for the Future, Kuala Lumpur, 27 juin-1er juillet 2011
- Colloque Eucarpia Genetic Resources, Wageningen, 5-7 avril 2011 (JLP rapporteur de session)

Posters exhibitions

- Conference TDWG (Biodiversity Information Standards), Montpellier, novembre 2009
- Science Forum. Wageningen - Pays-Bas, juin 2009
- Annual meeting Generation Challenge Programme, Bamako-Mali, septembre 2009
- Conférence Diversitas, Capetown, Afrique du Sud, octobre 2009
- Sciences en Fête, Montpellier, octobre 2010

Public animations

- Fête de la Biodiversité, animation scolaire + conférence, Montpellier, mai 2010
- Fête de la Biodiversité, animation scolaire, Montpellier, mai 2011, et mai 2012
- Festival Saperlipopettes, mai 2012 et avril 2013

Interview

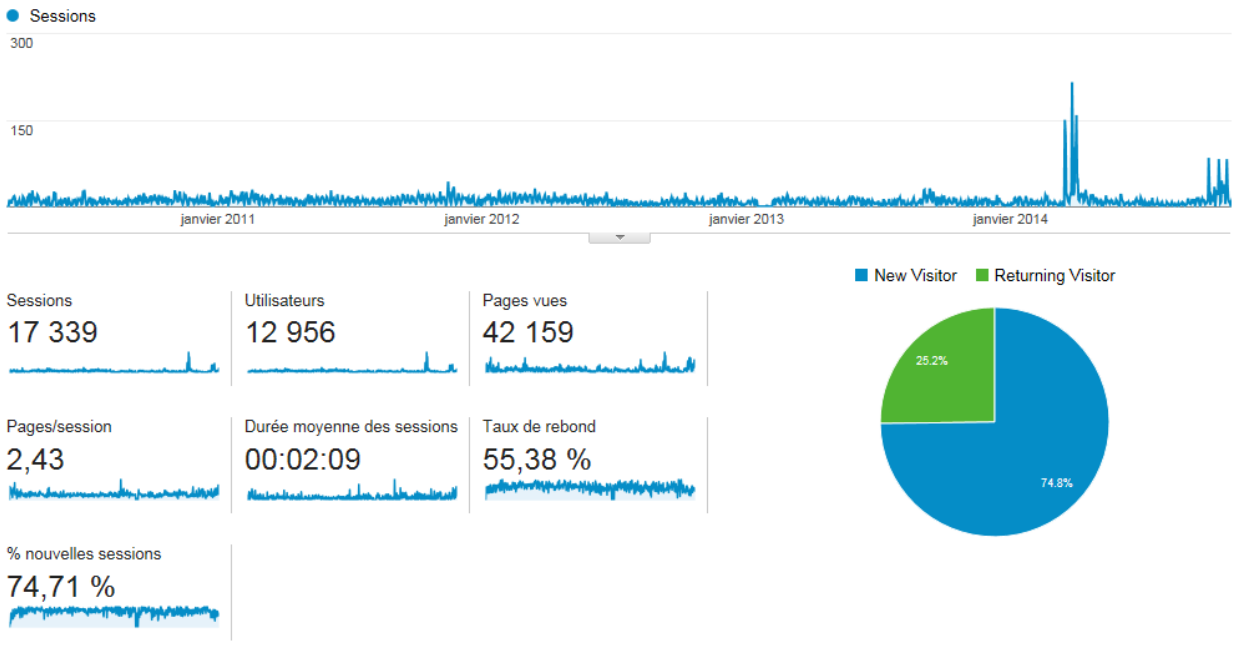
Interview JL Pham pour la Lettre Internationale d'Agropolis n°9 http://www.agropolis.fr/pdf/publications/lettre-internationale-agropolis-9.pdf
Reportage FR3-LR, 20 mai 2014
Interview de J-L Pham pour « Semences et Progrès », novembre 2014

Communication tools

- Rédaction de la charte de discours, conception du logo et charte graphique avec l'agence Wellcom (2009)
- Espace Intranet sous Quick-R réalisé (2009)

- Site Internet réalisé sous Ez publish (www.arcad-project.org). Mise en ligne le 15 mars 2010. Depuis cette date on a observé une moyenne de 3769 visites par an (dont 2816 visiteurs individuels) de 164 pays différents.

  Nombre de sessions du 15 mars 2010 au 1er décembre 2014



Top 10 des pays d'origine des visites

	Pays	Sessions	% Sessions
1.	France	8 708	50,22 %
2.	United States	1 386	7,99 %
3.	India	1 059	6,11 %
4.	Morocco	419	2,42 %
5.	United Kingdom	348	2,01 %
6.	Italy	271	1,56 %
7.	Brazil	266	1,53 %
8.	Canada	260	1,50 %
9.	Germany	248	1,43 %
10.	Spain	217	1,25 %

- Production à l'occasion du colloque FRB 2011 de matériel de communication présentant ARCAD : 1 kakemono, 1 plaquette (en français)
  PLAQUETTE ARCAD
  KAKEMONO ARCAD

- Production d'un dépliant à l'occasion des Journées de cloture d'ARCAD (Octobre 2014) DEPLIANT ARCAD

WP3 Activities on cross cutting issues

- Etude de l'accès aux ressources génétiques et leur circulation au sein du projet ARCAD et avec ses partenaires (Stage MSc2 Ashley Meter):
    - Travail bibliographique, enquêtes
    - Séminaire interne ARCAD le 26 mars 2010
    - Production d'un rapport de stage incluant des recommandations:
      Meter A. (2010) Projet ARCAD : Développement d'une démarche qualité pour la gestion des flux de ressources génétiques. Rapport de stage M2 Sciences, Technologies, Santé Mention : Biotechnologies-Droit, Université de Tours. IRD Montpellier. 45 pages + annexes
    - Restitution aux journées ARCAD le 18 novembre 2010

- Co-organisation de l'Atelier de réflexion et débat sur la mise en œuvre au niveau français du TIRRPA et du protocole de Nagoya (CDB) – 7 juin 2011

Annexe 3 – Training

Name/Content	**Participants (number, origin)**	**Duration (days)**	**Date**	**Location**
International Course on Agrobiodiversity « Des plantes et des hommes : outils et méthodes d'analyse)	25 – International (12 countries)	12	May 2010	Rabat (Morocco)
Sequences analyses for plant breeding	20-Agropolis scientists	5	6-10 June 2010	Montpellier
Sequences analyses for plant breeding	20- APIMET-SEPMET SupAgro students	5	21-25 February 2011	Montpellier
SP1-SP4 Internal training ARCAD: Analysis of polymorphism data	20 – mostly SP1 partners	5	9 – 13 May 2011	Montpellier
Sequences analyses for plant breeding	20-Montpellier SupAgro	5	4-9 February 2012	Montpellier
Training on Galaxy workflow system	20-Montpellier scientists	1	22 June 2012	Montpellier
Sequences analyses for plant breeding	20-Montpellier SupAgro	5	February 2013	Montpellier
Training for studying the evolutionary history of Amazonian crops and Amazonian biodiversity in Brazil and French Guiana	17 - Brazil & France	2 x 5	March 2013 July 2013	Cayenne (Guiana) Manaus (Brazil)
SP1-SP4 workshop in September 2013 on simple sequence analysis on the assembled data sets	SP1 partners	5	September 2013	Montpellier

Annexe 4. Students, post-doctoral fellows and other staff recruited in the framework of ARCAD project

Recruited staff in brief

1) Staff directly recruited on ARCAD budget

	Gender		Number and origin (or cursus) of recruited staff (May2013)		
	F	M	France	Foreign countries	**TOTAL**
PhD students	1	-	-	1	**1**
Post-doctoral fellows	1	4	-	5	**5**
Masters	6	14	13	7	**20**
Engineers	3	7	9	1	**10**
Technicians	5		5		**5**
TOTAL	**16**	**25**	**27**	**14**	**41**

2) Other (non permanent) scientific staff significantly involved in ARCAD activities but recruited on other budget

	Gender		Number and origin (or cursus) of recruited staff (May2013)		
	F	M	France	Foreign countries	**TOTAL**
PhD students	2	3	-	5	**5**
Post-doctoral fellows	3	2	1	4	**5**
Engineers	2	2	3	1	**4**
Masters	-	4	-	4	**4**
TOTAL	**7**	**9**	**4**	**14**	**18**

Grey background: Not funded by ARCAD

	Academic level ***Name, Surname***	***Foreign natio-nality***	***Origin***	***Contract period and recruiting institution***	***Main assignment or <u>title of the dissertation</u>***	***Position after ARCAD contract***
Coordination	Master **Ashley Meter**		Master Biotechnologie et Droit, Université de Tours	1/2010 - 7/2010 (IRD)	**<u>Meter A. 2010</u>**. <u>Projet ARCAD:Développement d'une démarche qualité pour la gestion des flux de ressources génétiques. Rapport de stage M2 Sciences, Technologies, Santé Mention : Biotechnologies-Droit, Université de Tours.</u>	Junior Brand Manager Abilify, Otsuka Europe, Otsuka Pharmaceutical Co., Ltd. Uxbridge, UK
SP1	Post-doctoral fellow 1 (WP3) **Benoit Nabholz**		Department of Evolutionary Biology, Uppsala University, Sweden	1/2011 - 1/2012 (Montpellier SupAgro) then 2/2012-7/2012 (INRA funded)	Analysis and comparison of the effect of domestication on genome evolution in different crop species with contrasted life-history traits, phylogenetic position, and domestication history.	Associated professor at ISEM
	Post-doctoral fellow 2 (WP5) **Iris Fisher**	Germany	University of Munich, Germany	4/2012 – 10/13 (IRD)	Documentation of genomic variations of the selective pattern among several crops species and their wild relatives in order to answer different questions related to (i) domestication, (ii) relations between life history traits or genomic environment and selective patterns and (iii) functional evolution.	Post-doctoral position at ISEM (2 years)
	Post-doctoral fellow 3 (WP4) **Yves Clément**		International Max Planck Research School for Computational Biology and Scientific Computing, Berlin, Germany	12/2012 – 6/2014 (INRA)	Analysis of the evolutionary forces affecting GC-content dynamics in several species distributed over the angiosperm phylogeny, including several monocot and eudicot species.	
	Post-doctoral fellow *(WP5)* **Jacques Dainat**		Evolution Biologique & Modélisation, Aix-Marseille University, France	*1/2013 – 12/2013 (SupAgro funded)*		
	Research Engineer (WP3) **Yan Holtz**		Montpellier Supagro, APIMET master, Montpellier, France	2/2013 – 12/2013 *(INRA funded) then 2/2014-6/2014 (ARCAD)*	Methodological development for the use of NGS on polyploid species – Application to the domestication of durum wheat	
	Post-doctoral fellow (WP4)	Italy	Arcad SP2	01/2013 – 06/2014 *Funded ANR TRANS (PI S.*	Analysis of the effect of mating systems on patterns of molecular evolution including species of the	

	Concetta Burgarella			*Glémin)*	Arcad (plants) and PoPhyl* (animals) projects * ERC project (PI N. Galtier)	
	Post doctoral fellow (WP5) **Iris Fisher**	Germany		2/2014-5/2014 (INRA)	WP5 - Comparative fonctional genomics (analysis of the DNA sequences)	Post-doctoral position at ISEM (2 years)
	Master student (WP3) **Emmanuel Reclus**		Montpellier SupAgro	3/2012 – 8/2012 (CIRAD)	**Reclus E, 2012**. Analyse de diversité dans les compartiments sauvages et cultivés du sorgho: identification des genes impliqués dans le processus de domestication. DAA-Ingénieur agronome. APIMET. Montpellier SupAgro	Bioinformaticia n at NINSAR (Private company in Spain)
	Master Student (WP3) **Alice Theisen**		Montpellier Supagro	3/2014 – 9/2014 (CNRS – Supagro) UMR ISEM	Génomique comparative de la domestication de trois céréales et de trois espèces pérennes	Recherche de thèse
	Master student (WP3) **Julia Morosini**	Brazil	BRAFAGRI Brasil-France exchange program, SEPMET Master, Montpellier Supagro, France	3/2013 – 7/2013 (INRA) UMR AGAP - DAAV	Domestication in grapevine : bio-statistical analysis of RNA-seq data.	Master in Brasil
	Master **Maxime De Sario**		Master BEE, Univ. Montpellier II;	1/2014 – 6/2014 INRA UMR AGAP GE2POP	Sequence polymorphism and demographic history of alfalfa (Medicago sativa) domestication.	Recherche de these
	Master 2 **Charlotte Aichholz**		Univ Angers Production et Technologie du Végétal	1/3/2014 – 31/8/2014 – INRA UR GAFL Avignon	Evolution moléculaire chez une espèce cultivée – Etude des processus à l'échelle du génome exprimé de la tomate.	
	Technician (WP1) **Laure Sauné**			1/2010 – 10/2011 (INRA)	Preparation of cDNA samples for sequencing	Technician at INRA Montpellier
SP2	Postdoctoral fellow 1 (WP1) **Stéphane de Mita**		Laboratory of Molecular Ecology, Wageningen, Netherlands	3/2010 - 9/2011 (IRD)	Methodological approaches to identify signature of natural selection along environmental gradient.	Researcher at INRA Nancy
	Post-doctoral fellow 2 (WP1) **Mathieu Siol**		Department of Ecology and Evolutionary Biology, University of Toronto, Canada	4/2011 - 11/2011 (INRA)	Analysis of available methodologies and development of new ones to detect footprints of selection in temporal samples and to determine the power of detection of a selected locus based on the number of individuals sampled and the mating system.	Researcher at INRA Dijon
	Research Engineer (WP2) **B. Rhone**		PhD Orsay	9/2013-6/2014	Association study in pearl millet	

	Post-doctoral fellow (WP2) **Concetta Burgarella**	Italy	Mediterranean Plant Evolution Laboratory, Forest Research Centre, INIA, Madrid, Spain	04/2011-12/2012 (*INRA funded)*	Spatial analysis of genome wide SNP variation in Medicago truncatula samples collected on climatic gradients. Search for footprints of selection on flowering genes.	Post-Doctoral fellowship at ISEM (ANR Trans S. Glémin) /ARCAD SP1
	Post-doctoral fellow (WP3) **Phinyarat Kongprakhon**	Thailand	Kasetsart university	09/2013-02/2014 (*French embassy funded)*	Temporal analysis of rice in Guinea	Kasetsart university
	Master 2 student **Emeric Figuet**		Master Biodiversité Ecologie Evolution, Université Montpellier II	5/2010 – 7/2010 (INRA)	**Figuet, E. 2010.** Structure génétique et évolution temporelle de populations autogames: le cas d'une légumineuse modèle *Medicago truncatula.*	PhD student at ISEM
	Master 1 student **Sarah Morris**	England	Erasmus student Montpellier SupAgro	10/2012-12/2012 (INRA)	Statistical analysis of the temporal evolution of a natural population of Medicago truncatula.	
	Master 1 student **Valerie Lemaire**		Ecole Supérieur d'Agriculture d'Angers	05/2011-08/2011 (INRA)	Population genetics and temporal evolution of natural populations of Medicago truncatula over 20 years.	
	Master 2 Student **Julien Dhinaut**		Master Biodiversité Ecologie Evolution Univ. Montpellier 2	01/2014 – 06/2014	Phenological changes and population dynamics under selfing	
	Engineer **Oscar Defrain**		IUT Informatique Montpellier	02/2014 – 05/2014	Building of a software for the genetic analysis of selfing populations	
	Master 2 **Issaka Salia Ousseini**	Niger	Université Montpellier2	01/2012-06/2012 (IRD)	Estimation of the coefficient of temporal selection of pearl millet	
	PhD **Issaka Salia Ousseini**			2013-2014 (IRD via Campus France)	Estimation of the coefficient of temporal selection of pearl millet	
	Technicians (WP2 & 3 Medicago) **Fanchon Mora**			5/2011 – 9/2011 (INRA)	Medicago genotyping	Université Abdou-Moumouni DE Niamey
	Master 2 student **Nicolas Fior**		Agrocampus Ouest-Rennes	2013 (CIRAD)	**Fior, N, 2013.** Analyse diachronique de la diversité de 2 espèces cultivées deu riz (*O. sativa et O. glaberrima*) en Guinée. Mémoire de Master.	

					AgroCampus Ouest- Université de Rennes.	
	Technician (WP2 & 3 Mil) **Buiron Marylène**			5/2010 - 11/2010 (IRD)	Pearl millet genotyping	
	Technician (WP2 & 3 Mil) **Carole Blay**			3/2011 - 10/2011 (IRD)	Pearl millet genotyping	Thèse IFREMER Tahiti
SP3	PhD student (WP 2) **Vanesse Labeyrie**		IRD, Nouvelle-Calédonie	11/2010 - 10/2013 (CIRAD) Ecole Sibaghe. Montpellier SupAgro.	**Labeyrie V,2013**. L'organisation sociale des plantes cultivées. Influence des échanges, représentations et pratiques sur la diversité du sorgho (*Sorghum bicolor* [L.] Moench) chez les peuples du mont Kenya . PhD dissertation. Ecole Sibaghe. Montpellier SupAgro.	Researcher at CIRAD Montpellier
	Master Student **Mamadou Tely Diallo**	Guinée	Université Montpellier II	4/2013 – (CIRAD funded)	Genetic diversity of fonio in Guinea	
	PhD Student **Sani Saidou**	Niger	Université Abdou-Moumouni Niamey, Niger	03/2013-07/2013 (co funding ARCAD/IRD/French Embassy)	Genetic diversity of fonio in Niger	
	Post-Doctoral Fellow **Adeline Barnaud**		Stellenbosch University, South Africa	12/2009-10/2011 (IRD funded)	Fonio diversity and evolution	Researcher at IRD Senegal
	PhD student **Ali Sarhi**	Morocco	IAV Hassan II - Rabat	02/ 2011-12/2014 (Co-funding ARCAD, PRAD project)	Durum wheat landraces in 2 traditional farming areas in Morocco : taxonomy, geographical structuration and genetic diversity	
	PhD student **Lamyae Chentoufi**	Morocco	IAV Hassan II - Rabat	02/2011-12/2014 (Co funding ARCAD, PRAD project, L'Oreal Fellowship)	Relationship between diversity of durum wheat landraces and farming practices or agro ecological factors in 2 traditional agro systems in Morocco	
	Master student (WP2- T4) **Leo Valette**		Montpellier SupAgro	7/2011 – 11/2011 (CIRAD)	Morphological characterization of sorghum landraces and soil sampling in order to describe the relation between sorghum intraspecific diversity, agro-ecological factors and anthropological factors.	
	Technicien WP2-T4) **Marylène Buiron**			11/2012 à 12/2012	Sorgho genotyping (V. Labeyrie study)	

	Master student (WP1) **Oscar Morais**	Angola	Montpellier SupAgro	3/2012 - 9/2012 (INRA)	Analysis and structuration of the genetic diversity of maize populations in Burkina Faso (in French).	Maize breeder in public institute in Angola
	Master student **Barbara Kaserer**	Germany	Institut des Régions Chaudes, Montpellier	4/2012 - 12/2012 (CIRAD)	**Kaserer, B. 2012.** Stratégies paysannes et dynamiques de l'espèce africaine de riz cultivé *Oryza glaberrima* Étude exploratoire sur le riz pluvial cultivé en abattis-brûlis dans le district de Balandougou, Guinée Conakry. Mémoire de Master en sciences et technologies "Agronomie et agroalimentaire". IRC Montpellier	
	Master student **Raphaelle Anginot**		Institut des Régions Chaudes, Montpellier	4/2012 – 10/2012 (CIRAD)	**Anginot R, 2012.** Stratégies paysannes et dynamique conservatoire de l'espèce africaine de riz cultivé *Oryza glaberrima*. Etude exploratoire dans 2 villages de la région des Cascades au Burkina Faso. Mémoire de Diplôme d'ingénieur SAADS. IRC Montpellier.	
	Master2 student **Simon Damien Ntab**	Senegal	Institut des Régions Chaudes, Montpellier	4/2013- 10/2013 (CIRAD)	**Ntab, SD, 2013** Conceptions Paysannes et dynamiques de conservation de l'espèce africaine de riz cultivé *Oryza glaberrima*. Cas des écosystèmes pluviaux inondés de la Casamance au Sénégal. Mémoire de Master en Sciences. IRC Montpellier.	
	Master 2 student **Roland Akakpo**	Benin	Agrocampus Ouest - Rennes	2/2013- 09/2013 (INRA)	Origin and genetic structuration of maize landraces originated from Burkina Faso and Africa analyzed from nuclear and cytoplasmic polymorphism	
	Engineer (WP2) **Youssouf Doumbia**	Mali	Institut Polytechnique Rural de Formation et Recherche Appliquée de Katibougou	6/2012- 12/2012 (IER)	Phenotypic characterization of wild and cultivation sorghums collected in Mali (in French)	
	Research Engineer **Julie Bouniol**			1/04-2012 – 1/7/2012 (IRD) *(GRiSP project funded)*	Development of *O. glaberrima* SNPs	
	Research assistant **Olufissayo Kolade**	Benin	AfricaRice (Cotonou, Bénin)	3 months/year (2011-2013) IRD. DSF/AIRD	Sequencing and assessment *of RYMV2* gene diversity in African rice species	
	Master Student **C. Monat**		Université Blaise Pascal (Clermont	1/3-30/6/2012 (M1) 1/1/2013 –	Sequencing *O. glaberrima* Genome	

			Ferrand)	30/62013 (M2) IRD		
	Master Student **G. Maillot**		Université Montpellier2	1/4/2013-30/6/2013 (M2) IRD	Phenotyping of African rice Collection for RYMV.	
SP4	Computer engineer (WP2) **Guilhem Sempere**			2/2010 – 7/2011 & 11/2012 – 8/2013 (CIRAD)	Management, collection, storage, and formatting of data produced by ARCAD. Adaptation of adapt components of existing databases (Chado, TropGene, GreenPhyl) and development of Web interfaces according to users' needs.	
	Engineer in bioinformatics (WP3) **Gautier Sarah**			11/2009 – 5/2011 (CIRAD)	Development of an integrated module for high-throughput sequence diversity analysis (assembling NGS sequencing outputs, SNP determination, haplotypes definition, and detection of intra-species selection)	Researcher at INRA Montpellier
	Research engineer (WP4) **Jean-François Dufayard**			9/2010 – 7/2011 (CIRAD)	Development of optimized methods for prediction of paralogue and orthologue genes using phylogenetic approaches.	Researcher at CIRAD
	Engineer in bioinformatics (WP3) **Felix Homa**			*11/2012 – 10/2013 (CIRAD)*	Development of an integrated module for high-throughput sequence diversity analysis	
	Research engineer **Stephanie Pointet (WP4)**			1/2013 – 3/2014 (CIRAD)	Development of optimized methods for prediction of paralogue and orthologue genes using phylogenetic approaches.	
	Engineer in bioinformatics **Sandy Contreras**			10/2012 – 3/2013 *(INRA funded)*	Olive tree data analysis	
SP5	Engineer in bioinformatics **Hajar CHOUIKI**			06/2013-06/2014 (INRA)	Sequence analysis produced by GbS for all the 20 species studied in the SP5	
SP7	Study Engineer **Isabelle Sylvestre**			09/10-08/13 (IRD)	Cryopreservation of tropical and Mediterranean species	
	Master student **Mohammad Salma**	Syria	University Montpellier 2	02/11 – 09/11	Cryopreservation of Rubia akane hairy roots	PhD UM2
	PhD student **Mohammad Salma**	Syria	University Montpellier 2	04/12 – 12/14	Cryopreservation of date palm	
	PhD student **Zvjezdana Markovic**	Croatia	Montpellier Supagro	09/10-12/13	Cryopreservation of grapevine	

Annexe 5 - Publications, communications

(Last update: December 2014)

Publications in brief

	Journal articles	Conferences	Posters	Book chapters
Coordination		3		
Comparative population genomics	10	7	3	
Adaptation to climate change	10	11		
Cereals in Africa	7	6		1
Bioinformatics	1	2		
DNA bank	1	1		
Cryopreservation	17	6		
TOTAL	**46**	**36**	**3**	**1**

Coordination

Conferences

Glaszmann J-C, Pham J-L, Labouisse J-P. 2013. **ARCAD, a structuring node in the French plant genetic resource conservation system.** 3rd International Symposium on Genomics of Plant Genetic Resources, 16-19 April 2013, Jeju, South Korea.

Louafi S, Arnaud E, Barthelemy D, Noyer J-L, Pham J-L. 2012. **Value, norms and practices in plant biodiversity-based research and innovation commons.** First thematic conference on the knowledge commons. Governing pooled knowledge resources: Building institutions for sustainable scientific, cultural and genetic resource commons. 12-14th September 2012, Université catholique de Louvain, Louvain-la-Neuve, Belgium. Abstract

Pham J-L, Labouisse J-P. 2009. **ARCAD- Agropolis Resource Center for Crop Conservation, Adaptation and Diversity: a new open multi-function platform devoted to agrobiodiversity. Objectives and challenges** [Abstract] In: Weitzman A. L. (Ed.). Proceedings of TDWG 2009 Annual Conference, Montpellier, France, 9-13 November 2009. Abstract.

SP1. Project *Comparative population genomics*

Papers

David J, Holtz Y, Ranwez V, Santoni S,Gautier S, Ardisson M, Poux G, Choulet F, Genthon G, Roumet P, Tavaud-Pirra M. 2014. **Genotyping by sequencing transcriptomes in an evolutionary pre-breeding durum wheat population** *Molecular Breeding*, 34(4), 1531-1548

Glémin, S., Clément, Y., David, J., & Ressayre, A. 2014. **GC content evolution in coding regions of angiosperm genomes: a unifying hypothesis**. *Trends in Genetics*, *30*(7), 263-270.

Nabholz, B., Sarah, G., Sabot, F., Ruiz, M., Adam, H., Nidelet, S., Ghesquière A, Santoni S, David J, Glémin, S. 2014. **Transcriptome population genomics reveals severe bottleneck and domestication cost in the African rice (Oryza glaberrima)**. *Molecular ecology*, *23*(9), 2210-2227.

Fischer, I., Dainat, J., Ranwez, V., Glémin, S., Dufayard, J. F., & Chantret, N. 2014. **Impact of recurrent gene duplication on adaptation of plant genomes.** *BMC Plant Biology*, *14*(1), 151.

Ranwez V, Holtz Y, Sarah G., Ardisson M. , Santoni S., Glemin S, Tavaud-Pirra M and J. David. 2013. **Disentangling homeologous contigs in tetraploid assembly: application to durum wheat.** BMC Bioinformatics, *14*(15), 1-11.

Gayral P, Melo-Ferreira J, Glémin S, Bierne N, Carneiro M, Nabholz B, Lourenco JM, Alves PC, Ballenghien M, Faivre N, Belkhir K, Cahais V, Loire E, Bernard A, and Galtier N. 2013. **Reference-free population genomics from next- generation transcriptome data and the vertebrate–invertebrate gap** *PLoS Genetics* 9(4): e1003457. doi: 10.1371/journal.pgen.1003457

Ronfort J, Glémin S. 2013. **Mating system, Haldane's sieve and the domestication process.** *Evolution*. 67(5):1518–1526. doi:10.1111/evo.12025

Serres-Giardi L, Belkhir K, David J, and Glémin S. 2012. **Patterns and evolution of nucleotide landscapes in seed plants.** *Plant Cell* 24:1379-1397. doi: 10.1105/tpc.111.093674

Ronfort J, Glémin S. 2013. **Mating system, Haldane's sieve and the domestication process.** *Evolution*. 67(5):1518–1526. doi:10.1111/evo.12025

Papers using ARCAD data

Sim S-C, Durstewitz G, Plieske J, Wieseke R, Ganal MW, Van Deynze A., Hamilton P.J., Buell, C.R, Causse M*., Wijeratne S. & D. M. Francis . 2012. **Development of a large SNP genotyping array and generation of high-density genetic maps in tomato.** PLoS ONE 7(7): e40563. doi:10.1371/journal.pone.0040563

Conferences and workshops

Fischer, I., Dainat, J., Ranwez, V., Glemin, S., Dufayard, J.-F., Chantret, N. **Impact of recurrent gene duplication on adaptation of plant genomes**; Ecological Genetics Group Meeting ; 14-16 Avril 2014 ; Newcastle ; UK ;

Tregear JW, Esbelin J, Adam H, Jouannic S, Richaud F, Sarah G, Sabot F. 2013. **Analysis of small RNAs and their potential role in the regulation of inflorescence development in palms**. XIIIème Colloque "European Network of Palm Scientists", Aarhus, Danemark, 10-12/5/2013.

Glémin S. 2013. **The ARCAD project, an NGS window to the domestication and adaptation of plants**. Workshop "Evolutionary Genomics and aquaculture". 31 May 2013. Sète, France (Invited speaker).

Glémin S. 2013. **Génomique des populations comparative du processus de domestication**. Journée Biodiversité et Bioinformatique. 7 June 2013. Paris, France (Invited speaker)

Nabholz B, Sarah G, Ruiz E, Santoni S, Sabot F, David J & Glémin S. 2012. **Population genomics in cultivated and wild populations of the African rice (O. glaberrima/O. barthii).** Jacques Monod Conference "Theoretical and empirical advances in evolutionary genomics", 31 March-4 April 2012, Roscoff, France. (Poster)

Glémin S.2012. **Génomique des populations comparatives d'espèces cultivées et de leurs apparentées sauvages.** Ecole Thématique Expert Génomique Environnementale, 23-27 April 2012. Aussois, France. (Invited speaker)

David J, Nabholz B & Glemin S. 2012. **Domestication: old questions and new data ,** 4 June 2012, Workshop Plant Science Student Conference, Leibniz Institute of Plant Genetics and Crop Research, Gatersleben, Germany (invited speaker)

Posters

Fischer, I., Dainat, J., Dufayard, J.-F., Ranwez, S., Chantret, N. **Selection positive chez les genes recemment dupliqués dans les genomes de plantes** ; Looking for positive selection in recently duplicated genes in plant genomes. JOBIM 1-4 juillet 2013, Toulouse, France.

Fischer, I., Dufayard, J.-F., Ranwez, V., Chantret, N. (2012). **Looking for positive selection in recently duplicated genes in plant genomes** . Presented at Population Genetics Group "PopGroup", 18-21 decembre 2012, Glasgow, Royaume Uni.

Clément Y, David J. and S. Glémin 2014 **Population genomics in flowering plants reveal the major importance of GC-biased gene conversion in shaping neutral polymorphism and base composition.** SMBE Puerto Rico June 8th -12 th 2014.

SP2. Project *Adaptation to climate change*

Papers

Saïdou AA, Thuillet AC, Couderc M, Mariac C, Vigouroux Y. 2014. **Association studies including genotype by environment interactions: prospects and limits**. *BMC Genetics.* **15**:3. doi: 10.1186/1471-2156-15-3

Saïdou AA, Clotault J, Couderc M, Mariac C, Devos KM, Thuillet AC, Amoukou IA, Vigouroux Y, 2013. **Association mapping, patterns of linkage disequilibrium and selection in the vicinity of the PHYTOCHROME C gene in pearl millet.** *Theor Appl Genet***.** doi: 10.1007/s00122-013-2197-3

Gay L., Siol M. and J. Ronfort. 2013. **Pedigree-free estimates of heritability in the wild: Promising prospects for selfing populations**. *PLoS One* **8**(6): e66983. doi: 10.1371/journal.pone.0066983.

De Mita S, Thuillet A-C, Gay L, Ahmadi N, Manel S, Ronfort J, Vigouroux Y. 2013. **Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations.** *Molecular Ecology* **22**:1383-1399 doi:10.1111/mec.12182

Loridon K, Burgarella C, Chantret N, Martins F, Gouzy J, Prospéri J-M and J Ronfort. 2013. **Single nucleotide polymorphism discovery and diversity in the model legume *Medicago truncatula*.** *Molecular Ecology Notes***. 13**:84-95. doi: 10.1111/1755-0998.12021

De Mita S, Siol M. 2012. **EggLib: processing, analysis and simulation tools for population genetics and genomics**. *BMC Genetics*. **13**:27. doi:10.1186/1471-2156-13-27

Clotault J, Thuillet AC, Buiron M, De Mita S, Couderc M, Haussmann BIG, Mariac C, and Vigouroux Y. 2012. **Evolutionary history of pearl millet (*Pennisetum glaucum* [L.] R. Br.) and selection on flowering genes since its domestication.** *Mol. Biol. Evol.* 29:1199-1212. doi: 10.1093/molbev/msr287

Mariac C, Jehin L, Saïdou A-A, Thuillet A-C, Couderc M, Sire P, Jugdé H, Adam H, Bezançon G, Pham J-L and Vigouroux Y. 2011. **Genetic basis of pearl millet adaptation along an environmental gradient investigated by a combination of genome scan and association mapping.** *Molecular Ecology,* 20:80–91. doi: 10.1111/j.1365-294X.2010.04893.x

Vigouroux Y, Mariac C, De Mita S, Pham J-L, Gérard B, et al. 2011. **Selection for earlier flowering crop associated with climatic variations in the Sahel.** *PLoS One* 6(5): e19563. doi:10.1371/journal.pone.0019563

Vigouroux, Y, Barnaud, A, Scarcelli, N, Thuillet, A-C 2011. **Biodiversity, evolution and adaptation of cultivated crops (Review)**. *Comptes Rendus - Biologies*. 334:450-457. doi:10.1016/j.crvi.2011.03.003

Conferences

Chantret N., Ronfort J., Burgarella C., David J., Ecarnot M., Fréville H., Gay L., Gouesnard B., Loridon K., Muller MH., Prosperi JM., Ranwez V., Roumet P., Santoni S., Tavaud-Pirra M., Vigouroux Y. 2014. **Exploiter la diversité génétique pour comprendre les mécanismes de l'adaptation**. Journées Scientifiques du Département INRA Biologie et Adaptation des Plantes, 14-16 avril 2014, Pont Royal, France.

Ahmadi N. Billot B., Droc G, Brunel D, Frouin J, Ramanantsoanirina A, McNally K, Courtois B, Glaszmann JC. 2013. **Patterns of rice diversity from SNP delineated the origin of the atypical *O. sativa* group in Madagascar from intermediary forms of the Indian sub-continent**. 7th International rice genetics symposium. Nov 5-8, Manila, Philippines.

Vigouroux Y. 2013. **Study of the genetic basis of plant adaptation to climate variation**. 19 février 2013. Conférence de l'UMR CBGP, Montpellier, France.

Vigouroux Y, S De Mita, N Ahmadi, AC Thuillet et J Ronfort. 2012. **Study of genetic basis of plant adaptation to climate variation.** 18 décembre 2012. Conférence Midipile d'evolution-ecologie de la region sud de Paris, Université Paris-Sud.

Burgarella C, Chantret N, Gay L, Prosperi J-M, De Mita S, Young N, Ronfort J. 2012. **Effet des variations climatiques sur la variabilité des gènes déterminant la date de floraison : une étude chez Medicago truncatula.** 34ème réunion annuelle du groupe de biologie et génétique des populations (le Petit Pois Déridé), 28-31 août 2012, Avignon France.

De Mita S, Thuillet AC, Gay L, Ahmadi N, Manel S, Ronfort J, Vigouroux Y. 2012. **Detecting selection along environmental gradients: analysis of seven genome scan methods and their effectiveness for outbreeding and selfing population.** Jacques Monod Conference "Theoretical and empirical advances in evolutionary genomics", 31 March-4 April 2012, Roscoff, France.

Vigouroux Y. 2011. **From genebank collections to functional variation and in situ evolution.** Genomics of Genebanks Workshop. Plant and Animal Genome XIX Conference. January 15-19, 2011. San Diego. USA.

Mariac C, Jehin L, Saïdou AA, Thuillet AC, Couderc M, Sire S, Jugdé H, Adam H, Bezançon G, Pham JL, Vigouroux Y. 2011. **Genetic basis of pearl millet adaptation along an environmental gradient investigated by a combination of genome scan and association mapping [oral presentation].** Sorghum & millet workshop. Plant and Animal Genome XIX Conference. January 15-19, 2011. San Diego. USA.

Ronfort J, Chantret N, Gay L, De Mita S, Loridon K, Prospéri J-M, Siol M, Bataillon T. 2011. **Naturally occurring variation in Medicago truncatula : what have we learn from population genetics studies**. Model Legume Congress, May 15-19, 2011, Sainte-Maxime (France).

Vigouroux Y. 2011. **Adaptation to a changing climate and the genetic basis of flowering time variation in pearl millet.** Colloque Des molécules aux écosystèmes. Faculté de Médecine, 13-14 septembre 2011, Montpellier, France.

Vigouroux Y. 2010. **Biodiversity, evolution and adaptation of cultivated crops.** Conférence débat de l'Académie des Sciences « La biodiversité face aux activités humaines », 9 février 2010, Institut de France, Paris.

SP3. Project *Cereals in Africa*

Papers

Chentoufi L, Sahri A, Arbaoui M, Belqadi L, Birouk A, Roumet P, Muller M-H. **Anchoring durum wheat diversity in the reality of traditional agricultural systems: varieties, seed management, and farmers' perception in two Moroccan regions**. *Journal of Ethnobiology and Ethnomedicine* 2014, **10**:58. doi:10.1186/1746-4269-10-58

Labeyrie V, Deu M, Barnaud A, Calatayud C, Buiron M, et al. 2014. **Influence of Ethnolinguistic Diversity on the Sorghum Genetic Patterns in Subsistence Farming Systems in Eastern Kenya.** *PLoS ONE* 9(3): e92178. doi:10.1371/journal.pone.0092178

Labeyrie, V., Rono, B., Leclerc, C. 2013. **How social organization shapes crop diversity: an ecological anthropology approach among Tharaka farmers in Kenya**. *Agric Hum Value*s doi: 10.1007/s10460-013-9451-9.

Barnaud A, Vigouroux Y, Barry B, Beavogui F, Camara M, Billot C, Noyer J-L, Pham J-L and Bakasso Y. 2013. **From advanced to underutilized crops: Making fonio benefit from research of major cereals in Africa.** Proceedings of the 2nd International Symposium on Underutilized Plant Species, Crops for the Future, Beyond Food Security, 27th June – 1st July 2011, Kuala Lumpur, Malaysia. Acta Hort. (ISHS) 979:421-430. http://www.actahort.org/books/979/979_45.htm

Barnaud A, Vignes H, Risterucci AM, Noyer JL, Pham JL, Blay C, Buiron M, Vigouroux Y, Billot C. 2012. **Development of nuclear microsatellite markers for the fonio, Digitaria exilis (Poaceae), an understudied West African cereal.** *Am. J. Bot.* March 2012 99:e105-e107; doi:10.3732/ajb.1100423

Barnaud, A., Billot, C. 2011. **Atelier international " De la connaissance à la valorisation du fonio " 2010, organisé par le Cirad, l'IRD, l'université Abdou Moumouni de Niamey, l'Irag et le projet ARCAD, Niamey, Niger, 9-11/12/2010.** *Cahiers Agricultures* 20 (4): 310-312. doi:10.1684/agr.2011.0501

Scarcelli N, Barnaud A, Eiserhardt W, Treier UA, Seveno M, D'anfray A, Vigouroux Y, Pintaud JC. 2011. **A set of 100 chloroplast DNA primer pairs to study population genetics and phylogeny in monocotyledons**. PlosOne, 6:e19954

Book chapters

Lorieux M., Garavito A., Bouniol J., Gutiérrez A., Ndjiondjop M-N., Guyot R., Martinez C.P., Tohme J. and Ghesquière A. 2013. **Unlocking the *O. glaberrima* treasure for rice breeding in Africa.** In Wopereis M, Johnson D, Ahmadi N, Tollens E, Jalloh A (Eds) *Realizing Africa's Rice Promise.* CABI, London, *In press*.

Conferences

Labeyrie V., Kamau J I., Leclerc C.2014. **The social diffusion pathways of sorghum varieties and associated knowledge in the Mount Kenya region**. **[Oral presentation]**: European Social Networks Conference. July 1- 4 2014.Barcelona, Spain.

Jouannic S. *et. al.* 2012. **African rice domestication associated to alteration of small RNA transcriptome**. 10th Inter. Symp. on Rice Functional Genomics. 26-29 November 2012, Chiang Mai, Thailand.

Labeyrie V, Leclerc C. 2012. **Does social organization shape crop diversity? A case study among Tharaka farmers in Kenya [Oral presentation].** 13th Congress of the International Society of Ethnobiology, 20-25 May 2012, Montpellier, France.

Barnaud A, Billot C, Vigouroux Y, Noyer J-L, Bakasso B, Barry B, Camara M, Beavogui F, Pham J-L. 2011. **Underutilized crops for the future, an untapped reservoir of genetic resources: the case of fonio [Poster]** Colloque FRB: Les Ressources Génétiques face aux nouveaux enjeux environnementaux, économiques et sociétaux. 20-22 September. 2011, Montpellier, France.

Barnaud A, Billot C, Vigouroux Y, Bakasso Y, Barry B, Beavogui F, Camara M, Noyer J-L, Pham J-L. 2011. **From advanced to underutilized crops: Making fonio benefit from research of major cereals in Africa. [Oral presentation]**. 2nd International Symposium on Underutilized Plant

Species, Crops for the Future, Beyond Food Security, 27th June – 1st July 2011, Kuala Lumpur, Malaysia.

Labeyrie V, Leclerc C, Barnaud A, Kamau J. 2011. **Influence des facteurs sociaux sur l'organisation de l'agrobiodiversité dans un milieu semi-aride du Kenya.** 5èmes Journées en Sciences sociales INRA-SFER-CIRAD, 2011.

SP4. Project *Bioinformatics*

Papers

Hamelin C, Sempere G, Jouffe V, Ruiz M. 2013. **TropGeneDB, the multi-tropical crop information system updated and extended** *Nucl. Acids Res.* 41(D1): D1172-D1175. doi:10.1093/nar/gks1105

Conferences

Dufayard J-F, Ruiz M. 2011. **Method for phylogenetic validation of mapped reads (resequencing)** [Poster] ISMB/ECCB 2011. Vienna, Austria

Maillol, V., et al. 2012. **Role of Galaxy in a bioinformatic plant breeding platform**, 2012 Galaxy Community Conference (GCC2012), 2012 July 25-27, Chicago, Illinois, United States.

SP6. Project *DNA Bank*

Papers

Clermont D, Santoni S, Saker S, Gomard MT, Gardais E, Bizet C. 2014. **Assessment of DNA encapsulation, a new DNA storage method at room temperature**. *Biopreservation and Biobanking*, 12 (3), 170-183

Conferences

Clermont D, Saker S, Santoni S, Martinet N, Zabot MT, Bizet C. 2011. **DNA Encapsulation: a new method for DNA long term preservation at room temperature.** [Poster] European Society for Biopreservation and Biobanking ESBB Inaugural conference. 16-19 November 2011. Marseille

Technical protocols

Santoni S, Weber A. 2011. **Protocole d'extraction d'ADN végétal (plantes herbacées simples) et purification sur fibre de silice**. INRA UMR AGAP. 6 pp.

SP7. Project *Cryopreservation*

Papers

Engelmann-Sylvestre I, Engelmann F. **Cryopreservation of in vitro-grown shoot tips of *Clinopodium odorum* using aluminium cryo-plates.** *Scientia Horticuturae*, submitted.

Marković Z, Preiner D, Bošnjak AM, Safner T, Stupić D, Andabaka Z, Maletić E, Chatelet P, Engelmann F, Kontić JK. 2014. **In vitro introduction of healthy and virus-infected genotypes of native Croatian grapevine cultivars.** *Central European Journal of Biology* 9:1087-1098 (DOI) 10.2478/s11535-014-0337-7.

Marković Z, Chatelet P, Preiner D, Sylvestre I, Engelmann F, Karoglan Kontić J. 2014. **Effect of shooting media and source of material of grapevine shoot tip (Vitis vinifera) recovery after cryopreservation**. *CryoLetters* 35: 40-47.

Barraco G, Sylvestre I, Collin M, Escoute J, Lartaud M, Verdeil JL, Engelmann F. 2014. **Histological study of yam shoot tips during their cryopreservation using the encapsulation-dehydration technique.** *Protoplasma* 251: 177-189. doi.org/10.1007/s00709-013-0536-5.

Marković Z, Chatelet P, Sylvestre I, Karoglan Kontić J, Engelmann F. 2014. **Cryopreservation of grapevine (*Vitis vinifera* L.) *in vitro* shoot tips using encapsulation-dehydration and droplet-vitrification**. *Central European Journal of Biology* 8: 993-1000.

Salma M, Sylvestre I, Collin M, Escoute J, Lartaud M, Yi JY, Kim HH, Verdeil JL, Engelmann F. 2014. **Effect of the successive steps of a cryopreservation protocol on the structural integrity of Rubia akane Nakai hairy roots.** *Protoplasma* 251:649-659. DOI 10.1007/s00709-013-0565-0.

Polzin F, Sylvestre I, Déchamp E, Ilbert P, Etienne H, Engelmann F. 2014. **Multiplication of African yam (*Dioscorea cayenensis-rotundata*) using a temporary immersion system**. *In Vitro Cellular and Developmental Biology – Plant* 50: 210-216. DOI 10.1007/s11627-013-9552-6.

Engelmann-Sylvestre I, Engelmann F. 2014. **Effect of various growth regulators on growth of yam (*Dioscorea trifida*) *in vitro* shoot tips**. *African Journal of Biotechnology* 13: 1645-1649.

Engelmann-Sylvestre I, Engelmann F. 2013. **Effect of brassinosteroids on growth of *in vitro* shoot tips of several yam (*Dioscorea* spp.) species**. *American Journal of Plant Sciences* 4: 2271-2274. doi.org/10.4236/ajps.2013.411280.

Marković Z, Chatelet P, Peyrière A, Preiner D, Engelmann-Sylvestre I, Karoglan Kontić J, Engelmann F. 2013. **Effect of proline pretreatment on grapevine shoot-tip response to a droplet-vitrification protocol**. *American Journal of Plant Sciences* 4:2414-2417.

Barraco G, Chatelet P, Balsemin E, Sylvestre I, Engelmann F. 2012. **Cryopreservation of *Prunus cerasus* using vitrification method and replacement of cold hardening with preculture on medium enriched with sucrose and/or glycerol.** *Scientia Horticulturae* 148:104-108. doi:10.1016/j.scienta.2012.09.034

Barraco G, Sylvestre I, Iapichino G, Engelmann F. 2013. **Investigating the cryopreservation of nodal explants of Lithodora rosmarinifolia (Ten.) Johnst., a rare, endemic Mediterranean species.** *Plant Biotechnology Report* 7:141-146. doi: 10.1007/s11816-012-0241-4

Yi JY, Sylvestre I, Collin M, Salma M, Lee SY, Kim HH, Park HJ, Engelmann F. 2012**. Improved cryopreservation using droplet-vitrification method and histological changes associated with cryopreservation of madder (*Rubia akane* Nakai**). *Korean Journal of Horticultural Science and Technology* 30:79-84. doi: 10.7235/hort.2012.11087

Barraco G, Chatelet P, Balsemin E, Sylvestre I, Engelmann F. 2012. **Cryopreservation of *Prunus cerasus* using vitrification method and replacement of cold hardening with preculture on medium enriched with sucrose and/or glycerol**. *Scientia Horticulturae* 148 :104-108.

Vujovic T, Sylvestre I, Ruzic D, Engelmann F. 2011. **Droplet-vitrification of apical shoot tips of Rubus fruticosus L. and Prunus cerasifera Ehrh**. *Scientia Horticulturae* 130:222-228. doi:10.1016/j.scienta.2011.06.049

Barraco G, Sylvestre I, Iapichino G, Engelmann F. 2011. **Cryopreservation of *Limonium serotinum* apical shoots from in vitro plantlets using droplet-vitrification.** *Scientia Horticulturae* 130:309-313. doi:10.1016/j.scienta.2011.07.001

Barraco G, Sylvestre I, Engelmann F. 2011**. Cryopreservation of sugarcane (*Saccharum* spp.) shoot tips using encapsulation and droplet-vitrification**. *Scientia Horticulturae* 130 : 320-324. doi:10.1016/j.scienta.2011.07.003

Conferences

Barraco G, Sylvestre I, Collin M, Escoute J, Lartaud M, Verdeil JL, Engelmann F. **Histological study of yam shoot tips during their cryopreservation using the encapsulation-dehydration technique**. *2nd Intl. Symp. on Plant Cryopreservation. Fort Collins, USA, 11-14 August 2013*, submitted.

Marković Z, Chatelet P, Sylvestre I, Karoglan Kontić J, Engelmann F. **Cryopreservation of grapevine (*Vitis vinifera* L.) *in vitro* shoot tips using encapsulation-dehydration and droplet-vitrification.** *2nd Intl. Symp. on Plant Cryopreservation. Fort Collins, USA, 11-14 August 2013*, submitted.

Salma M, Sylvestre I, Collin M, Escoute J, Lartaud M, Yi JY, Kim HH, Verdeil J L, Engelmann F. **Effect of the successive steps of a cryopreservation protocol on the structural integrity of *Rubia akane* Nakai hairy roots.** *2nd Intl. Symp. on Plant Cryopreservation. Fort Collins, USA, 11-14 August 2013*, submitted.

Barraco G, Sylvestre I, Engelmann F. 2012. **Cryopreservation of sugarcane (*Saccharum* sp.) shoot tips using encapsulation-dehydration and droplet-vitrification.** *In : A. Grapin, E.R.J. Keller, P.T. Lynch, B. Panis, A. Revilla Bahillo & F. Engelmann (eds.), Cryopreservation of crop species in Europe Proceedings of the final meeting*, Agrocampus Ouest INPH, Angers, France, 8-11 Feb. 2011, COST Office, Brussels, pp. 119-122.

Marković Z, Chatelet P, Sylvestre I, Karoglan Kontić J, Engelmann F. 2012. **Duration of culture of grapevine (*Vitis vinifera*) microcuttings on medium with zeatin riboside affects shoot tip recovery after cryopreservation**. *In : A. Grapin, E.R.J. Keller, P.T. Lynch, B. Panis, A. Revilla Bahillo & F. Engelmann (eds.), Cryopreservation of crop species in Europe Proceedings of the final meeting*, Agrocampus Ouest INPH, Angers, France, 8-11 Feb. 2011, COST Office, Brussels, pp. 145-147.

Vujović T, Sylvestre I, Ružić DJ, Engelmann F. 2012. **Cryopreservation of cherry plum and blackberry shoot tips by droplet-vitrification.** *In : A. Grapin, E.R.J. Keller, P.T. Lynch, B. Panis, A. Revilla Bahillo & F. Engelmann (eds.), Cryopreservation of crop species in Europe Proceedings of the final meeting*, Agrocampus Ouest INPH, Angers, France, 8-11 Feb. 2011, COST Office, Brussels, pp. 163-166.

Annexe 6. Submitted projects in connection with ARCAD

Acronym	*Name*	*Call - Year*	*Main partners*	*Leader (**in bold if involved in ARCAD**)*	*Period*	*Grant (obtained/requested)*	*Submitted/Accepted or not accepted*
TRANS	Les transitions de systèmes de reproduction chez les angiospermes et leurs conséquences génomiques	ANR Blanc 2011	ISEM Montpellier AGAP Montpellier Etc…	**S Glémin**	2011-2015	EUR 529 858	Accepted
CROP-DL	Analyse du déséquilibre de liaison chez plusieurs espèces d'intérêt agronomique	Métaprogramme INRA	INRA Clermont-Ferrand, INRA Le Moulon, INRA Dijon, INRA Toulouse				Accepted
EPO	Test of GBS using RNAseq in pre-breeding population of durum wheat	Initiative INRA	INRA	M. Tavaud	2011-2013		Accepted
RECIPE	Durum wheat genetic diversity and domestication	EU - 7th KBBE program	IPK Gatersleben, Univ Munich, Univ Bologna, Univ Bari, etc…	N. Stein (Germany) **J. David** for the evolutionary subtask		EUR 5 million	Submitted
SELFADAPT	Adaptation to climate change under selfing – experimental test and selection footprints in *Medicago truncatula*	MP INRA ACCAF	CBGP Montpellier & AGAP Montpellier	**L. Gay**	2013-2015	EUR 49700	Accepted
ADAPTINWILD	Identifying adaptive variation in the wild progenitors of two	ANR Bioadapt 2011	CNRS INRA IRD	M Tenaillon Co-PI **Y Vigouroux**	2012-2016	EUR 514 000	Accepted

	cereal crops, maize and pearl millet		ISRA (Sénégal) (Mexique)	(IRD)			
MILDIV	Phylogeography and adaptation of wild pearl millet	ANR RPOC 2012	IRD ISRA	C Berthouly	2012-2014	EUR 162 000	Accepted
GUYAMAZ	Training for studying the evolutionary history of Amazonian crops and Amazonian biodiversity in Brazil and French Guiana	AIRD/FAPEAM	IRD INRA CIRAD CNRS INPA UAM	C Clements**, Y Vigouroux** & C Campa	2012-2014	EUR 80 000	Accepted
AKIH	Agrobiodiversity knowledge and information hub	ANR Agrobiosphere 2013	AGAP, AMAP, GREEN, DIADE, G-EAU, Bioversity Int.	**S Louafi**	2014-2018	EUR 1 109 538	Not selected
HERBA-DIV	Exploration of historical herbaria resources: Geographical and temporal milestones to trace evolution and adaptation of cultivated plants under human and natural selection	Agropolis OpenScience	AGAP, DIADE BGPI MNHN Bioversity Int.	**C Jenny**	2014-2015	EUR 149 422	Not selected
CHLORODIV	Whole chloroplast genomes of crop species as a tool for population genetics, phylogeography and phylogeny in agrobiodiversity	Agropolis Fondation OpenScience	AGAP, DIADE ENS (Cameroun) ISRA (Sénégal)	T Couvreur	2013-2015	EUR 45 552	Accepted
PRAD Maroc	Genetic diversity and adaptation of durum wheat landraces to agricultural practices and climatic constraints in Moroccan agro-systems.	Campus France 2011-2013	IAV Hassan II (Maroc)	**P Roumet**	2011-2013	EUR 35 000	Accepted

SEAD	How selfing affects adaptation ? genetic and demographical consequences.	ANR BioAdapt	AGAP Montpellier ISEM Montpellier CEFE Montpellier	**L. Gay**	2013-2016	EUR 497 272	Accepted
ARCAD	ARCAD equipment and running	FEDER 2009-2013	AGAP	**INRA**	2013-2015	EUR 3.6 million	Accepted
CropAfrica	Documenting African Crop Domestication	ANR 2013	IRD CIRAD Université de Grenoble	**Y. Vigouroux**	2013-2017	EUR 698 000	Accepted
IBC	Setting up the « Institut de Biologie Computationnelle »	ANR Investissement d'Avenir Bioinformatique 2011	CNRS, INRIA, CIRAD, Université Montpellier 2, IRD, etc.	**O. Gascuel**	2012-2017		Accepted
ReNaBi-IFB	French Bioinformatics Institute	ANR Investissement d'Avenir Infrastructure 2010	INRA, CNRS, CEA, etc.	**J.F. Gibrat**	2012-2017		Accepted
Plant DNA Bank	Development of a methodological study for establishing a DNA bank for grape and Medicago,	IBiSA	AGAP (INRA, CIRAD)	**S. Santoni**	2009-2011	EUR 50 000	Accepted
CRYOVEG	Development/optimization of cryopreservation techniques in a range of selected species & establishment of a national scientific and technical network of plant BRCs using cryopreservation	CBRC/IBiSA	IRD INRA CIRAD	**F. Engelmann**	2009-2011	EUR 200 000	Accepted
EDGARR	Grape Exploitation De la sélection Génomique afin d'Accélérer la	CASDAR - Innovation	INRA IFV	**L. Le Cunff**	2015-2017	EUR 200 000	Accepted

	création de variétés Résistantes et qualitatives pour la filière viticole Rosé. (Genotyping based on GBS)						
FruitSelGen	Genomic selection for fruity perenial species (genotyping based on GBS)	Meta-programme INRA	INRA, AGAP	**T. Flutre**	2015-2016	EUR 114 018	Accepted
GCP BC-NAM	GBS genotyping of a BC-NAM population in Sorghum	Generation Challenge Program (GCP)	Cirad, IER, ICRISAT	J.-F. Rami	2013-2014	EUR 162 000	Accepted
SELVI	SELection génomique chez la Vigne (Genotyping based on GBS)	Initiative INRA	INRA	**T. Flutre**	2014-2015	EUR 25 000	Accepted
NA4MA	New Approaches for Management of Agrobiodiversity)	Agropolis Fonsation – Open Science Research - 2014	CIRAD IRD PAR GMSL (Sri Lanka)	**C. Billot**	2015-2016	EUR 200 000	Submitted
HERBA-Coffea	Exploration of herbaria resources to broaden knowledge of genetic diversity for breeding: the case of robusta coffee (*Coffea canephora*) in West Africa	Agropolis Fonsation – Open Science Research - 2014	CIRAD (UMR-AGAP) MNHN INRA (URFM) LEAE Laboratory (MNHN)	**J.P. Labouisse**	2015	EUR 30 000	Submitted

Annexe 7. Sequencing, genotyping and phenotyping data

SP	Plant species	Type of data/ used technology	Details	Type of files	Interest/Use	Accessibility		Comments
						Present	Upcoming	
SP1	**29 species**	RNASeq	Reads for all the individuals	FastQ	For new analyses	Public only for O Glaberrima at http://gohelle.cirad.fr/arcad-data/sp1_final/rice/raw_data/	All the reads should be available through http://arcad-bioinformatics.southgreen.fr	To be determined when all the reads will be public
			Assemblies of one individual by species	Fasta	Transcriptomes of reference	Freely available at http://arcad-bioinformatics.southgreen.fr		For coffee and coffee out-groups access is still private.
			Annotations of transcriptomes of reference	GFF, Blast XML, Annot (Blast 2GO)	Functional studies	Freely available at http://arcad-bioinformatics.southgreen.fr		For coffee and coffee out-groups access is still private.
			Mapping of reads against transcriptomes of reference	BAM			http://arcad-bioinformatics.southgreen.fr	Still in private access through http://gohelle.cirad.fr/arcad-data/sp1_final
			SNP for all the individuals	VCF	Polymorphisms studies		http://arcad-bioinformatics.southgreen.fr	Still in private access through http://gohelle.cirad.fr/arcad-data/sp1_final
SP2	**Pearl millet**	Sequencing RNAseq	85 accessions IRD	FastQ files	Can be used for extracting SNPs	Restricted access on ARCAD Bioinformatics Portal	To be decided after analysis	
		Genotyping /SNP	384 SNP	Excel			Free after publication	

	Medicago	Genotyping/ SSR	20 SSR/ 8 pop. x 2 years	Excel			Available in the upcoming publication then on ARCAD Bioinformatics Portal	
		Genotyping/ SNP	384 SNP/ 3 pop	Excel			Available in the upcoming publication then on ARCAD Bioinformatics Portal	
		Phenotyping	Flowering and production data	Excel			Data analysis available in the upcoming publication	
	Rice (O. sativa and O. glaberrima)	Sequencing/ GBS	**Guinea**: - 640 accessions x 15000 SNP and 15000 DArT -550 accessions x 640 accessions x 20000 SNP and 20000 DArT Madagascar: 640 accessions x 30000 SNP and 30000 DArt	Excel		Computer Team Nour Ahmadi	To be decided after analysis	
		Sequencing/ Sanger	150 accessions Madagascar/ 9 genes candidats ;	Excel		Computer Team Nour Ahmadi	To be decided after analysis	
		Genotyping/ SSR	820 accessions Guinea/13 loci ; 1120 accessions Madagascar /17 loci	Excel		Computer Team Nour Ahmadi	To be decided after analysis	
		Phenotyping	Flowering and photoperiod sensitivity 820 accessions (Guinea)	Excel		Computer Team Nour Ahmadi	To be decided after analysis	
SP3	**Maize**	Genotyping /nSSR	18SSR * 172 samples	Excel			Data analysis available in the	

							upcoming publication	
		Genotyping/ cpSSR	588 accessions * 26 cytotypes	Excel			Data analysis available in the upcoming publication	
		Phenotyping / Burkina	750 samples * 24 variables	Excel			Available on request	
		Phenotyping / Burkina	1250 samples * 4 variables	Excel			Available on request	
		Phenotyping / Mali	223 samples * 28 variables	Excel			Available on request	
		Phenotyping / NIRS	230 samples * 68 variables	Excel			Available on request	
	Durum wheat	Genotyping / nSSR Interpopulation genetic diversity	15 SSR * 667 samples	Excel		Available		
		Genotyping/ nSSR intrapop genetic diversity	15 SSR * 2000 Samples	Excel		Upcoming		
		Phenotyping / Morocco	4770 samples * 9 traits* 1 site	Excel		IAV Computer		
		Phenotyping Nirs / Inra	4770 samples* 2 years	Excel		Computer Team		
		Flowering date (INRA+ IAV)	300 accessions * 7 trials			Computer team for analysis		
	Sorghum bicolor	Genotyping/ SSR	297 accessions/Mont Kenya	International database		http://datadryad.org/resource/doi:10.5061/dryad.8ff54		
	Sorghum bicolor	Genotyping/ SSR (28 SSR)	300 accessions Mali, 289 accessions Guinea, 183 accessions Senegal	Excel		Team:Monique Deu	To be decided after analysis	

			(to be done)					
	Sorghum bicolor	Phenotyping	Morphological traits and flowering of Malian accessions	Excel		Team:Monique Deu	To be decided after analysis	
SP5	***Citrus* spp.**	Genomic DNA sequences /GBS	40 10^6 reads (x100 b)	FastQ files	Will be used for SNPs identification & genotyping	ARCAD/SP5 team* Patrick Ollitrault	To be decided after analysis	* SP5 team = Sylvain Santoni, Gautier Sarah, Patrice This, Marc Seguin
	Coconut (*Cocos nucifera*)	Genomic DNA sequences /GBS	12 10^6 reads (x100 b)	FastQ files	Will be used for SNPs identification & genotyping	ARCAD/SP5 team Luc Baudouin	To be decided after analysis	
	Coffee (*Coffea* spp.)	Genomic DNA sequences /GBS	64 10^6 reads (x100 b)	FastQ files	Will be used for SNPs identification & genotyping	ARCAD/SP5 team Thierry Leroy Fabien de Bellis	To be decided after analysis	
	Cotton (*Gossypium* spp.)	Genomic DNA sequences /GBS	64 10^6 reads (x100 b)	FastQ files	Will be used for SNPs identification & genotyping	ARCAD/SP5 team Jean-Marc Lacape Marc Giband	To be decided after analysis	
SP1 & SP5	**Olive tree (*Olea europaea*)**	cDNA sequences / RNAseq	20 10^3 DNA sequences (2 x 100 b)	Excell	Used for SNPs identification & genotyping	ARCAD/SP5 team Bouchaïd Khadari	To be decided after analysis	
SP5	**Olive tree (*Olea europaea*)**	SNPs markers / RNAseq	2 800 SNP markers	Excell	Placed in a genetic map	ARCAD/SP5 team Bouchaïd Khadari	To be decided after analysis	
	Rubber tree (*Hevea* spp.)	Genomic DNA sequences /GBS	111 10^6 sequences (tags) (x 100 b)	FastQ files	Used for SNPs identification & genotyping	ARCAD/SP5 team Pierre Mournet	To be decided after analysis	
	Rubber tree (*Hevea* spp.)	SNP markers / GBS	121 000 SNPs	Excel file	Genotyped in a segregating population	ARCAD/SP5 team Pierre Mournet	To be decided after analysis	
	Grapevine (*Vitis vinifera*)	Genomic DNA sequences /GBS	40 10^6 reads (x250 b) 200 10^6 reads (x100 b)	FastQ files	Used for SNPs identification & genotyping	ARCAD/SP5 team Agnès Doligez	To be decided after analysis	

	Grapevine (*Vitis vinifera*)	SNP markers / GBS	10^4 SNPs	Excel file	Genotyped in segregating populations	ARCAD/SP5 team Agnès Doligez	To be decided after analysis	
	Breadfruit (*Artocarpus altilis*)	Genomic DNA sequences /GBS	*Analysis in progress*	*Soon available*	Will be used for SNPs identification & genotyping	ARCAD/SP5 team Jean-Pierre Labouisse Fabien de Bellis	To be decided after analysis	

Annexe 8. Analysis tools and softwares, laboratory methods, and web sites

SP	Name and type of tools developed	Accessibility (present and upcoming)	Publication (existing or upcoming)	Comments
SP0	**ARCAD Internet** site	Free access www.arcad-project.org		
SP0	**ARCAD Intranet** site/Alfresco Share	Restricted access https://collaboratif.cirad.fr/share/page/site/ARCAD/dashboard		
SP2	**Egglib:** C++/Python library and program package for evolutionary genetics and genomics.	Free access http://egglib.sourceforge.net/	De Mita S. and M. Siol. 2012. *BMC Genet.*	
	Code of a R script performing LR and GEE tests for all loci of a given data set.	Free access in supplementary material of the publication	De Mita et al. 2013. Mol. Ecol.	
	Code Schell+Python for haplotypic analysis of autogamous populations	Free access in the publication (in progress) on the evolutionary dynamics of autogamous population		
	Array of 384 SNPs markers of M. truncatula	Described in the publication. Free access to DNA sequences in GenBank (accessions JX449155-JX451819)	Loridon et al. 2013. Mol Ecol Notes	
SP4	**ARCAD Bioinformatics Portal**	http://arcad-bioinformatics.southgreen.fr/	Sarah, G. et al., A large set of 26 new reference transcriptomes dedicated to comparative population genomics in crops and wild relatives, 2014, submitted.	
	arcad-hts, a package gathering scripts for the analysis of high-throughput sequencing data	Free access at https://github.com/SouthGreenPlatform/arcad-hts		
	South Green Galaxy, a workflow manager which permits to run several bioinformatic analyses using a simple Web interface	http://gohelle.cirad.fr/galaxy The South Green Galaxy instance is opened to anyone, but anonymous users are limited to 10 Mo data. The access to workflows developed for the ARCAD SP1 project is currently restricted to users with specific login.		
	GreenPhylDB *versions 3 then 4, Web resource designed for comparative*	http://www.greenphyl.org/		

	and functional genomics in plants.			
	GenTIC2, Web-based tool that provides the means to quickly build search interfaces over existing databases	Freely available at http://www.southgreen.fr/content/gentic2-tool	Hamelin C, et al. 2013. TropGeneDB, the multi-tropical crop information system updated and extended Nucl. Acids Res.	
	Training materials in Bioinformatics	http://www.southgreen.fr/content/teaching		
SP5	**ARCAD SP5 Pipeline for GBS data analysis and SNPs detection**	Publicly available on Git Hub : https://github.com/SouthGreenPlatform/arcad-hts/tree/master/sp5_gbs		Created by Hajar Chouiki & Gautier Sarah pers. comm (2014)
SP6	**DNA purification protocols**	**Restricted**	Clermont et al. 2013. Biopreservation and Biobanking	

Annexe 9– Biological material used in the project

Sous-projet	*Genre, espèce*	Nombre d'accessions reçues	Nature du MB reçu	Type de provenance du MB	Fournisseur du MB	Destinataire du MB	Date(s) de réception du MB	Statut juridique du MB- Propriété	Document de transfert (Signataires)	Devenir du MB
SP1	*Digitaria sanguinalis*	1	lot de graines	collecte privée in situ	Privé / Michel CANTAGREL	CIRAD, AGAP, France / Claire BILLOT	30-oct-11	?	Renseignements par email équivallent à une fiche de traçabilité	conservé au CRB-T CIRAD
SP1	*Digitaria sanguinalis*	1	lot de graines	banque de semences ex situ	Université Polytechnique de Catalogne,	CIRAD, AGAP, France / Claire BILLOT	20-nov-11	Système multilatéral (Traité)	SMTA	conservé au CRB-T CIRAD
SP1	*Digitaria longiflora*	2	lot de graines	Banque de semence ex-situ	USDA, ARS, PGRCU - USA / Tiffany FIELDS	CIRAD, AGAP, France / Claire BILLOT	21-janv-12	USDA	fiche de traçabilité d'envoi, engagement informatique	conservé au CRB-T CIRAD
SP1	*Digitaria exilis*	10	tissus (à partir de graines)	Banque de semence ex-situ	IRD, France	IRD/DIADE/ A. Barnaud & C. Billot	janv-12	IRD, France	néant (utilisation ressources internes du laboratoire)	sans objet
SP1	*Oryza spp.*	22 dont 10 (*O. glaberrima*) 10 (*O. barthii*) 1 (*O. sativa*) 1 (*O. meridionalis*)	tissus (à partir de graines)	Banque de semence ex-situ	IRD, France	IRD, DIADE/GGR/ F. Sabot – A. Ghesquière		IRD, France	néant (utilisation ressources internes du laboratoire)	sans objet
SP1	*Theobroma sp.*	12	fleurs et feuilles fixées dans RNA later	Collection ex situ	CRU/Trinidad Michel Boccara (Cirad)	CIRAD, AGAP, France / C. Lanaud	juin-10	CRU/Trinidad	néant	conservé au CIRAD/Montpellier
SP1	*Pennisetum glaucum*	20	graines	Banque de semence ex-situ	IRD, France	IRD, DIADE/DYNADIV/ Y Vigouroux	sept-09	IRD, France	néant (utilisation ressources internes du laboratoire collecté avant 1994)	

Sous-projet	*Genre, espèce*	Nombre d'accessions reçues	Nature du MB reçu	Type de provenance du MB	Fournisseur du MB	Destinataire du MB	Date(s) de réception du MB	Statut juridique du MB- Propriété	Document de transfert (Signataires)	Devenir du MB
SP1	*Dioscorea rotundata*	10	Fleurs et feuilles	Collecte in situ au Bénin (IRD)	IRD, UMR DIADE, Nora Scarcelli	IRD, UMR DIADE, Nora Scarcelli	août-11	Bénin	Convention avec l'université de Parakou (Bénin)	Conservé à l'IRD
SP1	*Dioscorea trifida*	5	Fleurs et feuilles	Collection ex situ INRA Guadeloupe	INRA, Guadeloupe, CRB Antilles Guyane	Cirad, UMR AGAP, Hana Chaïr	nov-11	INRA	MTA INRA-CIRAD	Conservé au Cirad
SP1	*Dioscorea alata*	5	Fleurs et feuilles	Collection ex situ Cirad Guadeloupe (CRB Plantes Tropicales)	Cirad Guadeloupe (sous MTA CIRAD/DARD Vanuatu)	Cirad, UMR AGAP, Hana Chaïr	déc-11	VARTC Vanuatu	Transfert simple dans le cadre de MTA CDB CIRAD/DARD Vanuatu	Conservé au Cirad
SP1	*Dioscorea abyssinica*	5	Fleurs et feuilles	Collecte in situ au Bénin (IRD)	IRD, UMR DIADE, Nora Scarcelli	IRD, UMR DIADE, Nora Scarcelli	août-11	Bénin	Convention avec l'université de Parakou (Bénin)	Conservé à l'IRD
SP1	*Dioscorea praehensilis*	5	Fleurs et feuilles	Collecte in situ au Bénin (IRD)	IRD, UMR DIADE, Nora Scarcelli	IRD, UMR DIADE, Nora Scarcelli	août-11	Bénin	Convention avec l'université de Parakou (Bénin)	Conservé à l'IRD
SP1	*Medicago sativa (sauvages)*	5	graines	Banque de semence ex-situ	CRB INRA, Montpellier, France	INRA, Jean-Marie Prosperi		INRA, (origine USDA 1990), non SMTA	aucun	conservé par le CRB Medicago Montpellier
SP1	*Medicago sativa (cultivées)*	5	graines	cultivars, banque de semences ex-situ	CRB INRA, Montpellier, France	INRA, Jean-Marie Prosperi		INRA , antérieur à 1980, non SMTA	néant (utilisation ressources internes du laboratoire)	conservé par le CRB Medicago Montpellier
SP1	*Medicago marina*	1	graines	prospection Languedoc Roussillon	CRB INRA, Montpellier, France	INRA, Jean-Marie Prosperi		INRA (1985), non SMTA	néant (utilisation ressources internes du laboratoire)	conservé par le CRB Medicago Montpellier
SP1	*Sorghum bicolor*	10	lot de graines	Banque de semence ex-situ	CRB-T CIRAD Montpellier	CIRAD, AGAP, France / David POT	mars-09	CIRAD	Fiche de traçabilité CRB-T	sans objet

Sous-projet	*Genre, espèce*	Nombre d'accessions reçues	Nature du MB reçu	Type de provenance du MB	Fournisseur du MB	Destinataire du MB	Date(s) de réception du MB	Statut juridique du MB- Propriété	Document de transfert (Signataires)	Devenir du MB
SP1	*Sorghum bicolor*	10	lot de graines	Banque de semence ex-situ	ICRISAT, Dir ICRISAT	CIRAD, AGAP, France / David POT	mars-09	Système multilatéral (Traité)	SMTA (Dir. ICRISAT/	conservé au CRB-T CIRAD
SP1	*Sorghum brachipodum*	1	lot de graines	Banque de semence ex-situ	DEEDI Australia/Sally Norton	CIRAD, AGAP, France / David POT	mars-09	DEEDI, Australia	Message mail de Sally Norton (Deedi) Pas de MTA. Restriction pour la fourniture à un tiers	conservé au CRB-T CIRAD
SP1	*Musa acuminata*	6	feuilles, fleurs, fruits	Collection ex - situ CARBAP, Cameroun	Collection CARBAP, Cameroun, Sebastien RICCI (Agent CIRAD en poste au CARBAP)	CIRAD, AGAP, France/Nabila Yahiaoui	jan -dec 2011	CARBAP, Cameroun	Mails de Sebastien Ricci.	déjà présent en collection CARBAP Cameroun, pour certains génotypes existent aussi au CRB-T Antilles
SP1	*Musa acuminata*	14	feuilles, fleurs, fruits	Collection ex - situ CIRAD Guadeloupe	CRB, Plantes tropicales,CIRAD	CIRAD, AGAP, France/Nabila Yahiaoui	nov-2010, oct-2011	CIRAD	néant et fiche de traçabilité CRB-T Antilles	déjà présent en collection CRB-T Antilles CIRAD
SP1	*Musa balbisiana*	1	feuilles, fleurs, fruits	Collection ex - situ CIRAD Guadeloupe	CRB, Plantes tropicales,CIRAD	CIRAD, AGAP, France/Nabila Yahiaoui	nov-2010, oct-2011	CIRAD	Fiche de traçabilité CRB-T Antilles	déjà présent en collection au CRB-T Antilles CIRAD
SP1	*Musa becarii*	1	feuilles, fleurs, fruits	Collection ex - situ CIRAD Guadeloupe	CRB, Plantes tropicales,CIRAD	CIRAD, AGAP, France/Nabila Yahiaoui	nov-2010, oct-2011	CIRAD	fiche de traçabilité CRB-T Antilles	déjà présent en collection au CRB-T Antilles CIRAD
SP1	Solanum lycopersicum	10	ADN	Ressources Genetiques INRA	UR GAFL INRA - Causse Mathilde	Sylvain Santoni	sept.-09	RG INRA	message Mail	Conservé INRA GAFL Avignon
SP1	Solanum pimpinelifolliu m	10	ADN	Ressources Genetiques INRA	UR GAFL INRA - Causse Mathilde	Sylvain Santoni	sept.-09	RG INRA	message Mail	Conservé INRA GAFL Avignon

Sous-projet	*Genre, espèce*	Nombre d'accessions reçues	Nature du MB reçu	Type de provenance du MB	Fournisseur du MB	Destinataire du MB	Date(s) de réception du MB	Statut juridique du MB- Propriété	Document de transfert (Signataires)	Devenir du MB
SP1	*Elaeis oleifera*	10	tissus	Collecte *in situ*	INRAB, Station de Pobè, Bénin	IRD/Cirad, UMR DIADE, France / James TREGEAR	oct-09	INRAB Bénin	MTA type CDB (DG INRAB/Dir UMR DAP)	Pas conservé (utilisé pour extraction d'ARN seulement)
SP1	*Elaeis guineensis*	10	tissus	Collecte *in situ*	INRAB, Station de Pobè, Bénin	IRD/Cirad, UMR DIADE, France / James TREGEAR	oct-09	INRAB Bénin	MTA type CDB (DG INRAB/Dir UMR DAP)	Pas conservé (utilisé pour extraction d'ARN seulement)
SP1	*Phoenix dactylifera*	2	plante entière	Collecte *in situ* à Tozeur et Deggache (Tunisie)	CRRAO Deggache F. Aberlenc, N. Chabrillange : achat de 2 palmiers avec facture	IRD DIADE, F. Aberlenc, N. Chabrillange	mars-11	IRD	Facture d'achat	Conservé à l'IRD
SP1	*Mauritia flexuosa*	2	tissus	Collecte *in situ*	Collecte en milieu sauvage près de Kourou, Guyane	IRD/Cirad, UMR DIADE, France / James TREGEAR + Jean-Christophe PINTAUD	juin-11	Pas connu (Guyane)	Pas de MTA	Pas conservé (utilisé pour extraction d'ARN seulement)
SP1	*Oleae europea sauvage (oleastre)*	12	fleurs et feuilles	Collection ex-situ	CBNMED Porquerolles	Bouchaib Khadari (CBNMED/INRA)	mai-11	France (CNBMED) et 3 autres pays d'origine des RB (Maroc, Turquie, Syrie) transférées après 1994	Message E-mail	ADN conservé par l'INRA en banque d'ADN
SP1	*Oleae europea cultivé*	5	fleurs et feuilles	Collection ex-situ	INRA Domaine de Melgueil (Montpellier)	Bouchaib Khadari (CBNMED/INRA)	mai-11	INRA	néant (utilisation ressources internes du laboratoire)	ADN conservé par l'INRA en banque d'ADN

Sous-projet	*Genre, espèce*	Nombre d'accessions reçues	Nature du MB reçu	Type de provenance du MB	Fournisseur du MB	Destinataire du MB	Date(s) de réception du MB	Statut juridique du MB- Propriété	Document de transfert (Signataires)	Devenir du MB
SP1	*Oleae europea cultivé*	3	fleurs et feuilles	Collection ex-situ	OWGB Olive World Germplasm Bnak Marrackech Maroc	Bouchaib Khadari (CBNMED/INRA)	juin-11	A définir (selon accord entre membres du Conseil Oléicole International)	Message E-mail	ADN conservé par l'INRA en banque d'ADN
SP1	*Olea europea ssp cuspoidata*	1	fleurs et feuilles	Collection ex-situ	CBNMED Porquerolles	Bouchaib Khadari (CBNMED/INRA)	mai-11	France (CNBMED)	Message E-mail	ADN conservé par l'INRA en banque d'ADN
SP1	*Phillyrea angustifolia*	3	fleurs et feuilles	Collection ex-situ	CNRS-UMR CEFE	Bouchaib Khadari (CBNMED/INRA)	mai-11	CNRAS-CEFE	Message E-mail	ADN conservé par l'INRA en banque d'ADN
SP1	*Coffea canephora sauvages*	13	fleurs et feuilles	Collection ex situ	CRB Cirad Guyane (9) et serre Montpellier (4)	Thierry Leroy	mai et juillet 2012	Cirad	sans objet	sans objet
SP1	*Coffea pseudozangue bariae*	1	fleurs et feuilles	Collection ex situ	CRB Coffea IRD La Réunion	A. de Kochko	juil-12	IRD	sans objet	sans objet
SP1	*Bertiera sp. & Empogona sp.*	3	fleurs et feuilles	Jardin botanique	Jardin Botanique National de Belgique (Meise)	A. de Kochko	juil-12	Jardin Botanique National de Belgique (Meise)	Echange de courriels. Pas de MTA requis par le fournisseur.	
SP2/WP2-Riz	*Oryza spp.*	400	graines	Collecte *in situ* (ARCAD)	FOFIFA, Madagascar	Cirad Agap, Nour Ahmadi	mai-12			
SP2/WP2-Riz	*Oryza spp.*	495	graines	Banque de semence ex-situ	IRD, France, Alain Ghesquière	Cirad Agap, Nour Ahmadi	févr-12	Non connu	néant	Conservé à l'IRD Une partie a été envoyé en Guinée pour phénotypage

Sous-projet	*Genre, espèce*	Nombre d'accessions reçues	Nature du MB reçu	Type de provenance du MB	Fournisseur du MB	Destinataire du MB	Date(s) de réception du MB	Statut juridique du MB- Propriété	Document de transfert (Signataires)	Devenir du MB
SP2/WP3-Riz	*Oryza spp.*	730	panicules / graines	Collecte *in situ* (ARCAD)	IRAG, Guinée, Mamadou Billo Barry	Cirad, UMR Agap Nour Ahmadi	nov-11	Système multilatéral (Traité)	SMTA	Echantillons envoyés à AfricaRice pour conservation. Une partie a été retourné en Guinée pour phénotypage
SP2/ WP2 & 4	Medicago truncatula	534	lots de 25 graines	Banque de semence ex situ	CRB INRA Medicago	UMR AGAP equipe DAVEM	40057	INRA	transfert interne à l'équipe _ rédaction d'un document de traçabilité	Conservé par le CRB
SP2	*Pennisetum glaucum*	480	graines	Banque de semence ex-situ	ICRISAT, Niger	IRD, DIADE/DYNADIV/ Y Vigouroux	sept-09	Système multilatéral (Traité)	SMTA (ICRISAT/IRD)	pas de conservation prévu
SP3/WP1	*Zea mays L.*	130	semences (50K)	collecte in situ (111 acc) + variétés améliorés (19 acc)	INERA (Burkina Faso) Dr Jacob Sanou	INRA, UMR AGAP, B. Gouesnard	fevr-2011 + sept- 2011		lettre d'accompagnement	les talons de semences seront détruits
SP3/WP1	*Pennisetum glaucum*	48	graines	Banque de semence ex-situ	IRD, France	IRD, DIADE/DYNADIV/ Y Vigouroux	sept-09	IRD, France	néant (utilisation ressources internes du laboratoire collecté avant la convention de RIO)	
SP3/WP2/T1	*Sorghum bicolor*	209	lot de graines	collecte in situ au Mali (ARCAD)	IER, Mali, Fournisseur : Amadou Sidibé (responsable RG IER)	CIRAD, AGAP, France / Monique DEU	avr-11	Système multilatéral (Traité)	SMTA (Dir IER/Dir BIOS-CIRAD)	conservé au CRB-T CIRAD
SP3/WP2/T1	*Sorghum bicolor*	93	lot de graines	collecte in situ au Mali (ARCAD)	IER, Mali, Fournisseur : Amadou Sidibé (responsable RG IER)	CIRAD, AGAP, France / Monique DEU	janv-12	Système multilatéral (Traité)	SMTA (Dir IER/Dir BIOS-CIRAD)	conservé au CRB-T CIRAD

Sous-projet	Genre, espèce	Nombre d'accessions reçues	Nature du MB reçu	Type de provenance du MB	Fournisseur du MB	Destinataire du MB	Date(s) de réception du MB	Statut juridique du MB- Propriété	Document de transfert (Signataires)	Devenir du MB
SP3/WP2/T 4	*Sorghum bicolor*	171	panicules	Collecte in-situ Kenya (ARCAD)	National Gene Bank of Kenya (KARI)	CIRAD, AGAP, France / C. Leclerc	févr-11	Système multilatéral (Traité)	SMTA	Retournés au KARI
SP3/WP2/T 4	*Sorghum bicolor*	316	lot de graines	Collecte in-situ Kenya (ARCAD)	National Gene Bank of Kenya (KARI)	CIRAD, AGAP, France / C.Leclerc	août-11	Système multilatéral (Traité)	SMTA	Conservé au CIRAD
SP3/WP2/T 4	*Sorghum bicolor*	3500	feuilles	Collecte in-situ Kenya (ARCAD)	National Gene Bank of Kenya (KARI)	CIRAD, AGAP, France / C. Leclerc	févr-12	Système multilatéral (Traité)	SMTA	Conservé au CIRAD
SP3/WP2	*Oryza spp.*	400 dont 300 (*O.glaberrima*) 100 (*O. barthii*)	graines	Banque de semence ex-situ	IRD, France	IRD, DIADE/GGR/ F. Sabot – A. Ghesquière		IRD, France	néant (utilisation ressources internes du laboratoire)	Transfert à l'AfricaRice (projet MENERGEP) pour redistribution aux partenaires. Une partie envoyée Burkina Faso pour phénotypage (juin 2012).
SP3/WP2	*Triticum turgidum ssp. durum (Blé dur)*	4820	lots de graines représentant 165 populations locales	Collecte in-situ Maroc (ARCAD)	IAV Rabat Maroc	INRA Agap, Montpellier France / Pierre Roumet	mars-12	Système multilatéral (Traité)	SMTA . En dépit du SMTA, engagement tacite stipulant que pendant la durée du projet ni l'inra ni l'iav ne diffuseront ce matériel à un tiers .	Conservation sur 2 sites INRA et Rabat
SP3/WP4	*Digitaria exilis*	10	lot de graines	banque de semence ex situ	IRAG, Guinée	CIRAD, AGAP, France / Claire BILLOT	21-févr-12	Système multilatéral (Traité)	SMTA (DG IRAG/Dir AGAP - Copie fournie)	conservé au CRB-T CIRAD
SP3/WP4	*Digitaria exilis*	71	lot de graines, panicules	Collecte in-situ Guinée (ARCAD)	IRAG, Guinée	CIRAD, AGAP, France / Claire BILLOT	01-sept-12	Système multilatéral (Traité)	SMTA (DG IRAG/Dir AGAP - Copie fournie)	conservé au CRB-T CIRAD

Sous-projet	*Genre, espèce*	Nombre d'accessions reçues	Nature du MB reçu	Type de provenance du MB	Fournisseur du MB	Destinataire du MB	Date(s) de réception du MB	Statut juridique du MB- Propriété	Document de transfert (Signataires)	Devenir du MB
SP7/WP1	*Dioscorea trifida*	5	vitroplants	Vitrothèque	CRB Antilles INRA Guadeloupe	IRD, DIADE, F. Engelmann	nov-10	INRA	MTA INRA CDB	pas décidé