

UNIVERSITE CHEIKH ANTA DIOP DE DAKAR



ECOLE DOCTORALE SCIENCES DE LA VIE DE LA SANTÉ ET DE
L'ENVIRONNEMENT

FACULTE DES SCIENCES ET TECHNIQUES

Année : 2018

N° d'ordre : 201866

THESE DE DOCTORAT UNIQUE

Spécialité : Biotechnologies Végétales et Microbiennes et Amélioration des Plantes

ETUDE DES BASES GENETIQUES DE L'ADAPTATION DU MIL (*Pennisetum glaucum* ssp. *glaucum*) DU SENEGAL ET IDENTIFICATION DE MARQUEURS MOLECULAIRES PAR GENETIQUE D'ASSOCIATION

Présenté et soutenu le 1^{er} Février 2019

par

Oumar DIACK

Devant la commission d'examen composée de :

Président :	Pr Kandiora	NOBA	Professeur Titulaire	UCAD
Rapporteurs :	Dr Laurent	LAPLAZE	Directeur de Recherches	IRD
	Dr Ndiaga	CISSE	Maître de Recherches	ISRA
	Dr Abdala Gamby	DIEDHIOU	Maître de Conférences	UCAD
Examineurs :	Dr Ndjido Ardo	KANE	Chargé de Recherches	ISRA
	Dr Marie Madeleine	SPENCER	Maître de Conférences	UCAD
	Dr Diaga	DIOUF	Professeur Titulaire	UCAD
Directeur de Thèse :	Dr Diaga	DIOUF	Professeur Titulaire	UCAD
Maître de stage :	Dr Ndjido Ardo	KANE	Chargé de Recherches	ISRA

ECOLE DOCTORALE SCIENCES DE LA VIE DE LA SANTÉ ET DE
L'ENVIRONNEMENT

FACULTE : SCIENCES ET TECHNIQUES

Année : 2018

N° d'ordre : 201866

THESE DE DOCTORAT UNIQUE

Spécialité : Biotechnologies Végétales et Microbiennes et Amélioration des Plantes

RESUME

Nom et prénom du candidat : Oumar DIACK

Titre de la thèse : Etude des bases génétiques de l'adaptation du mil (*Pennisetum glaucum* ssp. *glaucum*) du Sénégal et identification de marqueurs moléculaires par génétique d'association

Date et lieu de soutenance : 1^{er} Février 2019 à l'UCAD

Jury :

Président :	Pr Kandoura	NOBA	Professeur Titulaire	UCAD
Rapporteurs :	Dr Laurent	LAPLAZE	Directeur de Recherches	IRD
	Dr Ndiaga	CISSE	Maître de Recherches	ISRA
	Dr Abdala Gamby	DIEDHIOU	Maître de Conférences	UCAD
Examineurs :	Dr Ndjido Ardo	KANE	Chargé de Recherches	ISRA
	Dr Marie Madeleine	SPENCER	Maître de Conférences	UCAD
	Dr Diaga	DIOUF	Professeur Titulaire	UCAD
Directeur de Thèse :	Dr Diaga	DIOUF	Professeur Titulaire	UCAD
Maître de stage :	Dr Ndjido Ardo	KANE	Chargé de Recherches	ISRA

Dédicaces

Je dédie ce travail à :

-Ma maman, pour son amour son éducation et son soutien permanent. L'incarnation typique de l'expression « Main de fer dans un gant de velours ».

-Mon papa, pour son amour son éducation et son soutien permanent. Un homme d'une rare dignité, un modèle d'intégrité.

Quelle chance j'ai, de vous avoir comme parents !

Remerciements

Je remercie tout d'abord, mes deux principaux encadrants le Dr Ndjido Ardo Kane et Dr Diaga Diouf pour m'avoir encadré durant tout ce travail.

Dr Diaga Diouf, je tiens à vous remercier pour avoir accepté de m'encadrer et d'être mon répondant à l'UCAD depuis mon master 2 et pour avoir par la suite, accepté d'être le directeur de cette thèse. Les mots me manquent en ce moment, pour exprimer ma gratitude pour tout ce que vous m'avez apporté. Pour ce que je suis en mesure de coucher sur cette feuille, je vous remercie profondément pour votre encadrement, votre disponibilité, votre patience, pour vos conseils et suggestions, mais surtout pour votre sens de responsabilisation de l'étudiant. Je retiens de vous votre rigueur scientifique et votre sens des exigences de l'encadrement, couvrant jusqu'aux détails les plus fins.

Dr Ndjido Kane, je ne saurais traduire en écrits ou en paroles ce que j'ai dans le cœur, au moment où je suis en train de rédiger ces phrases. Vous m'avez pris sous votre aile alors que j'étais encore qu'un étudiant en master et m'avez donné ma chance dans le monde de la recherche, d'abord en tant que stagiaire en master, puis en tant que doctorant. Je vous remercie profondément pour votre encadrement, votre disponibilité, votre patience, vos nombreux conseils, votre sympathie à mon égard, votre financement, mais surtout pour m'avoir toujours accordé votre confiance. Vous m'avez beaucoup aidé tant sur le plan scientifique, institutionnel (administratif) qu'humain. Je retiens principalement de vous votre rigueur scientifique, votre sens de la responsabilité mais aussi votre sens des rapports humains.

Je remercie Pr Kandoura Noba, Dr Laurent Laplaze, Dr Ndiaga Cisse, Dr Abdala Gamby Diedhiou et Dr Marie Madeleine Spencer (membres du jury de cette thèse) pour avoir accepté de juger ce travail.

Je remercie également Dr Adeline Barnaud et Dr Cécile Berthouly-Salazar, pour avoir participé à l'encadrement de cette thèse.

Dr Barnaud je vous remercie profondément pour avoir encadré la partie diversité génétique de cette thèse, pour votre soutien, vos suggestions et conseils et pour m'avoir orienté vers les bonnes personnes.

Dr Berthouly-Salazar je vous remercie sincèrement d'avoir accepté de participer à l'encadrement de la partie génétique de cette thèse et d'avoir financé mon séjour à Vienne. Je vous remercie aussi pour vos corrections, suggestions, vos conseils et critiques constructives.

Je remercie Dr Ablaye Ngom, pour tous ses conseils et suggestions qui m'ont beaucoup apporté et pour toutes ses réponses favorables, à chaque sollicitation de ma part.

Je remercie Dr Adama Faye, pour nos discussions toujours intéressantes et pour toutes ses suggestions et réponses favorables.

J'adresse mes remerciements à l'équipe de génétique moléculaire du LNRPV (Adama Ndiaye, Bineta Ndao, etc...) pour l'aide qu'ils m'ont apportée au laboratoire. Je remercie aussi toute l'équipe du bureau des doctorants et masters du LNRPV (Sally Diata, Arona Sonko, etc..) pour leur sympathie et les moments d'humour quotidien, qui m'ont beaucoup aidé à supporter la pression et le stress au travail.

Je remercie tout le personnel du LNRPV (allant de la directrice jusqu'aux stagiaires) pour leur sympathie, leur bonne humeur quotidienne. Je remercie en particulier Mr Mawa Sène (l'ex directeur administratif du LNRPV, actuellement affecté au CERAAS) pour tout son soutien et ses conseils et Mme Seynabou Mbaye et Mme Mame Bolly Mané (secrétaires administratives au LNRPV) pour leur aide dans les procédures administratives et les nombreuses sollicitations de ma part, en termes d'impressions de documents, auxquelles, elles ont toujours répondu favorablement.

Je remercie le département de Biologie Végétale (BV) de l'Université Cheikh Anta Diop de Dakar. De même, je remercie tout le personnel du Laboratoire de Biotechnologies Végétales du département (professeurs, chercheurs, doctorants et stagiaires), pour les échanges constructifs et leur sympathie.

Je remercie toute l'équipe mil du Centre National de Recherche Agronomique (CNRA) de Bambey (Dr Amadou Fofana, Dr Ousmane Sy, Dr Ghislain Kanfany, Mr Khadim Diouf, etc...) pour toute leur collaboration et leurs réactions complaisantes à mes différentes sollicitations.

Je remercie Mr Siaka Ndiaye et Mr Tamsir Mané (techniciens à la station ISRA de Nioro) pour toute l'aide, qui est allée au-delà même du cadre professionnel, qu'ils m'ont apportée lors des évaluations agro-morphologiques en champ.

Je remercie Mr Macoumba Diop (le chef de la station ISRA de Nioro) pour tout son appui tant sur le plan professionnel que personnel, lors des essais agro-morphologiques à Nioro. Je remercie aussi tout le personnel de la station, en particulier Mr Youssou Sankaré, Mme

Seynabou Ka, Mme Diaw et Mr Mostapha Thioune, pour leur sympathie et leur bonne humeur quotidienne qu'ils ont toujours eu à mon égard.

Merci à tous ceux qui ont contribué directement ou indirectement à l'aboutissement de ce travail.

Je garde le meilleur des remerciements, pour ma famille. Je remercie mon papa, ma maman, ma sœur Mame Diarra, et mes deux frères Pape et Mohammed pour m'avoir toujours épaulé, accompagné, encouragé et soutenu, sous toutes les formes, tout au long de mes études. Merci à vous tous pour avoir été toujours présents pour moi.

Cette thèse fait partie des plus grandes aventures que j'ai vécues jusqu'ici. J'y ai eu beaucoup d'expériences positives et quelques expériences négatives. Les expériences positives m'ont grandement apporté et je suis sûr, continueront à m'apporter. Les expériences négatives m'ont appris de bonnes leçons de la vie. Merci à vous tous qui avez contribué, d'une manière ou d'une autre, à la réalisation de cette aventure.

Résumé

Ces dernières décennies ont vu l'accroissement considérable de la population mondiale, entraînant une demande accrue et une intensification de l'usage des ressources naturelles afin de satisfaire les besoins alimentaires dans un contexte de changements climatiques. Pour relever ce défi, les sélectionneurs doivent mieux connaître, préserver, exploiter et gérer les ressources naturelles disponibles qui constituent un réservoir dans lequel on peut puiser pour produire des variétés adaptées aux contraintes environnementales locales, qui répondent aux changements climatiques. Parmi ces ressources figure le mil qui occupe une place capitale dans la production céréalière sénégalaise et présente un grand intérêt pour la recherche du fait de sa diversité et de son adaptation aux zones sèches. Ainsi, l'objectif de ce travail est de contribuer à une meilleure gestion et valorisation des ressources génétiques du mil au Sénégal.

Il a été réalisé (i) une collection constituée de quatre cent soixante-dix-sept (477) accessions dont 353 accessions précoces (Souna), 112 accessions tardives (Sanio) et 12 variétés améliorées, couvrant tout le territoire national. (ii) Suivi d'une évaluation complète de sa diversité génétique et phénotypique, qui a permis de mettre en évidence une grande diversité et d'identifier 2 pools génétiques du mil en fonction des Souna et des Sanio et une subdivision en trois sous-groupes agro-morphologiques dans chaque pool génétique suivant des caractères phénotypiques d'intérêt comme le rendement ou la floraison. Cette évaluation a aussi mis en évidence des corrélations de marqueurs des gènes adaptatifs *PgPHYC* et *PgMADS11* au cycle de floraison du mil et à des facteurs climatiques. (iii) Ensuite, une mise en place d'une core collection qui a permis de sélectionner 91 accessions (couvrant 22% de la collection nationale) ayant capturer au mieux la diversité de cette collection. (iv) Enfin, une étude d'associations génotypes-phénotypes dans cette core collection qui a permis d'identifier 6 caractères phénotypiques de différenciation entre Souna et Sanio et 21 marqueurs SNPs et 10 gènes liés à ces traits de différenciation, dont certains (gènes) sont impliqués dans l'adaptation et la résistance aux stress abiotiques.

Ces résultats nous permettent de déposer les conclusions suivantes : (i) Le mil du Sénégal présente une large diversité et que les types variétaux Souna et Sanio qui le composent se différencient aussi bien génétiquement que sur le plan morphologique ; (ii) Cette grande diversité peut être exploitée par les programmes de sélection à travers la core collection ; (iii) Les marqueurs des gènes *PgPHYC*, *PgMADS11* et gènes sortis du GWAS peuvent être validés et utilisés pour appuyer ces programmes de sélection et d'amélioration du mil, suivant des traits adaptatifs ou des caractères d'intérêt.

En perspective, afin de vérifier les possibilités d'une perte de cette diversité en réponse aux prévision des changements climatiques, une modélisation de l'adaptation du mil, suivant les scénarii futurs, prédits par le GIEC, a été initiée. Il serait aussi intéressant de faire un suivi dynamique de cette diversité, grâce à des échantillonnages temporels.

Beaucoup de programmes de sélection ont été réalisés pour la précocité de floraison (Souma) laissant de côté les Sanio. Il serait intéressant de mieux utiliser cette diversité du mil en intégrant les Sanio dans les schémas de sélection considérant l'aspect cultures à double usage.

Mots clé : Ressource génétique, diversité, adaptation, valorisation, collection, mil.

Abstract

In recent decades, there has been an important increase in the World's population, leading to increased demand and growing use of genetic resources to meet food needs in the context of climate change. To cope with this challenge, breeders need to better understand, conserve, exploit and manage the available genetic resources that provide a reservoir from which to source for breeding varieties that are adapted to local environmental constraints and respond to climate change. Among these resources is pearl millet, which occupies a key place in Senegalese cereal production and is of great interest for research because of its diversity and its adaptation to drylands. Thus, the objective of this work is to contribute to better management and enhancement of the genetic resources of pearl millet in Senegal.

For this, (i) A collection consisting of four hundred and seventy-seven (477) accessions was completed, including 353 early flowering (Souna), 112 late flowering (Sanio) and 12 improved varieties, covering the entire national territory. (ii) Follow-up of a complete evaluation of its genetic and phenotypic diversity, which made it possible to highlight a great diversity and identify 2 genetic pools of millet according to Souna and Sanio and a subdivision into three phenotypic subgroups in each gene pool according to phenotypic traits of interest such as yield or flowering time. This assessment also revealed correlations of adaptive gene markers *PgPHYC* and *PgMADS11* to pearl millet flowering cycle and to climatic factors. (iii) Secondly, it was done a setting up of a core collection which made it possible to select 91 accessions (covering 22% of the national germplasm) that have best captured the diversity of this germplasm. (iv) Finally, a genome-wide association study was conducted in this core collection which allowed to identify 6 phenotype characters that differentiate Souna and Sanio and 21 SNPs and 10 genes related to these differentiation traits, some of which (genes) are involved in adaptation and resistance to abiotic stress.

These results allow us to file the following conclusions: (i) The diversity of Senegalese pearl millet is large and the Souna and Sanio varietal types that make it up are differentiated genetically as well as morphologically. (ii) This great diversity of Senegal's pearl millet can be tapped by breeding programs through the core collection. (iii) The markers from *PgPHYC*, *PgMADS11* and GWAS genes can be validated and used to support these pearl millet breeding programs, for adaptive or interest traits.

In prospect, in order to verify the possibilities of a loss of this diversity in response to climate change predictions, a modeling of pearl millet adaptation, according to the future scenarios predicted by the IPCC, was initiated. It would also be interesting to dynamically monitor the diversity of these resources by temporal samplings.

Many breeding programs were carried out for early flowering (Souana) leaving aside Sanio. It would be interesting for better tapping this diversity in breeding programs by integrating the Sanio into breeding systems taking into account, for example, the dual-purpose crops.

Key words: Genetic resources, diversity, adaptation, valorization, germplasm, pearl millet.

Table des matières

Liste des acronymes et des abréviations.....	xii
Liste des figures.....	xvi
Liste des tableaux	xx
CHAPITRE I : INTRODUCTION GENERALE.....	1
INTRODUCTION.....	2
Objectifs de la thèse	5
Plan de la thèse.....	6
SYNTHESE BIBLIOGRAPHIQUE.....	7
1. Le mil	7
1.1. Taxonomie et botanique du mil.....	7
1.2. Morphologie et développement du mil.....	8
1.3. Origine et évolution du mil.....	11
1.4. Diversité et adaptation du mil.....	13
1.5. Gestion des ressources phytogénétiques actuelles du mil	14
2. Les nouvelles technologies de séquençage et leurs applications.....	18
Chapitre II. Première partie : Evaluation génétique des ressources de mil du Sénégal.....	21
INTRODUCTION.....	22
Objectif.....	23
MATERIEL ET METHODES.....	24
1. Prospection et collectes	24
2. Evaluation génétique neutre (marqueurs SSR).....	24
3. Evaluation génétique adaptative.....	28
RESULTATS	30
1. Enrichissement de la collection de mil du Sénégal	30
2. Evaluation génétique neutre	32
3. Evaluation génétique adaptative.....	41
DISCUSSION.....	43
CONCLUSION.....	47
Chapitre II. Deuxième partie : Evaluation agro-morphologique de la collection de mil du Sénégal	48
INTRODUCTION.....	49
Objectif.....	49
MATERIEL ET METHODES.....	50
1. Évaluation agro-morphologique des variétés précoces	50
2. Évaluation agro-morphologique des variétés tardives.....	52
3. Analyse des données	54

RESULTATS	56
1. Diversité agro-morphologique des variétés précoces	56
2. Diversité agro-morphologique des variétés tardives	61
DISCUSSION	65
CONCLUSION	68
Chapitre III : Constitution d'une core collection de mil du Sénégal	69
INTRODUCTION	70
Objectif.....	72
MATERIEL ET METHODES	73
1. Constitution de la core collection de mil du Sénégal	73
2. Evaluation de la représentativité	73
2.1. Représentativité des « pré-core collections »	73
2.2. Représentativité de la core collection.....	75
RESULTATS	76
1. « Pré-core collection » du pool génétique des variétés précoces (Souna).....	76
2. « Pré-core collection » du pool génétique des variétés tardives (Sanio)	80
3. Core collection (collection de référence) du germoplasme des mils cultivés du Sénégal	84
DISCUSSION	88
CONCLUSION	92
Chapitre IV : Etudes d'associations génotypes-phénotypes sur la core collection de mil du Sénégal..	93
INTRODUCTION	94
Objectif.....	95
MATERIEL ET METHODES	96
1. Matériel végétal et phénotypage.....	96
2. Génotypage.....	98
3. Méthodes analytiques	101
RESULTATS	104
1. Evaluation de la qualité des reads	104
2. Analyses génétiques	104
3. Analyses phénotypiques	105
4. Analyses d'associations génotypes-phénotypes sur les caractères de différenciation entre Souna et Sanio	115
DISCUSSION	123
CONCLUSION	125
CHAPITRE V : CONCLUSION GENERALE ET PERSPECTIVES	126
REFERENCES BIBLIOGRAPHIQUES	129

Liste des acronymes et des abréviations

Liste des acronymes

CERAAS : Centre d'Etude Régional pour l'Amélioration de l'Adaptation à la Sécheresse

CNRA : Centre National de Recherche Agronomique

CRZ : Centre de Recherche Zootechnique

ED-SEV : Ecole Doctorale Sciences de la Vie, de la Santé et de l'Environnement

FAO : Food and Agriculture Organization

GIEC : Groupe d'experts intergouvernemental sur l'évolution du climat

ICRISAT: International Crops Research Institute for the Semi-Arid Tropics

IPGRI : International Plant Genetic Resources Institute

IRD : Institut de Recherche pour le Développement

ISRA : Institut Sénégalais pour la Recherche Agricole

LNRPV : Laboratoire National de Recherche sur les Productions Végétales

NBPGR : National Bureau of Plant Genetic Resources

NCBI : National Center for Biotechnology Information

NPGRC : National Plant Genetic Resources Center

UCAD : Université Cheikh Anta Diop

URCI : Unité de Recherche en Culture In Vitro

SGSV : Svalbard Global Seed Vault

WAAPP : West Africa Agricultural Productivity Program (Programme de Productivité Agricole en Afrique de l'Ouest : PPAAO)

WIEWS : World Information and Early Warning System on Plant Genetic Resources

Liste des abréviations

°C : Degré Celsius

ABI : Applied Biosystems

ACP : Analyse en composantes principales

ADN : Acide Désoxyribonucléique

AFC : Analyse Factorielle des Correspondances

AFD : Analyse Factorielle Discriminante

AFLP : Amplified fragment-length polymorphism (Polymorphisme de longueur des fragments amplifiés)

AMMI : Additive Main-effects and Multiplicative Interaction (Modèles des effets principaux additifs et interactions multiplicatives)

ANOVA : Analysis of variance (Analyse de variance)

ARN : Acide ribonucléique

ATM : Accord type de transfert de matériel

BET : Bromure d'éthidium

CHA : Classification Hiérarchique Ascendante

CO₂ : Dioxyde de carbone

CR : Coïncidence Rate (taux de coïncidence)

CV : Coefficient de Variation

EBD : Electrophorèse bidimensionnelle

EDTA : Acide éthylène diamine tétraacétique

eQTL : Expression Quantitative Trait Loci

EST : Expressed Sequence Tag (Marqueur de séquence exprimée)

F_{IS} : Fixation index of an individual relative to the subpopulation (coefficient de consanguinité entre individus de même sous-population)

F_{IT} : Fixation index of an individual relative to the total population (coefficient de consanguinité entre individus dans la population totale)

F_{ST} : Fixation index of subpopulations compared to the total population (indice de différenciation entre sous-populations)

g : Gramme

GIS : Geographic Information System (Système d'Information Géographique)

GWAS : Genome-wide association study

ha : Hectare

H_{EX} : Hétérozygotie attendue

H_O : Hétérozygotie observée

HWE : Hardy–Weinberg Equilibrium (Equilibre d'Hardy et Weinberg)

IBD : Isolation by distance (isolement par la distance)

InDel : Insertion Délétion

ISSR : Inter Simple Sequence Repeats

kb : Kilo base

L : Litre

Ln : Logarithme népérien

MCMC : Markov Chains Monte Carlo

MD : Mean Difference (différence de moyennes)

mM : Millimolaire

NGS : Next-generation sequencing

PCR : Polymerase Chain Reaction

pb : Paire de bases

QTL : Quantitative Trait Loci

RA : Richesse allélique

RAPD : Random Amplified Polymorphic DNA (Amplification aléatoire d'ADN polymorphe)

RFLP : Restriction Fragment Length Polymorphism (Polymorphisme de longueur des fragments de restriction)

RPGAA : Ressources phytogénétiques pour l'alimentation et l'agriculture

SNP : Single Nucleotide Polymorphisme

sPCA : Spatial Analysis of Principal Components (analyse spatiale des composantes principales)

SSR : Simple Sequence Repeat

STR : Short Tandem Repeats

UV : Ultraviolet

VNTR : Variable Number Tandem Repeats

μL : Microlitre

μM : Micromolaire

Liste des figures

Figure 1 : Différences morphologiques entre Souna et Sanio. (a) Port d'un Souna comparé à celui d'un Sanio. (b) Epi d'un Souna comparé à (c) celui d'un Sanio.

Figure 2 : Phases et stades successives du cycle de développement du mil

Figure 3 : Représentation des centres de recherches ou des banques de gènes ayant déposé un duplicata au SGSV (Svalbard Global Seed Vault)

Figure 4 : Schéma montant les différents domaines d'application et utilisations possibles des NGS

Figure 5 : Comparaison de la représentativité variétale et géographique de la nouvelle collection par rapport à l'ancienne collection de mil du Sénégal

Figure 6 : Valeurs de delta K (courbe rouge) et la courbe de régression de LnPD (courbe bleue) pour $K = 1$ à $K = 6$, selon la méthode de calcul d'Evanno et al (2005)

Figure 7 : Structure de la population pour $K = 2$ sur la base des 404 accessions analysées avec 12 SSR

Figure 8 : Analyse en composantes principales des variétés améliorées, des accessions précoces (Souna) et tardives (Sanio) de la collection

Figure 9 : Degrés de similarités entre les différentes accessions (précoces, tardives et variétés améliorées) de la collection de mil

Figure 10 : Analyse Spatiale des composantes principales (sPCA), des accessions précoces et tardives de la collection

Figure 11 : Comparaison de l'assignation génétique entre les résultats de STRUCTURE et ceux de la sPCA pour les accessions locales précoces et tardives

Figure 12 : Représentations graphiques du modèle d'isolement par la distance pour les accessions locales précoces et tardives

Figure 13 : Représentation graphique des probabilités que deux individus, ayant des liens de parenté ou non, présentent des génotypes identiques en fonction des marqueurs cumulés croissants

Figure 14 : Pluviométries journalières enregistrées à Bambey durant l'hivernage 2014 et l'hivernage 2015

Figure 15 : Pluviométries journalières enregistrées à Nioro durant l'hivernage 2014 et l'hivernage 2015

Figure 16 : Pluviométries journalières enregistrées à Sinthiou Malem durant l'hivernage 2014 et l'hivernage 2015

Figure 17 : Pluviométries journalières enregistrées à Kolda durant l'hivernage 2014 et l'hivernage 2015

Figure 18 : Représentation graphique de l'ACP du model AMMI des accessions précoces par rapport au paramètre de rendement en grains.

Figure 19 : Représentation graphique de l'ACP du model AMMI des accessions précoces par rapport au paramètre de date de floraison

Figure 20 : Analyse de Classification Hiérarchique Ascendante des Souna de la collection

Figure 21 : Analyse Factorielle discriminante des accessions précoces

Figure 22 : Représentation graphiques des barycentres des trois sous-groupes des accessions précoces par rapport aux caractères agro-morphologiques

Figure 23 : Représentation graphique de l'ACP du model AMMI des accessions tardives par rapport au paramètre de rendement en grains.

Figure 24: Représentation graphique de l'ACP du model AMMI des accessions tardives par rapport au paramètre de date de floraison.

Figure 25 : Analyse de Classification Hiérarchique Ascendante des Sanio de la collection

Figure 26 : Représentation graphique de l'Analyse Factorielle discriminante des accessions tardives

Figure 27 : Représentation graphiques des barycentres des trois sous-groupes des accessions tardives par rapport aux caractères agro-morphologiques

Figure 28 : Analyse en coordonnées principales de la « pré-core collection » des Souna

Figure 29 : Analyse en coordonnées principales de la « pré-core collection » des Sanio

Figure 30 : Analyses bayésiennes de la core collection de mil, en comparaison avec l'analyse bayésienne de la collection nationale

Figure 31 : Analyse Factorielle Discriminante des Souna et des Sanio, à partir des données agro-morphologiques.

Figure 32 : Carte de la core collection de mil du Sénégal, en comparaison avec celle de la collection nationale

Figure 33 : Schéma illustrant les mesures du dispositif expérimental sur un site d'essai

Figure 34 : Schéma illustrant les semis du dispositif expérimental d'un essai

Figure 35 : Diagramme du workflow de traitement bio-informatique de vérification de qualité, de trimming et de mapping des reads sur le génome de référence

Figure 36 : Diagramme du workflow d'analyse bio-informatique d'élimination des duplicatas techniques, de réalignement par rapport aux InDels et d'appel des SNPs.

Figure 37 : (a) Probabilité de subdivisions par rapport aux différent cas testés (la valeur de cross-entropy la plus basse est la meilleur probabilité) et (b) proportion d'appartenance de chaque individu aux différents K groupes, des accessions de la core collection

Figure 38 : Analyse en composantes principales des accessions de la core collection sur la base des données SNPs (Axe1= 7.15%, Axe2= 1.54%)

Figure 39 : Analyse en composantes principales montrant (a) les corrélations entre variables, (b) la structure entre Souna et Sanio et (c) les relations de corrélation entre variables phénotypiques et structure agro-morphologique des Souna et des Sanio (Axe1= 50.4%, Axe2= 18.6%)

Figure 40 : Box plots des caractères de (a) date d'épiaison, (b) date de floraison, (c) nombre d'entrenœuds, (d) de biomasse, (e) nombre de talles productives et (f) de hauteur de la plante en fonction des Souna (en rouge foncé) et des Sanio (en bleu).

Figure 41 : Analyse en composantes principales montrant (a) les corrélations entre variables, (b) la structure des sous-groupes (SG) de Souna et (c) les relations de corrélation entre variables phénotypiques et la structure agro-morphologique des Souna (Axe1= 39.6%, Axe2= 14.5%).

Figure 42 : Box plots des caractères de (a) date d'épiaison, (b) date de floraison et de (c) diamètre de la tige principale, en fonction des sous-groupes 1 (en bleu), sous-groupe 2 (violet orchidée) et sous-groupe 3 (en vert) des Souna.

Figure 43 : Analyse en composantes principales montrant (a) les corrélations entre variables, (b) la structure des sous-groupes (SG) de Sanio et (c) les relations de corrélation entre variables phénotypiques et la structure agro-morphologique des Sanio (Axe1= 42.5%, Axe2= 16.1%).

Figure 44 : Box plots des caractères de (a) date d'épiaison, (b) date de floraison et de (c) nombre d'entrenœuds, de (d) circonférence de l'épi, de la (e) biomasse et du (f) rendement, en fonction des sous-groupes 1 (en vert olive), sous-groupe 2 (en cyan) et sous-groupe 3 (en magenta) des Sanio.

Figure 45 : Manhattan plot par rapport à la date d'épiaison et la date de floraison et QQ plot par rapport aux mêmes caractères, respectivement.

Figure 46 : Manhattan plot par rapport au nombre d'entrenœuds et la biomasse et QQ plot par rapport aux mêmes caractères respectivement.

Figure 47 : Manhattan plot par rapport au nombre de talles productives et la hauteur de la tige principale et QQ plot par rapport aux mêmes caractères respectivement

Liste des tableaux

Tableau 1 : Statistiques des estimateurs de la diversité génétique pour les 12 SSR utilisés dans l'analyse des 404 accessions de mil.

Tableau 2 : Statistiques de la diversité génétique pour les variétés améliorées, les accessions locales précoces et tardives.

Tableau 3 : Décompte des génotypes pour les variétés améliorées, les accessions locales précoces et tardives par rapport aux valeurs de $q \leq 0,7$

Tableau 4 : Analyse de variance des variétés précoces (Souna).

Tableau 5 : Matrice de corrélation de Pearson entre les variables, à partir de l'Analyse en Composantes Principales (ACP) des accessions précoces (Souna)

Tableau 6 : Analyse de variance des variétés tardives (Sanio)

Tableau 7 : Matrice de corrélation de Pearson entre les variables, à partir de l'Analyse en Composantes Principales (ACP) des accessions tardives (Sanio)

Tableau 8 : Valeurs de nombre d'allèles, d'hétérozygotie attendue et observée des accessions précoces (Souna) et des pré-core collections du pool génétique des Souna

Tableau 9 : Statistiques de moyennes, d'écart type et de coefficients de variation des caractères agro-morphologiques des accessions précoces (Souna) et des pré-core collections du pool génétique des Souna

Tableau 10 : Valeurs de nombre d'allèles, d'hétérozygotie attendue et d'hétérozygotie observée des accessions tardives (Sanio) et des pré-core collections du pool génétique des Sanio

Tableau 11 : Statistiques de moyennes, d'écart type et de coefficients de des caractères agro-morphologiques des accessions tardives (Sanio) et des « pré-core collection » du pool génétique des Sanio

Tableau 12 : Liste de la collection de référence du germoplasme des mils cultivés du Sénégal

Tableau 13 : Récapitulatif des résultats d'analyse de variance et de calcul de l'héritabilité génétique (h^2) Significativité : * P-value < 0.05, ** P-value < 0.01 et *** P-value < 0.001, NS= Non significatif.

Tableau 14 : Résumé des statistiques de l'analyse d'association entre marqueurs SNPs et traits phénotypiques de différenciation entre Souna et Sanio

Tableau 15 : Liste des marqueurs SNPs et informations connexes

CHAPITRE I : INTRODUCTION GENERALE

INTRODUCTION

Ces dernières décennies ont vu l'accroissement considérable de la population mondiale, entraînant une demande accrue et une intensification de l'usage des ressources naturelles afin de satisfaire les besoins alimentaires. En effet la population mondiale a plus que doublé en cinquante (50) ans. Elle est passée de 3,25 milliards d'habitants en 1964 à 7,26 de milliards en 2014 (FAOSTAT, 2017). Cette croissance de la population continuerait dans les années à venir, car il est estimé que la taille de la population mondiale atteindra 9,72 milliards d'habitants en 2050 (FAOSTAT, 2017). Cet accroissement de la population, met à l'épreuve, la sécurité alimentaire, dans certaines zones pauvres du globe, comme en Afrique sub-saharienne. Il existe un déficit entre les besoins de nutrition et les disponibilités alimentaires, des ménages et au niveau global des pays (Dembélé, 2001).

En parallèle, les changements climatiques exposent l'agriculture à de nouveaux défis. Plusieurs modèles climatiques prédisent un réchauffement de la planète accompagné d'une augmentation des surfaces arides dans le monde et d'une baisse de la pluviométrie. En somme, de manière générale, ces modèles prédisent un accroissement de la sécheresse. Dans ces scénarios, le sahel fait partie des zones les plus touchées dans le monde. La production agricole se verra surement ralentie à cause de ces changements climatiques pouvant occasionner en plus de la sécheresse, des inondations et l'apparition de nouveaux ravageurs mettant ainsi en péril les espèces ou variétés locales. Cela peut conduire à long terme, à la perte de ces variétés, entraînant alors une érosion génétique, compromettant la sécurité alimentaire dans les zones où ces variétés constituent la base de l'alimentation.

Dans les 50 ans à venir, l'agriculture devra nourrir environ le double de la population actuelle, dans des conditions encore plus difficiles (FAO, 2010). Pour relever ce défi, l'agriculture devra s'appuyer sur les ressources naturelles disponibles. Elle devra non seulement exploiter au mieux les espèces majoritairement cultivées (riz, maïs, blé) mais également diversifier les cultures (mil, fonio, par exemple).

Les ressources naturelles constituent le réservoir à partir duquel on peut puiser pour relever ce défi de l'agriculture. Par conséquent, la conservation et la bonne gestion de ces ressources est indispensable. En effet, un des moyens d'avoir des variétés adaptées aux contraintes environnementales locales, qui répondent aux changements climatiques, est de produire ces variétés par le biais des programmes de sélections et d'améliorations variétales. Cependant ces programmes s'appuient sur les ressources disponibles et sur la bonne connaissance de ces

dernières. Il est donc important de disposer de ces ressources grâce à leur conservation dans les collections mais aussi d'évaluer ces ressources pour bien les connaître. D'après l'avis de chercheurs de plus en plus nombreux (Bezançon and Pham, 2004), les recherches sur la diversité, la conservation et la valorisation des ressources génétiques devraient être replacées dans un contexte plus large que celui de la recherche génétique simple (Bezançon and Pham, 2004). En effet une dimension intégrant les changements globaux est nécessaire. Car ces changements climatiques ont un impact sur la diversité génétique des plantes cultivées. D'autant plus que certaines adaptations de ces plantes sont précisément en réponses à ces changements globaux. En effet des études ont montré l'existence de changement de traits adaptatifs qui sont liés à ces changements climatiques. Ces études montrent à quel point le climat peut façonner la diversité génétique des plantes cultivées aussi bien sur le plan spatial que temporel. Elles montrent ainsi à quel point les changements climatiques peuvent impacter sur cette diversité. Par conséquent, une connaissance de la diversité de ces ressources génétiques intégrant une compréhension de leur adaptation en réponse aux changements climatiques et leur adaptation par rapport aux facteurs environnementaux est indispensable pour une bonne gestion et valorisation de ces ressources.

L'agriculture africaine est majoritairement céréalière. Parmi ces cultures, le mil occupe une place importante. Il représente la 3^{ème} céréale produite en Afrique de l'Ouest et constitue l'alimentation quotidienne de cinquante (50) millions d'habitants du Sahel (Saïdou et al., 2009). Au Sénégal, le mil occupe une place capitale avec 66 % de la production céréalière, et a une production en moyenne pour ces cinq dernières années de plus de 688 992 tonnes par an, et de 515 365 tonnes en 2013 (ANDS, 2014). En plus de son importance sur le plan alimentaire, le mil présente des intérêts pour la recherche car il présente une importante diversité agromorphologique (Akanvou et al., 2012), variétale (Bezançon et al., 2008) et génétique (Chowdari et al., 1998). C'est une espèce bien adaptée aux zones sèches comme le Sahel qui est aussi une zone caractérisée par une importante variabilité climatique (Kouressy et al., 2004). De ce fait, si les prédictions futures des modèles climatiques s'avèrent réelle, le mil pourrait être la céréale du futur car pouvant devenir un instrument majeur de la lutte contre les impacts de la sécheresse sur la sécurité alimentaire, du fait qu'il est cultivé dans des zones semi-arides et arides et du fait de son potentiel génétique.

Ces ressources génétiques de mil peuvent représenter ainsi un grand intérêt en termes de gestion de la biodiversité sans compter le fait qu'elles constituent un immense réservoir de gènes encore peu étudiés dans des pays comme le Sénégal et potentiellement intéressants pour l'agriculture

de demain. La connaissance de l'état de leur diversité et la compréhension de leur adaptation est fondamentale pour une bonne exploitation de ces ressources génétiques.

Objectifs de la thèse

Notre étude se déroule au Sénégal, pays non soustrait aux effets négatifs des changements climatiques sur l'agriculture, qui impactent sur toute la région Sahélienne, mais dont on a très peu de connaissances sur ses ressources génétiques de mil. A travers cette thèse, nous comptons utiliser des outils génétiques et génomiques disponibles comme les marqueurs moléculaires et des outils agro-morphologiques, à savoir les marqueurs phénotypiques pour évaluer la diversité et l'adaptation des ressources génétiques de mil du Sénégal afin de participer à l'effort de leur gestion et de leur valorisation.

Ainsi donc, l'objectif global de cette thèse est de participer à l'effort de gestion, de connaissance et de valorisation des ressources génétiques de mil du Sénégal. Ses objectifs spécifiques se déclinent en quatre points :

- Etoffer la collection de mil du Sénégal
- Evaluer la diversité génétique et agro-morphologique de cette collection
- Constituer une core collection représentative de cette collection nationale
- Identifier des marqueurs moléculaires liés à des traits phénotypiques ou adaptatifs de ces mils

A terme, l'atteinte de ces objectifs permettra d'avoir à notre disposition une collection nationale de mil bien connue, représentative des ressources de mil du pays et utilisable dans les programmes d'amélioration et de sélection variétale.

Plan de la thèse

Cette thèse sera organisée autour de trois (3) principaux chapitres qui seront encadrés par une introduction générale et une conclusion générale :

Dans l'introduction générale (premier chapitre de la thèse), nous introduirons la thèse, dégagerons ses objectifs et dresserons son plan avant de faire une synthèse de la bibliographie.

Un deuxième chapitre fera état de l'enrichissement et de l'évaluation des ressources génétiques de mil au Sénégal. Elle sera constituée de deux parties. (i) Après une brève description de l'étoffement de la collection de mil du Sénégal, une première partie traitera de la caractérisation génétique à l'aide de marqueurs moléculaires. Dans cette partie, nous ferons une caractérisation neutre des mils du Sénégal à l'aide de marqueurs de types microsatellites. Nous ferons ensuite une caractérisation adaptative, à l'aide de deux gènes liés à l'adaptation du mil à l'environnement (*PgMads11* et *PgPhyC*). (ii) Une deuxième partie de ce chapitre fera état d'une caractérisation à l'aide de descripteurs agro-morphologiques.

Après une bonne connaissance des ressources de mil du Sénégal, objet du second chapitre de cette thèse, un troisième chapitre traitera de la mise en place d'une core collection de mil du Sénégal et évaluera la représentativité de celle-ci par rapport à la collection nationale de mil.

Enfin, un quatrième chapitre s'intéressera à une étude d'association de marqueurs génétiques à des traits agro-morphologiques des accessions de cette core collection de mil du Sénégal.

Une conclusion générale fera une synthèse des conclusions partielles des différents chapitres et en dégagera des perspectives.

SYNTHESE BIBLIOGRAPHIQUE

1. Le mil

1.1. Taxonomie et botanique du mil

Le mil est un terme générique qui regroupe un ensemble de genres de plantes. D'après la classification de Cronquist, (1981), il appartient à la famille des Poaceae (Graminées), de l'Ordre des Cyperales, de la classe des Liliopsida, de la division des Magnoliophyta, du règne des Plantae. Le caractère commun de ces genres est la petite taille des graines (FAO et ICRISAT, 1997). Les différentes espèces de mils cultivés sont *Pennisetum glaucum*, *Eleusine coracana*, *Panicum miliaceum*, *Setaria italica*, *Eragrostis tef*, *Echinochloa crusgalli*, *Panicum sumatrense*, *Paspalum scrobiculatum*, *Coix lachryma-jobi*, *Digitaria exilis*, *Digitaria iburua* et *Brachiaria deflexa* (FAO et ICRISAT, 1997). Dans notre étude, nous allons nous intéresser à l'espèce *Pennisetum glaucum*.

Pennisetum glaucum

Pennisetum glaucum appelée mil pénicillaire ou mil à chandelle (Fig. 1-I), est une espèce appartenant au genre *Pennisetum*, à la tribu des Paniceae, à la sous-famille des Panicoideae, à la famille des Poaceae. C'est une plante diploïde à 14 chromosomes ($n = 7$), sexuée à fleurs hermaphrodites, à reproduction allogame favorisée par une protogynie prononcée et à pollinisation préférentiellement anémophile. C'est la céréale la plus tolérante à la sécheresse, cultivée dans des zones sous des isohyètes pouvant descendre jusqu'à 200 mm par an (Bezançon and Pham, 2004)

Cette espèce regroupe trois (3) sous-espèces ; *Pennisetum glaucum* ssp. *monodii*, *Pennisetum glaucum* ssp. *glaucum* et *Pennisetum glaucum* ssp. *stenostachyum*.

1.1.1. *Pennisetum glaucum* ssp. *monodii*

Elle est aussi appelée *Pennisetum glaucum* ssp. *violaceum* (Zon, 1992). C'est la sous-espèce sauvage. Elle pousse à l'état spontané. Sa distribution géographique est limitée à la zone Sahélienne de l'Afrique. Elle est caractérisée par une petite taille des épis, des épillets caducs à maturité et des graines très petites et enveloppées par les glumelles. C'est une plante herbacée, à tiges fines très ramifiées sur toute leur longueur, terminées chacune par un épi d'environ 8 cm de long.

1.1.2. *Pennisetum glaucum* ssp. *stenostachyum*

Elle regroupe l'ensemble des formes intermédiaires issues d'hybridations naturelles entre formes cultivées et formes sauvages (Marchais et al., 1993). Encore appelée *Pennisetum glaucum* ssp. *sieberianum* (Zon, 1992). Appelée en Wolof « N'douls » par les paysans Sénégalais (Bilquez and Lecomte, 1969). Elle est souvent rencontrée en périphérie des champs de mil cultivé et présente des épis de taille et de morphologie intermédiaire relativement par rapport à ceux des formes sauvages et cultivées.

1.1.3. *Pennisetum glaucum* ssp. *glaucum*

C'est la sous-espèce cultivée. Elle est rencontrée en Afrique, en Asie et en Amérique. Elle se différencie morphologiquement du type sauvage par la non-caducité des graines à maturité (Tostain, 1998), l'augmentation de la taille des graines, la perte de la dormance, la diminution de la longueur des soies de l'épi (Poncet et al., 1998). Génétiquement, elle se distingue par six gènes sous sélection, intervenant dans la date de floraison (Clotault et al., 2012).

Cette espèce cultivée se subdivise principalement en deux variétés distinguées par les paysans, en fonction de la longueur du cycle de culture, appelées différemment selon les endroits. Une variété précoce dont la date de floraison varie de 50 à 60 jours (Dussert et al., 2015a), appelée « Souna » au Sénégal et au Mali, « Hainis » au Burkina, « Haïni Kiré », « Guerguéra », « Zongo », « Ba-Angouré », « Ankoutess » et « Boudouma » au Niger (Bezançon et al., 1997a) et une variété tardive dont la date de floraison varie de 80 à 110 jours (Dussert et al., 2015a) appelée « Sanio » au Sénégal et au Mali, « Kazouyas » au Burkina et « Maïwa » au Niger (Bezançon et al., 1997a).

1.2. Morphologie et développement du mil

Le mil (*Pennisetum glaucum*) est une plante annuelle à port érigé dont la taille peut aller jusqu'à 4 mètres de haut. Le système racinaire est fasciculé et très développé. Les chaumes sont robustes et dressés. Le diamètre d'une tige varie de 1 à 4 cm. La plante présente des talles dont le nombre est souvent fonction de la variété. Les feuilles d'environ 100 cm de long et 10 cm de large, retombant le long des chaumes, ont une gaine glabre mais souvent ciliées au sommet. L'inflorescence est un épi cylindrique d'environ 5 à 40 cm de long et 1 à 4 cm de large composé d'un rachis qui porte les pédicelles à base desquelles il y a les involucre qui sont constitués par un bouquet de soies contenant les épillets. Chaque épillet porte deux fleurs. Une fleur inférieure mâle et une fleur supérieure femelle. Ceci dit, le mil est protogyne (la fleur femelle fleurit avant

la fleur mâle). Le fruit est un caryopse ovoïde, d'environ de 2 à 4,5 mm de long de couleur jaune ou brune.

Bien que formant une même sous-espèce, le Souna et le Sanio, au Sénégal, présentent quelques différences morphologiques observables à l'œil nu, en dehors de la différence du cycle de floraison (Figure 1). Parmi ces différences morphologiques, nous pouvons distinguer le tallage, la taille de la plant et l'aspect aristé des épis à maturité. En effet, le Sanio présente un tallage plus important avec un port plus développé que celui des Souna. Il a aussi généralement des arêtes au niveau des épis à maturité, ce qui n'est généralement pas observé chez le Souna.

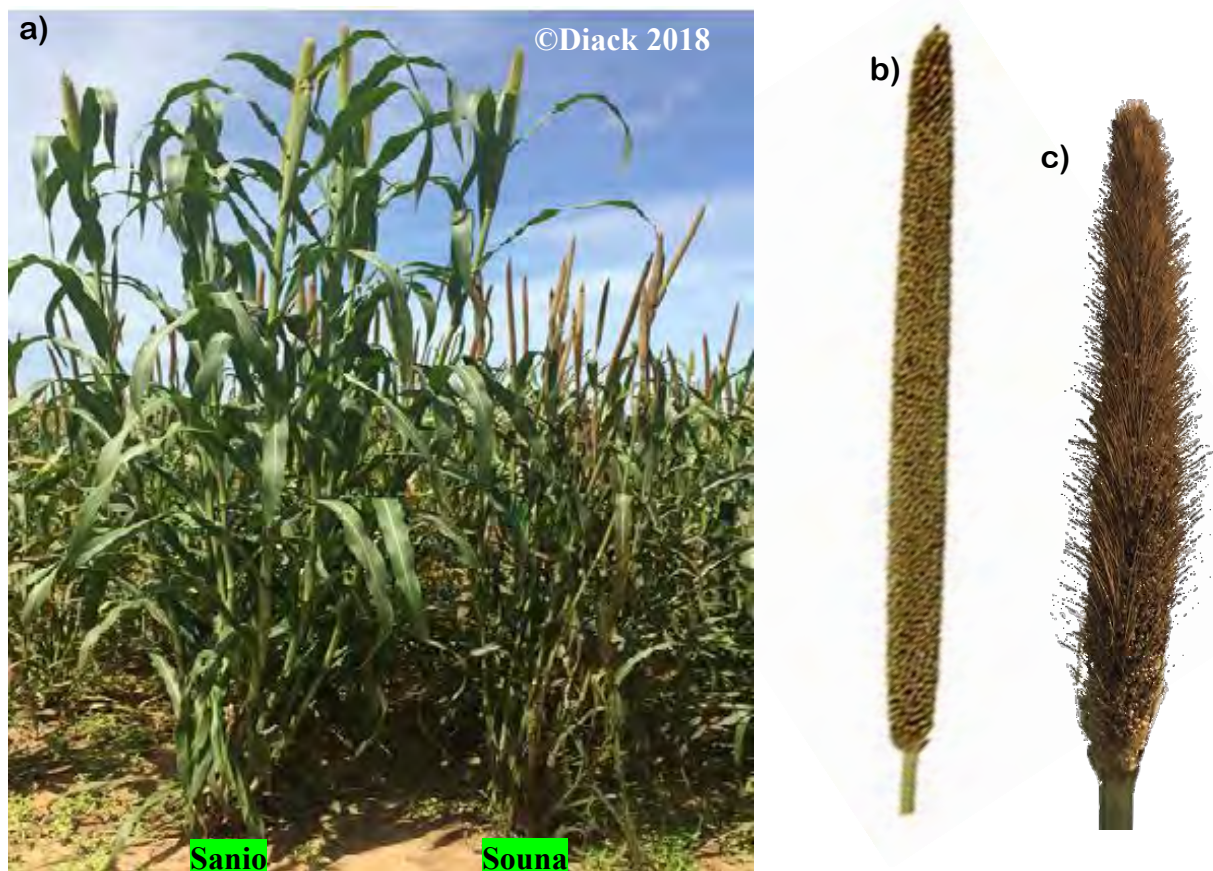


Figure 1 : Différences morphologiques entre Souna et Sanio. (a) Port d'un Souna comparé à celui d'un Sanio. (b) Epi d'un Souna comparé à (c) celui d'un Sanio.

Le cycle de développement du mil comporte principalement 3 phases (Fig. 2) : La phase végétative, la phase reproductive (ou phase de développement paniculaire) et la phase de maturation. Chaque phase comprend plusieurs stades successifs (Maiti and Bidinger, 1981) :

- **La phase végétative** : Elle comprend quatre (4) stades que sont : (i) L'émergence racinaire, ou germination de la graine. Durant ce stade, la racine établit le système racinaire primaire (racines séminales) et produit des racines adventives et initie les feuilles ; (ii) Le stade trois (3) feuilles durant lequel la racine principale augmente rapidement et développe des racines secondaires fines (dont une ou deux sont visibles) ; (iii) Le stade cinq (5) feuilles où la racine principale est déjà bien développée et possède plusieurs racines secondaires visibles ; (iv) L'initiation paniculaire qui est marquée par l'allongement du dôme apical et la formation d'une constriction à la base de l'apex. A la fin de cette phase, environ sept feuilles (y compris les feuilles embryonnaires) sont généralement déjà en place. Quant à la racine principale, elle produit un réseau de racines latérales qui se développent rapidement.

- **La phase reproductive** : Elle débute à l'initiation paniculaire et comprend trois (3) stades : (i) Le stade d'initiation de la feuille étendard ou feuille paniculaire durant lequel les feuilles non expansées qui étaient présentes se développent séquentiellement. Les inter-nœuds de la tige s'allongent en séquence à partir de la partie basale de la tige. Ce stade est accompagné par un assèchement des premières feuilles développées à la base de la tige. L'accumulation de matière sèche se produit dans les racines, les feuilles et la tige ; (ii) Le stade de développement de la feuille paniculaire durant lequel la panicule est enfermée dans la gaine de la feuille drapeau. La croissance de la panicule se fait en longueur comme en largeur. Puis, elle émerge du col de la feuille drapeau tandis que le pédoncule (l'inter-nœud supérieur) commence à s'allonger ; (iii) Le stade de floraison 50% où l'on observe une série de changements morphologiques et de développements distincts de la panicule, dont le développement des glumes, des fleurs, des styles, des anthères et du stigmate. La fin de ce stade est marquée par la pollinisation.

- **La phase de maturation** : Elle comprend trois (3) stades : (i) Le stade grains laiteux qui survient quelques jours après la fécondation de l'ovaire (6 à 7 jours) lorsque les grains grandissent suffisamment pour devenir visibles. À ce stade, ils sont composés d'une membrane remplie d'un liquide fluide qui se transforme ensuite en liquide laiteux. La présence cette dernière témoigne de l'accumulation d'amidons dans les cellules d'endosperme ; (ii) Le stade grains pâteux est identifiable par l'observation d'un changement dans l'endosperme allant d'un état laiteux à un état semi-solide puis solide. Il s'agit d'un changement progressif qui se produit lorsque la teneur en amidon dans l'endosperme augmente et que le pourcentage d'humidité diminue ; (iii) Le stade de maturité physiologique qui est caractérisé par la formation d'une sorte de teinte noirâtre au niveau du hile du grain. La formation de cette couche coïncide avec l'arrêt

de croissance du grain. À ce stade, le grain (partiellement sec) a atteint son poids maximal et son endosperme est devenu dur.

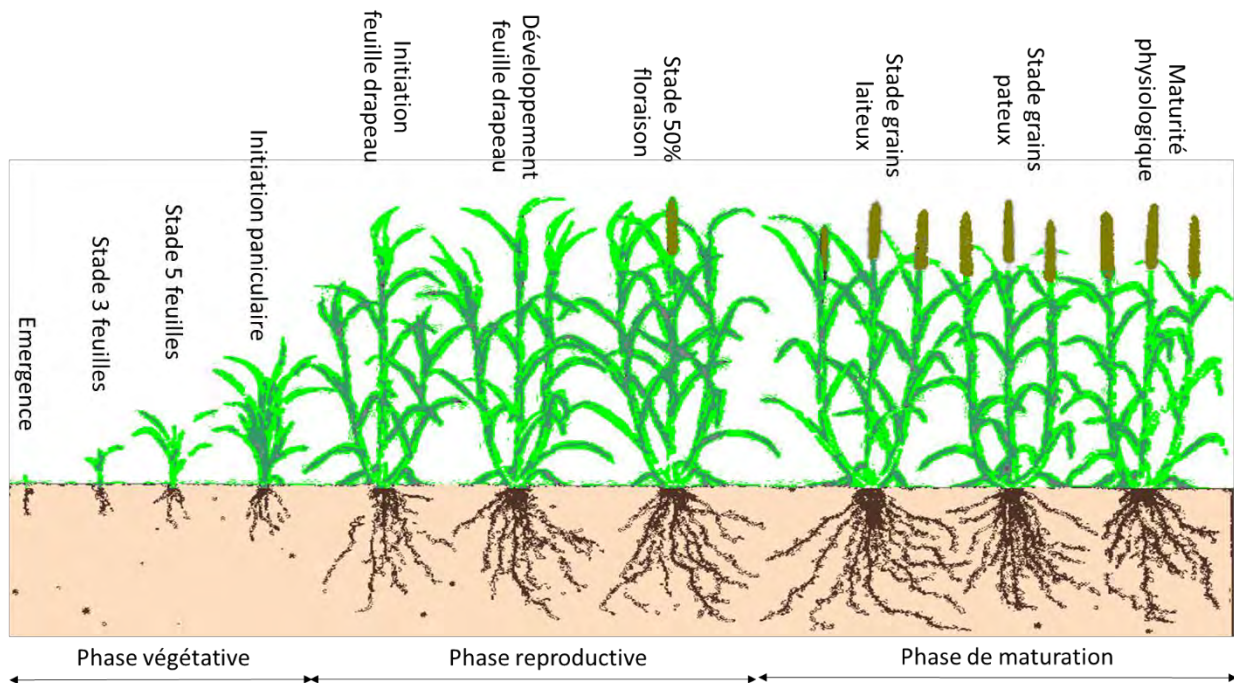


Figure 2 : Phases et stades successives du cycle de développement du mil (Maiti and Bidinger, 1981 modifié).

1.3. Origine et évolution du mil

A l'heure actuelle, les études ayant formulé des hypothèses sur l'origine et l'évolution du mil peuvent être classées en deux groupes. Ces deux groupes d'études n'ont en commun que l'affirmation selon laquelle le mil (*Pennisetum glaucum* ssp. *glaucum*) a été domestiqué en Afrique de l'Ouest à partir d'un seul centre de domestication.

(i) Le premier groupe d'étude plus ancien, se basant sur des données de marqueurs enzymatiques, situe la domestication du mil dans la zone constituée par le nord du fleuve Sénégal et le nord-ouest du Mali (Tostain, 1998). Cette hypothèse se base sur le fait que, la plus faible distance génétique entre mils sauvages et cultivés est observée dans cette zone (Tostain, 1992) et que la distribution géographique du type sauvage de mil est limitée à cette zone (Bezançon et al., 1997b). C'est à partir de ce foyer de domestication que *Pennisetum glaucum*

aurait évolué pour engendrer un groupe de mil (issu de la migration primaire) qui aurait migré secondairement vers trois grandes directions :

- une migration ancienne vers l’Afrique de l’Est et l’Inde ;
- une migration de mil tardif provenant du foyer secondaire de différenciation (de mil précoce en mil tardif) de mils issus du groupe primaire (à partir du foyer de domestication), situé autour du lac Tchad vers la zone soudanienne de l’Afrique de l’ouest et enfin ;
- une autre migration vers le centre et le sud de l’Afrique (Tostain, 1992).

(ii) Cependant, des études plus récentes utilisant des marqueurs microsatellites s’accordent sur une domestication qui aurait lieu dans une zone située entre le Mali et le Niger (Oumar et al., 2008 ; Dussert et al., 2015). Les travaux de Manning et al. (2011) ont montré, à partir de restes archéologiques, que la domestication du mil aurait déjà eu lieu il y a 4500 ans avec une diffusion rapide dans la région du Sahel. Ces données archéologiques sont en accord avec l’estimation d’une domestication datant de 4800 années (Clotault et al., 2012) et un centre d’origine de domestication unique (Dussert et al., 2015a ; Oumar et al., 2008). Néanmoins, deux centres de domestication avec un flux de gènes post-domestication important conduiraient au même signal évolutif qu’un centre unique de domestication, de sorte que l’hypothèse de deux centres ne peut être à ce jour entièrement écartée. Les études récentes de la structure génétique tendent à montrer l’existence de deux groupes génétiques de mil sauvage : l’un situé à l’Ouest (Sénégal-Mauritanie) et l’autre situé à l’Est (du Mali au Soudan) (Dussert et al., 2015a ; Oumar et al., 2008). C’est ce dernier qui aurait principalement contribué à la diversité du mil cultivé actuel. Cependant, Dussert et al. (2015) montrent des flux de gènes et donc une contribution néanmoins importante du groupe de mil sauvage de l’Ouest à la diversité du mil cultivé actuel. D’un point de vue de la structure génétique et de sa diversité, les études précédentes mettent en évidence un seul groupe génétique pour toute la région du Sahel qui, *a priori*, représenterait entre 70 - 90 % de la diversité sauvage (Dussert et al., 2015a ; Oumar et al., 2008). Cette théorie des études récentes ne prend pas en compte l’évolution du mil cultivé en mil précoce et tardif car, sur le plan génétique, il n’y aurait qu’un seul pool génétique de mil cultivé et non pas deux (Dussert et al., 2015).

En somme, le ou les centres d’origine de domestication d’une espèce cultivée ainsi que la date de la domestication sont des informations importantes qui permettent de comprendre en partie la diversité actuelle de l’espèce. Bien qu’il existe encore des incertitudes en ce qui concerne la

domestication du mil, le mil serait issu d'un seul centre de domestication et, *a priori*, d'un seul pool génétique sauvage, mais les flux de gènes importants au sein du compartiment cultivé ainsi qu'avec les deux pools sauvages font que le mil présente une diversité génétique neutre importante représentant jusqu'à 90 % de la diversité sauvage. Parallèlement, une différenciation tout du moins phénologique entre variétés précoces et tardives est visible mais cette différenciation ne semblerait pas correspondre à une réalité génétique (Dussert et al., 2015).

1.4. Diversité et adaptation du mil

Les études de diversité du mil tant antérieures (Marchais, 1982 ; Sarr and Pernès, 1988 ; Marchais et al., 1993 ; Tostain, 1992 ; Tostain and Marchais, 1993) que récentes (Hausmann et al., 2006 ; Kapila et al., 2008; Mariac et al., 2006a ; Stich et al., 2010) ont toujours fait état d'une grande diversité surtout à l'échelle régionale. Ces études qui ont aussi amélioré la connaissance de l'évolution de cette diversité à travers le temps (Chantereau et al., 2010) témoignent d'une absence d'érosion génétique de cette espèce au Niger.

En terme de structure de cette diversité, bien que génétiquement, à partir de marqueurs neutres (dont les allèles différents n'ont pas d'effet sur le phénotype), le mil cultivé s'organise en un groupe génétique lorsqu'il est comparé au mil sauvage (Dussert et al., 2015a ; Oumar et al., 2008), des variations phénotypiques importantes notamment phénologiques sont observées. En effet un gradient de précocité corrélé à la latitude comme chez de nombreuses espèces a été également identifié chez le mil (Hausmann et al., 2006). Dans la plupart des pays de la sous-région, y compris le Sénégal, deux grands groupes de mil sont reconnus par les paysans à savoir le mil précoce et le mil tardif. Néanmoins, Dussert et al. (2015) ont confirmé qu'à l'échelle de la sous-région il n'y avait pas de différence génétique entre les variétés de mil précoce et celles de mil tardif. Cependant, localement, une différenciation génétique neutre entre variétés précoces et tardives peut être observée comme au Niger (Lakis et al., 2012a). Il n'est pas clair à quel point cette différenciation du cycle précoce/cycle tardif est maintenue voir accentuée par les agriculteurs et l'effet de leur pratique sur la structuration de la diversité de ces variétés. Il est néanmoins fort probable que cette réduction du cycle soit une adaptation à des conditions environnementales plus arides et des pluies plus faibles (Kouressy et al., 2004). Les gènes mis en évidence par l'analyse des différences de précocité chez les mils cultivés sont essentiellement des gènes impliqués dans la régulation de la voie photopériodique de floraison. Des études menées sur 90 lignées originaires de l'Inde, de l'Afrique de l'Est et de l'Ouest ont permis de mettre en évidence des gènes associés à la précocité de la floraison (Saïdou et al., 2009). Ces études ont permis d'identifier un gène de la famille MADS-box qui code pour des

facteurs de transcription et qui joue un rôle majeur dans le développement végétatif et de la fleur (Mariac et al., 2011a). Cloutault et al., (2012) ont récemment identifié cinq gènes (*PgFY*, *PgGI*, *PgHD3a*, *PgPHYC* et *PgPRR73*) impliqués dans les voies de floraison qui sont associées à la domestication du mil. En comparant des échantillonnages des mêmes variétés de mil à 30 ans d'intervalle, Vigouroux et al. (2011) ont pu mettre en évidence une corrélation entre la réduction des cycles, et l'augmentation de la fréquence d'une mutation au niveau du gène *PgPHYC* au sein des variétés. Sachant que cet intervalle temporel de 30 ans correspond à une période de fort épisode de sécheresse, il est fort probable que ces modifications phénologiques et génétiques soient en partie une réponse d'adaptation aux conditions climatiques de cette période.

1.5. Les études sur le mil au Sénégal

Au Sénégal, le mil a été largement étudié dans beaucoup de domaines divers. Par exemple, sur le plan physiologique et culturale, des études des besoins en eau du mil au Sénégal (Dancette, 1980, 1983a) ont permis d'identifier les exigences hydriques du mil suivant différents cycles de floraisons. De même, une modélisation de ce bilan hydrique (Diop, 1999) du mil a montré que le taux de satisfaction des besoins en eau de la culture du mil (variant en fonction des régions suivant un gradient nord/sud), a diminué significativement après 1970 dans les régions du nord et du centre-nord du Sénégal. Les recherches sur l'apport du photopériodisme à l'adaptation du mil à la sécheresse ont montré que la plupart des accessions locales Souna et Sanio sont photosensibles (Kouakou et al., 2013) et la comparaison des indicateurs hydriques du rendement entre le mil photopériodique et non photopériodique, a révélé que les variétés de mil photopériodique ont de meilleurs indicateurs hydriques dans la moitié sud du Sénégal (Diop et al., 2005).

Par ailleurs, une caractérisation des réponses des racines du mil à faible teneur en phosphore (Faye et al., 2006) a montré que, sous déficit en phosphore, le mil présente une altération des paramètres radiculaires, par la formation de poils radiculaires et a identifié une variété (IBMV8402) comme présentant un meilleur comportement de croissance dans des conditions de stress hydrique indépendamment de la teneur en phosphore. Par contre, l'effet de la fertilisation à l'azote combinées à l'incorporation de paille sur la culture du mil (Ganry et al., 1978) a permis d'en déduire l'importance de la matière organique du sol pour l'efficacité de l'engrais azoté sur le mil.

Les études des facteurs biotiques sur la culture du mil ont déjà répertorié les espèces adventices dans le bassin arachidier, principale zone de production agricole du pays. Cette étude a permis d'enregistrer 125 espèces appartenant à 81 genres et 31 familles (Noba et al., 2004). La comparaison de 8 cultivars de mil homologués au cultivar Souna de Sénégal (Gahukari, 1990) par rapport à leur réaction à trois insectes ravageurs et deux maladies a déjà permis de recommander une variété (IBV-8001 qui a présenté plus de résistance à la chenille mineuse des épis, au mildiou et au charbon, avec des rendements plus élevés) dans les principales zones de production de mil au Sénégal.

Tandis que l'évaluation sur le plan morphologique de 234 accessions de mils précoces et de 9 accessions de mils tardives de mils du Sénégal (Sy et al., 2015) a permis de constater une grande diversité et une subdivision en trois groupes agro-morphologiques suivant le cycle végétatif. Toutes ces études ont exposé des résultats qui ont apporté de nouvelles connaissances sur le mil. Cependant, il est à noter l'absence d'évaluation du mil sur le plan génétique au Sénégal. Il est aussi à noter une évaluation sur le plan agro-morphologique principalement axée sur les accessions précoces et une quasi absence d'évaluation des accessions tardives du mil dans le pays. En complément à la grande connaissance du mil déjà apportée par les études précitées, il serait important de développer les études du mil du Sénégal, sur le plan génétique tout comme il serait important d'étendre l'évaluation agro-morphologique en plus de l'étude génétique sur les accessions tardives du mil du Sénégal. Ce qui serait important dans la gestion et la valorisation de ces mils.

1.6. Gestion des ressources phytogénétiques actuelles du mil

La gestion des ressources phytogénétiques actuelles du mil s'articule autour de la mise en place de collections nationales et internationales et de l'établissement de core collections représentatives. En effet à l'issue de la conférence de Rio en 1992 sur l'état des ressources naturelles, il a été souligné l'importance de conserver ces ressources génétiques. Cette conservation est un gage de sécurité contre les érosions génétiques et surtout devant les impacts négatifs des changements globaux. Elle est possible (cette conservation) à travers des banques de gènes qui gardent en collection des accessions d'espèces phytogénétiques. Le nombre de ces banques de gènes relativement faibles à l'époque poussait à tirer la sonnette d'alarme sur la nécessité de prendre des mesures urgentes pour progresser dans le développement des systèmes nationaux et internationaux de gestion des ressources phytogénétiques (Hammer, 2003). Ce qui a eu un impact positif car les années qui ont suivi ont vu le développement plus important de banques de gènes.

A ce jour, il y aurait au total 1750 banques de gènes à travers le monde (FAO, 2010) (Annexe1). Les données liées aux accessions et espèces conservées par ces banques de gènes, aux emplacements et contacts de ces banques de gènes, et toutes informations relatives à ces banques de gènes sont répertoriées dans une base de données gérée par la FAO du nom de World Information and Early Warning System on Plant Genetic Resources (WIEWS) (http://www.fao.org/wiews-archive/wiews.jsp?i_l=EN). Une décennie après la publication du premier rapport de la FAO sur l'état des ressources phytogénétiques pour l'alimentation et l'agriculture dans le monde en 1996, il a été créé la Svalbard Global Seed Vault (SGSV) qui est une sorte de chambre forte fournissant des installations afin de conserver un duplicata ou « échantillons de sécurité » d'accession provenant de banques de gènes du monde entier. En 2010, la SGSV conservait au total 412 446 entrées provenant de 204 pays dans le monde et couvrant 3286 espèces issues de 664 genres (FAO, 2010). En 2013 le nombre total d'entrées a été porté à 774 601 (Westengen et al., 2013).

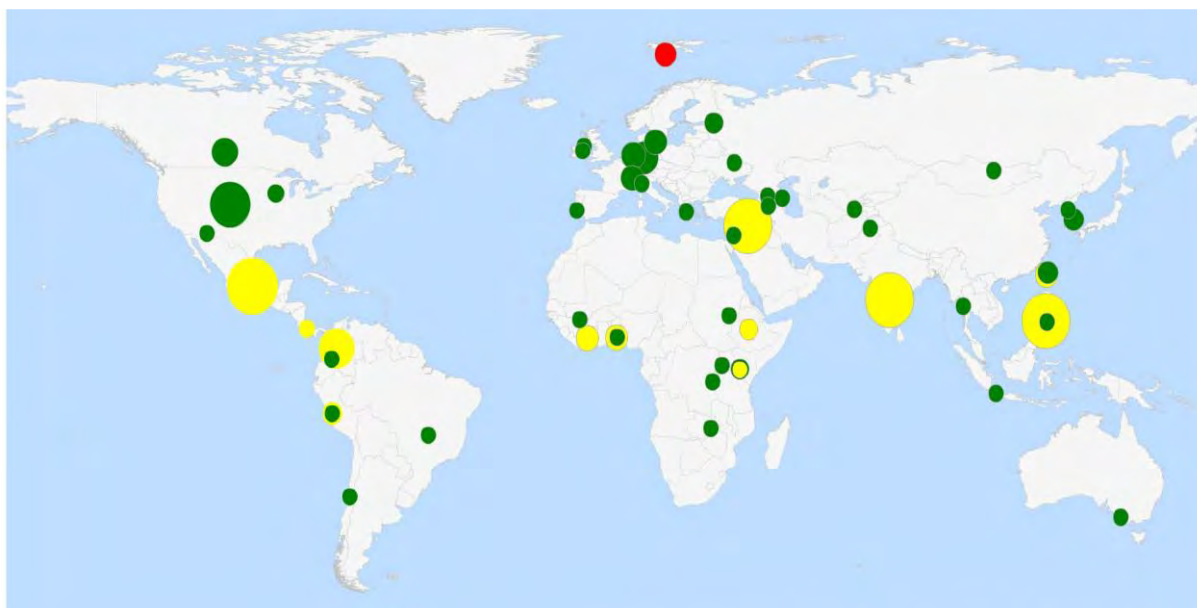


Figure 3 : Représentation des centres de recherches ou des banques de gènes ayant déposé un duplicata ou « échantillons de sécurité » de leurs accessions à SGSV (Svalbard Global Seed Vault) (représenté par le cercle rouge). Les cercles jaunes sont des centres internationaux de recherche agricole, et les cercles verts sont des banques de gènes régionales, nationales ou locales. Le rayon des cercles traduit le nombre d'échantillons déposés (Westengen et al., 2013).

A l'heure actuelle, il y aurait 65 447 accessions de mil conservées dans le monde dont les 33% sont conservées à l'ICRISAT (Institut International de Recherche sur les Cultures des Zones Tropicales Semiarides) qui est le détenteur de la plus grande collection de mil au monde suivi par la CNPMS de l'Institut Embrapa Milho e Sorgo qui en conserve 11% (FAO, 2010). Le reste des banques de gènes en conserverait 56%. De l'ensemble des 65447 accessions de mil conservées dans le monde, la SGSV en détiendrait 20 444 entrées en duplicata (Westengen et al., 2013).

Toutes ces collections ou banques de gènes qui conservent le mil, qu'ils aient ou non fourni de duplicata auprès de la SGSV, peuvent être classées en deux groupes :

(i) Les collections internationales qui renferment des accessions issues de plusieurs pays (ou même de plusieurs continents) qui ont l'avantage de permettre une conservation globale de la diversité du mil mais ayant souvent l'inconvénient de ne pas en garder la diversité fine. Parmi ces collections, nous pouvons citer celle de l'ICRISAT (code WIEWS : IND002) situé à Patancheru, Inde ; de la CNPMS de l'Institut Embrapa Milho e Sorgo (code WIEWS : BRA001) situé à Sete Lagoas au Brésil ; la Regional Station Jodhpur de la National Bureau of Plant Genetic Resources (NBPGR) (code WIEWS : IND064) située à Jodhpur en Inde et le Laboratoire des Ressources Génétiques et Amélioration des Plantes Tropicales de l'IRD (code WIEWS : FRA202) situé à Montpellier en France.

(ii) Les collections nationales qui renferment des accessions issues d'un seul pays. Ces collections ont souvent l'avantage de garder la diversité fine du mil ce qui très important compte tenu de la grande diversité de celui-ci mais présentent l'inconvénient d'en garder qu'au niveau d'une partie restreinte du globe. Parmi ces collection nous pouvons citer l'Unité de Recherche en Culture *In-vitro* (URCI) de l'ISRA (code WIEWS : SEN075) situé à Dakar au Sénégal, le NPGRC (National Plant Genetic Resources Centre) (code WIEWS : TZA016) situé à Arusha en Tanzanie, la NPGRC (National Plant Genetic Resources Centre) de la National Botanical Research Institute (code WIEWS : NAM006) situé à Windhoek en Namibie et la CENARGEN (Embrapa Recursos Genéticos e Biotecnologia) (code WIEWS : BRA003) situé à Brasília au Brésil.

Cependant, une bonne gestion et conservation de toutes ces ressources passe forcément par leur bonne connaissance à travers leurs évaluations tant sur le plan agro-morphologique que génétique. Car en effet, jusqu'en 2003, seulement 1/3 des accessions des banques de gènes dans le monde avait été complètement évaluées (Hammer et al., 2003). Il convient donc de s'assurer

d'une évaluation complète de toutes ces collections d'en ressortir les défauts s'il en existe et de les corriger. C'est uniquement par ce moyen que ces collections pourront être considérées comme de bon réservoir de la diversité du mil actuelle et des témoins de l'évolution la diversité du mil de demain.

2. Les nouvelles technologies de séquençage et leurs applications

La génétique et la génomique connaissent aujourd'hui une avancée considérable grâce aux nouvelles technologies de séquençage. L'identification et le suivi de la variation génétique sont maintenant efficaces et des milliers de variants peuvent être suivis au sein de grandes populations (Varshney et al., 2009). Ces nouvelles technologies (NGS) sont applicables dans différents domaines et permettent de résoudre de multiples problèmes (Varshney et al., 2009).

A titre d'exemples :

(ii) En biologie des populations, par la connaissance de la diversité et de la structure génétique des populations. Les données de marqueurs moléculaires de types SSRs, SNPs ou InDel, seuls permettent de faire des études d'introgression. Lorsqu'elles sont combinées aux données phénotypiques, elles permettent d'établir des cartes génétiques ou d'identifier des locis qui peuvent être utilisés soit pour la connaissance du génome ou pour des programmes de sélection et d'amélioration variétale.

(ii) L'utilisation de données de séquences exprimées du génome (ESTs) ou même de séquences complètes de génomes obtenus grâce aux NGS comparées à un génome de référence permet de connaître la structure et l'évolution des génomes des espèces.

(iii) L'utilisation de données d'expression de gènes combinées aux données agromorphologiques permet d'identifier des eQTLs, des marqueurs ou gènes candidats qui peuvent être utilisés pour des programmes de sélection et d'amélioration variétale.

La principale application de ces nouvelles technologies de séquençage sur le mil est le séquençage complet de son génome (Varshney et al., 2017), constituant un support de référence, qui rend désormais les analyses bio-informatiques des données issues de séquençages génomiques du mil, plus aisées.

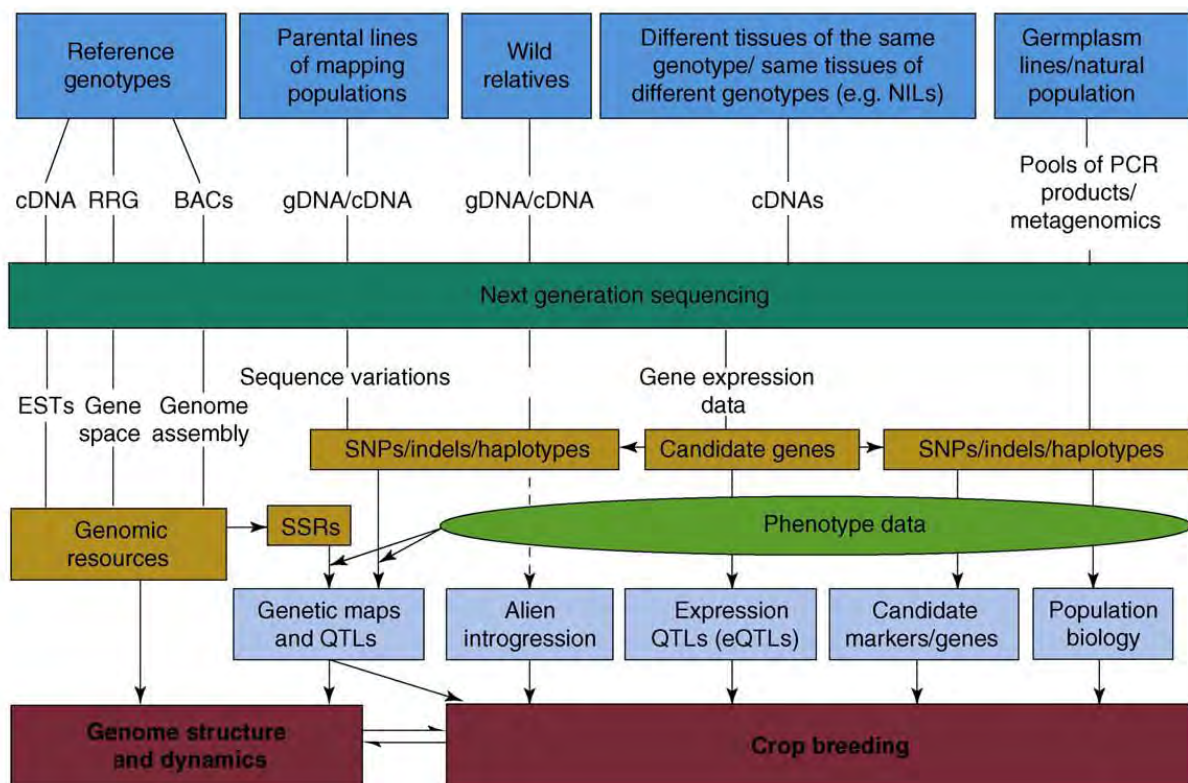


Figure 4 : Schéma montrant les différents domaines d'application et utilisations possibles des NGS (Next-generation sequencing) (Varshney et al., 2009)

L'analyse des données bioinformatiques générées grâce aux NGS a permis de découvrir de nombreux nouveaux gènes et des séquences régulatrices et de leurs positions. Elle a mis en disposition de grandes collections de marqueurs moléculaires comme les SNP (Perez-de-Castro et al., 2012). Ces marqueurs moléculaires ont déjà été utilisés pour valider ou réfuter des hypothèses (Johnson et al., 2009) sur l'évolution d'espèces. Elles ont permis de clarifier la structure de la diversité génétique dans un large éventail de cultures. Grâce à ces marqueurs, la description des matériels génétiques a augmenté la puissance des analyses sur les bases génétiques de la variation et de l'adaptation des espèces cultivées (Glaszmann et al., 2010). Ces ressources génomiques ont le potentiel d'accélérer la découverte des gènes et de lancer la sélection moléculaire chez les espèces cultivées sur lesquelles elles sont appliquées. Ceci améliorerait la productivité des cultures afin d'assurer la sécurité alimentaire dans les pays en développement (Varshney et al., 2010). Ces outils peuvent ainsi être appliqués à nos cas de figure pour apporter des réponses à nos questions.

Chapitre II : Evaluation des ressources phytogénétiques de mil du Sénégal



Chapitre II. Première partie : Evaluation génétique des ressources de mil du Sénégal

INTRODUCTION

Les ressources phytogénétiques sont le réservoir dans lequel on peut puiser pour faire face aux contextes de changements globaux actuels et assurer la sécurité alimentaire au Sahel. Pour ce faire, il convient d'avoir d'abord ces ressources à disposition et ensuite de bien les connaître. En effet, la connaissance des ressources génétiques est indispensable pour leur bonne gestion. C'est sur la base de cette connaissance que s'appuient les décisions dans le processus de conservation et c'est aussi sur la base de cette connaissance que s'appuient les acteurs de leur valorisation. Cette connaissance est possible grâce à une évaluation génétique permettant de connaître le niveau de diversité de ces ressources, de comprendre la structure de cette diversité et les facteurs qui expliquent cette structure, de constater la dynamique évolutive de cette diversité et plus encore.

A travers le monde, l'état de la diversité génétique du mil, la structure de cette diversité (Gepts and Clegg, 1989 ; Govindaraj et al., 2009 ; Mariac et al., 2006a), son évolution (Clotault et al., 2012 ; Dussert et al., 2015a ; Lakis et al., 2012a) et son adaptation ont été assez-bien étudiés. Ces études ont montré qu'au Niger et à l'échelle régionale de l'Afrique de l'Ouest, la sous-espèce cultivée du mil serait composée d'un groupe génétique unique. Or, dans la plupart des pays de la sous-région, y compris le Sénégal, les agriculteurs distinguent principalement deux grands groupes de mil, en fonction de la longueur du cycle de culture, appelés différemment selon les pays. Une variété précoce dont le temps de floraison varie de 50-60 jours (Dussert et al., 2015a), appelée "Souna" au Sénégal et au Mali, "Hainis" au Burkina "Haini Kire", "Guerguera", "Zongo", "Ba- Angouré ", " Ankoutess "et" Boudouma "au Niger (Bezançon et al., 1997a) et une variété tardive dont le temps de floraison varie de 80-110 jours (Dussert et al., 2015a) appelée "Sanio" au Sénégal et au Mali, "Kazouyas" au Burkina et "Maiwa" au Niger (Bezançon et al., 1997a). Néanmoins, Dussert et al. (2015) ont confirmé qu'à l'échelle de la sous-région il n'y avait pas de différenciation génétique entre les variétés de mil précoce et celles de mil tardif. Cependant, localement, une différenciation génétique neutre entre variétés précoces et tardives peut être observée au Niger, à l'échelle de villages (Lakis et al., 2012a).

Qu'en est-il au Sénégal ? En effet, il y a eu beaucoup d'études du mil au Sénégal (Angé, 1991 ; Dancette, 1983 ; Diakhaté et al., 2013 ; Diangar et al., 2004 ; Diop, 1999 ; Diop et al., 2005 ; Diouf, 1991 ; Kouakou et al., 2013 ; Mbaye, 1994 ; Noba et al., 2004 ; Sy et al., 2015). Cependant, presque aucune n'y a traité l'aspect génétique de cette céréale. Par conséquent, la diversité et la structure génétique du mil cultivé y est toujours inconnue.

Tandis qu'au de-là même de la simple connaissance de la diversité génétique, il a été établi au Niger, des évidences de sélection variétale associée à des variations climatiques qui suggèrent une possible adaptation du mil (Vigouroux et al., 2011b). Toute cette connaissance génétique du mil fait encore défaut au Sénégal. Des études adaptatives génétiques du mil n'ont pas été réalisées et il y a encore très peu d'évaluations génétiques neutres sur sa diversité au Sénégal. Pourtant, ces connaissances sont indispensables pour une bonne gestion et une meilleure exploitation de ces ressources à travers les programmes d'améliorations.

Objectif

Le but de cette partie est de répondre à tous ces questionnements. De manière globale, cette partie a pour objectif, de contribuer à une meilleure connaissance des ressources de mil. De façon spécifique, cette partie vise à :

- Etoffer la collection de mil du Sénégal ;
- Apprécier le niveau et la structure de diversité génétique des mils cultivés du Sénégal ;
- Evaluer la diversité adaptative de ces mils cultivés.

A terme, cette partie pourra apporter des explications à beaucoup de questions jusqu'ici restées sans réponses. Elle pourra fournir des connaissances d'un intérêt fondamental pour la bonne gestion et valorisation de ces ressources du Sénégal.

MATERIEL ET METHODES

1. Prospection et collectes

Dans le but d'avoir une meilleure représentativité de toutes les zones de culture du mil au Sénégal, et pour obtenir un effectif important pour chaque type variétal, il a été réalisé, des prospections dans les zones de production du mil, tout en y associant des enquêtes ethnobotaniques. Ces prospections ont été faites comme suit : Il a été prospecté le maximum de villages possibles en se focalisant sur les villages les plus reculés qui ne se trouvent pas près des routes et des marchés (qui favorisent les échanges) tout en respectant un pas de 20km. Une fois sur place, la diversité de chaque village a été collectée (Voire Annexe2 pour les données passeport).

2. Evaluation génétique neutre (marqueurs SSR)

2.1. Culture des plantes

Environ cinq (5) graines par accession ont été semées dans des gaines (une gaine par accession), contenant du sol « Dior ». Ces gaines ont été placées en serre pendant 15 jours et quotidiennement arrosées avec de l'eau de robinet, jusqu'à obtention de jeunes plants.

2.2. Extraction et quantification de l'ADN

Un prélèvement foliaire d'environ 200 g sur chaque accession (une plante par accession) a été effectué afin d'en extraire l'ADN sur la base de la méthode décrite par (Doyle, 1987) modifiée. Les feuilles ont été broyées au pilon dans des mortiers contenant 750 µL d'une solution tampon de lyse à CTAB 2%. Les broyats ont ensuite été incubés à 74°C pendant 20 min puis refroidis à température ambiante pendant 5 min. Un volume équivalent (750 µL) de chloroforme iso-amylque 24:1 a été ajouté puis l'ensemble homogénéisé et centrifugé à 10 000 rpm à 20°C pendant 2 min. Après centrifugation, environ 500 µl du surnageant ont été récupérés puis 500 µL d'isopropanol et 50 µL d'acétate de sodium 3 M à pH5 ont été rajoutés, et incubés à -20°C pendant 15 min. Une centrifugation à 10 000 rpm à 10°C pendant 10 min a été réalisée et le surnageant éliminé. Le culot d'ADN a été lavé avec 500 µL d'éthanol à 70°, puis centrifugé à nouveau à 10 000 rpm à 10°C pendant 10 min. L'ADN a été séché à la température ambiante pendant 15 heures de temps, puis dissout dans 100 µL d'une solution de Tris 10 mM - EDTA 1 mM.

Après l'extraction, pour vérifier la qualité de l'ADN, les échantillons ont été mis à migrer sur un gel d'agarose à 0,8% avec 6 µl d'ADN et 2 µl de bleu de bromophénol par échantillon. Le

gel a ensuite été coloré au Bromure d'éthidium (BET) (1 mg/L) pendant 15 min avant de le visualiser aux rayons UV. L'ADN a été quantifié au Nano Drop (Thermo) en déposant 1 µl de l'extrait. Les degrés de contamination en protéines et composés phénoliques ont été estimés à l'aide du même appareil.

Les solutions d'ADN ont été enfin diluées à 25 ng/µl (concentration de travail) ou conservées à -20°C.

2.3. Génotypage

2.3.1. Choix des marqueurs

Le génotypage neutre de la collection a été réalisé à l'aide de 15 marqueurs microsatellites (SSRs) (Tableau 2, Annexe 2) qui ont été utilisés auparavant dans plusieurs autres études de diversité génétiques sur le mil (Allouis et al., 2001 ; Budak et al., 2003 ; Mariac et al., 2006a ; Qi et al., 2001).

Réactions PCR et assignation des génotypes

L'amplification a été réalisée à l'aide d'un Kit Qiagen contenant 2 µL de solution Q 5X, 5 µL de master mix 2X, et du H₂O (RNase-free water) où on a ajouté 1.6 µL ou 1.4 µL de mix d'amorces (selon qu'il s'agisse du multiplex A ou du multiplex B, respectivement) et 1 µL d'ADN (25ng/µl) pour un volume total de 10 µl pour chaque puits. L'amplification a été réalisée à l'aide d'un thermocycleur TC-Plus (TECHNE). Les réactions d'amplifications des 15 marqueurs ont été multiplexées en deux PCR (Multiplex A et B ; 8 marqueurs pour le multiplex A et 7 marqueurs pour le multiplex B). Le programme PCR suivant a été utilisé : 95°C pendant 15 min pour le déblocage de la polymérase. Suivi de 35 cycles de dénaturation à 94°C pendant 30 s, hybridation des amorces à 55°C pendant 90 s, une élongation à 72°C pendant 60 s et une extension finale à 60°C pendant 30 min.

Les fragments amplifiés par PCR ont été détectés sur un séquenceur ABI3130 (Applied Biosystems) à la plateforme Génopole de Montpellier et les données de séquence ont été analysées avec le logiciel GeneMapper 3.7, qui a permis d'assigner les génotypes. L'assignation des pics correspondants aux allèles a été déterminée avec un standard de taille (LIZ 500). Les données d'assignation ont été exportées sous forme de matrice des individus en ligne avec leurs génotypes respectifs au niveau de chaque marqueur (en colonnes). Tous les individus et marqueurs présentant plus de 5% de données manquantes ont été supprimés afin d'éviter tout biais dans l'analyse des données. C'est sur la base de cette matrice que toutes les analyses statistiques ont été faites.

Analyses du niveau de diversité neutre des mils cultivés du Sénégal

Pour évaluer le niveau de diversité, il a été considéré l'indice de diversité de Nei (H_{EX}) et la richesse allélique tout en tenant compte de la taille de chaque population considérée. Ainsi donc, ont été calculées, la richesse allélique (RA), l'hétérozygotie Observée (H_O), l'hétérozygotie attendue (H_{EX}), pour chaque locus et dans chaque population considérée grâce au logiciel Genalex 6.5 (Peakall and Smouse, 2012). L'identité population a été considérée en fonction du type variétal. Il a aussi été calculé l'indice de différenciation entre populations (F_{ST}). Nous avons aussi vérifié l'équilibre d'Hardy et Weinberg (HWE) pour chaque locus.

L'hétérozygotie observée et l'hétérozygotie attendue ont permis d'obtenir l'indice de fixation (F_{IS}) ou coefficient de consanguinité qui reflète la différence des individus à l'intérieur des populations.

Afin d'avoir une comparaison du niveau de diversité génétique entre types variétaux, nous avons fait un test t de Student des moyennes de l'indice de diversité de Nei, par rapport aux différents locus de chaque population.

La richesse allélique étant dépendante de la taille de la population et n'ayant pas le même nombre d'individus dans chaque population considérée (le nombre d'accessions des variétés précoces, tardives et améliorées est différent), nous avons estimé la richesse allélique corrigée, par la méthode de raréfaction (Hurlbert, 1971), avec une taille minimale d'échantillonnage égale à sept (7), à l'aide de l'algorithme implémenté dans le programme ADZE version 1.0 (Szpiech et al., 2008) pour comparer la richesse allélique des différentes variétés grâce aux formules suivantes :

$$\alpha_g^{(j)} = \sum_{i=1}^I P_{ijg} \quad , \quad \left(P_{ijg} = 1 - \left[\frac{g}{\binom{N_j}{g}} \right] \right)$$

D'où, P_{ijg} est la probabilité de trouver au moins une copie de l'allèle i dans un échantillon de taille g de la population j . I représente les allèles distincts et N_{ij} , le nombre de copies de l'allèle i dans un échantillon de la population j .

Analyses de structure de diversité neutre des mils cultivés du Sénégal

L'analyse de la structure génétique des mils cultivés du Sénégal a été réalisée d'abord en composantes principales (ACP) à l'aide du package ade4 (Thioulouse et al., 1997) implémenté dans le logiciel R (R Development Core Team, 2008). La structure de la population a également été étudiée par approche bayésienne à l'aide du logiciel STRUCTURE 2.3.3 (Pritchard et al., 2000). L'analyse a été réalisée suivant le modèle d'admixture (Falush et al., 2003) avec des populations K allant de 1 à 12 (K étant le nombre de groupes simulés). Une période de burns de 500 000 et 10 runs de 700 000 itérations de tirages et remises par la méthode Chaines de Markov Monte Carlo (MCMC) pour chaque valeur de K a été utilisée. Les valeurs du nombre de clusters (K) ont été évaluées selon la méthode d'Evanno et al. (Evanno et al., 2005) par le critère (ΔK) et le logarithme de vraisemblance ($\ln P(D|K)$).

Afin de voir les relations entre accessions de la collection, il a été réalisé un dendrogramme, sur la base de la matrice d'analyse, par la méthode Neighbor Joining (Saitou and Nei, 1987) avec 1000 répliques à l'aide du logiciel Darwin 5.0.158 (Perrier et al., 1999).

Analyses spatiales de diversité neutre des mils cultivés du Sénégal

L'analyse spatiale n'a été réalisée que sur 367 accessions dont 281 variétés précoces et 86 variétés tardives. Car seules ces accessions présentaient des coordonnées GPS disponibles.

Une Analyse Spatiale des Composantes Principales (sPCA) pour étudier les profils spatiaux de la variabilité génétique en utilisant le package adegenet (Jombart, 2008) implémenté dans le logiciel R (R Development Core Team, 2008) a été utilisée. En choisissant comme réseau de connexion, la triangulation Delaunay (type 1), et en corrigeant les coordonnées géographiques doubles à l'aide de la fonction *jitter*. Les scores de l'Analyses Spatiales des Composantes Principales peuvent être informatifs à deux échelles : les structures globales (c'est-à-dire à grandes échelles) affichent une autocorrélation spatiale positive entre les individus, tandis que les structures locales (c'est-à-dire à échelles fines) affichent une autocorrélation spatiale négative.

L'isolement par la distance (IBD) a été vérifié grâce à un test de Mantel (Mantel, 1967) entre la matrice de distance génétique et la matrice de distance géographique suivant les méthodes de Smouse et collaborateurs (Smouse and Long, 1992 ; Smouse et al., 1986) avec un test de permutation de 999 itérations. La pente de régression a été représentée pour les accessions précoces d'une part et tardives d'autre part. Cette analyse a été réalisée à l'aide du logiciel GenAlex 6.5 (Peakall and Smouse, 2012).

Vérification de redondances de génotypes

Afin de tester la présence de doublons au sein de la collection, un calcul des probabilités d'identité des individus de la collection à l'aide du logiciel Genalex 6.5 (Peakall and Smouse, 2012) a été réalisé. En plus, il a été fait un alignement comparatif des génotypes des individus par rapport aux différents marqueurs microsatellites, à l'aide du logiciel Identity v 1.0 (Wagner and Sefc, 1999).

3. Evaluation génétique adaptative

3.1. Génotypage

3.1.1. Gènes d'adaptation

Amplification de l'ADN

L'amplification de l'ADN a été réalisée à l'aide d'un Kit Promega contenant une solution Tampon 5X, de la GoTaq G2 DNA Polymerase 2X, et du H₂O (RNase-free water) où sont ajoutés 1 µl de DNTP (2 mM), 0.2 µl d'amorce (Annexe 3) et 2 µl d'ADN (50 ng) pour un volume total de 10 µl pour chaque tube. Les PCR ont été réalisées avec un thermocycleur TC-Plus (TECHNE). Les amorces utilisées, de ces deux gènes, se réfèrent à celles de Mariac et al., (2011b) pour le gène *PgMADSII* et à Saïdou et al. (2009) pour le gène *PgPHYC*.

Pour le gène *PgPHYC*, le programme PCR suivant a été utilisé : 94°C pendant 3 min pour une dénaturation initiale. Suivi de 38 cycles comprenant une dénaturation à 94°C pendant 30 s, une hybridation à 55°C pendant 30 s, une élongation à 72°C pendant 30 s et une extension finale à 72°C pendant 10 min.

Et pour le gène *PgMADSII*, le programme PCR suivant a été utilisé : 94°C pendant 3 min pour une dénaturation initiale. Suivi de 38 cycles comprenant une dénaturation à 94°C pendant 30 s, une hybridation à 57°C pendant 30 s, une élongation à 72°C pendant 30 s et une extension finale à 72°C pendant 10 min.

Les fragments amplifiés ont été détectés par électrophorèse : 7 µl de produit PCR et 2 µl de bleu de bromophénol (fabricant SIGMA) ont été mis à migrer (dans un tampon contenant du TBE 1X) par électrophorèse pendant 1 h, sur gel d'agarose à 2% avec un marqueur de taille 1 Kb Ladder. Les gels ont ensuite été colorés au Bromure d'éthidium (BET, 1mg/L) pendant 15 min avant de les visualiser aux rayons UV (100nm - 400nm) à l'aide d'un Geldoc.

3.2. Analyses de la variabilité allélique par rapport aux gènes

***PgPHYC* et *PgMADS11* et études de corrélations**

Pour tester d'éventuelles corrélations de ces gènes par rapport à des données génétiques et géo-climatiques il a été procédé à des analyses de régressions logistiques entre les géotypes par rapport aux deux gènes et les pools génétiques, l'altitude, la latitude, la longitude, les températures et précipitations moyennes des mois de juin, juillet, août et septembre et 19 facteurs géo-climatiques. Les données géo-climatiques ont été obtenues à partir de la base de données WorldClim (<http://www.worldclim.org>). L'extraction des données a été réalisée à l'aide du logiciel DIVA-GIS v7.5. Toutes les analyses ont été faites à l'aide du logiciel R 3.3.1 (R Development Core Team, 2008).

Les 19 facteurs géo-climatiques sont : la température moyenne annuelle, l'écart diurne moyen, l'isothermalité, la saisonnalité de la température, la température maximale de la période la plus chaude, la température minimale de la période la plus froide, l'écart annuel de température, la température moyenne du trimestre le plus humide, la température moyenne du trimestre le plus sec, la température moyenne du trimestre le plus chaud, la température moyenne du trimestre le plus froid, les précipitations annuelles, les précipitations de la période la plus humide, les précipitations de la période la plus sèche, la saisonnalité des précipitations, les précipitations du trimestre le plus humide, les précipitations du trimestre le plus sec, les précipitations du trimestre le plus chaud, et les précipitations du trimestre le plus froid.

RESULTATS

1. Enrichissement de la collection de mil du Sénégal

Au total, cent soixante-cinq (165) villages ont été prospectés et deux cent vingt-six (226) accessions collectées. Ces accessions nouvellement collectées, ont été jointes à l'ancienne collection (Sy et al., 2015), ce qui a constitué la nouvelle collection de mil du Sénégal. Cette nouvelle collection est donc constituée de quatre cent soixante-dix-sept (477) accessions avec cent douze (112) variétés tardives, trois cent cinquante-trois (353) variétés précoces et 12 variétés améliorées. Désormais, nous avons une bonne couverture des régions, car cette nouvelle collection couvre toutes les régions où le mil est cultivé au Sénégal. En plus, nous avons une plus grande représentativité des variétés tardives (23%) dans la collection, comme le montre la Fig. 5.

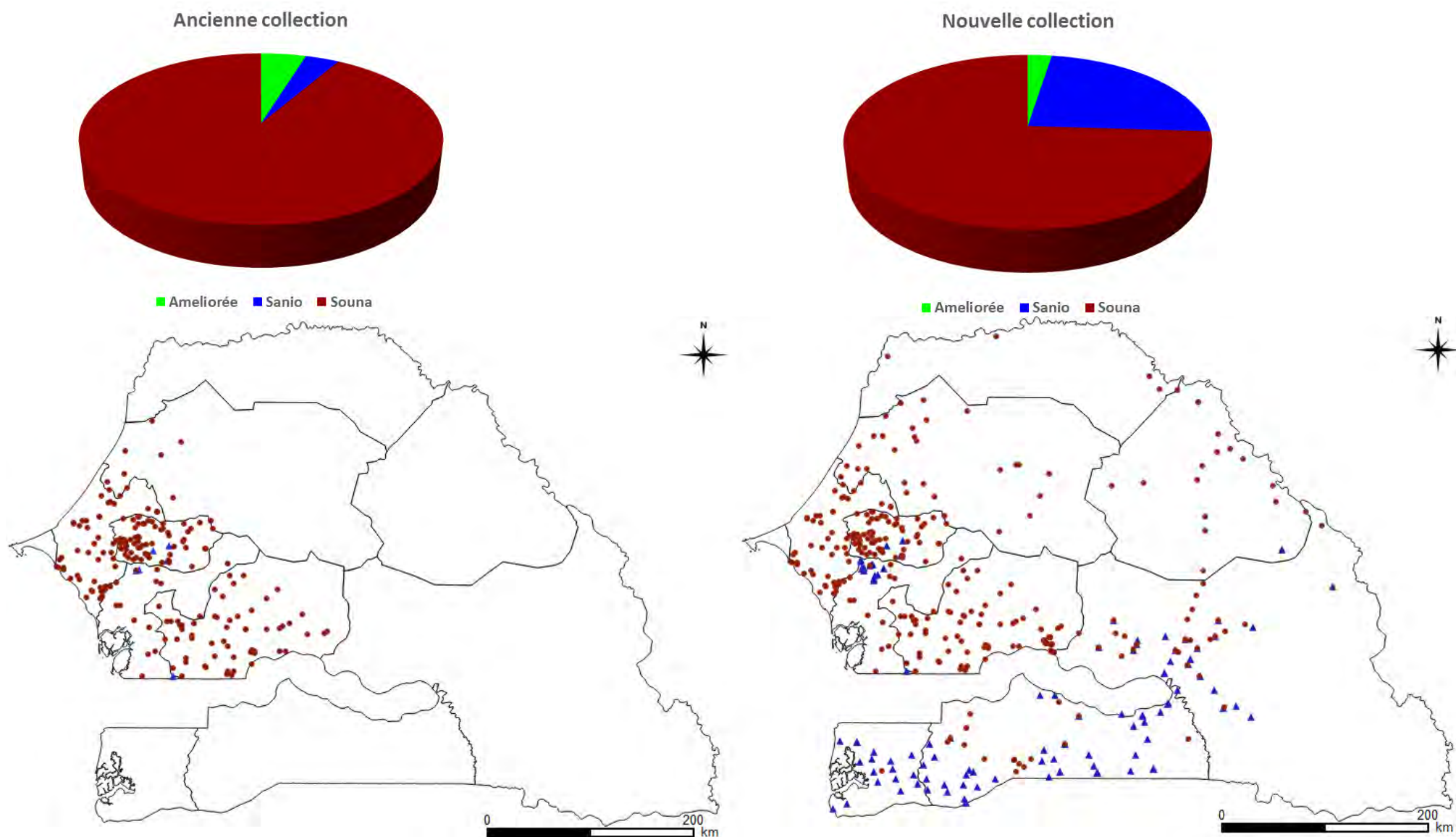


Figure 5 : Comparaison de la représentativité variétale et géographique de la nouvelle collection par rapport à l'ancienne collection de mil du Sénégal.

Cette collection est actuellement conservée sous forme de grains au Centre national de recherche agronomique de Bambey (CNRA) de l'Institut Sénégalais de Recherches Agricoles (ISRA) et au Laboratoire National de Recherches sur les Productions Végétales (LNRPV) de l'ISRA à Dakar. Une base de données en ligne nommée GENEDIV (www.genediv.com) a été créée (dans le cadre du projet WAAPP/PPAAO), dans laquelle toutes les données (génétiques, agro-morphologiques et géographiques) concernant cette collection, ont été déposées, pour permettre une meilleure valorisation de ce matériel génétique.

2. Evaluation génétique neutre

2.1. Niveau de diversité neutre des mils cultivés du Sénégal

Le tableau de données final contient 1,77% des données manquantes, avec au total 404 accessions dont 12 variétés améliorées, 306 Souna et 86 Sanio et 12 marqueurs microsatellites. Tous les 12 microsatellites ont été polymorphes. Les marqueurs ont présenté un total de 101 allèles avec en moyenne 8,4 allèles par locus. Le nombre d'allèles par marqueur varie de 2 (*Psm2202*) à 14 (*Psm2220*).

Le tableau 2 présente l'hétérozygotie observée et attendue, le nombre d'allèles, la richesse allélique corrigée avec une taille minimale d'échantillonnage égale à 7 et le F_{IS} pour les accessions locales précoces et tardives. Nous observons un niveau élevé de diversité génétique pour toutes les accessions locales. Les variétés précoces présentent un niveau de diversité génétique légèrement plus élevé par rapport à l'hétérozygotie observée H_O (0,481) et attendue H_{EX} (0,567), tandis que les variétés améliorées montrent le plus faible H_O (0,391). Cependant, les tests statistiques de la diversité montrent un niveau de diversité similaire entre les accessions locales précoces et tardives. Le test t de Student entre les accessions locales précoces et tardives n'indique pas de différences significatives pour l'hétérozygotie observée (P-valeur = 0.8839, t = 0.14782), l'hétérozygotie attendue (valeur P = 0.772, t = 0.29381), le F_{IS} (valeur P = 0.9459, t = 0.068834) et la richesse allélique corrigée (valeur P = 0.9008, T = 0.12787), ce qui suggère que le niveau de diversité génétique est égal.

Un seul locus (*Psm2249*) se trouve être en déséquilibre d'HWE chez les accessions précoces, contre deux loci (*Psm2249* et *Psm2246*) chez les accessions tardives. Les groupes présentent de faibles taux de consanguinité avec une valeur F_{IS} de 0,160, 0,128 et 0,258 respectivement pour les accessions locales précoces, tardives et les variétés améliorées. Les accessions locales précoces, tardives et les variétés améliorées présentent respectivement 22 allèles (avec une moyenne de fréquences alléliques de 0.01), 8 allèles privés (avec une moyenne de fréquences

alléliques de 0.015) et 1 allèle privé (avec une moyenne de fréquences alléliques de 0.09). De faibles différentiations (F_{ST}) ont été notées entre les variétés améliorées et les accessions locales précoces (0.004), entre les accessions locales précoces et tardives (0.052) et entre les accessions locales tardives et les variétés améliorées (0.063).

Tableau 1 : Statistiques des estimateurs de la diversité génétique pour les 12 SSR utilisés dans l'analyse des 404 accessions de mil.

Locus	Tailles allèles	Na	H _{EX}	H _O	F _{IT}	F _{ST}	F _{IS}
<i>PsmP2202</i>	146-163	2	0.495	0.470	0.167	0.196	-0.037
<i>PsmP2206</i>	197-213	9	0.793	0.710	0.133	0.051	0.086
<i>PsmP2214</i>	234-244	10	0.666	0.504	0.254	0.006	0.250
<i>PsmP2219</i>	227-291	12	0.612	0.591	0.086	0.082	0.004
<i>PsmP2220</i>	118-149	14	0.868	0.789	0.112	0.033	0.081
<i>PsmP2227</i>	192-246	7	0.654	0.565	0.161	0.039	0.126
<i>PsmP2237</i>	233-265	9	0.318	0.236	0.408	0.171	0.286
<i>PsmP2247</i>	199-206	4	0.523	0.383	0.485	0.473	0.023
<i>PsmP2249</i>	130-155	3	0.388	0.154	0.608	0.011	0.604
<i>PgIRD12</i>	121-147	13	0.637	0.290	0.605	0.114	0.554
<i>PgIRD25</i>	160-184	11	0.494	0.376	0.400	0.340	0.091
<i>PgIRD46</i>	84-107	7	0.588	0.552	0.112	0.086	0.028

Tableau 2 : Statistiques de la diversité génétique pour les variétés améliorées, les accessions locales précoces et tardives. Nb : nombre d'entrées ; Na : nombre moyen d'allèles, R_{Ac} : Richesse allélique corrigée avec une taille minimale d'échantillonnage égale à 7, H_O : hétérozygotie observée ; H_{EX} : hétérozygotie attendue ; F_{IS} : coefficient de consanguinité

	Nb	Na	R _{Ac} (Taille =7)	H _O	H _{EX}	F _{IS}
Variétés améliorées	12	4,4	2,55	0,350	0,513	0,258
Accessions locales précoces	306	7,7	2,81	0,446	0,542	0,160
Accessions locales tardives	86	5,8	2,77	0,434	0,494	0,128

2.2. Structure de diversité neutre des mils cultivés du Sénégal

Les analyses bayésiennes montrent une structure claire entre les accessions locales précoces et tardive (Fig. 7). La valeur du critère delta K (ΔK) d'Evanno utilisé, est plus élevée pour $K = 2$ (Fig. 6). Cela indique que le nombre le plus probable de groupes (K) identifiés par STRUCTURE est $K = 2$. La plupart des accessions (89% des accessions précoces et 90% des accessions tardives) ont été attribuées à leurs clusters respectifs avec une valeur $q \geq 0,7$ (q est le degré d'appartenance à un cluster. Il varie entre 0 et 1. Plus q est proche de 1 pour un cluster donné, plus l'accession considérée, appartient au cluster en question). Trente-quatre (34) accessions précoces et neuf (9) accessions tardives ont été mal assignées ($q < 0,7$), mais chacune a présenté une part du génome intermédiaire. Considération faite des 10 villages au niveau desquels des accessions précoces et tardives ont été à la fois collectées, toutes les accessions ont été assignées à leur groupe respectif avec des valeurs de q supérieures à 0,75.

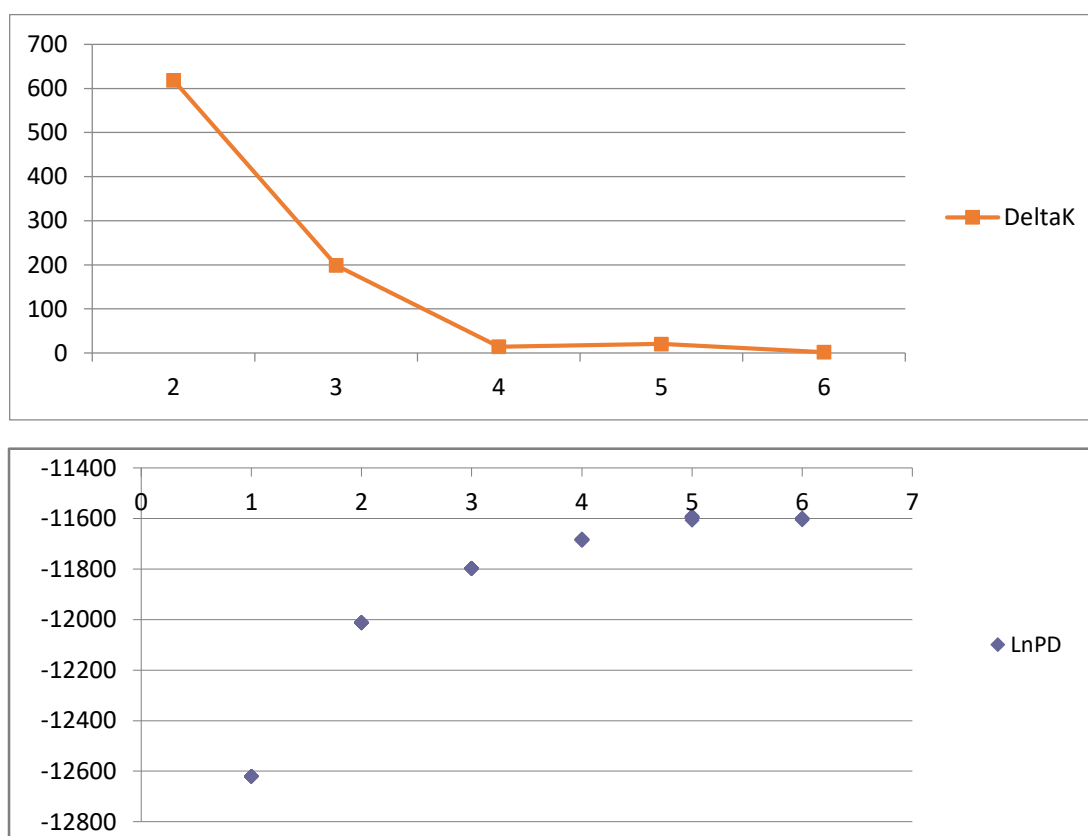


Figure 6: Valeurs de delta K (courbe rouge) et la courbe de régression de LnPD (courbe bleue) pour $K = 1$ à $K = 6$ (nombre de K groupes) selon la méthode de calcul d'Evanno et al (2005) Ces résultats ont été obtenus pour 404 accessions analysées avec 12 marqueurs SSR.

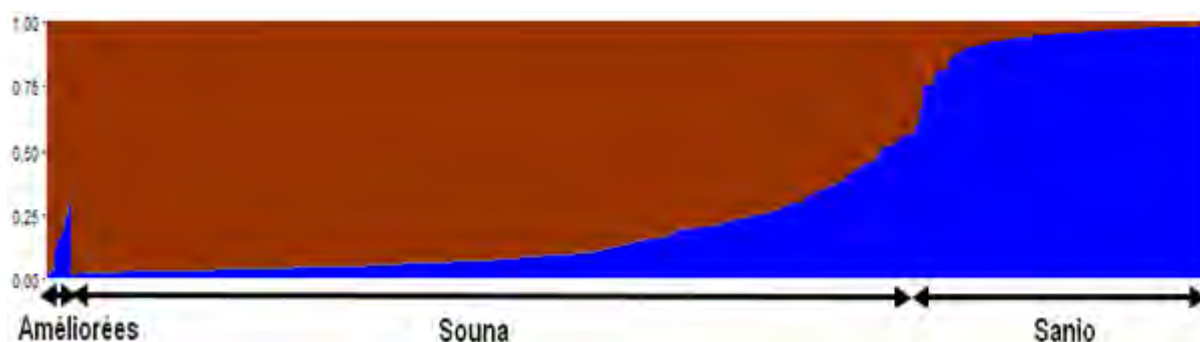


Figure 7 : Structure de la population pour $K = 2$ sur la base des 404 accessions analysées avec 12 SSR.

Des résultats similaires ont été observés avec l'Analyse en Composantes Principales, montrant une distinction claire entre les accessions locales précoces et tardives (Fig. 8). Les deux premières composantes principales expliquent respectivement 5,2% et 2,8% de l'inertie. Les marqueurs *PsmP2247* et *PsmP2202* ont eu la contribution la plus élevée au premier axe de l'ACP. En effet, deux allèles, *PsmP2247_199* et *PsmP2202_146*, ont montré des fréquences très élevées chez les accessions locales tardives (0,92 et 0,83, respectivement). Toutefois, l'élimination de ces deux loci ne modifie pas les résultats de l'ACP.

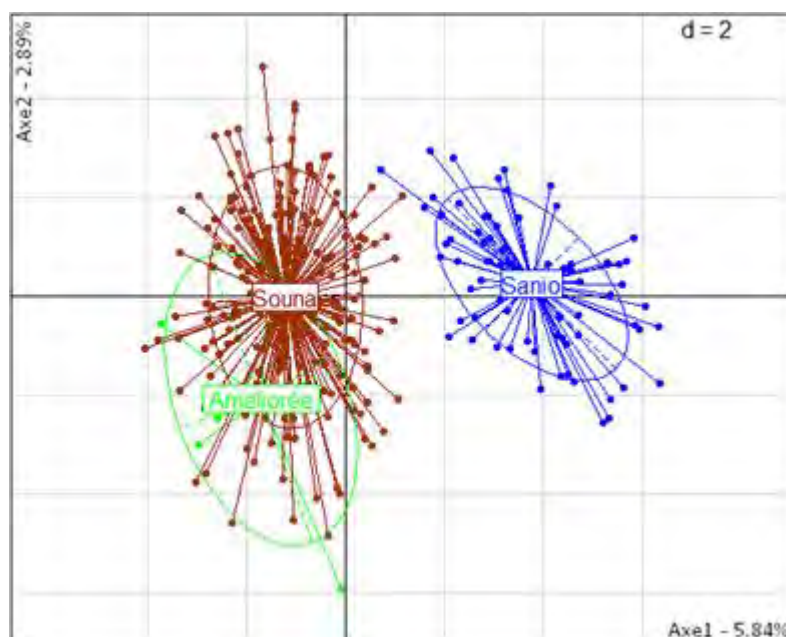


Figure 8: Analyse en composantes principales des variétés améliorées, des accessions précoces (Souna) et tardives (Sanio) de la collection (Axe1= 5.2, Axe2= 2.8).

Le dendrogramme (Fig. 9) présente les mêmes observations faites avec l'ACP et le l'analyse bayésienne avec d'un côté les précoces et améliorées et de l'autre, les variétés tardives.

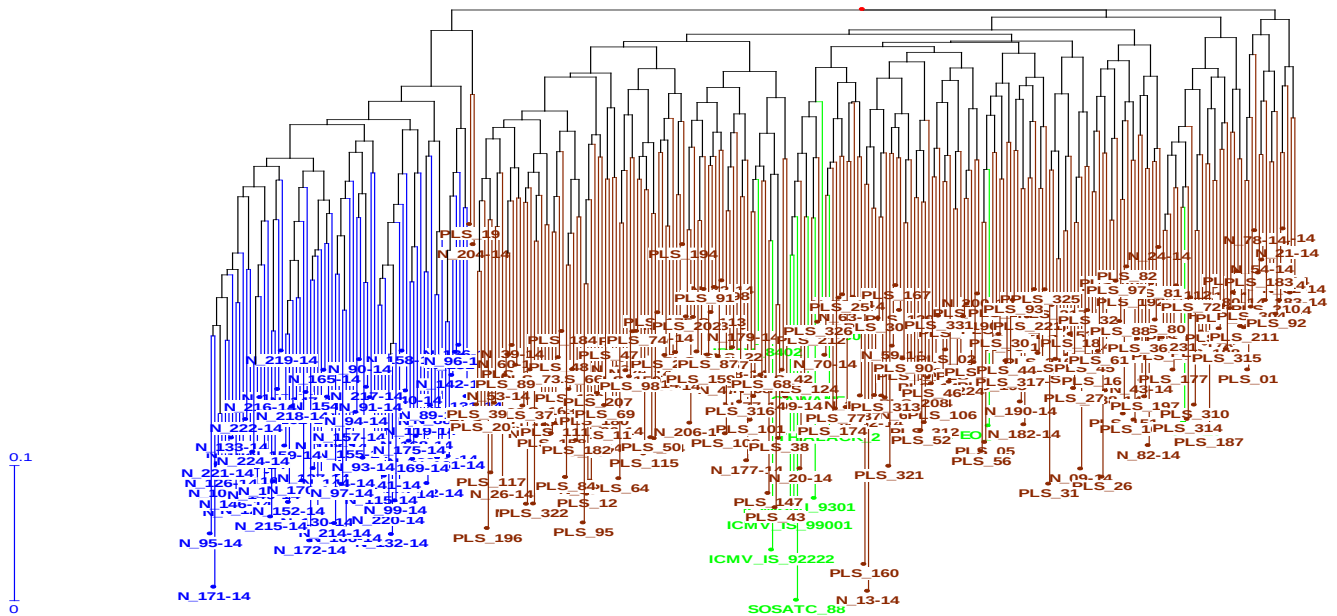


Figure 9 : Degrés de similarités entre les différentes accessions (précoces, tardives et variétés améliorées) de la collection de mil. Vert : variétés améliorées ; Rouge : accessions précoces ; Bleu : accessions tardives

A travers toutes ces analyses, aucune différenciation génétique entre les variétés améliorées et les accessions précoces n'a été observée, ce qui prouve que ces variétés améliorées provenaient des variétés précoces.

2.3. Analyses spatiales de diversité neutre des mils cultivés du Sénégal

L'analyse spatiale des composantes principales (sPCA) met en évidence, l'existence d'un pattern spatial de la diversité génétique des mils du Sénégal. Elle montre des résultats similaires aux analyses précédentes, confirmant les deux groupes génétiques sur la base des accessions locales précoces et tardives et la répartition spatiale nord-sud à travers le Sénégal. La valeur absolue propre du premier axe représentant la structure globale est considérablement supérieure à la somme des deux premiers axes représentant la structure locale (Fig. 10). Le test de permutation Monte Carlo pour 999 itérations, des structures locales (Geary) et globales (Moran)

indiquent une structure globale significative ($p\text{-value} = 0,0001$) et une structure locale non-significative ($p\text{-valeur} = 0,77$). Compte tenu de ces observations, la visualisation de cette analyse a été faite sur le premier axe (de la structure globale), pour lequel, I de Morran = 0,50. La projection des scores d'autocorrélations (valeurs du premier axe représentant la structure globale, ou à grande échelle spatiale) montre l'existence d'un motif spatial divisé en deux groupes ; Un groupe prédominant au sud et un autre prédominant au nord et au central-nord du Sénégal (Fig. 10).

L'analyse comparée des résultats de la sPCA par rapport à ceux de STRUCTURE indique que l'analyse spatiale des composantes principales (sPCA) a révélé une structure génétique plus cachée (Fig. 11). En effet, si l'on considère l'information spatiale, les accessions locales précoces du sud du Sénégal ont présenté un profil plus atténué que celui observé avec STRUCTURE, alors qu'aucune différence n'a été observée en considérant les accessions locales tardives. Les deux approches confirment bien les attributions génétiques des variétés locales tardives du centre Sénégal.

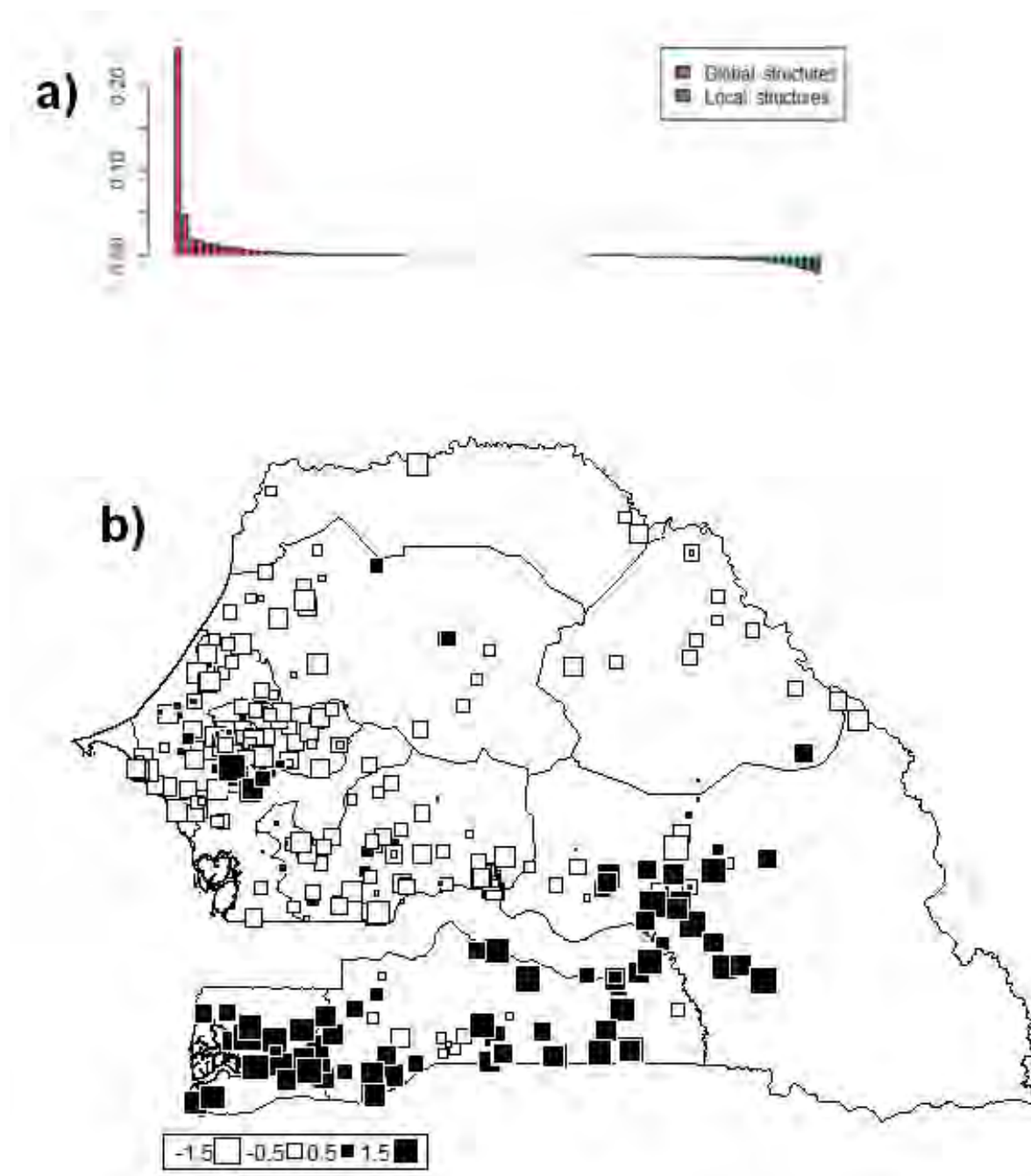


Figure 10 : l'Analyse Spatiale des composantes principales (sPCA), des accessions précoces et tardives de la collection. a) Représentation graphique des axes de la structure globale et locale. b) Représentation graphique des scores d'autocorrélations, du premier axe de la sPCA.

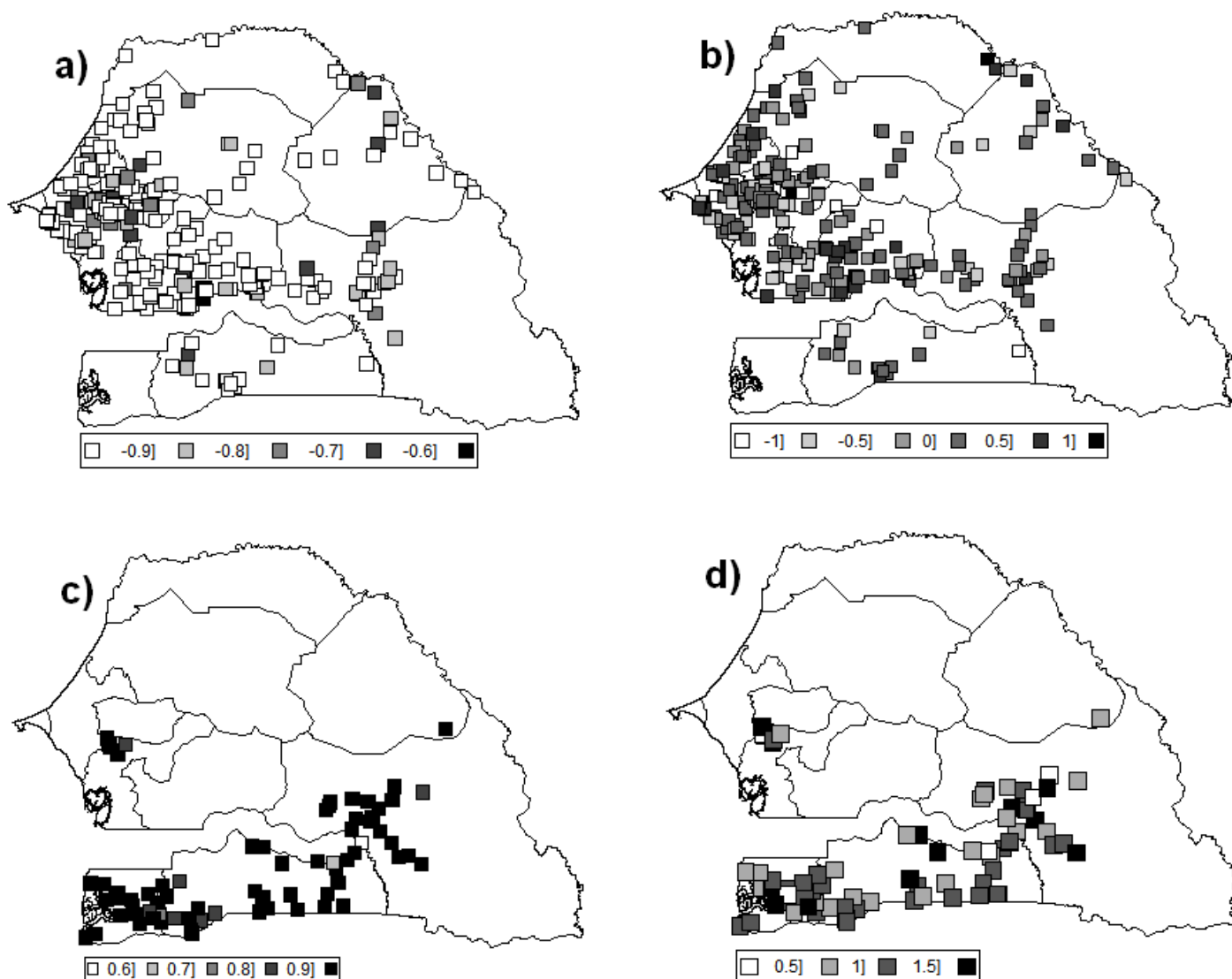


Figure 11 : Comparaison de l'assignation génétique entre les résultats de STRUCTURE et ceux de la sPCA pour les accessions locales précoces et tardives. a) : Représentation graphique des valeurs q obtenues avec l'analyse STRUCTURE pour les accessions précoces (Souana), b) : Représentation graphique des scores d'autocorrélations obtenus avec l'analyse sPCA pour les accessions précoces (Souana), c) : Représentation graphique des valeurs q obtenues avec l'analyse STRUCTURE pour les accessions tardives (Sanio), d) : Représentation graphique des scores d'autocorrélations obtenus avec l'analyse sPCA pour les accessions tardives (Sanio)

L'analyse du modèle d'isolement par la distance (IBD) grâce au test de Mantel révèle de faibles pentes d'IBD pour les accessions précoces ($R^2 = 0.0029$) et tardives ($R^2 = 0.0092$) (Fig. 12). En effet la valeur du coefficient de corrélation entre les deux matrices comparées (R^2) doit être proche de 1 ou -1 pour indiquer l'existence d'un motif d'isolement par la distance. Dans notre cas de figure, les valeurs de R^2 sont très proches de 0. Cela ne suggère aucun pattern significatif d'isolement génétique par la distance.

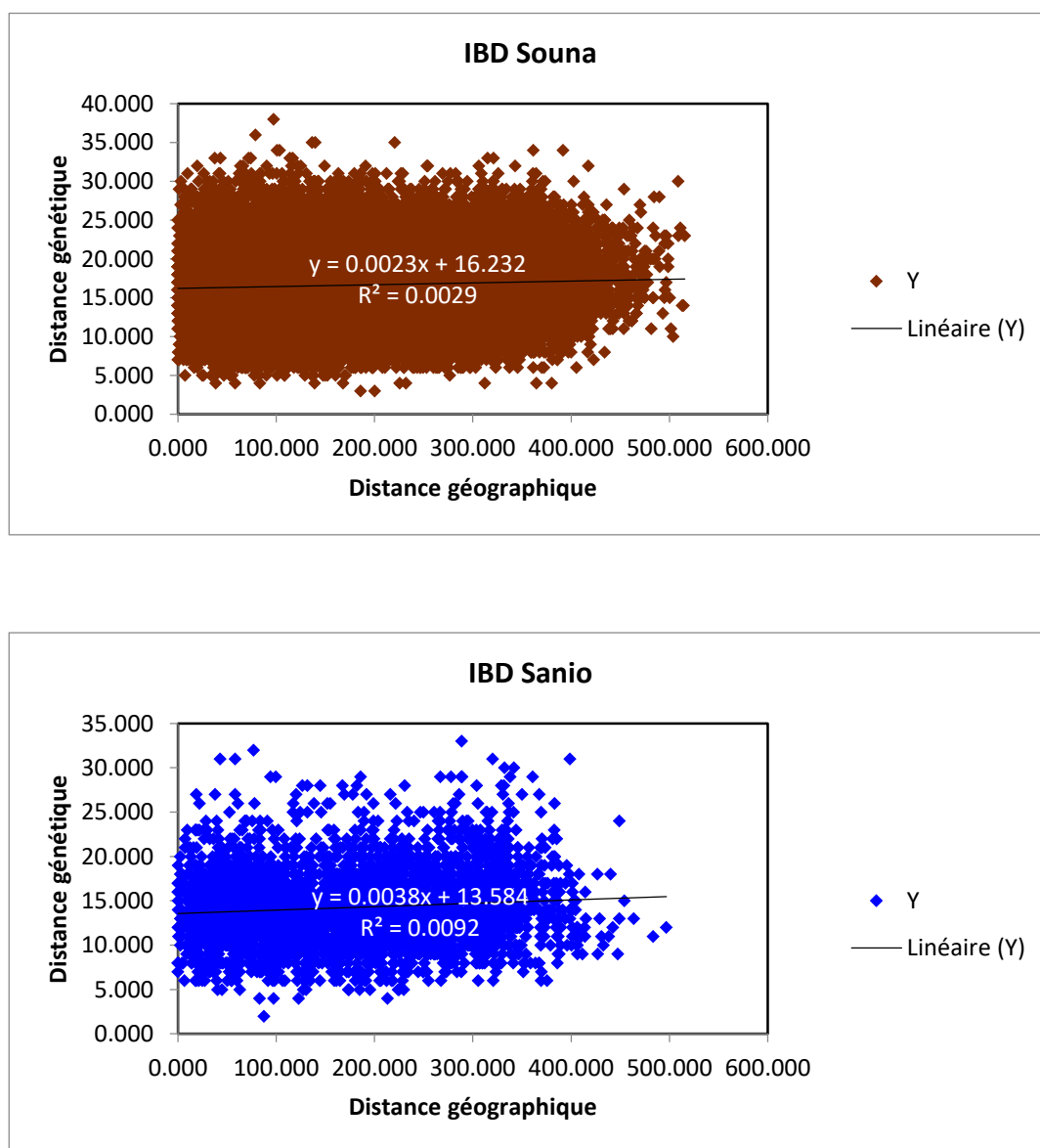


Figure 12 : Représentations graphiques du modèle d'isolement par la distance pour les accessions locales précoces (rouge) et tardives (bleu), obtenues grâce au test de Mantel entre la matrice de distance génétique et la matrice de distance géographique.

2.4. Vérification de redondances de génotypes

La vérification de redondances de génotypes par le calcul des probabilités d'identité des individus de la collection en fonction des marqueurs cumulés croissants montre que la probabilité que deux individus ayant des liens de parenté ou non soit identiques est nulle à partir de sept (7) loci cumulés croissants. Ce qui veut dire que lorsque l'on considère tous les 12 marqueurs SSR, il n'y a pas dans la collection deux individus présentant les mêmes allèles sur tous les 12 marqueurs. Autrement dit, il n'y pas de doublons dans la collection. Ce résultat est confirmé par un alignement comparé de tous génotypes de tous individus par rapport aux 12 marqueurs.

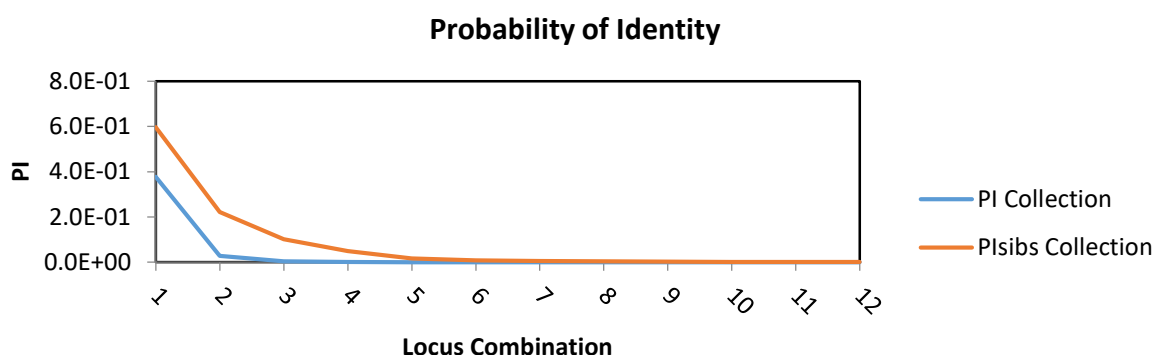


Figure 13 : Représentation graphique des probabilités que deux individus, ayant des liens de parenté (trait rouge) ou n'ayant aucun lien de parenté (trait bleu), présentent des génotypes identiques en fonction des marqueurs cumulés croissants.

3. Evaluation génétique adaptative

Dans le but d'évaluer la diversité génétique en relation avec les cycles de floraison, les accessions ont été génotypées pour les allèles des gènes *PgPHYC* et *PgMADS11*.

Le tableau 3 indique le nombre de génotypes pour les accessions locales précoces et tardives. Nos résultats montrent une spécificité de l'allèle G du gène *PgPHYC* aux accessions précoces, qui ont été les seules à présenter cette allèle de floraison précoce (G), ce qui a entraîné une différence significative des fréquences des génotypes (test de F, $P = 0.006$). Pour *PgMADS11*, une différence significative des fréquences des génotypes (test de F, $P = 0.004$) a été observée. Les accessions locales précoces ont une fréquence plus élevée d'allèles associés à la floraison précoce.

Une corrélation significative entre l'allèle G du gène *PgPHYC* avec la latitude (p-valeur = 0,0348) a été observée. Cependant, la significativité disparaît en considérant le regroupement génétique ($q \leq 0,7$). Pour l'allèle de délétion du gène *PgMADS11*, nos données révèlent une corrélation significative par rapport à la longitude (valeur $p < 0,001$) même si l'on considère le modèle de groupement génétique.

La régression logistique appliquée aux gènes *PgPHYC* et *PgMADS11* comme variables réponses par rapport aux données bioclimatiques révèle que :

- Le gène *PgPHYC* est corrélé aux variables d'isothermalité c'est-à-dire le rapport de la température moyenne annuelle sur l'écart annuel de température ($P = 0,02$), la saisonnalité de la température ($P < 0,0001$), et la température minimale de la période la plus froide ($P < 0,0001$), la température minimale au mois d'octobre ($P < 0,0001$), la température maximale au mois d'octobre ($P < 0,0001$), la température minimale au mois d'août ($P < 0,0001$), la température minimale au mois de septembre ($P < 0,0001$), la température maximale au mois de juin ($P < 0,0001$) et les précipitations au mois d'août ($P < 0,0001$).

- Tandis que le gène *PgMADS11* est corrélé aux variables de température minimale au mois d'avril ($P = 0,01$), la saisonnalité des précipitations ($P < 0,0001$), les Précipitations du trimestre le plus sec ($P < 0,0001$), les précipitations du trimestre le plus chaud ($P < 0,0001$) et les précipitations du trimestre le plus froid ($P < 0,0001$).

Tableau 3 : Décompte des génotypes pour les variétés améliorées, les accessions locales précoces et tardives par rapport aux valeurs de $q \leq 0,7$

	Génotypes de <i>PgPHYC</i>			Génotypes de <i>PgMADS11</i>		
	C/C	C/G	G/G	363/363	363/387	387/387
Variétés améliorées	8	3	1	1	2	9
Accessions précoces	239	21	0	18	65	185
Accessions tardives	79	0	0	2	16	62

DISCUSSION

A travers cette étude, une nouvelle collection de mil du Sénégal a été mise en place. Une collection qui représente mieux les ressources génétiques de mil du Sénégal que l'ancienne collection. En effet cette collection prend en compte toutes les régions du pays où le mil est cultivé et prend aussi en compte toute la diversité variétale qui est cultivée dans ces différentes régions. Ainsi cette collection représente bien le mil cultivé du Sénégal, tant sur le plan géographique que sur le plan variétal.

L'évaluation de l'existence de doublons dans la collection actuelle de mil du Sénégal a montré qu'il n'y a pas de redondances génotypiques dans la collection. Le mode de reproduction du mil est un système allogame. Les espèces allogames sont souvent moins enclines à montrer des redondances de génotypes dans les collections que les espèces autogames. En outre, ces résultats témoignent de l'efficacité de la méthode de prospection lors des collectes de la nouvelle collection et de la bonne épuration de l'ancienne collection.

Une grande diversité génétique (H_{EX} : 0,516) du mil cultivé au Sénégal a été observée dans cette étude. Ce niveau de diversité est inférieur à celui rapporté précédemment par Stich et al. (2010) (Stich et al., 2010) où ses échantillons de mil du Sénégal ont eu un niveau de diversité (H_{EX}) de 0,69. Cette différence s'expliquerait en partie par le fait que les deux études n'ont utilisé que deux SSR en commun. Cependant, comparé à Mariac et al. (2006a), où les 12 SSR sont utilisés sans compter le fait que Mariac et al. (2006a) ont utilisé un plus grand nombre de marqueurs (25 SSR), nous avons trouvé une plus grande diversité génétique des mils du Sénégal que ceux du Niger (H_{EX} = 0,49) et un nombre d'allèles par locus similaire (N_a Sénégal: 6 contre N_a Niger: 6,2). En outre, la valeur de la F_{IS} pour le mil du Niger (0,30) était plus élevée que pour le Sénégal (0,18). La diversité génétique élevée trouvée dans notre étude est en accord avec d'autres études menées à l'échelle régionale (Dussert et al., 2015a ; Hu et al., 2015 ; Oumar et al., 2008). La diversité génétique élevée du mil peut s'expliquer par son fort taux de d'allogamie (75%), (Sandmeier, 1993) et les flux génétiques existant avec son parent sauvage (Lewis, 2010 ; Mariac et al., 2006a). Cette grande diversité génétique du mil, est en accord avec sa forte diversité agromorphologique observée aussi bien au Sénégal (Sy et al. 2015 ; Kane, 2014) qu'en Afrique de l'Ouest (Pucher et al., 2015) et met en évidence un potentiel inexploité pour la sélection.

Selon Dussert et al. (2015), un seul événement de domestication aurait conduit à l'obtention des variétés locales précoces et tardive. Ce qui expliquerait en partie la différenciation génétique limitée observée à l'échelle régionale (Dussert et al., 2015a), même si la floraison est un trait

majeur utilisé par les agriculteurs pour définir leurs variétés locales. Une étude comparée des valeurs F_{ST} entre accessions locales précoces et tardives montre de légères différentiations (0,052), qui est de la même ampleur (0,053) que celle observée au Niger (Lakis et al., 2012b). Cependant, les résultats bayésiens et les analyses multivariées ont montré une différenciation claire et nette avec deux pools génétiques. Le pool génétique des variétés précoces et le pool génétique des variétés tardives.

Cette structure génétique pourrait être expliquée en partie par la répartition géographique des accessions locales comme le montre l'analyse sPCA, où il est difficile de dissocier complètement la structure spatiale et génétique. A une échelle plus fine, on a observé des modèles d'admixture plus contrastés. Dans le centre du Sénégal, on a pu constater une nette différenciation entre les variétés précoces et tardives. En revanche, des schémas plus ou moins nets ont été observés dans le sud du Sénégal. Les résultats pourraient être liés à des pratiques agricoles des agriculteurs et la circulation des semences (Kouakou et al., 2013 ; Lakis et al., 2012b ; Mariac et al., 2006b).

La proximité génétique entre les accessions locales précoces et les variétés améliorées met en évidence une histoire de sélection variétale et des pratiques agricoles. Au Sénégal, peu de programmes de sélections ont été entrepris sur le mil, mais tous ceux qui ont été menés, ont été axés sur la réduction du cycle de floraison. Les parents géniteurs des variétés produites proviennent d'accessions locales. En plus, les agriculteurs cultivent encore dans leurs champs, des variétés améliorées conjointement avec les accessions locales précoces, augmentant ainsi les flux de gènes et donc la proximité génétique. Jusqu'à présent, les programmes de sélection ont été mis en place en utilisant des ressources génétiques restreintes aux variétés locales précoces, comme le prouvent nos résultats avec un groupage des accessions précoces et des variétés améliorées d'une part et les accessions tardives d'autre part, dans toutes les analyses incluant ces trois groupes. En effet, l'Afrique subsaharienne a enregistré de longues périodes de sécheresse dans les années 70 et 80 qui ont conduit à la production de variétés améliorées de cycles courts. Cependant, la diversité génétique élevée du mil cultivé observée, à travers cette étude, pourrait être mieux utilisée lorsqu'on la considère (cette diversité génétique) globalement.

La sensibilité à la photopériode des accessions locales de mil et donc la variation du cycle de floraison représentent une réponse clé pour l'adaptation aux conditions climatiques futures (Sultan et al., 2013). En effet, les cultivars de mil du Niger auraient réduit leur cycle de floraison pendant les épisodes de sécheresse de 1976 à 2003 (Vigouroux et al., 2011b). Cette réduction

a été corrélée avec le changement des fréquences alléliques pour le gène *PgPHYC*. Une corrélation similaire a été trouvée avec *PgMADS11* (Mariac et al., 2011a ; Saïdou et al., 2009). Nos données ont révélé des différences génétiques significatives des types variétaux locaux pour les deux gènes avec des accessions locales précoces plus riches en allèles dites « de précocité ». Le temps de floraison a été corrélé avec la latitude, les accessions précoces étant cultivés plus fréquemment dans des latitudes septentrionales où les conditions environnementales sont plus arides (Hausmann et al., 2006 ; Pucher et al., 2015). Dans notre étude, l'effet de latitude a été confondu avec la structure génétique spatiale pour le gène *PgPHYC*. D'autre part, une corrélation avec la longitude a été trouvée pour le gène *PgMADS11*. Même si la longitude est corrélée à un gradient de température au Sénégal, des recherches plus approfondies seraient nécessaires pour y répondre de manière plus complète et démêler cette corrélation par rapport à la longitude et la température. En attendant, des gènes de floraison tels que *PgPHYC* et *PgMADS11* pourraient être utilisés dans des programmes de sélection assistée par marqueurs.

Nos résultats montrent clairement une corrélation de ces gènes par rapport à des facteurs bioclimatiques liés à la température et aux précipitations. Ce qui suggère qu'il y aurait très certainement une adaptation de ces gènes des mils cultivés du Sénégal, par rapport aux facteurs climatiques.

En effet, nous observons que les fréquences alléliques du gène *PgPHYC* bien que lié à quelques facteurs de précipitation serait plus influencée par des facteurs liés à la température tandis que les fréquences alléliques du gène *PgMADS11* serait plus influencée par des facteurs liés à la précipitation, même si ce gène est corrélé à quelques facteurs de la température. On pressent une certaine dualité parallèle des deux gènes en réponse aux facteurs liés d'une part à la température et d'autre part à la précipitation.

Ce qui est en quelque sorte une confirmation des observations de Vigouroux et al. (2011a) et de Mariac et al. (2011b). Une étude plus poussée sur ces gènes avec séquençage complet et prise en compte de l'aspect fonctionnel serait nécessaire pour une validation définitive de ces observations. En attendant, nos résultats confirment des observations précédentes à leur sujet et suggèrent que ces gènes peuvent d'ores et déjà être utilisés comme outils dans la recherche et les programmes de sélection.

Les corrélations observées de ces gènes en fonction des facteurs enregistrés durant certains mois sont souvent des mois de la période hivernale au Sénégal (qui dure de juin à juillet à septembre

ou octobre en fonction des années), ou des mois qui sont autour de cette période. Même s'il semble relativement hasardeux, il serait donc tout à fait naturel d'imaginer l'influence du fait que le mil est une plante annuelle qui est cultivée par les paysans durant la période hivernale. Il est cependant impossible d'être catégorique avec nos données. Néanmoins, nos résultats montrent bien les corrélations et suggèrent qu'il existerait une adaptation de ces gènes par rapport à des facteurs bioclimatiques liés à la température et à la précipitation.

CONCLUSION

Dans cette partie, la diversité génétique du mil a été étudiée, et tous les objectifs de départ, de cette étude ont été atteints.

En effet, cette étude a permis de mettre en place une nouvelle collection plus représentative des ressources de mils cultivés du Sénégal. Une collection qui représente bien les différentes variétés de mil cultivés et qui représente aussi toutes les zones de culture du mil au Sénégal.

Et l'évaluation de cette collection a montré qu'il n'y a pas de redondances génotypiques en son sein. Parmi les résultats observés à travers cette évaluation, nous pouvons retenir que les mils du Sénégal présentent une grande diversité génétique et qu'ils sont constitués de deux pools génétiques. Un pool génétique à floraison précoce et un second pool génétique à floraison tardive. Cette structure génétique de ces mils est accompagnée par une structure spatiale similaire, en outre il existe un pattern spatial de la diversité du mil au Sénégal.

L'évaluation de la diversité adaptative de cette collection, a montré une spécificité de l'allèle G d'un SNP du gène *PgPHYC* aux accessions précoces et a montré qu'il existe une différenciation significative des fréquences des génotypes de ce gène *PgPHYC* et du gène *PgMADS11* par rapport aux pools génétiques précoces et tardives. Ces gènes de floraison (*PgPHYC* et *PgMADS11*) pourraient être utilisés dans des programmes de sélection assistée par marqueurs. Cette évaluation a aussi montré qu'il y a une corrélation de ces deux gènes par rapport à des facteurs de pluviométrie et de température. Corrélations qui devraient être étudiées de manière plus poussée en vue d'une validation.

Cette étude a largement statué sur la diversité génétique du mil au Sénégal en y apportant des éclaircissements et en traitant des questions qui jusqu'ici étaient restées sans réponses. Cependant, une caractérisation génétique seule n'est pas suffisante en termes d'évaluation des ressources phytogénétiques. Il convient d'y associer aussi une évaluation agro-morphologique en vue d'une bonne gestion et valorisation de ces ressources.

Chapitre II. Deuxième partie :

Evaluation agro-morphologique de la collection de mil du Sénégal

INTRODUCTION

La bonne connaissance des ressources génétiques est indispensable pour leur bonne gestion et valorisation. En effet, c'est sur la base de cette connaissance que s'appuie toute activité de conservation, d'amélioration, ou de valorisation quelconque. Ainsi, les efforts d'évaluation des ressources génétiques sont un point de départ crucial et indispensable pour bien gérer ces ressources génétiques.

Dans le chapitre précédant, nous avons évalué, sur le plan génétique, les accessions de la collection nationale, qui représentent les mils du Sénégal, en vue d'une bonne gestion et valorisation. Cette évaluation a mis en évidence la grande diversité génétique des accessions de mil du Sénégal et l'existence de deux pools génétiques en fonction des accessions précoces d'une part et tardives d'autre part. Cette évaluation a donc fourni une meilleure connaissance de ces ressources sur le plan génétique. Cependant, une évaluation génétique seule ne suffit pas pour une bonne connaissance des ressources génétiques en vue de leur bonne valorisation. Il convient d'y associer aussi une évaluation agro-morphologique.

Sy et al. (2015) ont étudié la diversité agro-morphologique d'accessions de mils de la collection. Cependant, cette étude n'est pas complète. Elle n'a pas couvert toutes les accessions de la collection car les nouvelles accessions (celles collectées en 2014) de la collection n'ont pas été prises en compte. En effet, parmi les 404 accessions qui constituent la collection nationale de mil, seul 243 accessions locales précoces et 9 accessions locales tardives collectées entre 1993 et 2010 y ont été étudiées.

Afin de compléter cette évaluation des ressources génétiques de mil du Sénégal, dans cette partie nous avons fait la caractérisation agro-morphologique de ces dernières au complet.

Objectif

L'objectif global de cette étude est de contribuer à la connaissance des ressources génétiques de mil du Sénégal par l'évaluation agro-morphologique des différentes variétés dans leurs zones de culture. De façon spécifique, cette partie vise à :

- Evaluer le niveau et la structure de diversité agro-morphologiques des mils cultivés du Sénégal ;
- Identifier des caractères discriminants de ces mils cultivés.

MATERIEL ET METHODES

Les variétés précoces ont été évaluées sur deux sites, au centre-nord et centre-sud du pays. Les variétés tardives ont été évaluées sur deux sites, au sud et centre-sud du pays.

1. Évaluation agro-morphologique des variétés précoces

1.1. Matériel végétal et sites d'expérimentation

Le matériel végétal est constitué de trois cent cinq (305) accessions de mil Souna de la collection du Sénégal. Ces accessions ont été collectées entre 1993 et 2014, à travers le pays. Les accessions précoces ont été semées dans deux sites d'expérimentations de l'ISRA : Bambey et Nioro, durant deux années correspondant aux périodes hivernales de 2014 et de 2015.

Le site de Bambey se trouve au Centre National de Recherche Agronomique (CNRA). Il est situé entre 14°32'12" de latitude Nord et 16°36'41" de longitude Ouest. Le sol est de type ferrugineux tropical faiblement lessivé (« Dior »). Le climat est de type sahélien, la période hivernale y dure environ 4 mois, entre juillet et octobre. La pluviométrie annuelle moyenne de 1940 à 2013 varie entre 370 mm et 728 mm pour une moyenne de 549mm (Bodian, 2014). Durant l'hivernage 2014, Bambey a enregistré un cumul de 407 mm de pluies et 692,2 mm durant l'hivernage 2015. La figure 14 représente les pluviométries journalières enregistrées à Bambey durant l'hivernage 2014 et l'hivernage 2015.

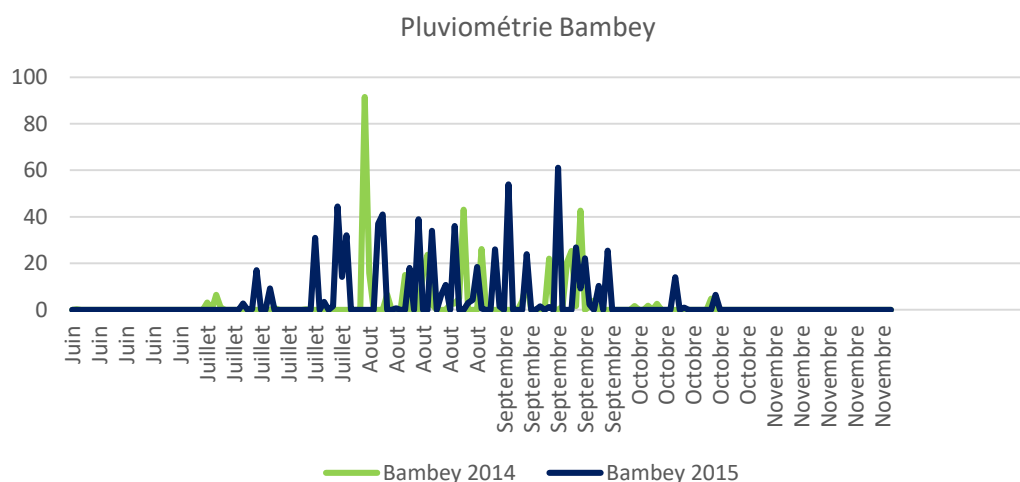


Figure 14 : Pluviométries journalières enregistrées à Bambey durant l'hivernage 2014 (vert) et l'hivernage 2015 (bleu).

Le site de Nioro est une annexe du Centre National de Recherche Agronomique (CNRA). Il est situé entre 13°45'00" de latitude Nord et 15°45'00" de longitude Ouest. Le sol est de type ferrugineux tropical lessivé (« Dior »). Le climat est de type soudano-sahélien, la période hivernale y dure environ 4 mois, principalement entre juillet et octobre. La pluviométrie annuelle moyenne de 1940 à 2013 est entre 594 mm et 1012 mm pour une moyenne de 803 mm (Bodian, 2014). Durant l'hivernage 2014, Nioro a enregistré un cumul de 541,8 mm de pluies et 872,7 mm durant l'hivernage 2015. La figure 15 représente les pluviométries journalières enregistrées à Nioro durant l'hivernage 2014 et l'hivernage 2015.

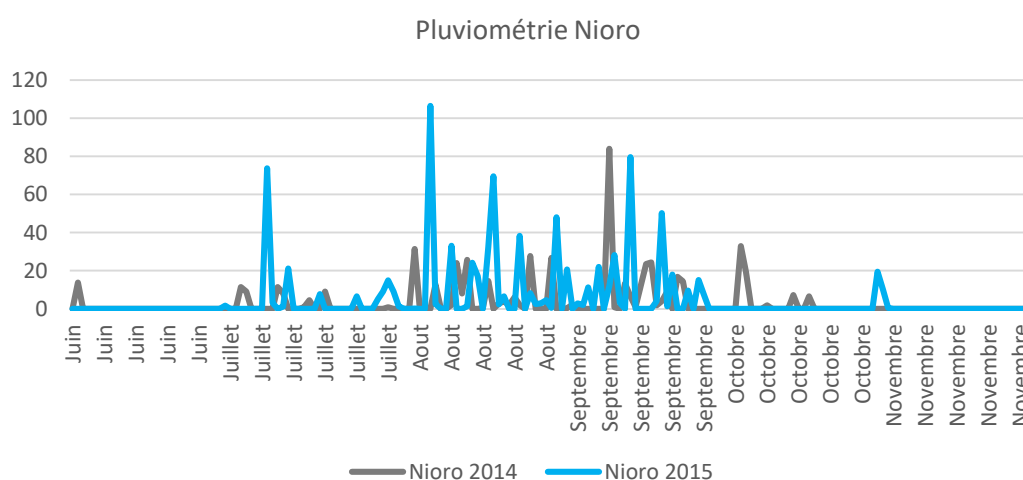


Figure 15 : Pluviométries journalières enregistrées à Nioro durant l'hivernage 2014 (vert) et l'hivernage 2015 (bleu)

1.2. Dispositif expérimental

Les essais ont été conduits selon un dispositif en blocs complètement randomisés avec trois répétitions. Chaque parcelle élémentaire était constituée d'une rangée de huit poquets avec un espacement entre lignes et entre poquets de 90 cm. Le numérotage des parcelles élémentaires a été serpenté afin de faciliter le suivi des mesures. Sur chaque essai, deux lignes de bordure ont été prévues de part et d'autre du champ pour éviter l'effet de bordure.

1.3. Conduite des essais

Il a été opéré un labour et un ameublissement du sol avant les semis. Puis, un premier sarclage a été effectué suivi d'un binage et démariage en laissant un seul plant par poquet 15 jours après les semis. Ensuite, un deuxième sarclage et binage (à cheval) au 25ème jour après le démariage a été fait. Un apport d'engrais NPK (15-15-15) à hauteur de 150 Kg/ha, a été fait avant

l'ameublissement. Deux apports d'urée à hauteur de 100 kg/ha, ont été faits après le démariage et à la montaison. En dehors de cet itinéraire, il a été procédé au binage (désherbage) à chaque fois qu'il a été nécessaire pour éradiquer les mauvaises herbes.

1.4. Mesures des paramètres agro-morphologiques

Les douze (12) paramètres suivants ont été observés : l'incidence du mildiou (DM), date de 50% de floraison (FLO), longueur de la feuille paniculaire (LOFP), largeur de la feuille paniculaire (LAFP), nombre d'entreœuds (NEN), hauteur de la plante (HTR), nombre de talles productives (NTP), circonférence de l'épi (CIR), longueur de l'épi (LEP), poids de 1000 grains (PMG), le rendement en grains (RDT_kg_ha) et le rendement en épis (Rdt_Epis_kg_ha). Les mesures de ces paramètres ont été réalisées suivant les recommandations de l'IPGRI et de l'ICRISAT, (1993).

2. Évaluation agro-morphologique des variétés tardives

2.1. Matériel végétal et sites d'expérimentation

Le matériel végétal est constitué de 86 accessions de mil Sanio de la collection du Sénégal. Ces accessions ont été collectées en 2014, à travers le pays. Elles ont été semées dans deux sites de l'ISRA : Kolda et Sinthiou Malem, durant les périodes hivernales de 2014 et 2015. Les essais ont comporté un seul facteur étudié : ces accessions. Le nombre d'accessions (86) étant le nombre de traitements.

Le site de Sinthiou Malem est situé entre 13°49'01" de latitude Nord et 13°55'03" de longitude Ouest. Le sol est de type ferrugineux tropical (« Deck »). Le climat est de type soudanien, la période hivernale y dure environ 4 mois, principalement entre juillet et octobre. La pluviométrie annuelle moyenne varie entre 581 et 969 mm pour une moyenne de 775 mm (Bodian, 2014). Sinthiou Malem a enregistré un cumul de 608,5 mm de pluies durant l'hivernage 2014, et 648,5 mm durant l'hivernage 2015. La figure 16 représente les pluviométries journalières enregistrées à Sinthiou malem durant l'hivernage 2014 et l'hivernage 2015.

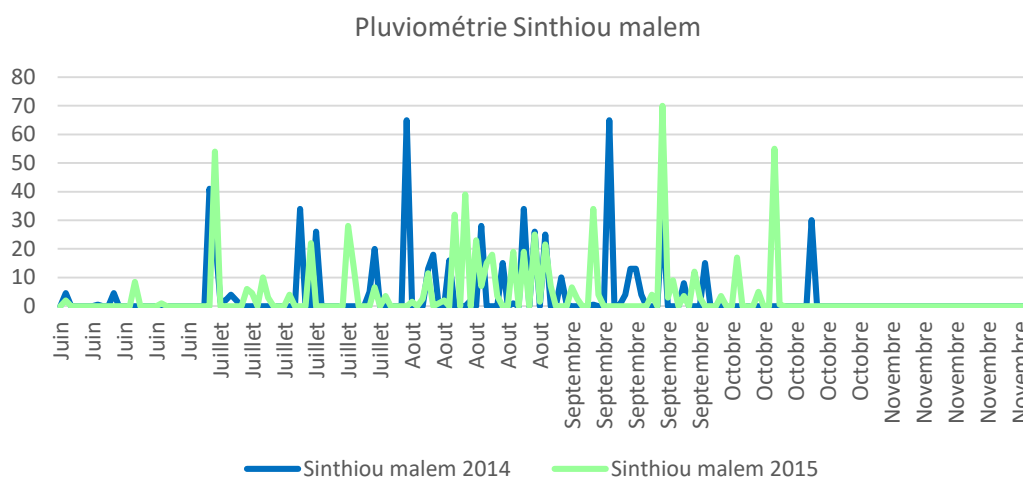


Figure 16 : les pluviométries journalières enregistrées à Sinthiou Malem durant l’hivernage 2014 (bleu) et l’hivernage 2015 (vert)

Le site de Kolda se trouve au Centre de Recherche Zootechnique (CRZ). Il est situé entre 12°53'02" de latitude Nord et 14°57'05" de longitude Ouest. Le sol est de type ferrugineux tropical lessivé (« Dior »). Le climat est de type soudano-guinéen, la période hivernale y dure environ 4 mois, principalement entre juillet et octobre. La pluviométrie annuelle moyenne est entre 849 et 1369 mm pour une moyenne de 1109 mm (Bodian, 2014). Durant l’hivernage 2014, Kolda a enregistré un cumul de 788,15 mm de pluies et 1065 mm durant l’hivernage 2015. La figure 17 représente les pluviométries journalières enregistrées à Kolda durant l’hivernage 2014 et l’hivernage 2015.

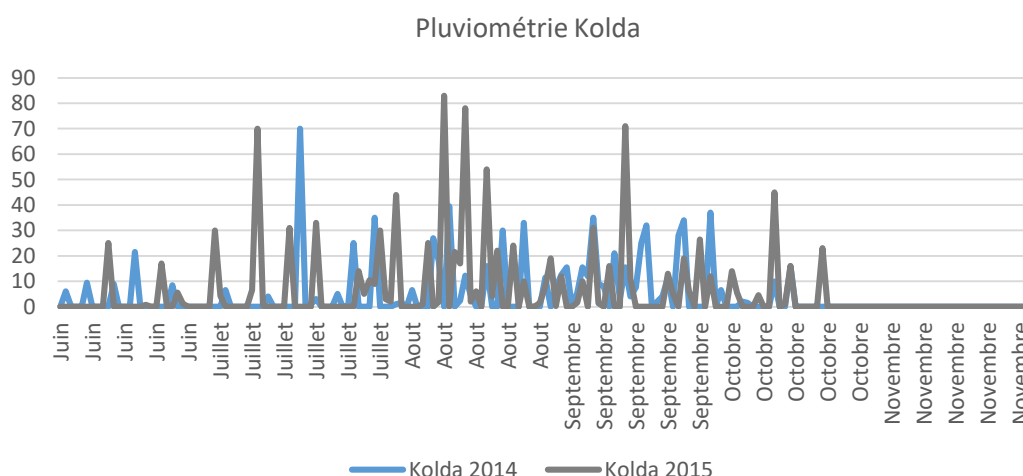


Figure 17 : les pluviométries journalières enregistrées à Kolda durant l’hivernage 2014 (bleu) et l’hivernage 2015 (gris)

2.2. Dispositif expérimental

Les essais ont été conduits selon un dispositif en blocs complètement randomisés avec trois répétitions. Chaque parcelle élémentaire était constituée de deux rangées de huit poquets avec un espacement entre lignes et entre poquets de 90 cm. Le numérotage des parcelles élémentaires était serpenté, pour faciliter le suivi des mesures. Sur chaque essai, deux lignes de bordure avaient été prévues de part et d'autre du champ pour éviter l'effet de bordure.

2.3. Conduite des essais

Le sol a été ameubli et labouré avant les semis de chaque essai. Puis un premier sarclage et binage et démariage en laissant un seul plant par poquet 15 jours après semis a été fait. Ensuite, un deuxième sarclage et binage (à cheval) au 25ème jour après le démariage. Un apport d'engrais NPK (15-15-15) à hauteur de 150 Kg/ha, a été fait avant l'ameublissement. Deux apports d'urée à hauteur de 100 kg/ha, ont été faits après le démariage et à la montaison. En dehors de cet itinéraire, il a été procédé au binage (désherbage) à chaque fois qu'il a été nécessaire pour éradiquer les mauvaises herbes.

2.4. Paramètres mesurés

Les paramètres suivants ont été observés : l'incidence du mildiou (DM), date de 50% de floraison (FLO), longueur de la feuille paniculaire (LOFP), largeur de la feuille paniculaire (LAFP), nombre d'entrenœuds (NEN), hauteur de la plante (HTR), nombre de talles productives (NTP), circonférence de l'épi (CIR), longueur de l'épi (LEP), poids de 1000 grains (PMG), le rendement engrains (RDT_kg_ha) et le rendement en épis (Rdt_Epis_kg_ha). Les mesures de ces paramètres ont été prises suivant les recommandations de l'IPGRI et de l'ICRISAT, (1993).

3. Analyse des données

Les analyses des variables quantitatives ont été réalisées suivant deux approches.

(1) Des analyses de variances, ont d'abord été faites : Afin de vérifier l'effet site et l'effet saison, une analyse de variance (ANOVA) a été faite, à l'aide du logiciel XLStat 2014.

Les interactions génotype environnement (GxE) ont aussi été vérifiées, sur la base du modèle multiplicatif (AMMI), pour observer au niveau individuel les accessions qui présentent une certaine stabilité par rapport à différents environnements et/ou conditions pluviométriques (saisons) et celles qui présentent une « adaptation spécifique » à des environnements donnés.

L'interaction génotype-environnement (GxE) est définie par Romagosa et al. (1993) comme étant l'expression génotypique différentielle à travers les environnements.

Cette analyse a été faite à l'aide du package Agricolae v 1.2-4 (2016) du logiciel R v 3.3.1 (R Development Core Team, 2008). L'équation du model est la suivante :

$$Y \sim \text{ENV} + \text{REP} + \text{GEN} + \text{ENV:GEN}$$

Y étant la variable étudiée, ENV étant le facteur environnement, REP sont les répétitions dans chaque environnement et GEN les génotypes.

(2) Ensuite, des analyses multivariées ont été faites afin d'étudier la structure et les relations entre accessions à travers une Analyse en Composantes Principales (ACP) et une Classification Hiérarchique Ascendante (CHA) par la méthode de Ward, grâce à la matrice de dissimilarité obtenue par les distances euclidiennes des moyennes agro-morphologiques ajustées. Puis, une analyse factorielle discriminante a été faite pour vérifier la pertinence des possibles subdivisions en groupes suggérées par la Classification Hiérarchique Ascendante et pour identifier les caractères qui discriminent ces groupes. Enfin un test sur tableaux de contingence (χ^2) a été réalisé pour vérifier un possible lien entre les sous-groupes agro-morphologiques et leurs origines par rapport aux zones agro-écologiques du Sénégal (Fall, 2008). Ces analyses ont été faites à l'aide du logiciel XLStat 2014 (Addinsoft, 2014) et le logiciel R v 3.3.1 (R Development Core Team, 2008).

RESULTATS

1. Diversité agro-morphologique des variétés précoces

Les analyses de variances (Tableau 4) ont détecté l'existence d'un effet environnement sur tous les caractères agro-morphologiques sauf pour la largeur de la feuille paniculaire et la circonférence de l'épi. Ce qui traduit un certain contraste entre les sites d'évaluation. Il existe aussi un effet génotypes hautement significatif sur l'ensemble des caractères agro-morphologiques. Ce qui signifie une certaine variabilité des accessions. Cependant, les interactions génotypes - environnements ne sont observées que sur les variables suivantes : date de floraison, nombre de talles productives, hauteur de la plante, nombre d'entrenœuds et rendement en grain et en épi.

Tableau 4 : Analyse de variance des variétés précoces (Souna). * signifie seuil de significativité à $\alpha < 0.05$, ** $\alpha < 0.01$ et *** $\alpha < 0.001$

	DF	DM	FLO	NTP	HTP	NEN	LOFP	LAFP	LEP	CIR	PMG	RDT_Grain	RDT_Epis
Site	3	***	***	***	**	*	**	NS	**	*	NS	***	***
REP within Site	8	***	***	***	***	***	***	***	**	***	***	***	***
Genotype	304	***	***	***	***	***	***	***	***	***	***	***	***
Site~Genotype	912	NS	***	***	***	**	NS	NS	NS	NS	NS	***	***
Mean		15,08	57,09	4,46	235,84	7,98	47,41	4,71	53,06	8,44	6,95	1537,51	2557,74
CV(%)		94,29	6,06	29,06	7,73	10,26	15,13	15,61	12,03	10,41	12,67	33,75	30,24

L'analyse de la matrice de corrélation de Pearson de l'Analyse en Composantes Principales (ACP) (Tableau 5), entre les variables montre globalement, une corrélation positive du rendement (RDT_Grain, RDT_Epis) aux paramètres de développement végétatif (nombre d'entrenœuds (NEN), la hauteur de la plante (HTP), longueur et largeur de la feuille paniculaire (LOFP, LAFP) et aux paramètres de développement de l'épi (longueur de l'épi (LEP) et de circonférence de l'épi (CIR)). Cependant, la plupart de ces paramètres de développement végétatif (HTP, NEN, LOFP et CIR), sont eux positivement corrélés à la date de floraison (FLO). L'incidence du mildiou elle, est négativement corrélée aux paramètres de date de floraison (FLO), de développement végétatif (HTP, NEN) de développement de l'épi (LEP, CIR) et au rendement (RDT_Grain et RDT_Epis).

Tableau 5 : Matrice de corrélation de Pearson entre les variables, à partir de l'Analyse en Composantes Principales (ACP) des accessions précoces (Souna). Les valeurs en gras sont significatives au seuil de 0,001

Variables	DM	FLO	NTP	HTP	NEN	LOFP	LAFP	LEP	CIR	PMG	RDT_Grain	RDT_Epis
DM	1	-0,3474	0,0680	-0,3919	-0,3340	-0,1394	-0,1544	-0,2826	-0,2541	0,0491	-0,2058	-0,2716
FLO	-0,3474	1	-0,3737	0,4324	0,5850	0,3068	0,0709	0,1455	0,4339	-0,3406	-0,0701	0,0987
NPT	0,0680	-0,3737	1	-0,0996	-0,1720	-0,1931	-0,0992	-0,0947	-0,3204	-0,0320	0,2474	0,1746
PH	-0,3919	0,4324	-0,0996	1	0,6336	0,4145	0,2506	0,6730	0,2617	0,0044	0,3927	0,4539
NTN	-0,3340	0,5850	-0,1720	0,6336	1	0,3609	0,2658	0,2651	0,4626	-0,1279	0,2124	0,3353
FLL	-0,1394	0,3068	-0,1931	0,4145	0,3609	1	0,4411	0,3371	0,3192	0,0230	0,2692	0,3410
FLW	-0,1544	0,0709	-0,0992	0,2506	0,2658	0,4411	1	0,2976	0,2762	0,1088	0,4012	0,4032
SL	-0,2826	0,1455	-0,0947	0,6730	0,2651	0,3371	0,2976	1	0,0589	0,1307	0,4239	0,4372
ST	-0,2541	0,4339	-0,3204	0,2617	0,4626	0,3192	0,2762	0,0589	1	-0,0226	0,2059	0,2943
SW	0,0491	-0,3406	-0,0320	0,0044	-0,1279	0,0230	0,1088	0,1307	-0,0226	1	0,1454	0,0840
GY	-0,2058	-0,0701	0,2474	0,3927	0,2124	0,2692	0,4012	0,4239	0,2059	0,1454	1	0,8597
YS	-0,2716	0,0987	0,1746	0,4539	0,3353	0,3410	0,4032	0,4372	0,2943	0,0840	0,8597	1

L'analyse des interactions GxE par le model multiplicatif (AMMI) pour les accessions précoces révèle tout d'abord des interactions significatives entre géotypes et environnements aussi bien pour le rendement en grains que pour la date de floraison. En effet Lorsque les interactions sont significatives, la somme du pourcentage de variances exprimées par les deux premiers axes de l'Analyse en composantes principales est supérieure à 50%. Les deux premiers axes des ACP pour le rendement en grains et pour la date de floraison ont exprimé au total 80.2% et 87.9%, respectivement. Elle révèle aussi que le paramètre de rendement des individus est généralement plus stable par rapport aux différents environnements que le paramètre de date de floraison. Autrement dit, les individus tendent à plus se grouper à l'origine pour le paramètre de rendement que pour le paramètre de date de floraison. En effet quelques individus (PLS_54 et N_203-14, principalement) ont graphiquement manifesté une certaine instabilité par rapport aux quatre environnements pour le paramètre date de floraison. Ce qui n'est pas le cas pour le rendement en grains. Cette analyse a permis de dresser une liste des accessions précoces par ordre de stabilité du rendement et de la date de floraison (Annexe 1). En termes de rendements, l'accession N_68-14 est la plus stable et N_69-14 est la moins stable. En terme date de floraison, l'accession PLS_158 est la plus stable PLS_87 est la moins stable.

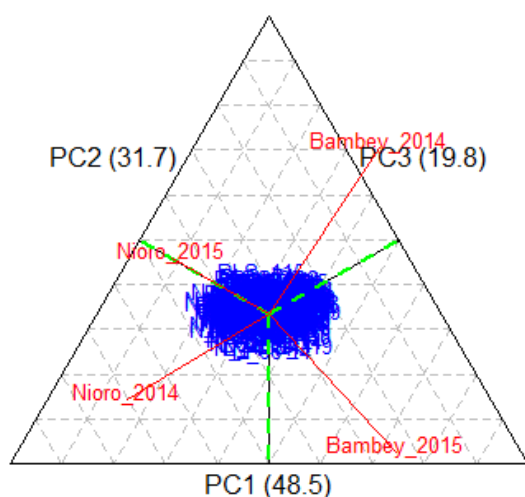


Figure 18: Représentation graphique de l'ACP du model AMMI des accessions précoces par rapport au paramètre de rendement en grains.

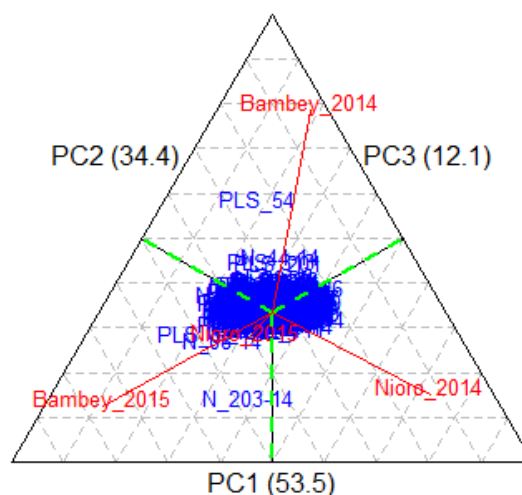


Figure 19: Représentation graphique de l'ACP du model AMMI des accessions précoces par rapport au paramètre de date de floraison.

La Classification Hiérarchique Ascendante a permis de regrouper les accessions en trois groupes (Fig. 20). L'analyse factorielle discriminante confirme bien cette subdivision en trois groupes (Fig. 21). Le test du lambda de Wilks est significatif au seuil de $\alpha = 0.05$ ($P\text{-value} < 0.0001$). Ces trois groupes des variétés précoces se différencient par rapport au premier axe de l'analyse factorielle discriminante. Les paramètres significativement discriminants sont le rendement en grains et en épi (RDT_Grain et RDT_Epis) ($\alpha = 0.05$, $P\text{-value} < 0.0001$ pour les deux caractères qui sont tous positivement corrélés au premier axe de l'AFD), la longueur de l'épi (LEP) et la circonférence de l'épi (CIR) ($\alpha = 0.05$, $P\text{-value} < 0.0001$ et $P\text{-value} = 0.0001$, respectivement. Ces deux caractères sont positivement corrélés au premier axes de l'AFD), la hauteur de la plante (HTR) et le nombre d'entrenœuds (NEN) ($\alpha = 0.05$, $P\text{-value} < 0.0001$ pour les deux caractères qui sont positivement corrélés au premier axe de l'AFD) et l'incidence du mildiou (DM) ($P\text{-value} = 0.003$, $\alpha = 0.05$) (qui est négativement corrélé à l'axe 1 de l'AFD). La représentation des barycentres des trois groupes par rapport aux paramètres agro-morphologiques significativement discriminants (Fig. 22), nous permet d'observer que les paramètres du rendement (RDT_Grain et RDT_Epis), de l'incidence du mildiou (DM) et de la longueur de l'épi sont les plus discriminants. Les individus du groupe 1 sont ceux qui présentent un plus faible rendement et des épis plus courts. Ils sont plus sensibles au mildiou et sont moins développés sur le plan végétatif. Les individus du groupe 3 sont moins sensibles au mildiou, sont plus développés et ont un rendement plus élevé. Les individus du groupe 2 sont intermédiaires par rapport à tous ces traits.

Le test de χ^2 a mis en évidence une absence de relation ($P\text{-value} = 0.164$) entre les sous-groupes agro-morphologiques des Souna et les zones agro-écologiques du pays (Fall, 2008).

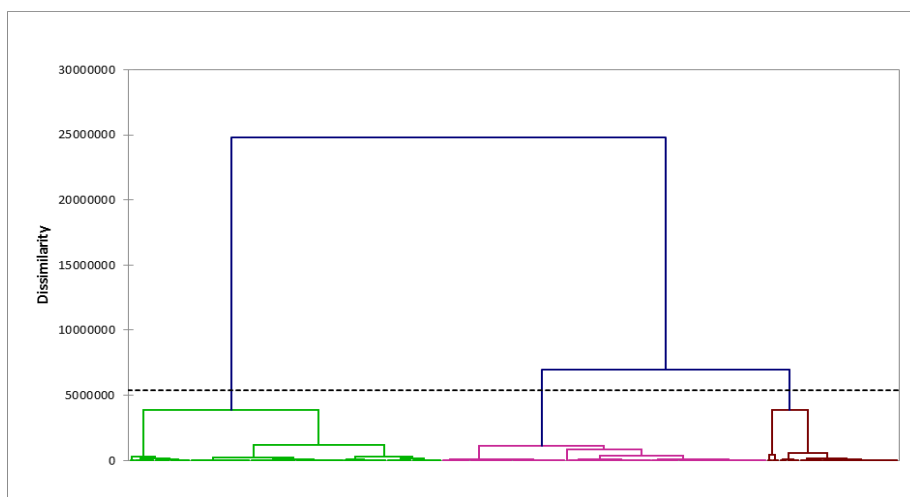


Figure 20: Analyse de Classification Hiérarchique Ascendante des Souna de la collection

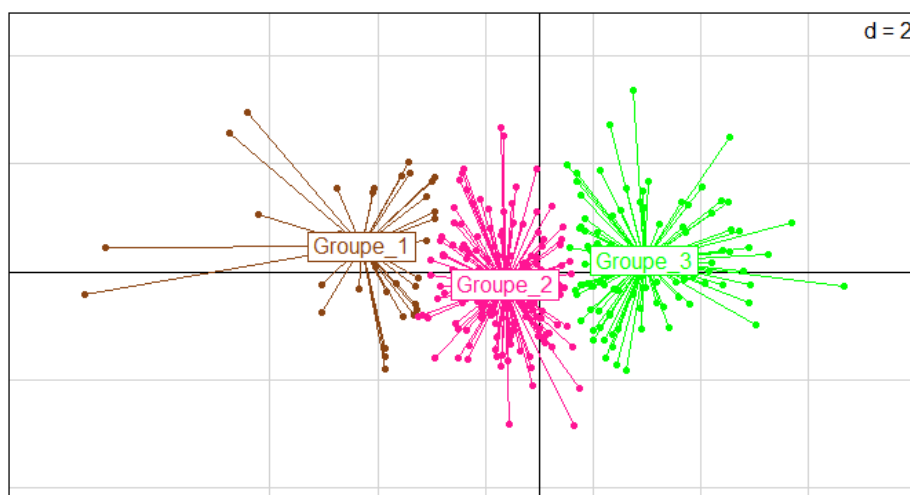


Figure 21 : Analyse Factorielle discriminante des accessions précoces (Axis 1 = 98.49%, Axis 2= 1.51%)

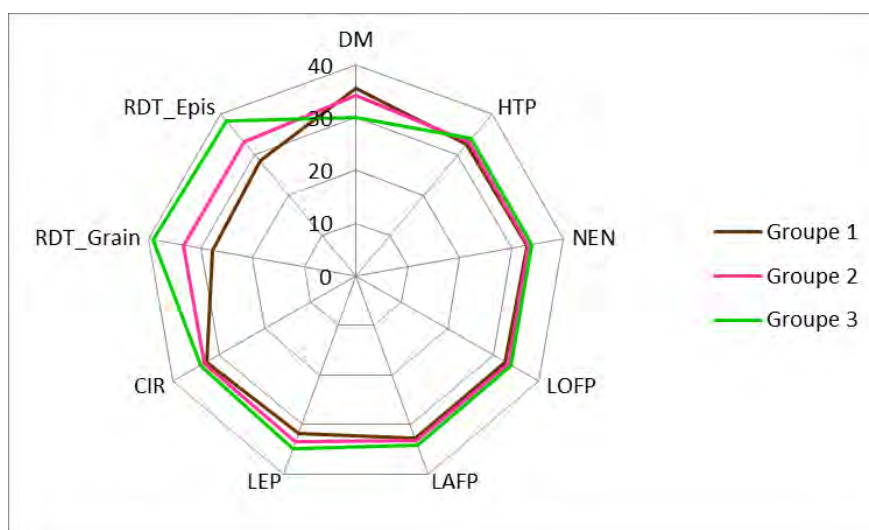


Figure 22 : Représentation graphiques des barycentres des trois groupes par rapport aux caractères agro-morphologiques

2. Diversité agro-morphologique des variétés tardives

Les analyses de variances (ANOVA) des variétés tardives (Tableau 6) ont montré des différences entre les sites avec un effet environnement très significatif sur tous les caractères agro-morphologiques sauf pour la longueur de l'épi. Il existe aussi une assez forte variabilité des accessions tardives sur tous les caractères. Les interactions entre génotypes et environnements sont aussi hautement significatives sur tous les caractères sauf la hauteur de la plante.

Tableau 6 : Analyse de variance des variétés tardives (Sanio). * signifie seuil de significativité à $\alpha < 0.05$, ** $\alpha < 0.01$ et *** $\alpha < 0.001$

	DF	DM	FLO	NTP	HTP	NEN	LOFP	LAFP	LEP	CIR	PMG	RDT_Grain	RDT_Epis
Site	3	***	***	***	**	***	***	***	NS	***	***	***	***
REP within Site	8	*	**	***	***	*	**	*	NS	***	***	***	***
Genotype	73	*	***	***	***	***	**	*	***	***	**	**	***
Site~Genotype	219	***	***	***	NS	***	***	***	***	***	***	***	***
Mean		24,17	86,23	9,81	335,63	13,48	52,33	3,89	40,13	1,55	6,2	853,57	2248,88
CV(%)		54,84	4,76	20,06	7,22	6,7	9,77	10,67	12,27	11,66	16,76	40,92	30,42

L'analyse de la matrice de corrélation de Pearson des variables, issue de l'ACP révèle que la date de floraison (FLO) est positivement corrélée aux paramètres de développement végétatif de nombre d'entreœuds (NEN) et de hauteur de la plante (HTP) (qui sont corrélés entre eux) mais aussi de rendement en épi (RDT_Epis). Tandis que le rendement en grains (RDT_Grain) est négativement corrélé à la longueur de l'épi (LEP).

Tableau 7 : Matrice de corrélation de Pearson entre les variables, à partir de l'Analyse en Composantes Principales (ACP) des accessions tardives (Sanio). Les valeurs en gras sont significatives au seuil de 0,001

Variables	DM	FLO	NTP	HTP	NEN	LOFP	LAFP	LEP	CIR	PMG	RDT_Grain	RDT_Epis
DM	1	0,0886	-0,0200	-0,0251	0,1008	-0,0033	0,2345	0,0713	-0,1229	0,1907	-0,0746	-0,1034
FLO	0,0886	1	0,0021	0,5202	0,5539	-0,2313	-0,0652	-0,0855	0,1908	0,0877	0,3215	0,4089
NPT	-0,0200	0,0021	1	-0,0508	0,3552	0,1193	-0,2325	-0,1186	-0,1047	-0,0591	0,0696	0,0481
PH	-0,0251	0,5202	-0,0508	1	0,4699	-0,1640	-0,2221	0,0637	0,0758	0,1637	0,0886	0,2291
NTN	0,1008	0,5539	0,3552	0,4699	1	-0,0257	-0,0247	-0,1562	0,0655	0,1559	0,2391	0,2025
FLL	-0,0033	-0,2313	0,1193	-0,1640	-0,0257	1	0,1991	0,2522	-0,1741	-0,0819	-0,1128	-0,2915
FLW	0,2345	-0,0652	-0,2325	-0,2221	-0,0247	0,1991	1	0,0466	0,2152	0,0213	0,0081	-0,1377
SL	0,0713	-0,0855	-0,1186	0,0637	-0,1562	0,2522	0,0466	1	0,0973	0,3098	-0,4982	-0,1032
ST	-0,1229	0,1908	-0,1047	0,0758	0,0655	-0,1741	0,2152	0,0973	1	0,1278	0,0996	0,2416
SW	0,1907	0,0877	-0,0591	0,1637	0,1559	-0,0819	0,0213	0,3098	0,1278	1	-0,1170	0,1257
GY	-0,0746	0,3215	0,0696	0,0886	0,2391	-0,1128	0,0081	-0,4982	0,0996	-0,1170	1	0,4512
YS	-0,1034	0,4089	0,0481	0,2291	0,2025	-0,2915	-0,1377	-0,1032	0,2416	0,1257	0,4512	1

L'analyse des interactions GxE par le model multiplicatif (AMMI) pour les accessions tardives a mis en évidence des interactions significatives entre géotypes et environnements aussi bien pour le rendement en grains que pour la date de floraison (Somme des pourcentages de variances des deux premiers axes égale à 87.6% et 86.5%, respectivement). Le paramètre de rendement est moins sensible aux environnements que le paramètre de date de floraison. Une liste classant les accessions tardives de la plus stable à la moins stable par rapport aux différents environnements pour les deux paramètres, a été établie. En termes de rendements, l'accession N_155-14 est la plus stable et N_96-14 est la moins stable. En termes date de floraison, l'accession N_152-14 est la plus stable N_214-14 est la moins stable.

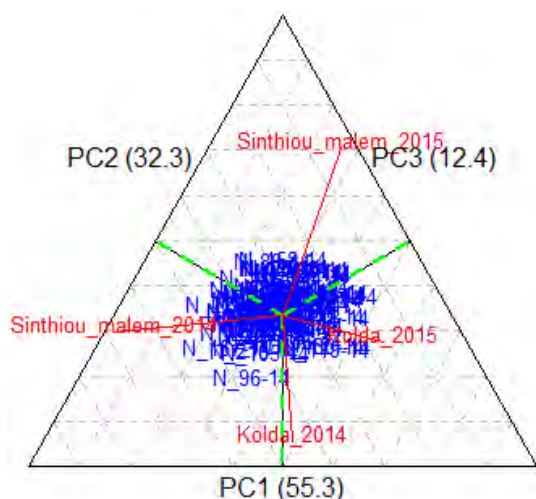


Figure 23: Représentation graphique de l'ACP du model AMMI des accessions tardives par rapport au paramètre de rendement en grains.

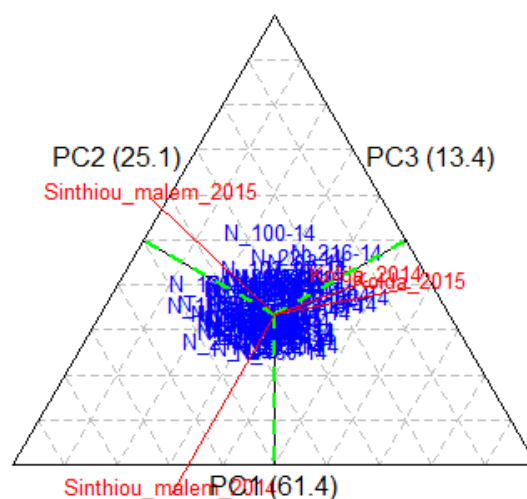


Figure 24: Représentation graphique de l'ACP du model AMMI des accessions tardives par rapport au paramètre de date de floraison.

L'analyse de la Classification Hiérarchique Ascendante a permis de regrouper les accessions en trois groupes (Fig. 25) qui ont aussi été confirmés par l'analyse factorielle discriminante (Fig. 26) avec un test du lambda de Wilks significatif au seuil de $\alpha = 0,05$ ($P\text{-value} < 0,0001$). La différenciation de ces trois groupes des variétés tardives se fait par rapport au premier axe de l'analyse factorielle discriminante. Les caractères agro-morphologiques significativement discriminants sont le rendement en épi (RDT_Epis) ($P\text{-value} < 0,001$, $\alpha = 0,05$) (qui est positivement corrélé à l'axe1 de l'AFD), la date de floraison ($P\text{-value} = 0,01$, $\alpha = 0,05$) (positivement corrélé au premier axe de l'AFD) et la longueur de la feuille paniculaire ($P\text{-value} = 0,005$, $\alpha = 0,05$) (qui est négativement corrélé à l'axe1 de l'AFD). La représentation graphique des barycentres des trois groupes par rapport aux caractères agro-morphologiques significativement discriminants (Fig. 27), a montré que le rendement en panicule est le paramètre le plus discriminant. Les individus du groupe 1 des variétés tardives sont en moyenne ceux qui fleurissent le plus tôt chez les Sanio mais ont un rendement plus faible et une feuille paniculaire plus longue. Les individus du groupe 3 sont par contre ceux qui fleurissent le plus tard chez les Sanio, ont un rendement plus élevé et une feuille paniculaire plus courte. Les individus du groupe 2 sont eux intermédiaires par rapport à ces trois discriminants caractères agro-morphologiques. Le test de χ^2 a mis en évidence une absence de relation ($P\text{-value} = 0.072$) entre les sous-groupes agro-morphologiques des Sanio et les zones agro-écologiques du pays (Fall, 2008).

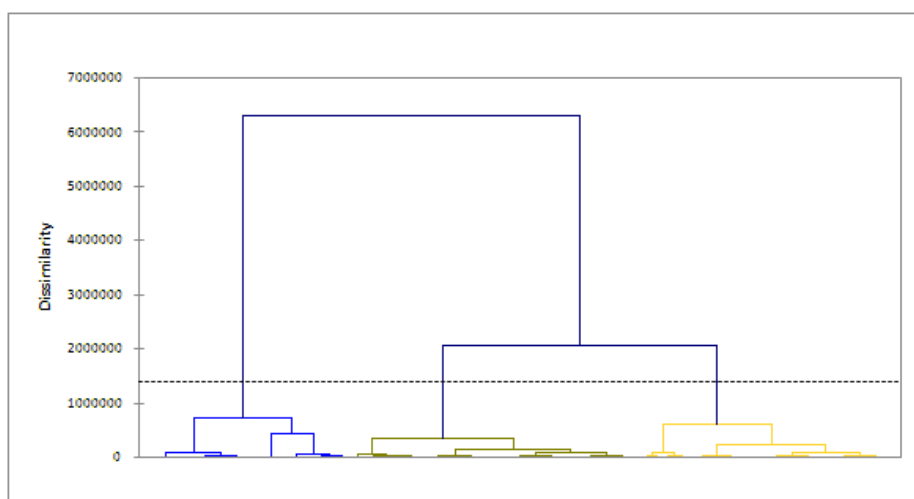


Figure 25 : Analyse de Classification Hiérarchique Ascendante des Sanio de la collection

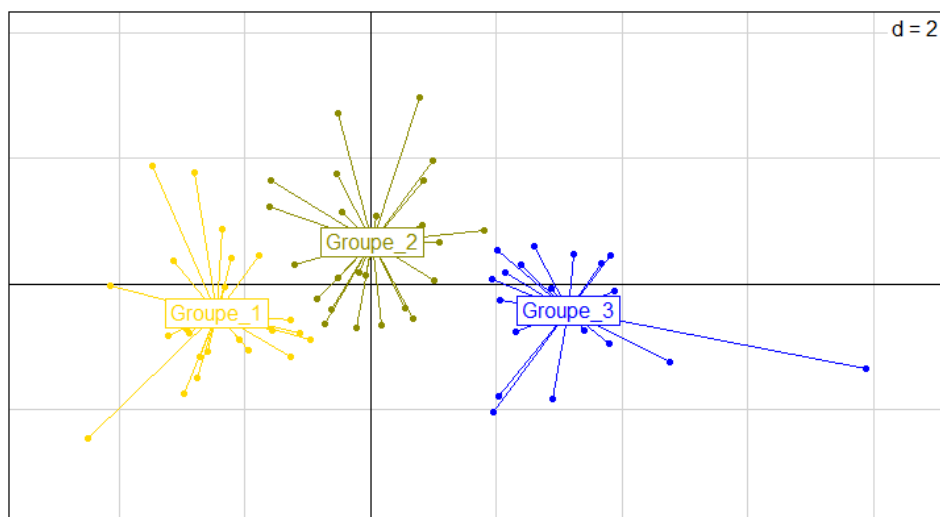


Figure 26 : Représentation graphique de l'Analyse Factorielle discriminante des accessions tardives (Axis 1 = 98,49%, Axis 2= 1,51%)

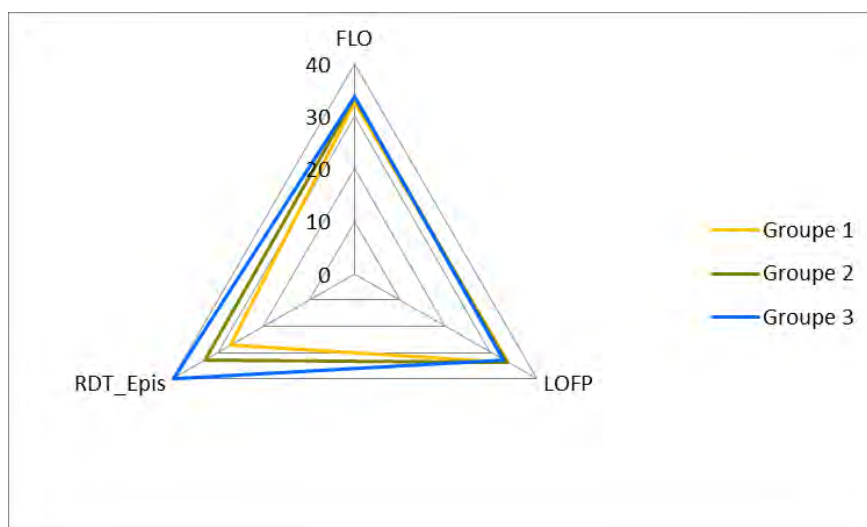


Figure 27 : Représentation graphiques des barycentres des trois groupes par rapport aux caractères agro-morphologiques

DISCUSSION

L'analyse de la diversité agro-morphologique des mils du Sénégal a révélé une grande variabilité des mils cultivés du Sénégal. Ce qui est en accord avec les observations faites sur le plan génétique. La subdivision en clusters dans les différentes formes de mil dans cette étude est assez proches des observations agro-morphologiques réalisées en Côte d'Ivoire (Akanvou et al., 2012), au Bénin (Dagba et al., 2015) et au Sénégal (Sy et al., 2015), avec dans chaque étude une subdivision en trois clusters. La subdivision en clusters se faisant suivant des caractères agro-morphologiques mais aussi suivant les origines agro-écologiques. Les clusters obtenus dans notre étude se discriminent bien suivant des caractères agro-morphologiques mais la vérification de relations entre les groupes obtenus et leur répartition suivant les zones agro-écologiques du Sénégal n'indique pas de relation chez les Souna et faible corrélation chez les Sanio.

Cette étude a aussi révélé que la plupart des caractères qui interviennent dans la différenciation des variétés tardives, interviennent aussi dans la différenciation des variétés précoces. Ces caractères sont le rendement en épi et la longueur de la feuille paniculaire. Cependant il existe des différences qui sont la date de floraison qui intervient uniquement dans la différenciation des variétés tardives et non chez les variétés précoces et d'autres paramètres qui sont le rendement en grains, l'incidence du mildiou, la hauteur de la plante, le nombre d'entrenœuds, la largeur de la feuille paniculaire, la longueur et la circonférence de l'épi, qui discriminent uniquement les accessions précoces.

La subdivision des accessions précoces en trois groupes est en accord avec l'étude de Sy et al. (2015) où, trois sous-groupes ont été trouvés. Parmi les caractères qui permettent la différenciation des accessions précoces dans notre étude, quatre (4) ont été cités comme caractères discriminants dans l'étude de Sy et al. (2015), qui sont le rendement en épi et en grains, la longueur de l'épi et la hauteur de la plante. Cependant dans notre étude, l'incidence du mildiou intervient dans la différenciation des accessions précoces. Ce qui n'est pas le cas dans l'étude de Sy et al. (2015) qui en plus ont cité la date de floraison comme un caractère discriminant. Dans notre étude, la date de floraison n'est pas impliquée dans la différenciation des accessions précoces mais elle intervient dans celle des accessions tardives. Or l'étude de Sy et al. (2015) incluait 9 accessions tardives. Ceci pourrait certainement expliquer la différence observée entre notre étude et celle de Sy et al. (2015), compte tenu de la date de floraison. Sy et al. (2015) ont aussi cité deux autres caractères (nombre de tiges productives et la biomasse) qui ne font pas parties des caractères agro-morphologiques que nous avons utilisés dans notre

étude. Mis à part quelques différences observées entre notre étude et celle de Sy et al. (2015), que sont les deux paramètres (sensibilité au mildiou et la date de floraison) non partagés dont l'un intervient dans la différenciation en groupes dans notre étude et non dans celle de Sy et al. (2015) et l'autre inversement. Il y a beaucoup de similarités entre les deux études, à savoir la subdivision en trois groupes et les caractères agro-morphologiques partagés, qui interviennent dans la différenciation de ces groupes. Cependant, il est tout à fait normal de noter des différences entre les deux études du fait que la résolution des deux études n'est pas la même. L'étude de Sy et al. (2015) s'est focalisée sur une part des accessions qui ont fait l'objet de notre étude. En plus de cela, ce ne sont pas les mêmes caractères agro-morphologiques qui ont été utilisés dans les deux études. Sans compter le fait que l'étude de Sy et al. (2015) incluait 9 accessions tardives.

Dans notre étude, le caractère de date de floraison ne s'est pas particulièrement différencié. Il s'est différencié uniquement au niveau des accessions tardives. Car ce paramètre est le caractère principal du cycle de développement. Or dans notre étude, les accessions à cycle précoce et celles à cycle tardif ont été évaluées séparément. L'objectif était d'évaluer les Souna et les Sanio dans leurs différentes zones de production. Les variétés précoces de mil au Sénégal sont principalement produites entre le nord et le centre sud du pays, tandis que les variétés tardives sont essentiellement produites du sud au centre sud. Cependant, il est évident que le paramètre de date de floraison se serait significativement différencié, si les différentes variétés avaient été évaluées ensembles.

L'étroitesse géographique du Sénégal n'a pas empêché d'observer une grande variabilité des variétés précoces et tardives du mil, même lorsqu'elles sont évaluées séparément. Ce qui témoigne de la grande diversité du mil au Sénégal et confirme ainsi les mêmes observations déjà faites sur le plan génétique. Ce qui est tout à fait compréhensible étant donné que le Sénégal serait proche de l'origine géographique de domestication du mil cultivé (Dussert et al., 2015a). Cette diversité doit être exploitée en réponse aux besoins des agriculteurs à travers les programmes de sélection et d'amélioration variétales (Pucher et al., 2015). Ce qui justifie l'importance de la valoriser après une bonne connaissance de celle-ci en la mettant à disposition et en facilitant son utilisation par la mise en place de collections de références.

L'analyse des interactions GxE et de la stabilité du rendement et du paramètre de date de floraison par rapport aux différents environnements testés, a révélé que le paramètre de rendement est moins sensible aux variations d'environnements que la date de floraison aussi bien chez les accessions précoces que tardives. Certainement à cause du fait que, cette variation

de la date de floraison, en fonction des environnements (un cas de plasticité phénotypique), pourrait avoir pour finalité d'accroître la stabilité du rendement de la plante (Hausmann et al., 2007). En effet, si l'on considère un cas de figure où la plante fleurit plus tôt que d'habitude, du fait d'un changement d'un facteur quelconque (durée d'ensoleillement journalier par exemple) c'est juste pour pouvoir boucler son cycle à temps et donc produire ses graines. Dans ce supposé cas de figure, la plante adapte ainsi sa date de floraison (qui change) pour maintenir sa productivité.

Concernant ce paramètre de date de floraison, on note des accessions qui présentent particulièrement plus de sensibilité par rapport aux différents environnements, que d'autres. C'est-à-dire que la date de floraison de ces accessions varie plus fortement, en fonction de l'environnement, que les autres accessions. Cela pourrait être expliqué par le fait que ces accessions soient probablement sensibles à la photopériode. D'autant plus que cette instabilité est d'avantage notée chez les accessions tardives (Sanio) qui tendent généralement à être plus sensibles à la photopériode que les accessions précoces (Souma) (Hausmann et al., 2007)

Cette analyse AMMI a permis de dresser une liste permettant de classer les accessions de la plus stable à celle la moins stable par rapport aux différents environnements. Ce qui peut s'avérer très utile en termes de support pour la sélection d'individus stables « à large adaptation ». D'ailleurs, la stabilité est l'un des indices les plus précieux dans le cadre de programmes de sélection variétale (Sabaghnia et al., 2008). En effet, les accessions présentant une grande stabilité pourraient mieux convenir dans des environnements où les conditions climatiques sont incertaines, comme présentant par exemple, de fortes variations interannuelles de la pluviométrie. D'ailleurs cette variabilité interannuelle du climat est caractéristique même de la zone Sahélienne.

CONCLUSION

Cette étude a statué sur la diversité agro-morphologique des mil du Sénégal. Elle a permis de constater une grande diversité agro-morphologique des mils cultivés du Sénégal et une structure en trois sous-groupes agro-morphologiques des variétés précoces et trois sous-groupes aussi des variétés tardives. Elle a permis d'identifier 9 caractères agro-morphologiques qui permettent la subdivision des accessions précoces et trois caractères qui discriminent les accessions tardives. Elle a permis d'évaluer les performances des différents sous-groupes des accessions précoces et tardives par rapport aux caractères agro-morphologiques discriminants. Elle a aussi permis de différencier les accessions à large adaptation du reste des accessions, par rapport aux différents environnements testés, en dressant une liste des accessions classées par ordre de stabilité par rapport aux différents environnements, pour chaque type variétal.

Cette étude a donc permis de mieux connaître le matériel génétique des mils cultivés du Sénégal. Elle fournit des informations intéressantes sur ce matériel. Informations qui lorsqu'elles sont combinées aux résultats des données génétiques du chapitre précédant, constituent un support d'une grande valeur en vue d'une bonne gestion et valorisation de ce matériel. Support qui pourrait par exemple servir à assister les programmes de sélection variétale dans le cadre d'une valorisation des ressources de mil du Sénégal.

Cependant la grande taille de cette collection pourrait constituer un facteur pouvant compliquer cette valorisation. Il serait donc intéressant de développer une collection de référence qui renfermerait un plus faible nombre d'accessions mais qui serait représentative de cette collection nationale.

Chapitre III : Constitution d'une core collection de mil du Sénégal



INTRODUCTION

La valorisation est le but ultime de toute activité de gestion et conservation des ressources génétiques. En effet, il y a une succession d'étapes dans la gestion des ressources phytogénétiques, qui va de la collecte à la valorisation, en passant par l'évaluation de ces ressources. Cependant, l'un des facteurs limitants dans ce continuum de gestion des ressources (collection-évaluation-valorisation) est la taille des collections. En effet il existe un déphasage entre le but des gestionnaires des ressources et celui de certains acteurs de leur valorisation ou leurs utilisateurs potentiels, comme par exemple les sélectionneurs. Ces derniers ont besoin d'un accès facile à un nombre réduit d'accessions réunissant les caractères d'intérêt pour leurs programmes de sélection. Tandis que le responsable de collection a pour mission de conserver le maximum de variabilité possible de ces ressources.

C'est l'une des raisons pour lesquelles Frankel et Brown, en 1984 ont proposé l'élaboration de core collections qui regroupent le plus faibles nombres possibles d'accessions, non redondantes, et qui soient assez représentatives de la diversité génétique de toute la collection d'origine.

Ainsi donc, plusieurs stratégies d'échantillonnages d'accessions dans les collections pour la constitution de core collections, ont été proposées. Brown (1989a) a d'abord proposé l'échantillonnage randomisé (R-strategy), basé sur le model d'allèles neutres (Kimura and Crow 1964) qui a montré qu'un choix au hasard de 10% d'accessions d'une collection, peut exprimer jusqu'à 80% de sa diversité, plus ou moins 5%. Puis d'autres stratégies ont suivi au fil des années. Brown, durant la même année a proposé une manière plus élaborée d'échantillonner au sein des collections pour constituer les core collections. Il s'agit de la S-Strategy (Brown 1989b) qui est une procédure de stratification (groupage) arbitraire suivant des données agromorphologiques, ou des données d'origines géographiques, ou autres, avant l'échantillonnage de la core collection. Autrement dit, il s'agit d'aller échantillonner dans la structure des collections pour constituer les core collections. Il y a eu plusieurs familles de S-Strategy qui se différencient par la manière de l'échantillonnage après stratification, qui ont été proposées. On peut citer par exemple, l'échantillonnage randomisé, proportion du nombre d'accessions par groupe (P-Strategy), l'échantillonnage randomisé d'un nombre constant d'accessions dans chaque groupe (C-Strategy), l'échantillonnage fonction du logarithme du nombre d'individus par groupe (L-Strategy) et l'échantillonnage sur la base d'un scoring en composantes principales (PCS-Strategy) (Noirot et al., 1996).

Jusqu'en 1993, les stratégies d'échantillonnages étaient principalement basées sur le nombre d'individus par groupe dans la collection. A partir de 1993, un nouveau concept qui prend en compte essentiellement la diversité a été développé. A savoir, l'échantillonnage en fonction de la diversité de Nei (H-Strategy) et l'échantillonnage de combinaison d'accessions permettant de maximiser le nombre d'allèles observés au niveau des marqueurs tout en gardant le plus faible nombre d'individus possibles (M-Strategy) proposés par Schoen et Brown, (1993).

Diverses études sur l'efficacité de ces différentes méthodes d'échantillonnages ont été menées. La quasi-totalité de ces études a montré que la stratégie de maximisation (M-Strategy) est la plus puissante pour sélectionner des accessions qui gardent le plus de diversité et le moins de redondance (Bataillon et al., 1996 ; Hamon et al., 1994; Kim et al., 2007 ; Schoen and Brown, 1993).

Cette méthode (M-Strategy) a été programmée dans des algorithmes implémentés dans des logiciels comme MStrat (Gouesnard et al., 2001) pour l'établissement de core collection. Cependant son application sur des collections posait parfois un problème de reproductibilité, jusqu'à plus récemment. En effet lors qu'on applique la M-Strategy sur une collection donnée à un temps T_0 , on tombe sur un résultat R_0 . À un temps T_1 , si on applique cette M-Strategy sur cette même collection, on tombe sur un résultat R_1 qui peut être différent de R_0 .

En 2007, Kim et al. (2007) ont proposé un algorithme utilisant la stratégie de maximisation (M-Strategy) avancée, avec comme nouveauté une recherche heuristique lors du choix des accessions de la core collection. Cet algorithme testé par (Kim et al. 2007) a été plus performant que celui de la stratégie de maximisation simple, car permettant une meilleure représentativité de la core collection et générant une liste d'accessions reproductible pour un même jeu de données. De sorte que si on applique l'approche heuristique sur une collection donnée à un temps T_0 , on tombe sur un résultat R . À un temps T_1 , T_2 ou T_3 si on applique cette approche heuristique sur cette même collection, on tombe toujours sur le même résultat R . Durant ces dernières années, cet algorithme a été utilisé pour l'élaboration de core collections de plusieurs spéculations dont le haricot (Moe et al. 2012), le blé (Dutta et al. 2015), le riz (Tiwari et al. 2015), l'éléusine (Chandrashekhar et al. 2012), le sésame (Frery et al., 2015) et le millet des oiseaux (Gowda et al. 2013).

Le deuxième chapitre de cette thèse a fait état de l'évaluation de la collection de mil du Sénégal avec la mise en évidence de deux pools génétiques et une subdivision en trois sous-groupes agro-morphologiques de chaque pool.

Objectif

L'objectif global de ce chapitre est de faciliter la valorisation des ressources génétiques de mil, avec comme objectif spécifique :

- d'établir une core collection de mil nationale, en utilisant la stratégie de maximisation avancée proposée par Kim et al. (2007) ;
- d'évaluer la représentativité de cette core collection par rapport à la collection nationale de mil.

MATERIEL ET METHODES

1. Constitution de la core collection de mil du Sénégal

Matériel végétal

Le matériel végétal a été constitué de 305 accessions dans le pool des mils précoces (Souna) et 74 accessions dans le pool des mils tardifs (Sanio). Ces accessions ont été échantillonnées à travers tout le pays (Fig.1) entre 1993 et 2014.

Méthodes

La core collection de mil du Sénégal a été construite à l'aide des données génétiques microsatellites (12 marqueurs moléculaires) et des données agro-morphologiques (12 caractères agro-morphologiques). Les procédures de laboratoire et de terrain qui ont permis d'obtenir ces données ont été décrites dans le chapitre précédant. Afin d'éviter tout biais dans les analyses, tous les échantillons présentant plus de 5% de données manquantes ont été enlevés. Une stratégie d'échantillonnage de maximisation (M-Strategy), avancée (Kim et al., 2007), a été appliquée sur les différents pools génétiques de mil de la collection, à l'aide du logiciel PowerCore 1.0 (Kim et al., 2007).

Dans chaque pool génétique, Il a d'abord été défini, trois « pré-core collections ». Une « pré-core collection » en se basant sur les données génétiques uniquement, puis une autre en se basant sur les données agro-morphologiques seules et enfin, une dernière, en se basant sur les données génétiques et agro-morphologiques combinées. Puis une comparaison a été faite dans le but de sélectionner la « pré-core collection » qui permet une meilleure représentativité de la collection d'origine.

2. Evaluation de la représentativité

2.1. Représentativité des « pré-core collections »

Afin de pouvoir choisir la meilleure « pré-core collection » dans chaque pool génétique, nous avons procédé à une évaluation de la représentativité de chaque « pré-core collection » définie.

Pour évaluer la représentativité des core collections définies, par rapport à la collection d'origine sur le plan génétique, la vérification du captage de la totalité des allèles de la collection de départ a été faite, puis une vérification de la significativité statistique des différences de l'indice de diversité génétique de Nei pour chaque « pré-core collection » par rapport à celle de la collection d'origine a été faite grâce à un test de Student. Enfin, une matérialisation de la couverture de l'échantillonnage de chaque « pré-core collection » par rapport à la collection

d'origine a été faite grâce à une analyse en coordonnées principales, à l'aide du logiciel Genalex 6.5 (Peakall and Smouse, 2012).

L'évaluation de la représentativité sur le plan agro-morphologique a été faite par le calcul du pourcentage de différence des moyennes (%MD) et du taux de coïncidence (%CR), suivant les formules ci-après :

$$\%MD = \left(\frac{M_s}{n} \right) \times 100$$

$$\%CR = \frac{1}{n} \sum_{i=1}^n \frac{Rcc(i)}{Rec(i)} \times 100$$

M_s : Le nombre de caractères agro-morphologiques de la core collection, présentant une moyenne significativement différente au seuil de 5%, de celle de la collection d'origine. La significativité est établie à l'aide d'un test t de Student.

n : Le nombre de caractères agro-morphologiques.

$Rcc(i)$: L'amplitude de la core collection par rapport aux caractère agro-morphologique i .

$Rec(i)$: L'amplitude de la collection de départ par rapport aux caractère agro-morphologique i .

En effet, du point de vu des caractères quantitatifs, Hu et al. (2000) ont proposé des critères d'évaluation qui considèrent qu'une core collection est représentative de la collection d'origine lorsque moins de 20% de ces caractères quantitatifs présentent des moyennes significativement différentes au seuil de 0,05 (%MD), entre la core collection et la collection d'origine et que le taux de coïncidence (%CR) retenu pour la core collection est supérieur à 80%. Nous avons aussi calculé le pourcentage de CV (%) pour chaque caractère agro-morphologique dans chaque « pré-core collection » définie, en comparaison à celui dans la collection d'origine, pour voir si les dispersions des valeurs de chaque caractère autour des moyennes sont biens reproduites.

2.2. Représentativité de la core collection

Après avoir évalué la représentativité de chaque « pré-core collection » définie, celle qui présente une meilleure représentativité a été choisie dans chaque pool génétique. Les deux « pré-core collections » choisies au niveau des deux pools génétiques ont été jointes pour constituer la core collection nationale de mil du Sénégal.

Par la suite, il a été vérifié la reconstitution des différents pools génétiques de la collection nationale dans cette core collection par analyse bayésienne à l'aide du logiciel STRUCTURE 2.3.3 (Pritchard et al., 2000), suivant le modèle d'admixture (Falush et al., 2003) pour un nombre de groupes simulés allant de 1 à 5. Avec une période de burns de 100 000 et 10 runs de 200 000 itérations de tirages et remises par la méthode Chaines de Markov Monte Carlo (MCMC) pour chaque valeur de K. Les valeurs du nombre de clusters (K) ont été évaluées selon la méthode d'Evanno et al. (Evanno et al., 2005) par le critère (ΔK) et le logarithme de vraisemblance ($\ln P(D|K)$).

Ensuite, pour vérifier la couverture des différents sous-groupes agro-morphologiques, il a été procédé à une projection de ces différents sous-groupes obtenus grâce à l'analyse factorielle discriminante tout en identifiant dans le plan factoriel les différentes accessions de la core collection. Cette projection a été réalisée à l'aide du Package ade4 (Thioulouse et al., 1997) avec le logiciel R (R Development Core Team, 2008).

Enfin, il a été vérifié la représentativité géographique de la core collection grâce à une projection sur la carte administrative du Sénégal en comparaison à la carte de la collection nationale de mil. Cette projection a été faite à l'aide du logiciel Quantum GIS v 2.10.1 (QGIS Development Team, 2012).

RESULTATS

La stratégie d'échantillonnage de maximisation, appliquée aux accessions précoces d'une part et aux accessions tardives d'autre part sur la base des données agro-morphologiques, des données génétiques et la combinaison des données agro-morphologiques et génétiques a abouti à l'obtention de « pré-core collection » dans chaque pool génétique. Ces « pré-core collection » ont servi de base de sélection de la core collection de mil du Sénégal, en fonction de leurs représentativités.

1. « Pré-core collection » du pool génétique des variétés précoces (Souna)

La « pré-core collection » Souna définie à l'aide des données génétiques et agro-morphologiques combinées est constituée de 60 accessions soit 19,67% de la collection de départ. L'analyse de la représentativité de cette « pré-core collection » par rapport à la collection de départ montre que la « pré-core collection » a reproduit 100% de la richesse allélique de la collection de départ. L'indice de diversité de Nei de la pré-core collection est de 0,62 contre 0,56 pour la collection de départ. Le test (test de Student) de comparaison des indices de diversité a révélé qu'il n'y avait pas de différence significative entre la collection de base et cette « pré-core collection » ($p\text{-value} = 0,398$). Ce qui témoigne de sa bonne représentativité sur le plan génétique. Cette représentativité est matérialisée par l'analyse en coordonnées principales des données génétiques qui montre un échantillonnage qui couvre très bien la collection d'origine (Fig. 28-a). La même observation est faite, compte tenu des paramètres agro-morphologiques, où la valeur de pourcentage de différence des moyennes (%MD) est égale à 1,66% et celle du taux de coïncidence (%CR) est égale à 96,74%. Ce qui veut dire que dans cette « pré-core collection », il y a aussi une bonne représentativité des paramètres agro-morphologiques de la collection d'origine.

La « pré-core collection » Souna définie à l'aide des données génétiques seules est constituée de 45 accessions soit 14,75% de la collection d'origine. L'analyse de la représentativité de cette « pré-core collection » par rapport à la collection de départ montre qu'elle a reproduit 100% de la richesse allélique de la collection de départ. L'indice de diversité (H_{EX}), de la « pré-core collection » est de 0,63 contre 0,56 pour la collection de départ. Le test de comparaison des indices de diversité a révélé qu'il n'y avait pas de différence significative entre la collection de base et cette « pré-core collection » ($p\text{-value} = 0,282$). Il y a donc une bonne représentativité de cette « pré-core collection » sur le plan génétique. L'analyse en coordonnées principales des données génétiques montre un échantillonnage qui couvre assez-bien la collection d'origine

(Fig. 28-b) mais pas autant que la « pré-core collection » définie à l'aide des données génétiques et agro-morphologiques combinées. Le pourcentage de différences des moyennes et le taux de coïncidence (% MD = 0%, % CR = 67,14%) indiquent une non représentativité des caractères agro-morphologiques.

La « pré-core collection » Souna définie à l'aide des données agro-morphologiques seules, est constituée de 25 accessions soit 8,19% de la collection de base. L'analyse de la représentativité de cette « pré-core collection » par rapport à la collection de départ montre qu'elle a reproduit 67,81% de la richesse allélique de la collection de départ. Ce qui n'est pas assez-représentatif comparée aux autres « pré-core collection ». L'indice de diversité (H_{EX}), de cette « pré-core collection » est de 0,56 contre 0,56 pour la collection de départ. Le test de comparaison des indices de diversité a montré qu'il n'y avait pas de différence significative (p -value= 0,775). L'analyse en coordonnées principales des données génétiques montre un échantillonnage qui ne couvre pas assez la collection d'origine (Fig. 28-c). Cependant, nous notons une bonne représentativité des caractères agro-morphologique (%MD= 3,2%, %VR= 96,08%). Cette « pré-core collection » permet une bonne représentativité agro-morphologique de la collection d'origine, mais elle n'est pas représentative de sa diversité génétique.

Comparaison faite des trois « pré-core collections », nous pouvons facilement observer que la « pré-core collection » définie à partir des données agro-morphologiques et génétiques combinées est la meilleure. Elle permet une représentativité à la fois des paramètres génétiques et agro-morphologiques.

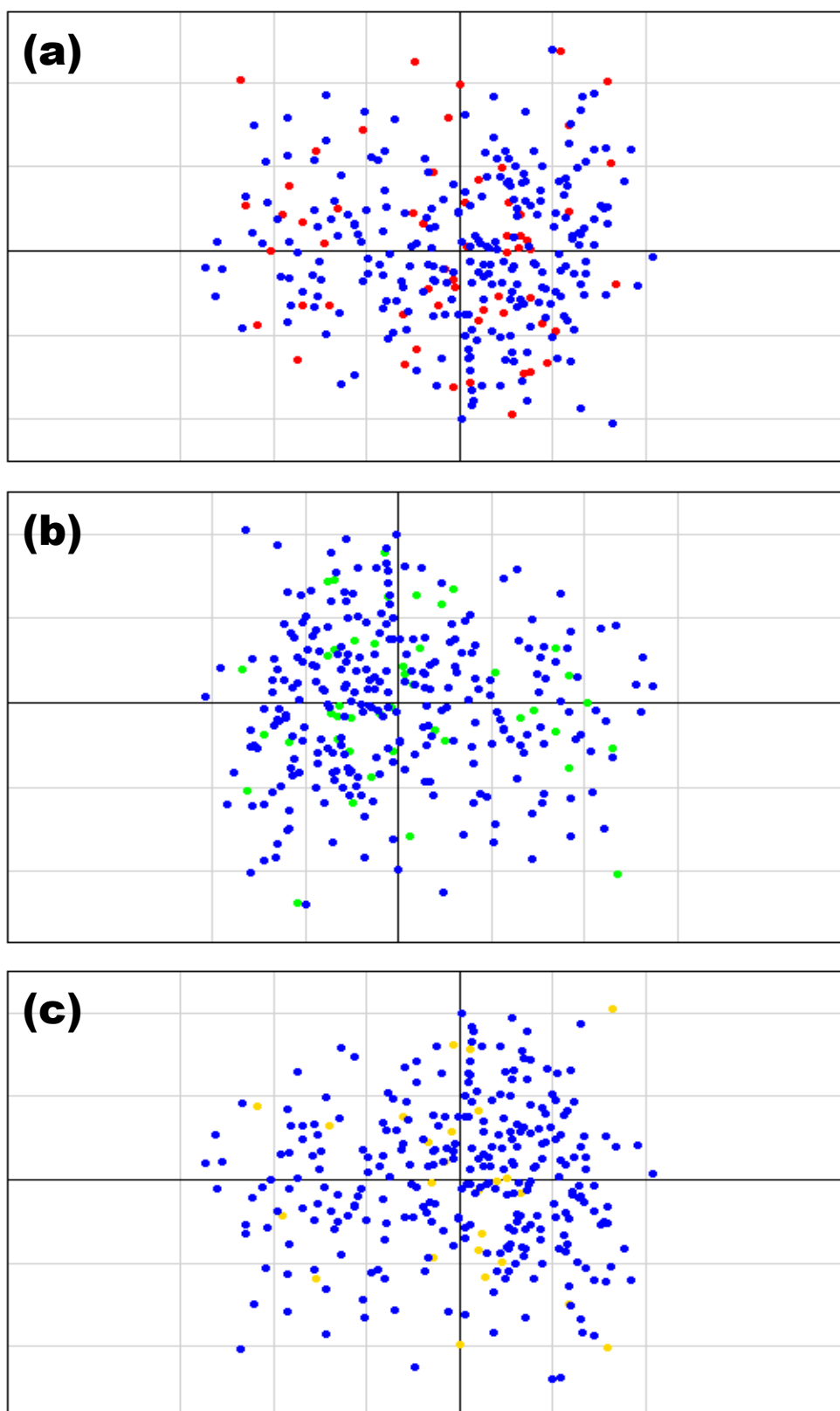


Figure 28 : Analyse en coordonnées principales de la « pré-core collection » définie sur la base de données génétiques et morphologiques combinées (a), sur la base de données génétiques seules (b) et sur la base de données morphologiques seules (c), en comparaison au pool génétique des Souna (Axe 1 = 25.08%, Axe 2 = 17.89%).

Tableau 8: Valeurs de nombre d'allèles (Na), d'hétérozygotie attendue (H_{EX}) et observée (H_O) des accessions précoces (Souna) et des pré-core collections du pool génétique des Souna

Marqueurs	Toutes les accessions précoces			Pré-core collection Morpho			Pré-core collection Génétique			Pré-core collection Génétique & Morpho		
	Na	H _O	H _{EX}	Na	H _O	H _{EX}	Na	H _O	H _{EX}	Na	H _O	H _{EX}
<i>PsmP 2202</i>	2	0,528	0,499	2	0,400	0,499	2	0,644	0,496	2	0,550	0,499
<i>PsmP 2206</i>	9	0,741	0,772	8	0,720	0,783	9	0,644	0,809	9	0,633	0,801
<i>PsmP 2214</i>	10	0,515	0,677	5	0,480	0,650	10	0,556	0,740	10	0,567	0,735
<i>PsmP 2219</i>	11	0,557	0,550	7	0,560	0,531	11	0,578	0,603	11	0,633	0,627
<i>PsmP 2220</i>	13	0,826	0,861	10	0,800	0,858	13	0,750	0,863	13	0,814	0,874
<i>PsmP 2227</i>	5	0,574	0,636	4	0,625	0,593	5	0,568	0,689	5	0,627	0,657
<i>PsmP 2237</i>	8	0,293	0,373	4	0,261	0,367	8	0,350	0,563	8	0,309	0,498
<i>PsmP 2247</i>	4	0,459	0,459	4	0,560	0,544	4	0,511	0,487	4	0,467	0,474
<i>PsmP 2249</i>	3	0,132	0,406	3	0,080	0,386	3	0,067	0,466	3	0,117	0,496
<i>IRD12</i>	12	0,283	0,669	7	0,333	0,749	12	0,308	0,790	12	0,300	0,779
<i>IRD25</i>	10	0,328	0,358	5	0,160	0,187	10	0,422	0,537	10	0,333	0,436
<i>IRD46</i>	5	0,531	0,533	4	0,720	0,593	5	0,511	0,510	5	0,617	0,553

Tableau 9 : Statistiques de moyenne, d'écart type (E) de coefficients de variation (CV) des caractères agro-morphologiques des accessions précoces (Souna) et des pré-core collections du pool génétique des Souna

Caractères	Toutes les accessions précoces				« Pré-core collection » Génétique & Morpho				« Pré-core collection » Génétique				« Pré-core collection » Morpho			
	Min - Max	Moyenne	E	CV	Min - Max	Moyenne	E	CV	Min - Max	Moyenne	E	CV	Min - Max	Moyenne	E	CV
DM	2,08-42,67	15,08	5,53	0,37	2,08-42,67	15	7,39	0,49	5,51-27,48	15,1	5,73	0,38	2,08-42,67	14,97	8,53	0,57
FLO	52,83-65,92	57,09	2,23	0,04	52,83-65,92	57,77	2,93	0,05	52,83-61,75	56,92	2,37	0,04	53,92-65,92	58,8	3,18	0,05
NTP	2,59-6,42	4,46	0,5	0,11	2,59-6,42	4,51	0,64	0,14	3,41-5,59	4,56	0,47	0,1	2,59-6,42	4,29	0,8	0,19
HTP	193,24-270,47	235,84	10,83	0,05	193,24-267,64	234,41	14,99	0,06	214,47-267,64	234,14	10,37	0,04	193,24-265,08	233,68	19,33	0,08
NEN	7,11-9,06	7,98	0,35	0,04	7,22-9,06	8,04	0,44	0,05	7,22-8,87	7,97	0,38	0,05	7,27-8,91	8,1	0,5	0,06
LOFP	40,09-57,74	47,41	2,57	0,05	40,09-57,74	47,57	3,27	0,07	40,09-54,08	47,11	2,96	0,06	41,8-57,74	47,81	3,68	0,08
LAEP	3,82-5,59	4,71	0,26	0,06	4-5,42	4,68	0,32	0,07	4-5,42	4,72	0,29	0,06	4,01-5,59	4,7	0,41	0,09
LEP	34,91-70,9	53,06	4,59	0,09	34,91-70,9	51,73	5,77	0,11	41,93-61,15	52,11	4,02	0,08	34,91-70,9	50,81	7,72	0,15
CIR	6,61-10,64	8,44	0,44	0,05	6,61-10,64	8,45	0,65	0,08	7,74-9,26	8,41	0,39	0,05	6,61-10,64	8,62	0,89	0,1
PMG	5,63-8,34	6,95	0,37	0,05	5,78-8,34	6,92	0,42	0,06	5,78-7,8	6,94	0,37	0,05	5,79-8,34	6,94	0,55	0,08
RDT_G	440,72-2195,73	1537,51	227,4	0,15	440,72-2161,27	1427,39	300,2	0,21	815,09-2161,27	1502,72	229,91	0,15	440,72-2016,33	1414,28	416,5	0,29
RDT_E	1174,56-3499,66	2557,74	319,79	0,13	1174,56-3499,66	2453,56	430,69	0,18	1850,58-3499,66	2534,13	336,78	0,13	1174,56-3302,48	2417,42	571,28	0,24

2. « Pré-core collection » du pool génétique des variétés tardives (Sanio)

La « pré-core collection » Sanio définie à l'aide des données génétiques et agro-morphologiques combinées est constituée de 31 accessions soit 40,25% du pool génétique de départ. L'analyse de la représentativité de cette « pré-core collection » par rapport à la collection de départ montre que la « pré-core collection » a reproduit 100% de la richesse allélique de la collection de départ. L'indice de diversité de Nei, de la « pré-core collection » est de 0,54 contre 0,50 pour la collection de départ. Le test (test de Student) de comparaison des indices de diversité a révélé qu'il n'y avait pas de différence significative entre la collection de base et cette « pré-core collection » ($p\text{-value} = 0,774$). Ce qui témoigne de sa bonne représentativité sur le plan génétique. De même l'analyse en coordonnées principales des données génétiques montre un échantillonnage qui couvre bien la collection d'origine (Fig. 29-a). Nous observons aussi une bonne représentativité des caractères agro-morphologiques de cette « pré-core collection » par rapport à la collection d'origine (%MD= 1,28%, %CR= 93,77%).

La « pré-core collection » Sanio définie à l'aide des données génétiques seules est constituée de 27 accessions soit 36,48% de la collection d'origine. Cette « pré-core collection » a reproduit 100% de la richesse allélique de la collection de départ. L'indice de diversité (H_{ex}), de la « pré-core collection » est de 0,53 contre 0,50 pour la collection de départ. Le test de comparaison des indices de diversité a révélé qu'il n'y avait pas de différence significative entre la collection de base et cette « pré-core collection » ($p\text{-value} = 0,747$). Il y a donc une bonne représentativité de cette « pré-core collection » sur le plan génétique. L'analyse en coordonnées principales des données génétiques montre un échantillonnage qui couvre assez-bien la collection d'origine (Fig. 29-b). Cependant, la « pré-core collection » définie à l'aide des données génétiques et agro-morphologiques combinées offre une meilleure couverture. Le pourcentage de différence des moyennes (% MD = 0%) et le taux de coïncidence (%CR = 87,63%) indiquent une représentativité acceptable des caractères agro-morphologiques.

La « pré-core collection » Sanio définie à l'aide des données agro-morphologiques seules, est constituée de 17 accessions soit 22,97% de la collection de base. L'analyse de la représentativité de cette « pré-core collection » par rapport à la collection de départ montre qu'elle a reproduit 77,61% de la richesse allélique de la collection de départ. Ce qui n'est pas assez-représentatif comparée aux autres « pré-core collections ». L'indice de diversité, de cette core collection est de 0,49 contre 0,50 pour la collection de départ. Le test de comparaison des indices de diversité

a montré qu'il n'y avait pas de différence significative ($p\text{-value} = 0,954$). L'analyse en coordonnées principales des données génétiques montre un échantillonnage qui ne couvre pas assez la collection d'origine (Fig. 29-c). Cependant, nous notons une bonne représentativité des caractères agro-morphologiques ($\%MD = 1,88\%$, $\%CR = 90,76\%$). Nous notons donc dans cette « pré-core collection », une bonne représentativité des paramètres agro-morphologiques de la collection d'origine, mais pas assez de représentativité de sa diversité génétique.

En comparaison des trois « pré-core collections », il ressort que la « pré-core collection » définie à partir des caractères agro-morphologiques et données génétiques combinés est la meilleure et permet une bonne représentativité des paramètres génétiques et agro-morphologiques.

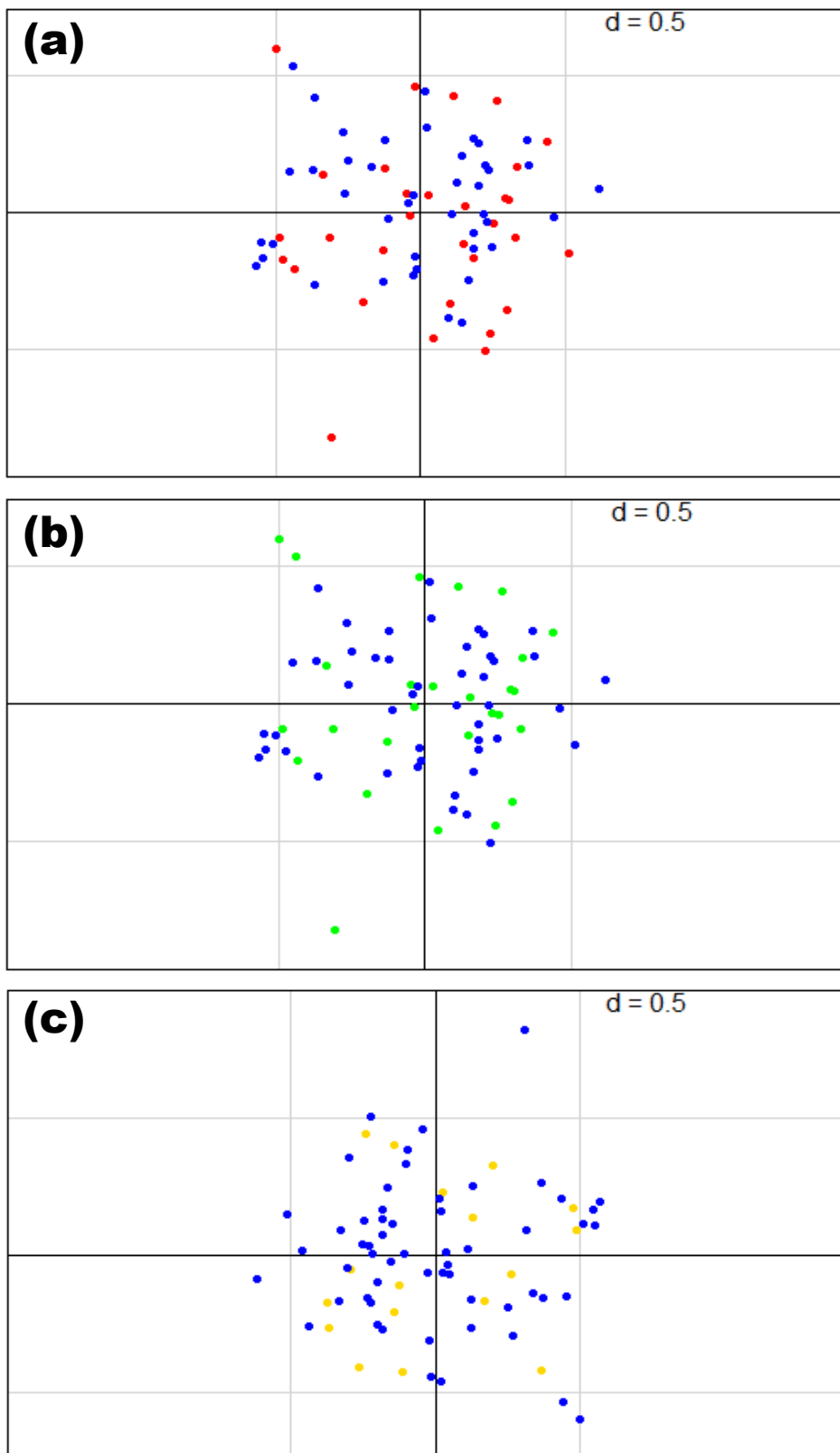


Figure 29 : Analyse en coordonnées principales de la « pré-core collection » définie sur la base de données génétiques et morphologiques combinées (a), sur la base de données génétiques seules (b) et sur la base de données morphologiques seules (c) en comparaison au pool génétique des Sanio (Axe 1 = 22.24%, Axe 2 = 18.24%).

Tableau 10 : Valeurs de nombre d'allèles (Na), d'hétérozygotie attendue (H_{EX}) et d'hétérozygotie observée (H_O) des accessions tardives (Sanio) et des pré-core collections du pool génétique des Sanio

	Toutes les accessions précoces			« Pré-core collection » Morpho			« Pré-core collection » Génétique			« Pré-core collection » Génétique & Morpho		
Marqueurs	Na	H _O	H _{EX}	Na	H _O	H _{EX}	Na	H _O	H _{EX}	Na	H _O	H _{EX}
<i>PsmP 2202</i>	2	0,243	0,272	2	0,235	0,208	2	0,222	0,302	2	0,323	0,383
<i>PsmP 2206</i>	6	0,622	0,785	6	0,824	0,747	6	0,407	0,765	6	0,452	0,779
<i>PsmP 2214</i>	6	0,452	0,595	6	0,588	0,637	6	0,556	0,694	6	0,516	0,653
<i>PsmP 2219</i>	8	0,712	0,727	5	0,813	0,666	8	0,769	0,764	8	0,700	0,766
<i>PsmP 2220</i>	12	0,685	0,840	10	0,813	0,840	12	0,731	0,849	12	0,733	0,844
<i>PsmP 2227</i>	6	0,608	0,702	4	0,529	0,701	6	0,593	0,704	6	0,581	0,707
<i>PsmP 2237</i>	2	0,014	0,013	1	0,000	0,000	2	0,037	0,036	2	0,032	0,032
<i>PsmP 2247</i>	3	0,149	0,161	2	0,294	0,251	3	0,148	0,201	3	0,129	0,177
<i>PsmP 2249</i>	2	0,176	0,263	2	0,118	0,208	2	0,185	0,278	2	0,161	0,248
<i>IRD12</i>	4	0,301	0,367	3	0,353	0,344	4	0,385	0,459	4	0,400	0,454
<i>IRD25</i>	9	0,568	0,620	5	0,588	0,663	9	0,630	0,709	9	0,613	0,693
<i>IRD46</i>	7	0,635	0,691	6	0,765	0,696	7	0,519	0,713	7	0,484	0,687

Tableau 11: Statistiques de moyenne, d'écart type (E) de coefficients de variation (CV) des caractères agro-morphologiques des accessions tardives (Sanio) et des « pré-core collections » du pool génétique des Sanio

	Toutes les accessions précoces				« Pré-core collection » Génétique & Morpho				« Pré-core collection » Génétique				« Pré-core collection » Morpho			
Caractères	Min - Max	Moyenne	E	CV	Min - Max	Moyenne	E	CV	Min - Max	Moyenne	E	CV	Min - Max	Moyenne	E	CV
DM	12,61-35,98	24,17	4,27	0,18	12,61-35,98	24,33	4,65	0,19	12,61-32,03	23,77	4,23	0,18	12,61-33,46	23,34	5,88	0,25
FLO	80,08-91,75	86,23	3,36	0,04	80,5-91,25	86,23	3,44	0,04	80,75-91,25	86,48	3,26	0,04	80,25-91,5	86,28	3,55	0,04
NTP	08-déc	9,81	0,79	0,08	8-11,58	9,94	0,87	0,09	8-11,58	9,91	0,89	0,09	8,4-11,58	9,66	0,87	0,09
HTP	309,58-361,42	335,63	10,76	0,03	309,58-361,42	335,22	11,16	0,03	309,58-361,42	336,03	10,93	0,03	310,05-361,42	336,4	11,67	0,03
NEN	12,48-14,45	13,48	0,46	0,03	12,6-14,45	13,56	0,48	0,04	12,83-14,45	13,64	0,46	0,03	12,48-14,38	13,42	0,5	0,04
LOFP	48,7-58,12	52,33	1,81	0,03	49,57-58,12	52,42	2	0,04	49,57-58,12	52,35	2,02	0,04	49,7-57,77	52,79	2,02	0,04
LAFP	3,56-4,2	3,89	0,14	0,04	3,56-4,2	3,87	0,16	0,04	3,56-4,2	3,86	0,16	0,04	3,56-4,14	3,92	0,16	0,04
LEP	34,82-48,25	40,13	2,88	0,07	35,53-46,52	40,54	2,68	0,07	35,53-45,3	40,14	2,32	0,06	35,58-46,53	40,41	3	0,07
CIR	1,38-1,72	1,55	0,07	0,04	1,38-1,68	1,54	0,06	0,04	1,42-1,68	1,55	0,06	0,04	1,42-1,72	1,56	0,07	0,05
PMG	5,35-7,08	6,2	0,37	0,06	5,52-6,87	6,15	0,33	0,05	5,52-6,87	6,22	0,33	0,05	5,49-6,85	6,17	0,4	0,07
RDT_G	511,37-1104,86	853,57	128,57	0,15	511,37-1102,32	809,02	136,69	0,17	511,37-1102,32	835,84	129,77	0,16	511,37-1104,86	837,04	161,65	0,19
RDT_E	1443,45-3550,03	2248,88	375,65	0,17	1443,45-3550,03	2197,09	456,56	0,21	1656,13-3550,03	2279,38	415,28	0,18	1462,77-3550,03	2264,47	476,57	0,21

3. Core collection (collection de référence) du germoplasme des mils cultivés du Sénégal

Les « pré-core collections » définies sur la base des données agro-morphologiques et génétiques combinées, sont les mieux représentatives de chaque pool génétique. Par conséquent, nous les avons combinées et considérées comme étant la collection de référence du germoplasme des mils cultivés du Sénégal (Tab. 12), définie à partir des données génétiques et agro-morphologiques combinées dans chaque pool génétique. Cette core collection est constituée de 91 accession soit 22% de la collection nationale de mil du Sénégal. Elle est constituée de 60 accessions précoces (Souna) et 31 accessions tardives (Sanio).

L'analyse bayésienne a montré que cette core collection reconstitue bien les différents pools génétiques de mil de la collection nationale (Fig. 30). La projection des analyses factorielles discriminantes des différents pools génétiques de collection révèle que la core collection couvre bien tous les sous-groupes agro-morphologiques de la collection car chaque sous-groupe est bien représenté dans cette core collection (Fig. 31).

Une projection de la distribution géographique de cette core collection montre aussi une bonne représentativité de l'échantillonnage géographique (Fig. 32), avec une couverture de toutes les régions où le mil est cultivé dans pays, avec leur diversité variétale. Toutes les zones de production du mil au Sénégal sont représentées dans la collection nationale, elles sont toutes aussi représentées dans la core collection nationale. Chaque type variétal de chaque région de la collection est aussi retrouvée dans le core collection.

Tableau 12 : Liste de la collection de référence du germoplasme des mils cultivés du Sénégal

Entrée	Echantillon	Groupe génétique	Sous-groupe agro-morphologique	Entrée	Echantillon	Groupe génétique	Sous-groupe agro-morphologique
1	N_101-14	Sanio	Sub-group_1	66	N_07-14	Souna	Sub-group_2
4	N_113-14	Sanio	Sub-group_1	67	N_09-14	Souna	Sub-group_2
9	N_121-14	Sanio	Sub-group_1	68	N_11-14	Souna	Sub-group_2
21	N_170-14	Sanio	Sub-group_1	82	N_177-14	Souna	Sub-group_2
23	N_175-14	Sanio	Sub-group_1	83	N_180-14	Souna	Sub-group_2
24	N_214-14	Sanio	Sub-group_1	85	N_187-14	Souna	Sub-group_2
25	N_223-14	Sanio	Sub-group_1	87	N_192-14	Souna	Sub-group_2
26	N_88-14	Sanio	Sub-group_1	88	N_196-14	Souna	Sub-group_2
27	N_91-14	Sanio	Sub-group_1	90	N_203-14	Souna	Sub-group_2
29	N_96-14	Sanio	Sub-group_1	91	N_211-14	Souna	Sub-group_2
31	N_99-14	Sanio	Sub-group_1	70	N_33-14	Souna	Sub-group_2
2	N_103-14	Sanio	Sub-group_2	79	N_71-14	Souna	Sub-group_2
3	N_106-14	Sanio	Sub-group_2	81	N_83-14	Souna	Sub-group_2
7	N_118-14	Sanio	Sub-group_2	32	PLS_01	Souna	Sub-group_2
11	N_132-14	Sanio	Sub-group_2	33	PLS_05	Souna	Sub-group_2
12	N_138-14	Sanio	Sub-group_2	34	PLS_100	Souna	Sub-group_2
15	N_146-14	Sanio	Sub-group_2	36	PLS_112	Souna	Sub-group_2
16	N_153-14	Sanio	Sub-group_2	42	PLS_160	Souna	Sub-group_2
20	N_167-14	Sanio	Sub-group_2	45	PLS_204	Souna	Sub-group_2
22	N_171-14	Sanio	Sub-group_2	46	PLS_209	Souna	Sub-group_2
28	N_95-14	Sanio	Sub-group_2	49	PLS_23	Souna	Sub-group_2
30	N_98-14	Sanio	Sub-group_2	50	PLS_31	Souna	Sub-group_2
5	N_115-14	Sanio	Sub-group_3	51	PLS_312	Souna	Sub-group_2
6	N_117-14	Sanio	Sub-group_3	52	PLS_318	Souna	Sub-group_2
8	N_119-14	Sanio	Sub-group_3	54	PLS_326	Souna	Sub-group_2
10	N_131-14	Sanio	Sub-group_3	56	PLS_48	Souna	Sub-group_2
13	N_139-14	Sanio	Sub-group_3	57	PLS_49	Souna	Sub-group_2
14	N_140-14	Sanio	Sub-group_3	58	PLS_50	Souna	Sub-group_2
17	N_155-14	Sanio	Sub-group_3	61	PLS_57	Souna	Sub-group_2
18	N_156-14	Sanio	Sub-group_3	62	PLS_61	Souna	Sub-group_2
19	N_160-14	Sanio	Sub-group_3	64	PLS_95	Souna	Sub-group_2
69	N_31-14	Souna	Sub-group_1	84	N_182-14	Souna	Sub-group_3
71	N_34-14	Souna	Sub-group_1	86	N_191-14	Souna	Sub-group_3
72	N_42-14	Souna	Sub-group_1	89	N_198-14	Souna	Sub-group_3
73	N_44-14	Souna	Sub-group_1	74	N_52-14	Souna	Sub-group_3
77	N_66-14	Souna	Sub-group_1	75	N_56-14	Souna	Sub-group_3
35	PLS_104	Souna	Sub-group_1	76	N_58-14	Souna	Sub-group_3
38	PLS_121	Souna	Sub-group_1	78	N_69-14	Souna	Sub-group_3
39	PLS_138	Souna	Sub-group_1	80	N_81-14	Souna	Sub-group_3
44	PLS_175	Souna	Sub-group_1	37	PLS_115	Souna	Sub-group_3
48	PLS_228	Souna	Sub-group_1	40	PLS_151	Souna	Sub-group_3
53	PLS_319	Souna	Sub-group_1	41	PLS_158	Souna	Sub-group_3
55	PLS_38	Souna	Sub-group_1	43	PLS_163	Souna	Sub-group_3
59	PLS_54	Souna	Sub-group_1	47	PLS_224	Souna	Sub-group_3
63	PLS_94	Souna	Sub-group_1	60	PLS_56	Souna	Sub-group_3
65	N_03-14	Souna	Sub-group_2				

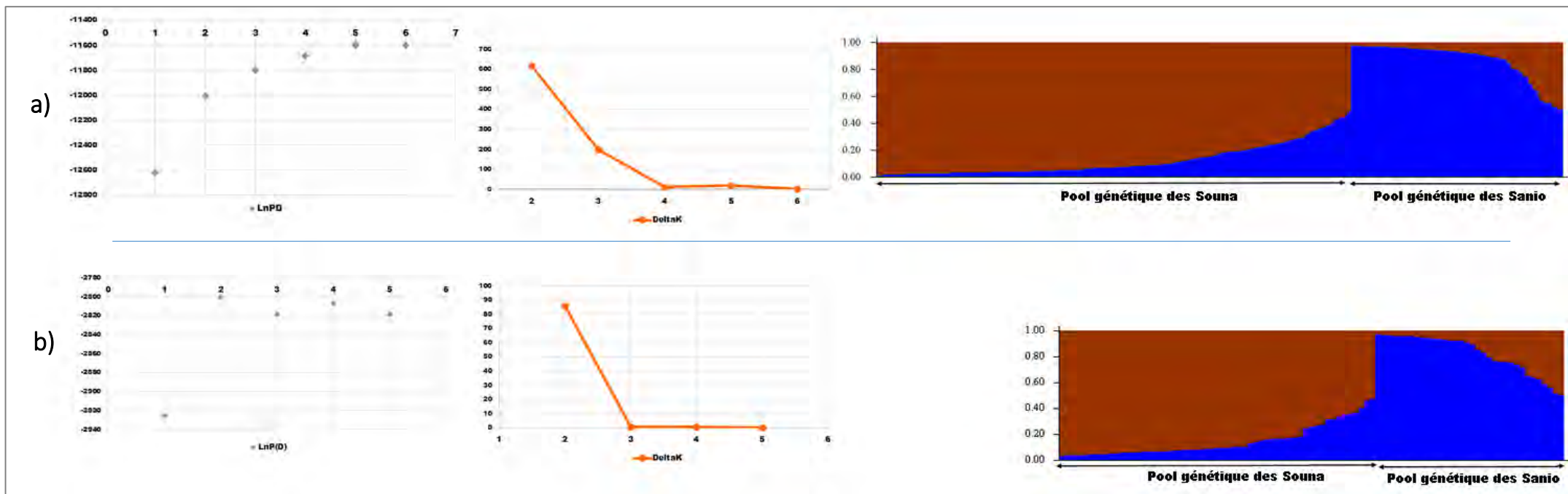


Figure 30: Analyses bayésiennes de la core collection de mil (b), en comparaison avec l'analyse bayésienne de la collection nationale (b).

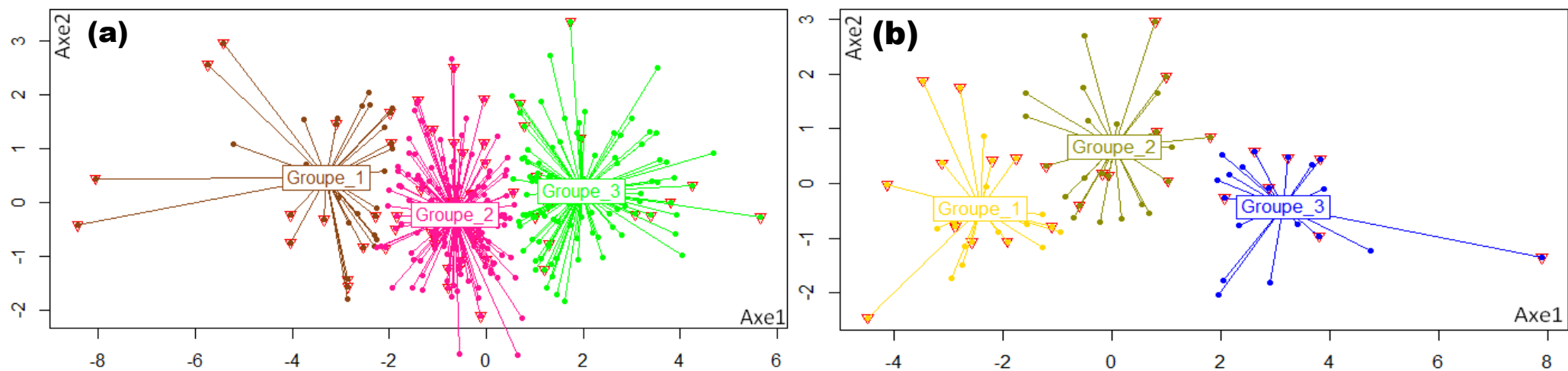


Figure 31 : AFD des (a) Souna et des (b) Sanio, à partir des données agro-morphologiques. Les accessions de la core collection sont encadrées par un triangle rouge

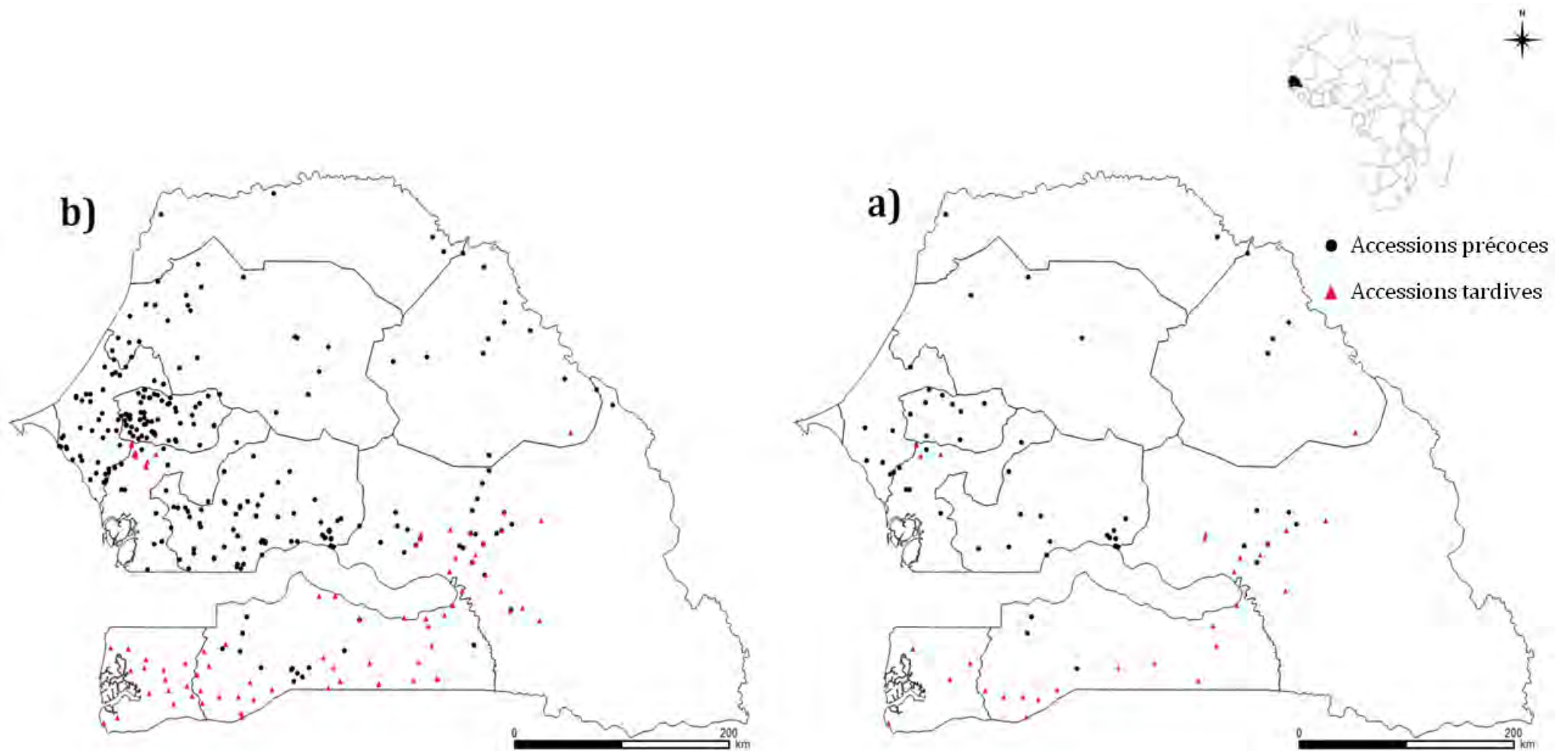


Figure 32 : Carte de la core collection de mil du Sénégal (a), en comparaison avec celle de la collection nationale (b)

DISCUSSION

Les « pré-core collections » définies sur la base des données agro-morphologiques et génétiques combinées ont été les mieux représentatives de chaque pool génétique de la collection de mil du Sénégal. Cette étude confirme davantage que les core collection définies sur la base des données génétiques et agro-morphologiques sont mieux représentatives des collections de départ que celles définies uniquement sur la base de données génétiques ou agro-morphologiques. En effet des résultats similaires ont été obtenus en comparant la représentativité de core collection définies sur la base de données moléculaires uniquement, de données agro-morphologiques uniquement et de données génétiques et agro-morphologiques combinées, sur le sésame (Frary et al., 2015).

La core collection définie sur la base des données agro-morphologiques et génétiques couvre 22,5% de la collection nationale de mil du Sénégal. Cette proportion est similaire à celle obtenue par Frary et al. (2015) (21%), lors de la mise en place d'une core collection de sésame en utilisant des données agro-morphologiques et moléculaires (marqueurs AFLP) et en appliquant la même stratégie que la nôtre. Selon Hintum et al. (2000) , aucune proportion de taille unique ou fixe pour une core collection n'est appropriée pour tous les cas, même s'il y a eu des études sur la bonne proportion d'une core collection. En effet, la fraction d'échantillons optimale dépend en grande partie du degré de redondance génétique des accessions, des ressources disponibles pour la maintenance des entrées de la collection de base et de la fréquence de régénération des entrées (Hintum et al., 2000). Il est cependant, à noter qu'il n'existe aucune redondance génétique des accessions de la collection de mil du Sénégal, qui plus est, a une taille relativement conséquente par rapport à une simple collection nationale, du fait de sa stratégie d'échantillonnage fine lors de sa constitution.

En plus, Hintum et al. (2000) ont publié une représentation de la taille de 37 collections à travers le monde et du pourcentage de leurs core collection respectives, à travers laquelle on peut y remarquer un grand nombre de collections dont le nombre d'entrées est compris entre 100 et 1000 accessions (comme dans notre cas), qui ont des « core collection » qui renferment plus de 30% voire même 40% d'accessions.

Par ailleurs, Selon Charmet et Balfourier, (1995) et Bisht et al. (1998) les tailles de 5 à 10% sont optimales, pour capturer 75 à 90% de la diversité. Il est donc certain qu'une proportion plus grande est nécessaire pour capturer 100% de la diversité. D'un autre côté, Noirot et al. (1996) ont suggéré que des pourcentages plus élevés que ceux de Charmet et Balfourier, (1995)

(20-30%) sont nécessaires, en particulier lorsque l'objectif est de capturer la diversité génétique des caractères hérités quantitativement (Hintum et al., 2000). Or, si nous considérons le fait que notre core collection ainsi définie pourrait être destinée à des études plus approfondies sur le mil, comme par exemple, une étude d'association, la diversité génétique des caractères hérités quantitativement serait alors plus qu'importante.

Compte tenu de toutes ces observations ci-dessus, notre core collection (qui a reproduit 100% de la diversité en gardant 22% de proportion) est dans la bonne fourchette d'un pourcentage optimum.

Cette core collection est représentative de la collection nationale de mil du Sénégal. L'évaluation de représentativité de la core collection a révélé une bonne représentativité de la collection de mil nationale, tant sur le plan agro-morphologique que génétique. En effet le captage des allèles de la collection par la core collection et le niveau de diversité génétique de la collection est statistiquement égal à celui de la core collection. Les indices de représentativité des caractères agro-morphologiques (%MD et %CR) sont conformes à une bonne représentativité, sans parler de la reconstitution des pools génétiques de la collection dans la core collection et de la couverture des différents sous-groupes agro-morphologiques de la collection de mil.

Cette core collection est représentative aussi sur le plan géographique, malgré la non prise en compte de l'origine géographique des accessions lors de son établissement. Balfourier et al. (1998) ont observé lors d'une étude de comparaisons de stratégies de mise en place de core collections testées sur deux espèces (*Lolium perenne* et *Medicago truncatula*), que les stratégies qui avaient utilisé des informations de structure spatiale, avaient produit des core collections les plus représentatives. Il semble qu'il y ait une relation entre la bonne représentation spatiale et la bonne représentation génétique. Ceci n'est en aucun cas surprenant car la génétique est souvent façonnée suivant la géographie. En effet il est tout à fait normal que des individus plus proches géographiquement soient plus enclins à partager plus d'allèles que des individus plus éloignés. Même si les tests d'isolement par la distance sur les Souna et les Sanio ne soient pas statistiquement significatifs, il n'empêche qu'il existe un pattern spatial de diversité des mils cultivés du Sénégal. Cette structure génétique, étroitement liée à la structure spatiale qui est dévoilée jusqu'à un certain point ou à un autre, suivant l'analyse faite (Bayésienne ou sPCA) est certainement responsable ou pour une bonne partie de la représentativité géographique, sans sa prise en compte lors de l'établissement de la core collection. Cette représentativité

géographique de la core collection est très intéressante, compte tenu de l'existence de ce pattern géographique de la diversité du mil cultivé au Sénégal.

A notre connaissance, cette étude est la première qui utilise la stratégie de maximisation avancée pour la mise en place d'une core collection de mil cultivé. Bhattacharjee et al, (2007) ont mis en place une core collection de mil de l'ICRISAT en utilisant la Random P-Stratégie à l'aide de données agro-morphologiques et des données d'origines géographiques. Cette core collection a par la suite été augmentée toujours sur la base de données agro-morphologiques et géographiques (Upadhyaya et al., 2009). Puis cette core collection nouvellement obtenue a été ensuite réduite en « mini-core collection » (Upadhyaya et al., 2011) toujours sur la base des données et stratégies qui avaient servi à mettre en place la première core collection. Toutes les évaluations faites de ces core collections les ont considérées comme représentatives.

Nous avons appliqué la M-Strategy sur la collection de mil du Sénégal. La core collection qui a été ainsi obtenue a été représentative du germoplasme des mils cultivés du Sénégal, sur les plans génétique, agro-morphologique et géographique. Cette étude montre ainsi l'efficacité de l'approche heuristique et ouvre une porte pour son application future pour la mise en place de core collections représentatives de mil cultivés.

A présent, il existe grâce à cette étude, une core collection nationale de mil du Sénégal, qui est représentative sur le plan génétique, agro-morphologique et géographique. Si l'on considère le fait que la conduite des collectes de cette collection a été réalisée suivant une stratégie à fine échelle et que les prospections se sont déroulées de sorte à pouvoir collecter la diversité même des ressources de mil du Sénégal. Ce qui a conduit à une bonne représentation de ces ressources de mil du pays dans la collection nationale. On peut aisément dire que la core collection est assez représentative des ressources de mil cultivé du Sénégal. A ce titre il est important de la valoriser et de l'utiliser dans les recherches plus approfondies sur le mil au Sénégal que ça soit dans le domaine de la génétique, génomique, ou autre, pour approfondir les connaissances sur l'histoire évolutive du mil. D'autant plus que l'on sait que les mils du Sénégal présentent une certaine particularité génomique (Hu et al., 2015) et génétique (Diack et al., 2017) par rapport à d'autres mils étudiés jusqu'à présent (Dussert et al., 2015b ; Mariac et al., 2006a). Des particularités qui confèrent à ces mils du Sénégal la propriété d'être un matériel intéressant pour l'étude de l'histoire évolutive des différentes formes de mils cultivés.

Afin que la mise en place de la core collection puisse avoir un sens et qu'elle puisse être valorisée, il serait aussi intéressant d'utiliser cette core collection dans les programmes de

sélection et d'amélioration variétale du mil au Sénégal d'autant plus que notre étude a permis de mieux connaître ces ressources.

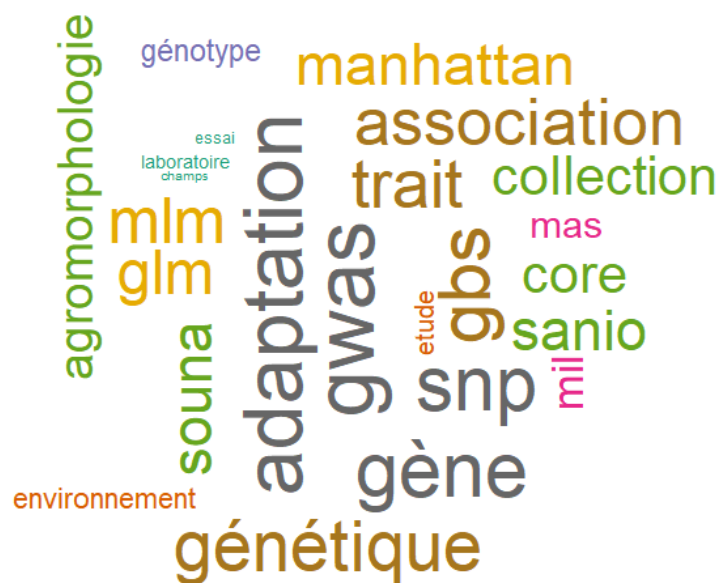
CONCLUSION

Cette étude Confirme que les core collections définies sur la base des données génétiques et agro-morphologiques sont meilleures et sont représentatives de la collection de mil du Sénégal. Elle a permis d'obtenir une core collection de mil dont l'évaluation a montré qu'elle est représentative de la collection de mil du Sénégal, tant sur le plan génétique, agro-morphologique que géographique.

Ainsi les sélectionneurs et utilisateurs potentiels (chercheurs ou autres) disposent d'un nombre réduit d'accessions représentatives de la collection pour des activités de recherches approfondies sur le mil, ou des programmes de sélection et d'amélioration du mil au Sénégal. Il serait donc intéressant d'utiliser cette core collection dans la recherche et/ou dans les programmes de sélections pour donner un sens à sa mise en place et dans le but d'assurer sa valorisation.

Ce pendant une bonne gestion des ressources et leur valorisation à travers un programme d'amélioration efficace, doivent aussi disposer d'outils génomiques, pour les appuyer, en plus de la connaissance du matériel génétique. Ainsi donc, il serait intéressant d'adjoindre à ce travail, une étude d'association génotype-phénotype, qui pourrait permettre d'identifier des régions génomiques associées à des traits phénotypiques ou adaptatifs importants. Ce qui sera l'objet de notre prochain chapitre.

Chapitre IV : Etudes d'associations génotypes-phénotypes sur la core collection de mil du Sénégal



INTRODUCTION

L'adaptation des plantes constitue une réponse par rapport aux changements environnementaux. Cette réponse est souvent perceptible sur le plan phénotypique. Une des réponses adaptatives les plus importantes, phénotypiquement perceptible, est la modification dans la date de floraison chez certaines plantes cultivées, outre la plasticité phénotypique. Ces différences de date floraison, surviendraient après sélections différentielles suivant des latitudes et des climats dont la réponse adaptive serait une modification de celle-ci (Olsen and Wendel, 2013). Les conditions climatiques locales peuvent donc constituer une pression sélective sur les gènes qui sous-tendent des traits comme cette date de floraison. En effet, chez plusieurs espèces de plantes, la phénologie de floraison varie fortement en fonction de gradients climatiques (Burgarella et al., 2016).

Il existe au Sénégal un gradient nord-sud de pluviométrie (Diop, 1999 ; Negre et al., 1988). Les régions du sud du pays ont une pluviométrie plus abondante que celles du nord où nous avons de très faibles valeurs pluviométriques. Ce gradient nord-sud serait aussi corrélé à la répartition géographique des variétés de mil cultivés dans le pays, qui se différencient principalement par leur cycle de floraison (Rao et al., 1986). Les variétés tardives (Sanio) sont rencontrées au sud, à la limite au centre-sud du pays et les variétés précoces se retrouvent plus au nord et au centre du pays.

L'identification des caractères phénotypiques à l'origine de la différenciation de ces morphotypes et l'association de ces caractères à des régions génomiques, permettrait d'identifier, par ricochet, les bases génétiques responsables de la différenciation entre Souna et Sanio. Ce qui constituerait un apport considérable dans la connaissance de l'évolution et de l'adaptation du mil au Sénégal et permettrait, à terme une sélection moléculaire efficace du mil au Sénégal, améliorant ainsi sa productivité et contribuer à assurer la sécurité alimentaire.

Le développement des nouvelles technologies de séquençage (NGS) a permis des avancées notoires dans le domaine de la génétique et de la génomique. Les données génomiques obtenues grâce à ces NGS, peuvent être combinées à des données phénotypiques, pour identifier leurs éventuelles associations à ces caractères phénotypiques, grâce à des modèles développés à cet effet. Cette technique est appelée étude d'association génotypes phénotypes ou GWAS (Genome-wide association study). Le GWAS utilise les déséquilibres de liaison pour associer des allèles de marqueurs à des traits phénotypiques. Cependant, il existe un certain nombre de phénomènes qui entraînent aussi des déséquilibres de liaison dont leurs non prise en compte

dans les analyses GWAS peuvent entraîner des estimations non réelles (biaisées) d'associations génotypes-phénotypes. Ce sont principalement l'existence de structure dans la population étudiée et l'apparentement entre individus de la population. Pour gérer ces phénomènes, on utilise des modèles destinés à corriger ces biais. Parmi ces modèles on peut citer principalement le modèle linéaire généralisé (GLM) qui permet de prendre en compte l'effet de la structure dans la population ou le modèle linéaire mixte (MLM) qui peut prendre en compte l'effet de la structure de la population et l'effet de l'apparentement entre individus de la population. Il existe d'autres modèles (comme EMMA : Efficient Mixed Model Association, CMLM : Compressed MLM, EMMAx : EMMA expedited, etc..) qui ne sont en réalité que des extensions de ces modèles précités, soit pour améliorer le temps d'exécution et/ou la puissance statistique.

Dans les chapitres précédents de cette thèse, il a été fait une évaluation complète des accessions de mil du Sénégal et une mise en place d'une core collection représentative de ces dernières dans le but de leur bonne gestion et valorisation. Identifier, au sein de cette core collection, des marqueurs moléculaires et/ou régions génomiques associées à des traits agro-morphologiques ou adaptatifs de ces mils serait intéressant en termes de connaissance de l'évolution et de l'adaptation du mil et pour leur bonne gestion, mais surtout pour leur valorisation. À terme, ces outils génétiques pourraient être utilisés pour assister les programmes de sélection et d'amélioration du mil au Sénégal.

Objectif

L'objectif global de ce chapitre est de contribuer à la bonne gestion des ressources génétiques de mil du Sénégal.

Spécifiquement, ce chapitre a pour but :

- d'identifier des traits agro-morphologiques de différenciation des morphotypes du mil au Sénégal (Souma et Sanio)
- de produire de nouveaux marqueurs moléculaires (SNP par exemple) chez le mil.
- de génotyper en même temps les accessions de la core collection en fonction de ces marqueurs
- d'identifier des régions génomiques associées à des traits agro-morphologiques ou adaptatifs du mil au Sénégal.

MATERIEL ET METHODES

1. Matériel végétal et phénotypage

1.1. Matériel végétal

Le matériel végétal est constitué de 91 accessions représentant la core collection de mil du Sénégal. La procédure ayant permis d'obtenir cette core collection a été décrite dans le chapitre précédant. Le matériel est donc constitué de 60 accessions précoces et 31 accessions tardives.

1.2. Site, facteurs et traitements

Les essais se sont déroulés au niveau de la station expérimentale de Nioro, de l'ISRA, durant l'hivernage 2016 et l'hivernage 2017. Chaque essai comportait un seul facteur : accessions. Le nombre d'accessions était de 91 (traitements).

1.3. Dispositif expérimental

Le dispositif expérimental pour chaque essai, était un 7 x 13 alpha lattice design, avec trois répétitions (Fig.33 et 34). Chaque parcelle élémentaire était constituée d'une ligne de 8 poquets. L'écartement entre lignes et entre poquets sur la ligne était de 90 cm. Deux lignes de bordure avaient été prévues de part et d'autre du champ pour éviter l'effet de bordure, à chaque essai.

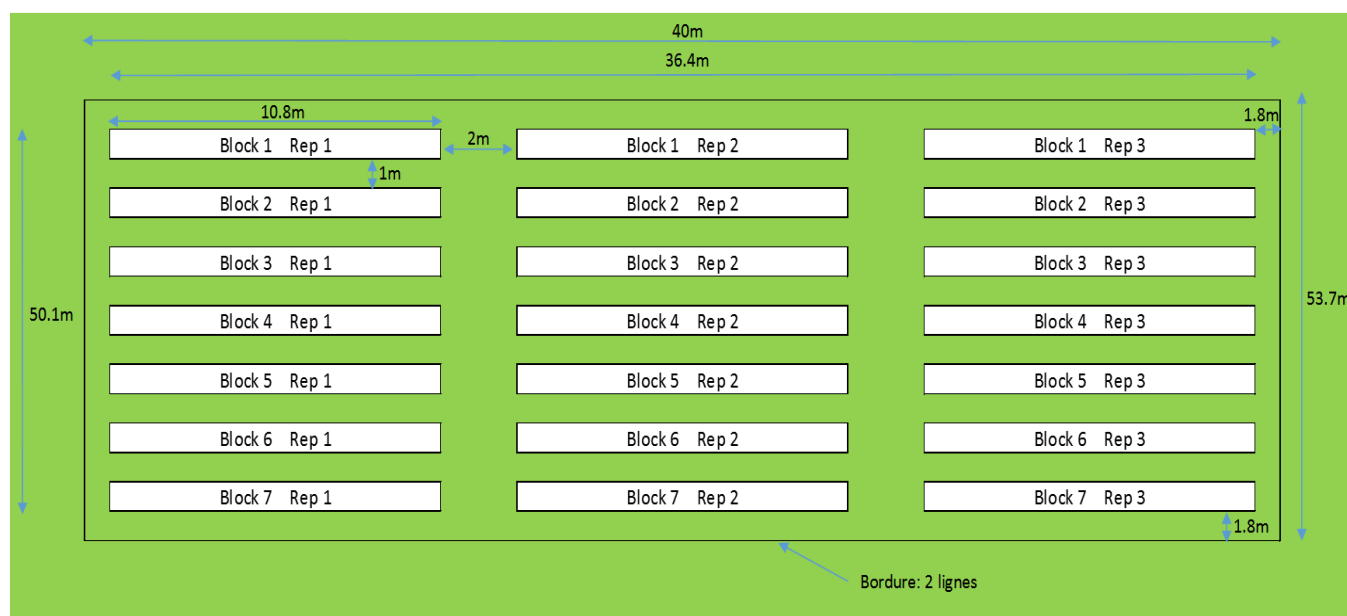


Figure 33 : Schéma illustrant le dispositif expérimental sur un site d'essai

		R E P E T I T I O N 1													R E P E T I T I O N 2													R E P E T I T I O N 3														
		P L O T													P L O T													P L O T														
		1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	3	4	5	6	7	8	9	10	11	12	13		
B	1	1	2	3	4	5	6	7	8	9	10	11	12	13	92	93	94	95	96	97	98	99	100	101	102	103	104	183	184	185	186	187	188	189	190	191	192	193	194	195		
		65	79	80	44	38	48	19	90	51	7	68	40	30	12	61	40	56	63	7	26	25	20	8	66	87	3	90	9	61	71	64	86	76	39	30	49	10	43	25		
		---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
L	2	14	15	16	17	18	19	20	21	22	23	24	25	26	105	106	107	108	109	110	111	112	113	114	115	116	117	196	197	198	199	200	201	202	203	204	205	206	207	208		
		31	53	17	45	61	50	49	77	32	58	52	16	59	75	21	80	30	31	32	18	47	71	24	2	5	81	26	31	11	22	45	91	38	34	66	35	36	67	75		
		---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
O	3	27	28	29	30	31	32	33	34	35	36	37	38	39	118	119	120	121	122	123	124	125	126	127	128	129	130	209	210	211	212	213	214	215	216	217	218	219	220	221		
		29	2	66	20	24	23	74	1	55	27	35	10	86	79	34	45	6	54	48	59	14	82	43	1	62	10	12	27	1	50	41	28	80	6	13	51	15	21	4		
		---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
C	4	40	41	42	43	44	45	46	47	48	49	50	51	52	131	132	133	134	135	136	137	138	139	140	141	142	143	222	223	224	225	226	227	228	229	230	231	232	233	234		
		56	42	9	34	37	84	81	83	13	11	12	78	43	85	39	4	42	52	58	23	90	84	88	27	67	38	42	7	81	68	59	3	89	82	52	70	72	69	2		
		---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
K	5	53	54	55	56	57	58	59	60	61	62	63	64	65	144	145	146	147	148	149	150	151	152	153	154	155	156	235	236	237	238	239	240	241	242	243	244	245	246	247		
		54	8	47	6	41	36	69	63	76	71	70	46	67	60	29	72	69	17	50	46	55	64	37	22	51	9	85	8	62	73	48	23	17	20	44	18	78	46	16		
		---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
	6	66	67	68	69	70	71	72	73	74	75	76	77	78	157	158	159	160	161	162	163	164	165	166	167	168	169	248	249	250	251	252	253	254	255	256	257	258	259	260		
		15	57	25	22	73	3	82	75	4	88	5	60	62	53	41	78	89	68	83	28	19	86	36	35	77	73	79	60	74	84	32	88	29	63	33	40	47	77	83		
		---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
	7	79	80	81	82	83	84	85	86	87	88	89	90	91	170	171	172	173	174	175	176	177	178	179	180	181	182	261	262	263	264	265	266	267	268	269	270	271	272	273		
		64	33	39	87	14	21	85	28	26	91	18	89	72	76	74	13	11	57	91	33	16	49	15	44	70	65	5	14	57	55	65	58	54	19	37	56	24	87	53		
		---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Figure 34 : Schéma illustrant les semis du dispositif expérimental d'un essai (Les numéros en vert sont les numéros des entrées et les numéros en noir sont les numéros des parcelles)

1.4. Déroulement de chaque essai

Les essais ont été conduits suivant l'itinéraire ci-après :

Préparation du terrain avant semis : Labour et ameublissement du sol

Semis : Après la première pluie de chaque saison hivernale

Premier sarclage et binage : 7 jours après la levée

Sarclage et démariage 15 jours : 15 jours après semis

Deuxième sarclage et binage (avec traction à cheval) : 25 jours après démariage

Pour chaque essai, un apport d'engrais NPK (15-15-15) à hauteur de 150 Kg/ha, a été appliqué avant l'ameublissement. Deux apports d'urée à hauteur de 100 kg/ha, ont été appliqués après le démariage et à la montaison. En dehors de cet itinéraire, il a été procédé au binage (désherbage) à chaque fois que cela a été nécessaire pour éradiquer les mauvaises herbes. Il a aussi été procédé à l'irrigation des essais à chaque fois qu'il s'est avéré nécessaire (longue pause pluviométrique).

1.5. Phénotypage

Les paramètres suivants ont été relevés: nombre de poquets à la levée (NPL), Vigueur juvénile (VJ), Incidence du mildiou (IM) 30 jours après semis (jas), date d'épiaison (EPI), date de 50% de floraison (FLO), longueur de la feuille paniculaire (LOF), largeur de la feuille paniculaire (LAF), nombre d'entrenœuds (NEN), longueur moyenne des entrenœuds (LEN), diamètre de la tige principale (DTP), hauteur de la plante (HTR), nombre de talles productives (NTP), circonférence de l'épi (CEP), longueur de l'épi (LEP), poids des épis (PEP), poids paille (Biomasse), poids des grains (PG) et poids de 1000 grains (PMG).

2. Génotypage

2.1. Extraction de l'ADN

Les ADN des échantillons de la core collection ont été extraits à partir des feuilles échantillonnées au stade 5 feuilles de la phase végétative lors de la première année d'essai (2016), suivant la technique décrite dans la partie (b) *Extraction et quantification de l'ADN*, du deuxième chapitre de cette thèse. L'ADN ainsi extrait a été conservé dans une solution de Tris-HCl, puis envoyé à séquencer à la Plateforme de Séquençage de Nouvelle Génération du Centre de recherche du CHU, de l'Université de Laval (Québec).

2.2. Séquençage

Les banques ont été générées en deux multiplexes de 45 et 46 échantillons. Une double digestion *PstI-MspI* a été appliquée sur chaque ADN, puis des adaptateurs ont été liés à chaque échantillon suivie d'un mélange et d'une amplification. Le séquençage de ces banques a été fait à l'aide d'un séquenceur Illumina HiSeq2500. La génération des banques et le séquençage ont été entièrement réalisés à la Plateforme NGS du Centre de recherche du CHU, de l'Université de Laval (Québec).

2.3. Analyses bio-informatiques

L'évaluation de la qualité des séquences « reads » a été faite à l'aide de FastQC v 0.72 et de MultiQC v 1.6 et le nettoyage des séquences a été fait grâce à FastQ Trimmer v 1.0.0. Seules les séquences présentant une valeur moyenne de la qualité Q supérieure ou égale à 30 (format sanger) ont été retenues et les 7 premières bases (coté 5') de chaque read ont été enlevées. Les séquences ont été alignées sur la version 2 du génome de référence du mil ASM217483v2, de numéro d'accessions GenBank GCA_002174835.2, à l'aide de BWA v 1.2.3, avant le réaligement des séquences par rapport aux insertions et délétions, grâce à RealignerTargetCreator v 0.0.4 et IndelAligner v 0.0.6. Ces procédures sont illustrées par le workflow suivant.

Les fichiers au format BAM issus des procédures décrites ci-dessus, ont été joints à l'aide de l'outil MergeBAM v 1.2.0 et les appels des SNPs ont été réalisés grâce à UnifiedGenotyper v0.0.6. Au total, 502 382 variants ont été appelés dont 545 834 SNPs. Le filtrage des SNPs a d'abord été fait suivant la qualité du Mapping et la profondeur, en appliquant un filtre sévère (Hard filtering) à l'aide de l'outil Variant filtration v0.0.5 (MQ supérieur ou égale à 40 et de MQ divisée par la profondeur des échantillons non filtrés supérieur à 0.1). Puis un second filtre a été réalisé suivant la fréquence des allèles mineurs (MAF) supérieur à 0.05 et une proportion admise des données manquantes de 0.1, à l'aide de l'outil VCFtools Filter v 1.2.0. Les marqueurs multialléliques ont en suite été éliminés, à l'aide de Tassel v5.2.48. Le fichier de sortie comportait au final 19 470 marqueurs SNPs de haute qualité. Les analyses subséquentes ont été réalisées sur ces SNPs.

Toute l'analyse bio-informatique a été réalisée sur la plateforme Galaxy v18.0.5 (Afgan et al., 2018), implémentée dans le système d'exploitation Bio-Linux 8 (Field et al., 2006).

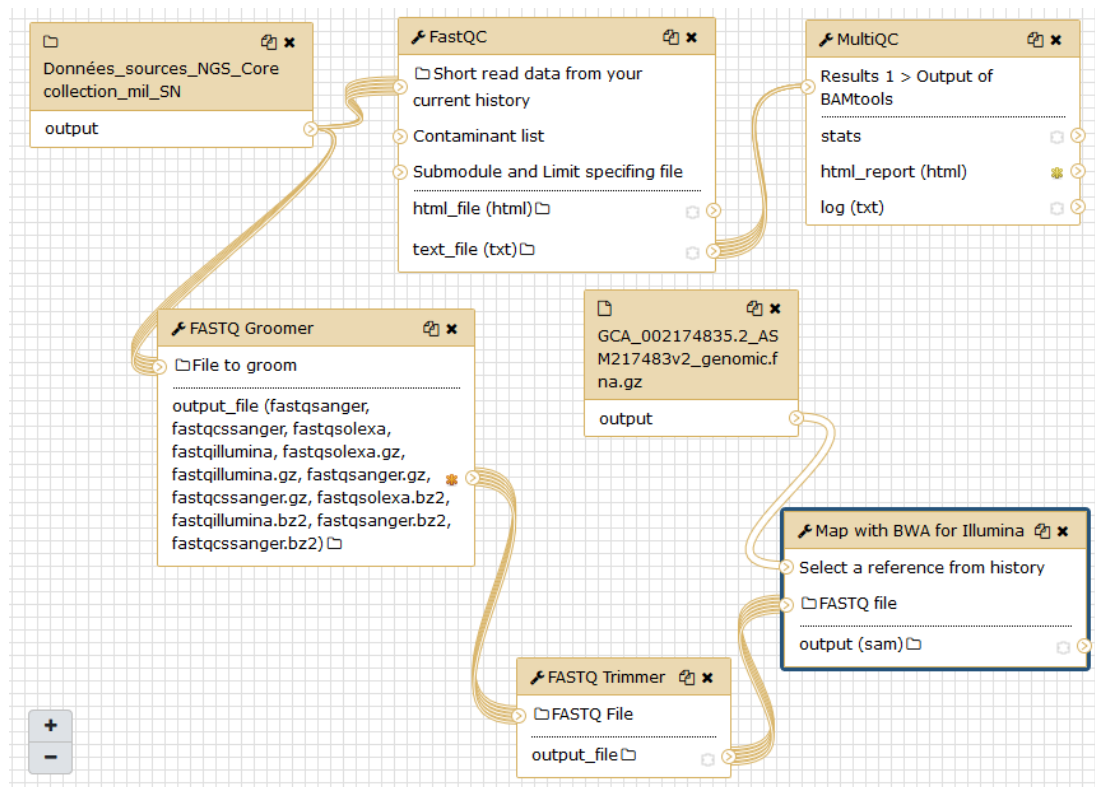


Figure 35 : Diagramme du workflow de traitement bio-informatique de vérification de qualité, de trimming et de cartographie des séquences sur le génome de référence.

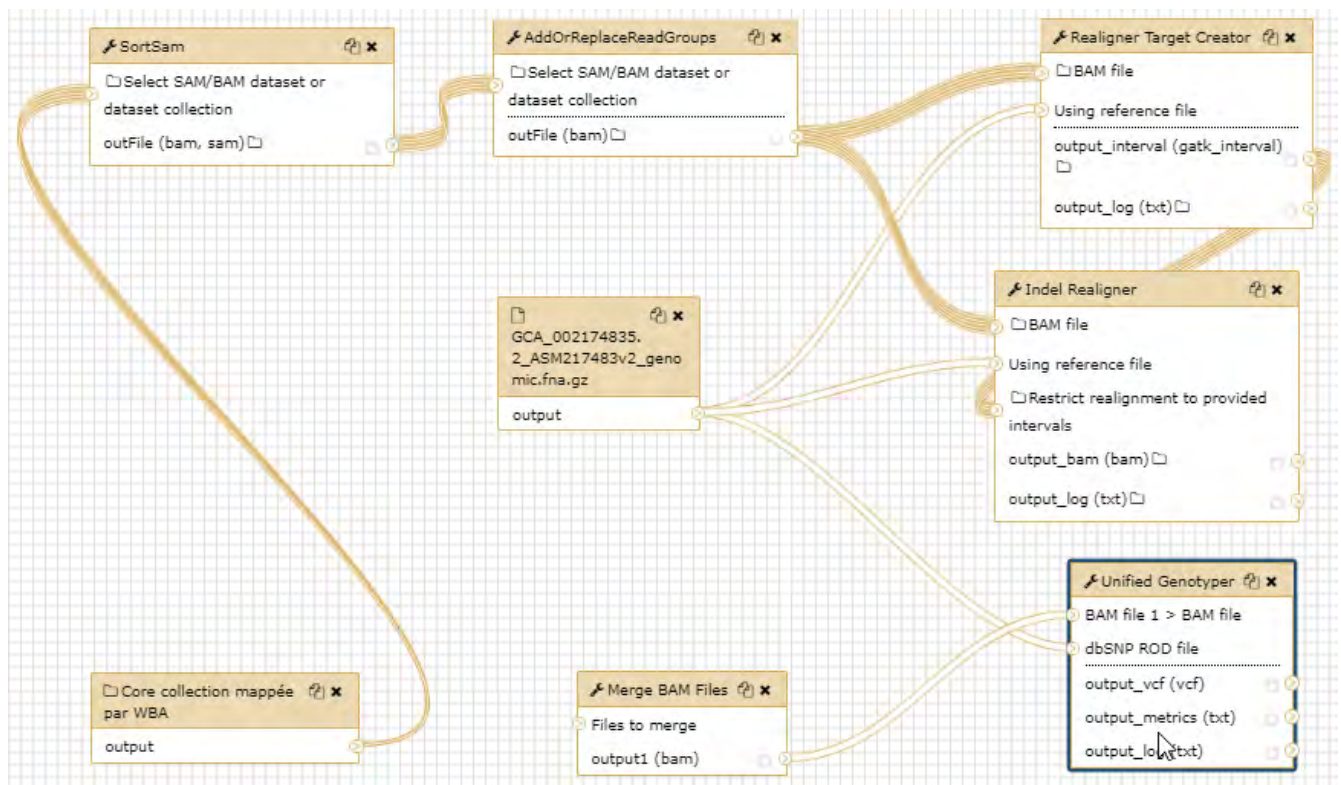


Figure 36 : Diagramme du workflow d'analyse bio-informatique d'élimination des duplicatas techniques, de réalignement par rapport aux InDels et d'appel des SNPs.

3. Méthodes analytiques

3.1. Analyses phénotypiques

Il a d'abord été fait des analyses de variances sur les différents paramètres phénotypiques, afin de tester l'existence d'effet génotype, d'effet année et les interactions entre génotypes et années. Puis il a été vérifié l'héritabilité des caractères agro-morphologiques grâce au modèle linéaire mixte à effet aléatoire pour les individus, à l'aide du logiciel Plant Breeding Tools v 1.3, suivant l'équation :

$$Y \sim 1 + (1|\text{Génotype}) + (1|\text{Année}) + (1|\text{REP:Année}) + (1|\text{REP:BLOCK:Année}) + (1|\text{Génotype:Année})$$

Les analyses qui ont suivi ont été réalisées à deux niveaux. (i) Un premier niveau regroupant les Souna et les Sanio et (ii) un second niveau qui opère au sein des Souna et au sein des Sanio séparément. Nous avons d'abord réalisé une analyse en composantes principales grâce aux moyennes ajustées par individus pour chaque variable, centrées et réduites, afin de vérifier l'effectivité de la différenciation agro-morphologique des accessions de Souna et de Sanio et les potentielles subdivisions des Souna et des Sanio, à l'aide du package adegenet v2.1.1 (Jombart, 2008), du logiciel R (R Development Core Team, 2008). Il a ensuite été fait une analyse factorielle discriminante pour identifier les caractères discriminants dans la différenciation des accessions précoces (Souna) et tardives (Sanio) et les caractères discriminant les sous-groupes des accessions précoces et tardives. Les caractères qui ont présenté une significativité avec une P-value < 0.001 au seuil de $\alpha = 0.001$, dans la discrimination et qui ont présenté un degré de corrélation (r) supérieur ou égal à 0.7, par rapport à l'axe de différenciation, ont été sélectionnés et considérés comme caractères hautement discriminants des Souna et des Sanio ou des sous-groupes de Souna et de Sanio. Cette analyse a été réalisée à l'aide du logiciel XLStat 2014.

Puis il a été réalisé des tests de Mann-Whitney pour confirmer la nature différente des caractères choisis en fonction des Souna et des Sanio. Des tests de Kruskal-Wallis ont été réalisés, pour confirmer la nature différente des caractères choisis en fonction des sous-groupes de Souna et de Sanio. En fin il a été fait des box plots, afin d'illustrer la répartition des caractères agro-morphologiques discriminants, par rapport aux Souna et Sanio et aux sous-groupes de Souna et de Sanio de la core collection. Ces dernières analyses ont été faites à l'aide du logiciel Minitab v. 17.

3.2. Analyses génétiques

Pour choisir un modèle pour le GWAS, nous avons étudié la structure de la population et évalué l'apparentement entre individus. L'étude de la structure a été réalisée à l'aide de l'algorithme sNMF, à travers le package LEA v2.2.0 implémenté dans le logiciel R, en testant un nombre de K populations allant de 1 à 10 pour 10 répétitions et par Analyses en composantes principales à l'aide du logiciel Tassel v5.2.48 et du package Ade4 v 1.7-13 du logiciel R (R Development Core Team, 2008).

Les relations entre individus ont été évaluées grâce au calcul de la matrice d'identité par la distance (IBS) Kinship, normalisée à l'aide du logiciel Tassel v5.2.48.

3.3. Analyses d'associations génotypes-phénotypes

Afin d'identifier les SNPs liés aux caractères agro-morphologiques de différenciation au sein de la core collection, nous avons testé leurs associations grâce à un modèle linéaire mixte (MLM).

Le modèle linéaire mixte prend en compte l'effet de la structure de la population et l'effet de relations de parenté (cryptiques) entre individus de la core collection. Ces analyses d'associations ont été réalisées à l'aide du logiciel Tassel v5.2.48. La représentation graphique des résultats a été faite sous forme de (i) Manhattan plot, qui illustre la valeur des p-values des SNPs en fonction de leurs positions sur les différents chromosomes et sous forme de (ii) diagramme quantile-quantile (QQ plot) qui permet d'apprécier la capacité du modèle à gérer les effets de la structure et de relations de parenté cachées, en comparant la distribution des statistiques de tests observée à celle attendue sous l'hypothèse nulle, c'est-à-dire en l'absence d'association.

Le seuil de significativité (α) de l'association des marqueurs SNPs aux différents traits a été calculé en utilisant la correction de Bonferroni, suivant la formule ci-dessous :

$$\alpha = \frac{0.05}{NH} \quad \text{soit} \quad NH = NMark \times NPheno$$

α = Seuil de significativité réajusté

NH : Nombre d'hypothèses testées

NMark : Nombre de marqueurs

NPheno : Nombre de phénotypes

Un SNP est retenu comme significativement associé à un caractère agro-morphologique, lorsque le test est significatif au seuil de α , en d'autres termes, lorsque la p-value du test est strictement inférieure à α . Il a ensuite été réalisé le calcul des q-value (ou FDR pour False Discovery Rate), sur chaque SNPs significativement associé à un caractère agro-morphologique. La q-value ou FDR mesure la proportion de risque de faux positif encouru pour un test donné, lorsque ce test en question est significatif.

Pour identifier les régions génomiques liées à ces associations, les SNPs significativement associés à des caractères agro-morphologiques et retenus, ont été localisés dans les intervalles du génome du mil. Cette localisation a été réalisée avec le package `sqldf` v0.4-11 de R. Les marqueurs significativement associés à des traits et situés dans des régions géniques ont été considérés comme marqueurs causals (association directe) et les gènes qui les portent, aussi comme gènes causatifs. Les marqueurs significativement associés à des traits et situés dans des régions inter-géniques ont été considérés comme marqueurs en déséquilibre de liaison avec les locus causatifs (association indirecte). Les séquences peptidiques des gènes causatifs ont été blastées dans la base de données NCBI (National Center for Biotechnology Information) (<http://blast.ncbi.nlm.nih.gov>) à des fins descriptives des gènes.

RESULTATS

1. Evaluation de la qualité des reads

Les bibliothèques générées et séquencées ont été démultiplexées et les code-barres éliminés, avant réception des données. L'évaluation de la qualité a permis de remarquer la bonne qualité des reads, avec de très bons scores de qualités comme le montre le tableau 1 annexe 4. Les séquences étaient à bouts uniques et la longueur des reads variait de 35 à 40 bases. Le point alarmant résidait au niveau d'un taux élevé en %GC. Ce qui est normal, le mil a naturellement un taux de GC élevé, de l'ordre de 47.9% (Debieu et al., 2017). Le fichier d'analyse comportait 99.63% de données présentes. Avec en moyenne 0.37% de données manquantes par individu et par marqueur.

2. Analyses génétiques

L'analyse de la structure par sNM et l'ACP a reconstitué les groupes génétiques déjà observés avec des marqueurs microsatellites (SSRs) dans les chapitres précédants, en mettant en évidence une forte structure constituée d'une part par les accessions précoces et d'autre part par les accessions tardives (Figure 43 et 44).

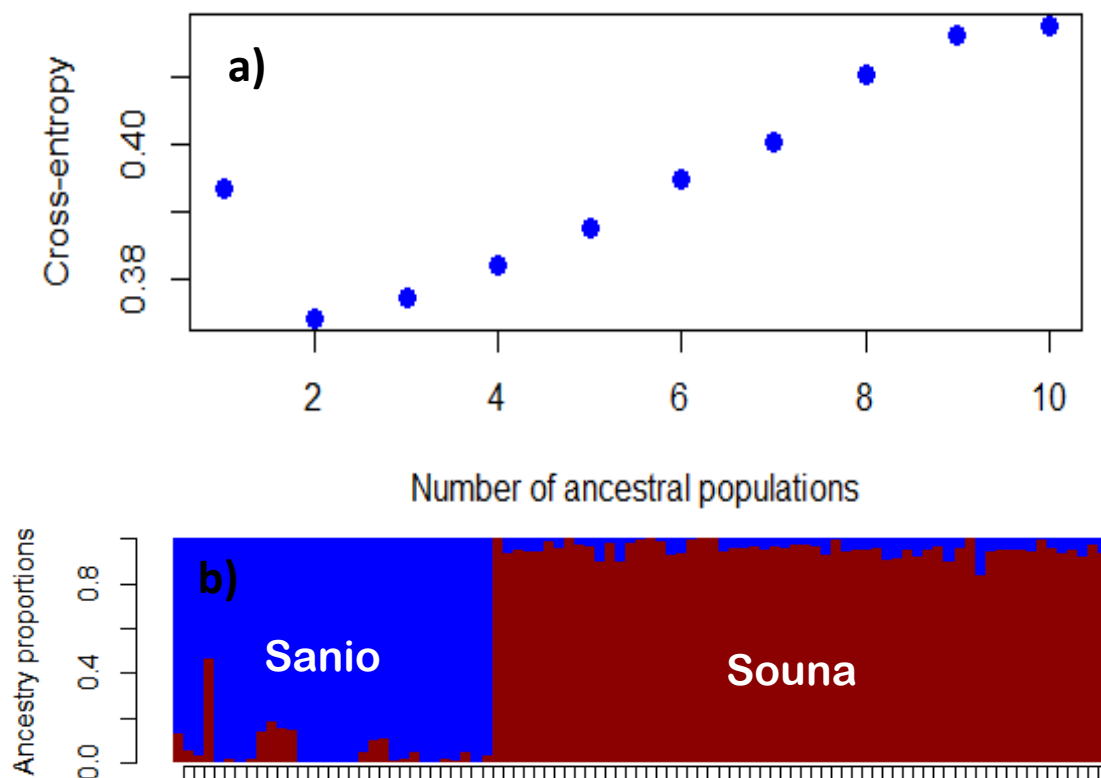


Figure 37 : (a) Probabilité de subdivisions par rapport aux différent cas testés (la valeur de cross-entropy la plus basse est la meilleur probabilité) et (b) proportion d'appartenance de chaque individu aux différents K groupes, des accessions de la core collection

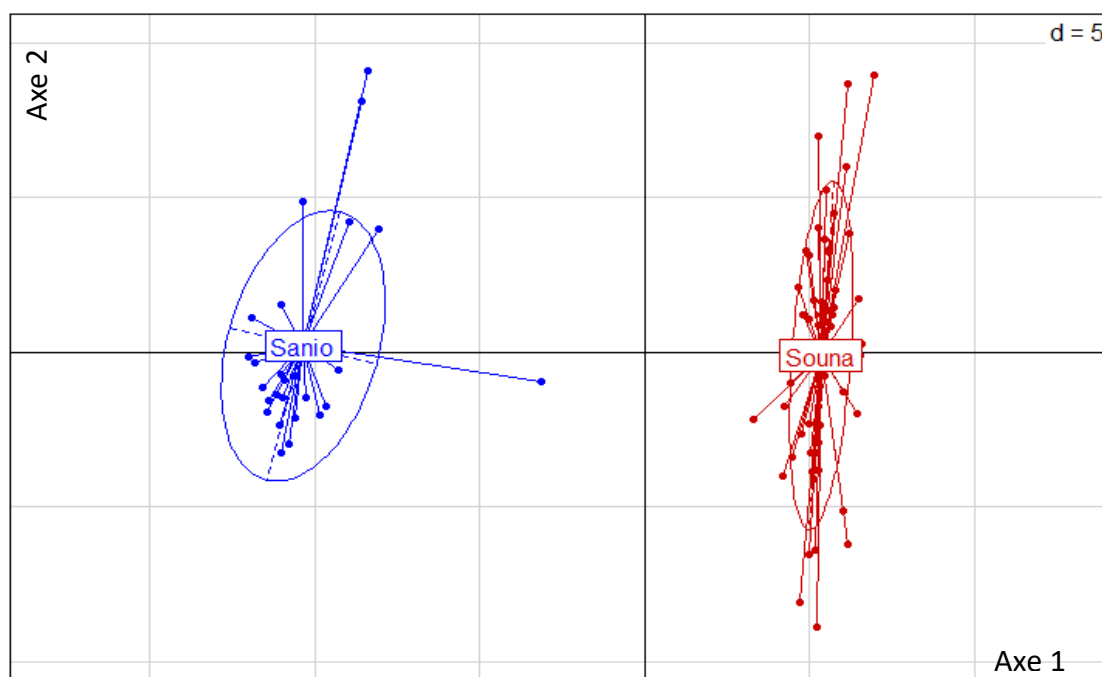


Figure 38 : Analyse en composantes principales des accessions de la core collection sur la base des données SNPs (Axe1= 7.15%, Axe2= 1.54%)

3. Analyses phénotypiques

Chaque mesure phénotypique a été prise sur trois poquets par accession, au niveau de chaque répétition. La moyenne des trois données a constitué la valeur de l'individu, au niveau de la répétition. Le fichier d'analyse de variance comportait pour chaque individu une valeur au niveau de chaque répétition pour chaque année d'évaluation. Ce fichier ne comportait aucune donnée manquante.

3.1. Analyses de variances

Les analyses de variances ont montré un effet génotype fortement significatif presque pour tous les caractères sauf pour le poids 1000 grains ; qui n'est pas significatif. Ce qui dénote d'une importante variabilité phénotypique des individus. L'effet année cependant n'a été significatif que pour quelques caractères. Les interactions entre génotypes et années sont significatives à des degrés différents sur tous les caractères agro-morphologiques sauf pour trois caractères.

Tableau 13 : Récapitulatif des résultats d'analyse de variance et de calcul de l'héritabilité génétique (h^2) Significativité : * P-value < 0.05, ** P-value < 0.01 et *** P-value < 0.001, NS= Non significatif.

	DF	IM	EPI	FLO	NTP	NEN	LEN	LOFP	LAFP	DTP	HTP	LEP	CIR	EXR	PMG	Biomasse	RDT
Effet	90	***	***	***	***	***	***	***	**	NS	***	***	***	***	NS	*	**
Génotype																	
Effet Année	1	NS	NS	NS	**	***	NS	*	NS	***	NS	NS	***	NS	***	***	**
Génotype x																	
Année		NS	***	***	NS	*	**	*	*	***	***	*	*	NS	***	***	***
Moyenne		0.03	56.37	60.68	6.36	9.33	21.11	46.09	4.59	4.19	241.62	48.9	8.43	3.15	5.84	1528.39	1637.49
Ecart type		0.07	12.27	12.68	2.16	2.19	1.93	6.53	0.66	1.6	39.11	9.35	1.21	3.16	1.44	1114.7	1112.15
h^2		0.69	0.97	0.98	0.86	0.96	0.68	0.62	0.6	0.23	0.94	0.9	0.84	0.71	0	0.42	0.55

3.2. Analyses multivariées regroupant Souna et Sanio

3.2.1. Analyse en Composantes Principales (ACP)

L'analyse en composantes principales confirme la forte structure agro-morphologique des Souna et des Sanio avec une subdivision claire et nette. Elle révèle aussi les caractères agro-morphologiques suivant lesquelles, cette subdivision entre Souna et Sanio est faite. Ces caractères sont d'une part, la hauteur de la plante, la biomasse, le nombre d'entrenœuds, la longueur des entrenœuds la date de floraison, la date d'épiaison, le nombre de talles productives et le rendement qui sont positivement corrélés au groupe des Sanio et négativement aux Souna. D'autre part, l'incidence du mildiou qui est positivement corrélée au groupe des Souna et négativement aux Sanio.

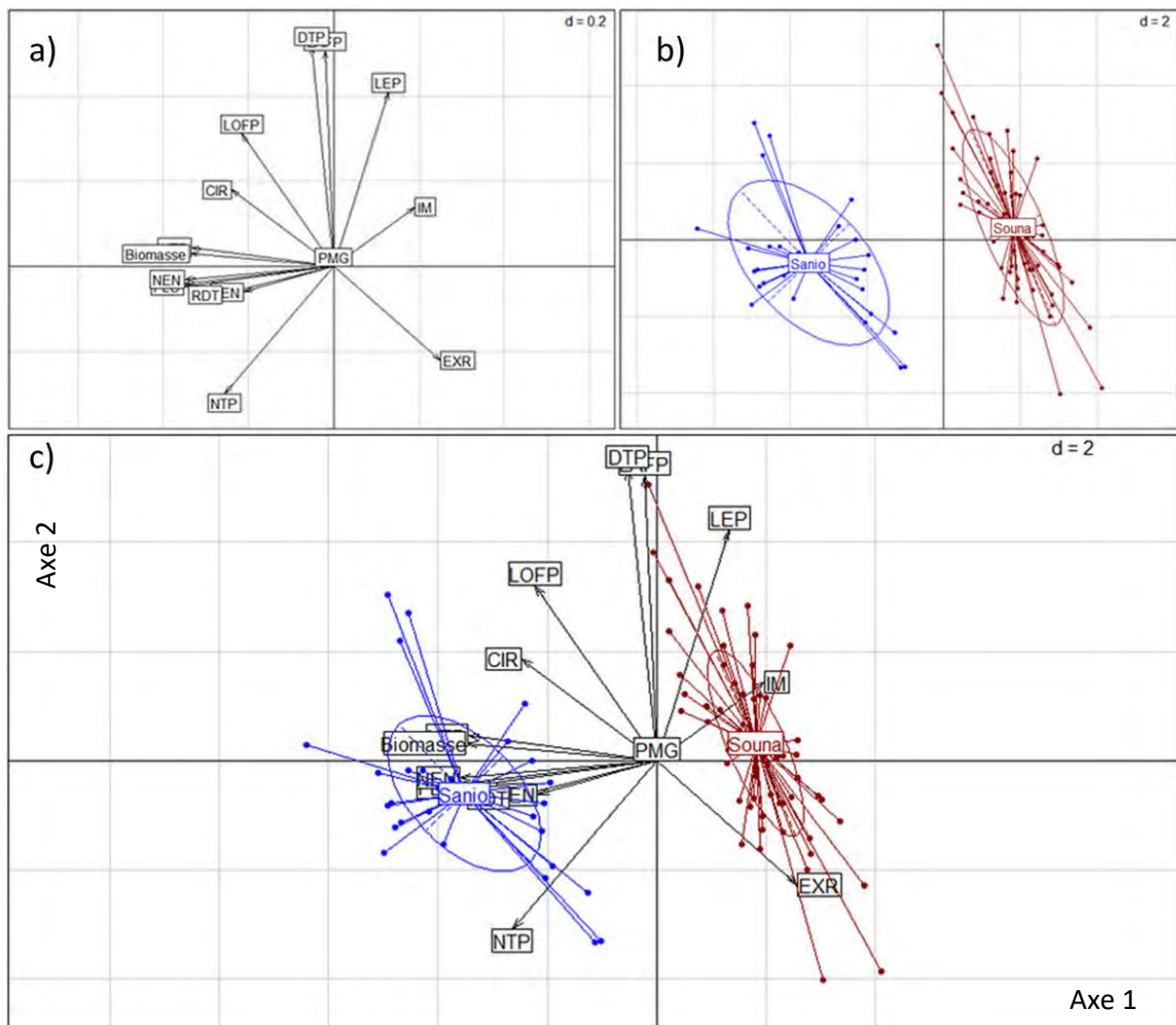


Figure 39 : Analyse en composantes principales montrant (a) les corrélations entre variables, (b) la structure entre Souna et Sanio et (c) les relations de corrélation entre variables phénotypiques et structure agro-morphologique des Souna et des Sanio (Axe1= 50.4%, Axe2= 18.6%)

3.2.2. Analyse Factorielle Discriminante (AFD)

Parmi les 16 caractères évalués par l'analyse factorielle discriminante, 13 caractères se sont révélés statistiquement significatifs (au seuil de $\alpha = 0.01$ et tous avec une p-value < 0.0001), dans la différenciation des Souna et des Sanio. Ces caractères sont l'incidence du mildiou, la date d'épiaison, la date de floraison, le nombre de talles productives, le nombre d'entrenœuds, la longueur moyenne des entrenœuds, la longueur de la feuille paniculaire, la hauteur de la plante, la longueur et la circonférence de l'épi, l'exsertion paniculaire, la biomasse et le rendement. Parmi ces 13 caractères, 6 caractères ont présenté un degré de corrélation à l'axe de différenciation des Souna et des Sanio, supérieur ou égal à 0.7, et ont donc été considérés comme étant des caractères hautement discriminants des Souna et des Sanio. Ces caractères peuvent être classés en deux groupes. D'une part, le groupe des caractères phénologiques qui regroupe la date d'épiaison et la date de floraison et d'autre part, le groupe des caractères de développement végétatif qui regroupe la biomasse, la hauteur de la plante, le nombre moyen d'entrenœuds et le nombre de talles productives. Ces observations obtenues grâce à l'analyse factorielle discriminante sont parfaitement en accord avec l'analyse en composantes principales. Tous ces caractères hautement discriminants des Souna et des Sanio sont des caractères qui ont présenté de hautes héritabilités. En d'autres termes, les variabilités observées par rapport à ces différents caractères sont expliquées principalement par une variabilité au niveau génétique.

Les tests de Mann-Whitney sur ces caractères ont montré des différences significatives au seuil de $\alpha = 0.05$, des Souna et des Sanio par rapport à ces six caractères avec une P-Value < 0.0001 pour chaque caractère. Ce qui a confirmé les observations précédentes.

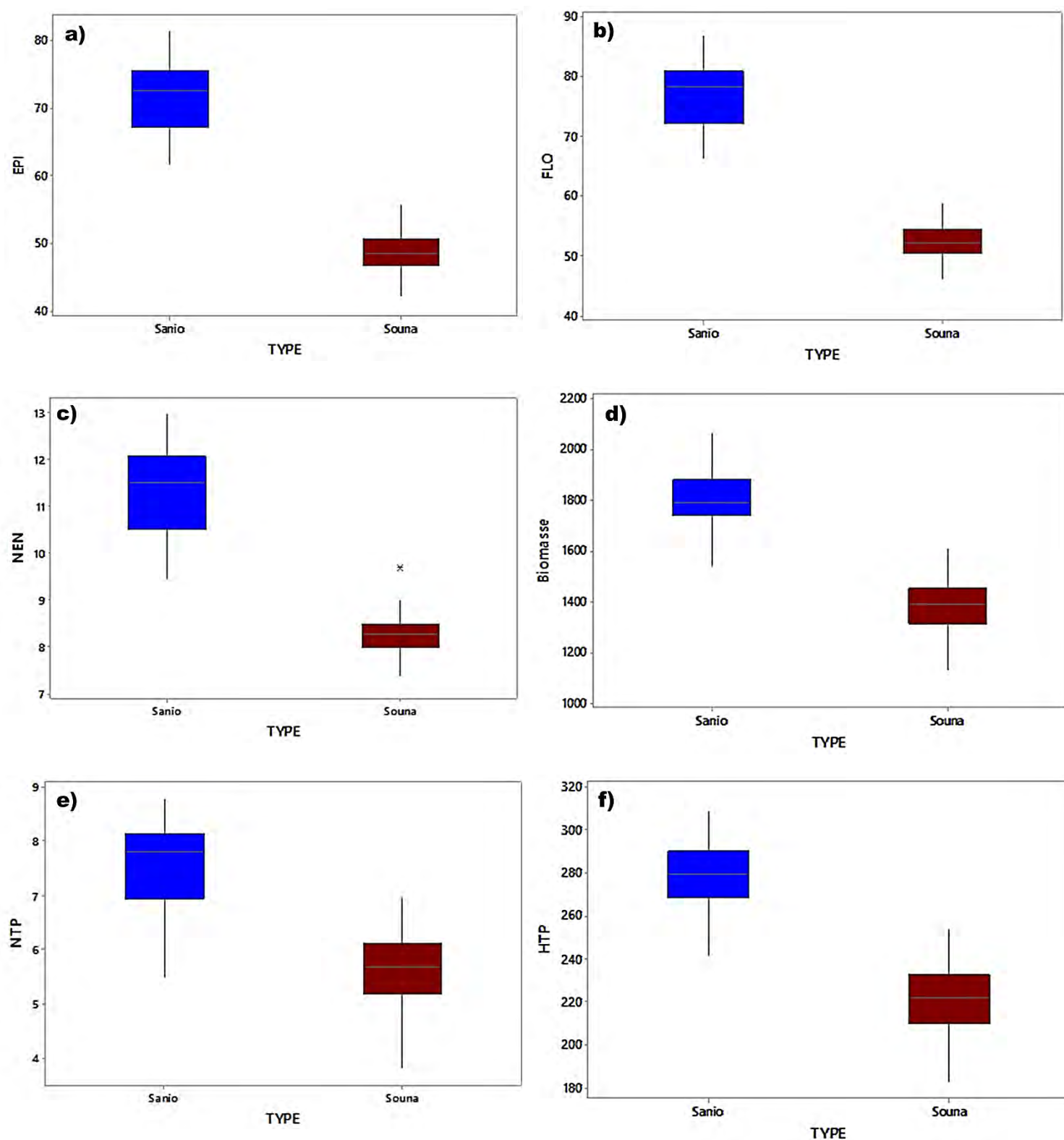


Figure 40 : Box plots des caractères de (a) date d'épiaison, (b) date de floraison, (c) nombre d'entrecœuds, (d) de biomasse, (e) nombre de talles productives et (f) de hauteur de la plante en fonction des Souna (en rouge foncé) et des Sanio (en bleu).

3.3. Analyses multivariées au sein des Souna

3.3.1. ACP des Souna

L'analyse en composantes principales des Souna montre une subdivision en trois sous-groupes. Les caractères agro-morphologiques suivant lesquels se fait cette subdivision sont la date d'épiaison et de floraison, le diamètre de la tige principale et la largeur de la feuille paniculaire qui sont positivement corrélés au sous-groupe 3 et négativement corrélés au sous-groupe 1.

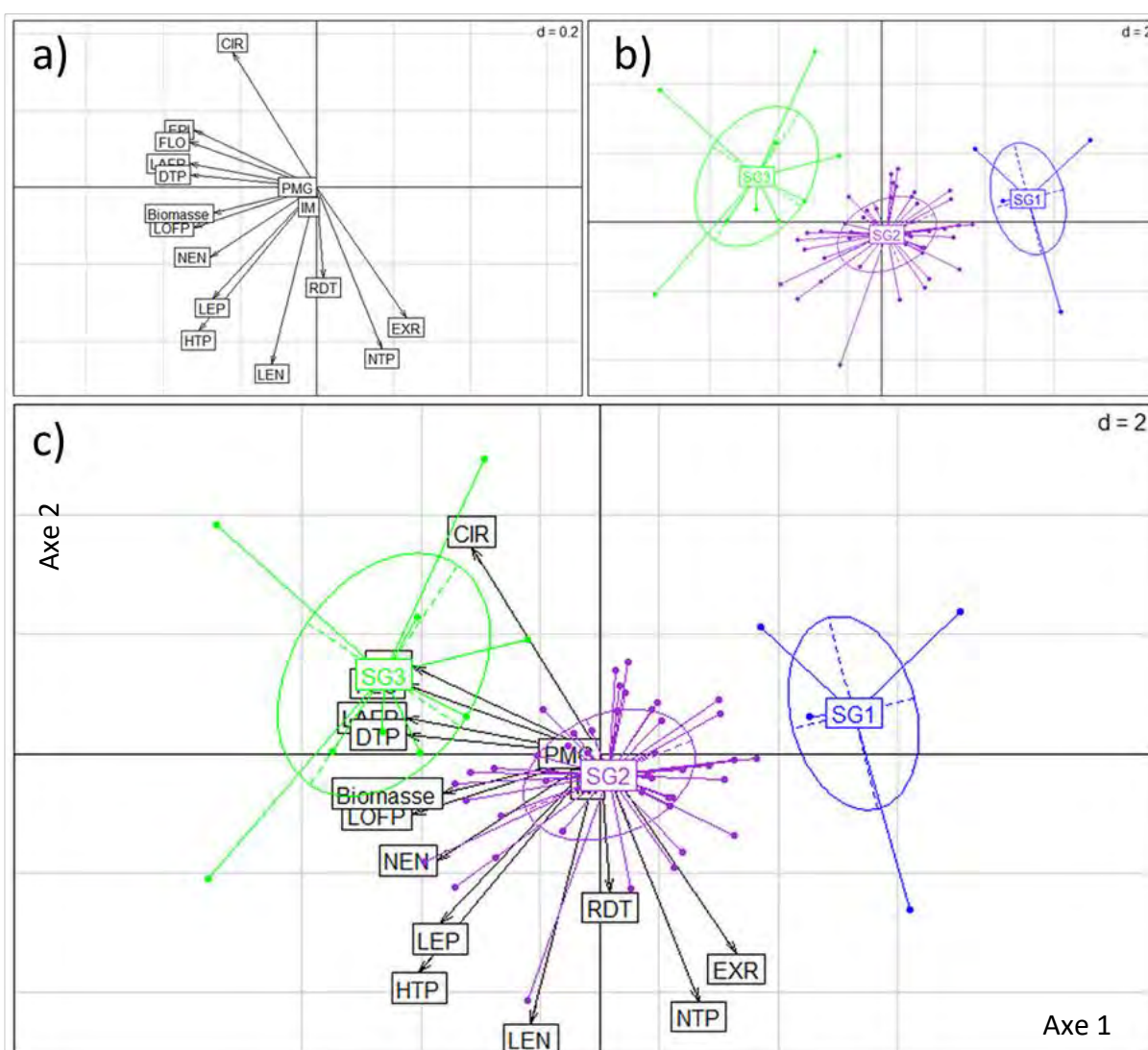


Figure 41 : Analyse en composantes principales montrant (a) les corrélations entre variables, (b) la structure des sous-groupes (SG) de Souna et (c) les relations de corrélation entre variables phénotypiques et la structure agro-morphologique des Souna (Axe1= 39.6%, Axe2= 14.5%).

3.3.2. AFD des Souna

Neuf (9) caractères ont été statistiquement significatifs (au seuil de $\alpha = 0.01$ et tous avec une p-value < 0.0001), dans la différenciation en sous-groupes des Souna. Ces caractères sont la date d'épiaison, la date de floraison, le nombre de talles productives, la longueur de la feuille paniculaire, la largeur de la feuille paniculaire, le diamètre de la tige principale, la longueur et la circonférence de l'épi, l'excertion paniculaire et la biomasse. Parmi ces caractères, 3 caractères ont présenté un degré de corrélation à l'axe principale de différenciation (Axe 1) des sous-groupes de Souna supérieur ou égal à 0.7, et ont donc été considérés comme étant des caractères hautement discriminants des sous-groupes de Souna. Ces caractères sont la date d'épiaison, la date de floraison et le diamètre de la tige principale. Ces résultats de l'AFD sont parfaitement en accord avec l'ACP. Les tests de Kruskal-Wallis ont confirmé que les sous-groupes des Sanio sont significativement différents au seuil de $\alpha = 0.05$ par rapport à ces trois caractères, avec pour chaque caractère une P-value < 0.001 .

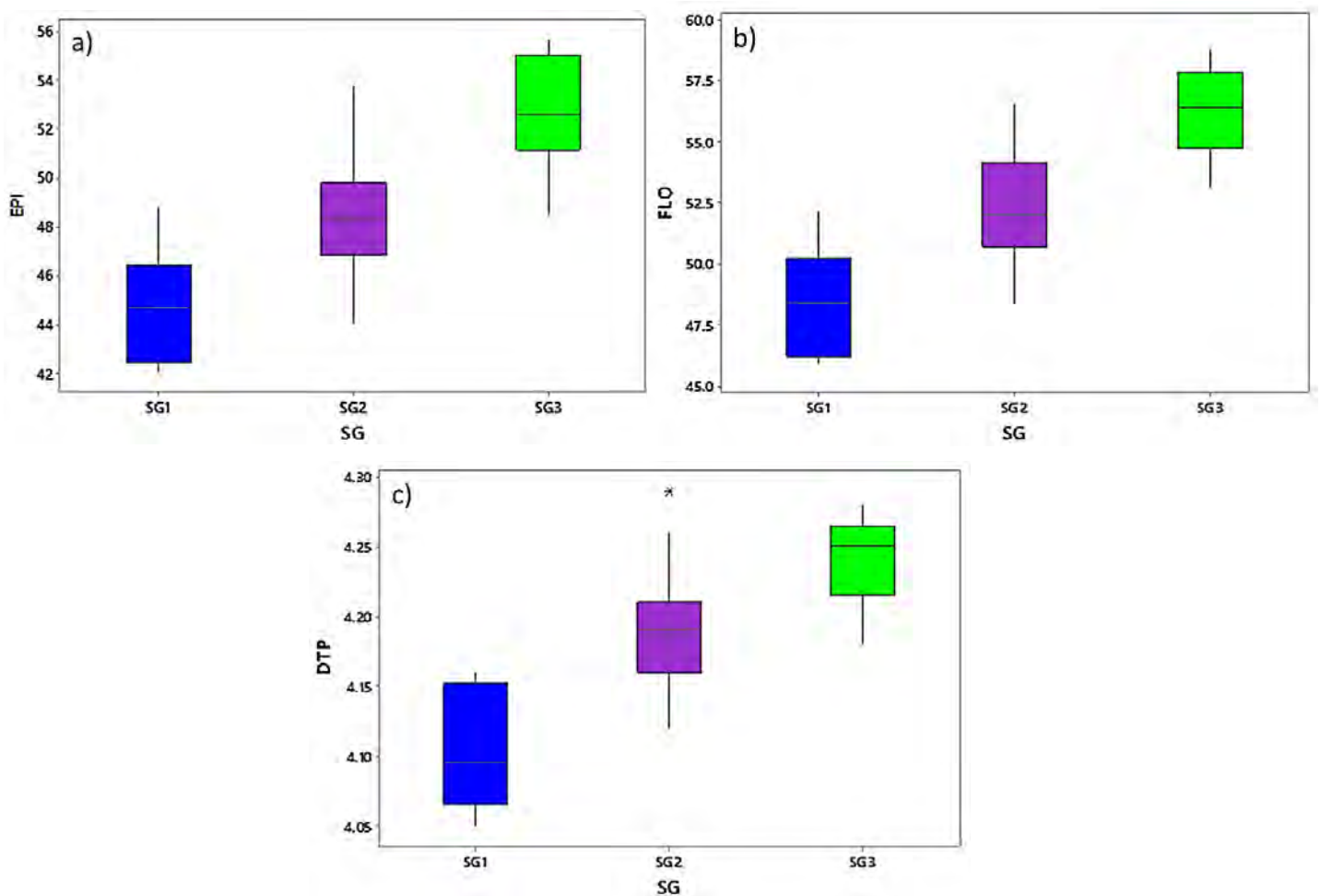


Figure 42 : Box plots des caractères de (a) date d'épiaison, (b) date de floraison et de (c) diamètre de la tige principale, en fonction des sous-groupes 1 (en bleu), sous-groupe 2 (violet orchidée) et sous-groupe 3 (en vert) des Souna.

3.4. Analyses multivariées au sein des Sanio

3.4.1. ACP des Sanio

L'analyse en composantes principales des Sanio montre une subdivision en trois sous-groupes. Les caractères agro-morphologiques suivant lesquels se fait cette subdivision sont la date d'épiaison et de floraison, le nombre d'entrenœuds, la biomasse, la circonférence de l'épi et le rendement, qui sont positivement corrélés au sous-groupe 3 et négativement corrélés au sous-groupe 1.

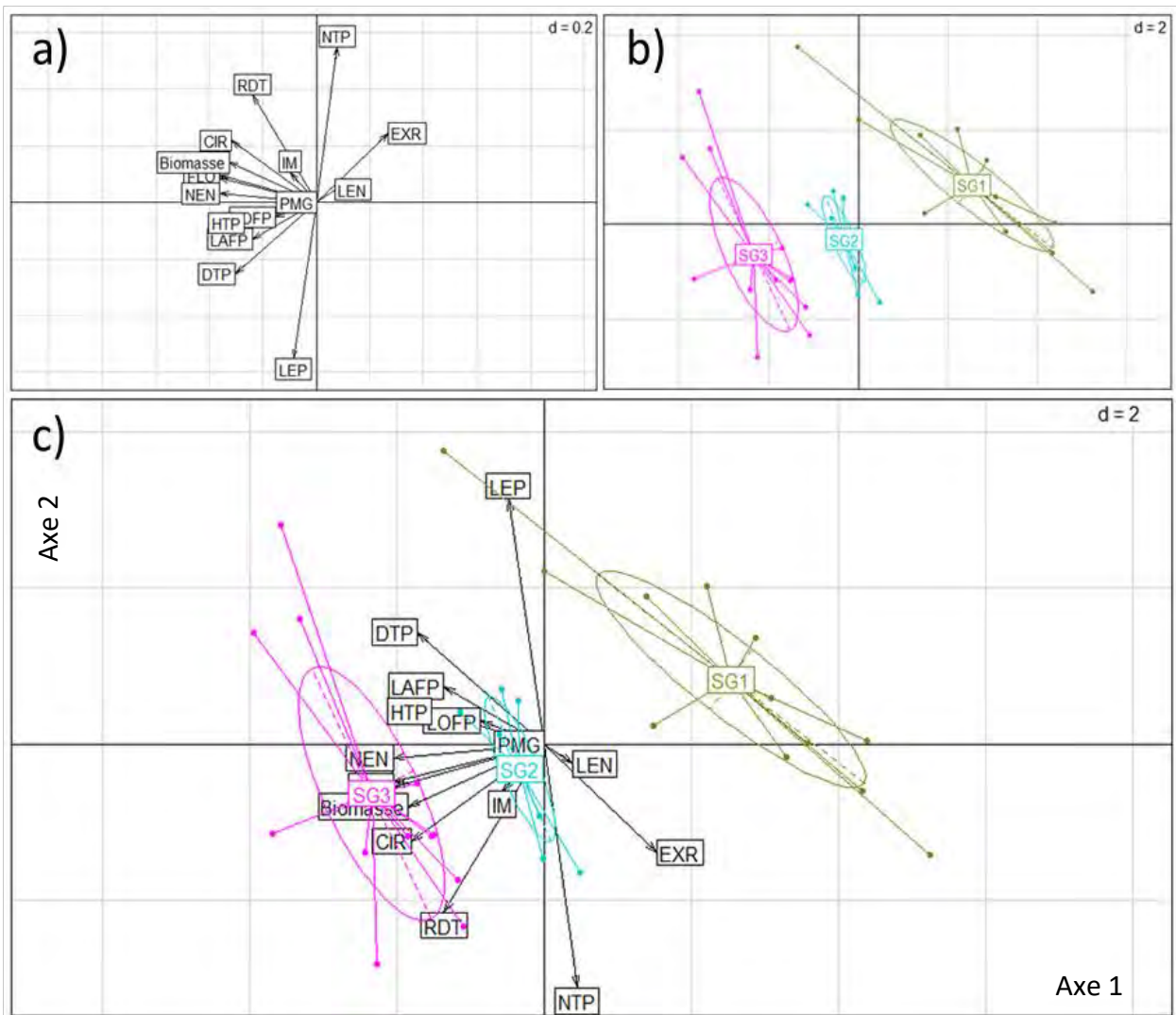


Figure 43 : Analyse en composantes principales montrant (a) les corrélations entre variables, (b) la structure des sous-groupes (SG) de Sanio et (c) les relations de corrélation entre variables phénotypiques et la structure agro-morphologique des Sanio (Axe1= 42.5%, Axe2= 16.1%).

3.4.2. AFD des Sanio

L'AFD a révélé que la date d'épiaison, la date de floraison, le nombre d'entrenœuds, la circonférence de l'épi, la biomasse et le rendement sont statistiquement significatifs (au seuil de $\alpha = 0.01$ et tous avec une p-value < 0.0001) dans la différenciation en sous-groupes des Sanio. Tous ces caractères ont présenté un degré de corrélation à l'axe de différenciation (Axe 1) des sous-groupes de Sanio, supérieur ou égal à 0.7 et ont donc été considérés comme étant des caractères hautement discriminants des sous-groupes de Sanio. Ces résultats sont en accord avec l'ACP.

Les tests de Kruskal-Wallis ont confirmé que les sous-groupes des Sanio sont significativement différents au seuil de $\alpha = 0.05$ par rapport à ces caractères, avec pour chaque caractère une P-value < 0.001 , sauf pour le nombre d'entrenœuds où la P-value = 0.001.

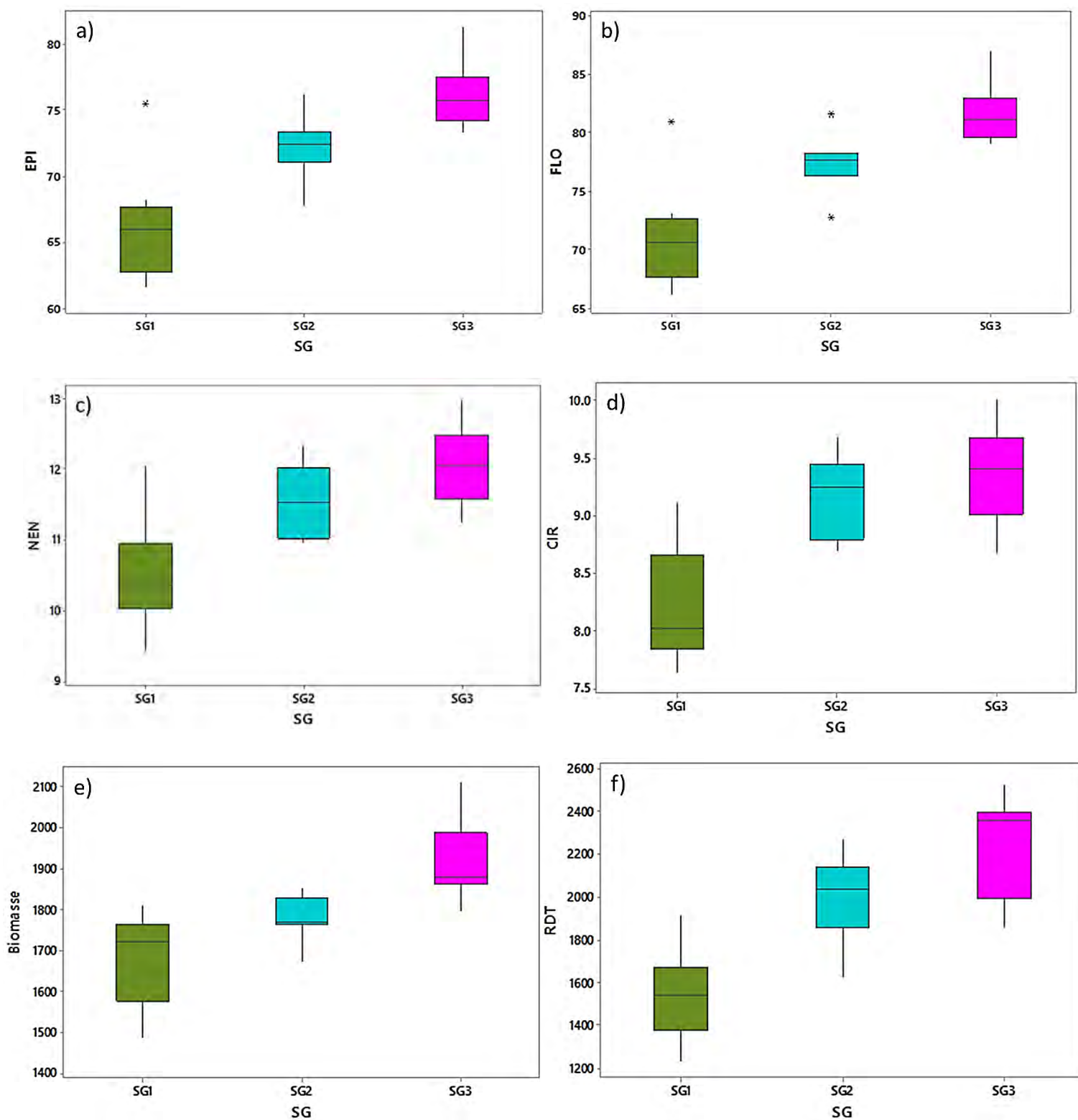


Figure 44 : Box plots des caractères de (a) date d'épiaison, (b) date de floraison et de (c) nombre d'entrenœuds, de (d) circonférence de l'épi, de la (e) biomasse et du (f) rendement, en fonction des sous-groupes 1 (en vert olive), sous-groupe 2 (en cyan) et sous-groupe 3 (en magenta) des Sanio.

4. Analyses d'associations génotypes-phénotypes sur les caractères de différenciation entre Souna et Sanio

Les diagrammes quantile-quantile de chaque modèle pour les différents caractères montrent que les p-values observées sous l'hypothèse nulle, sont relativement bien en accord avec celles attendues. En d'autres termes, le modèle a été approprié pour l'analyse d'association.

Les Manhattan plots montrent que beaucoup de SNPs sont associés (p-values faibles) aux différents traits étudiés, sauf pour la hauteur de la tige principale qui n'est significativement associé à aucun SNP (Figures 45, 46 et 47).

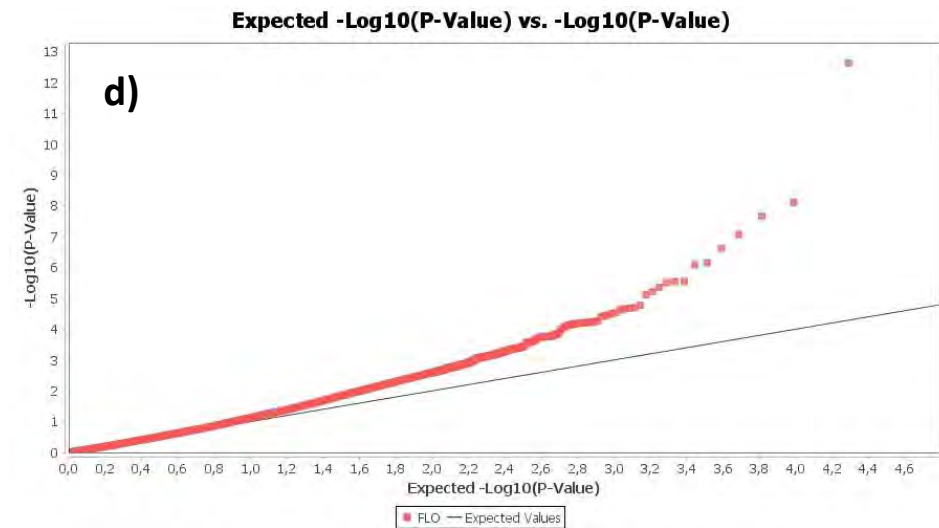
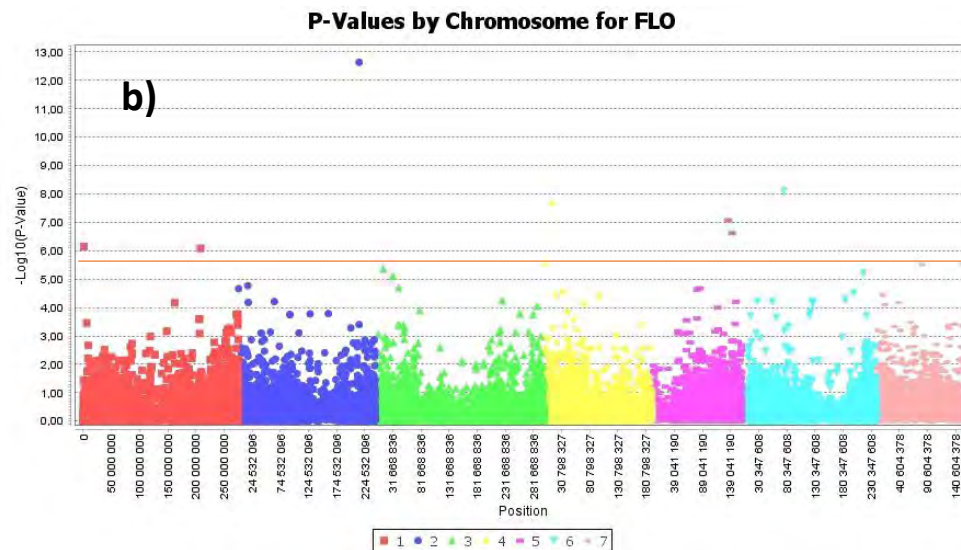
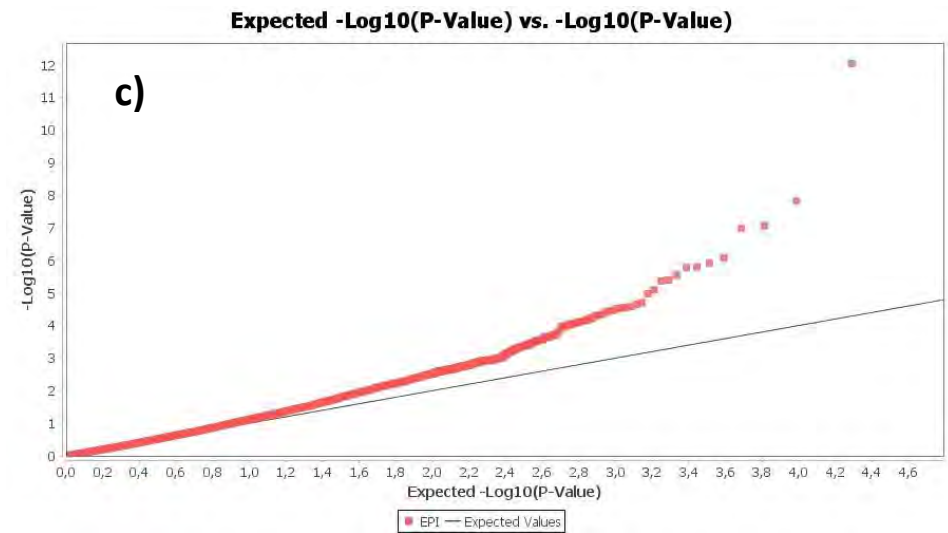
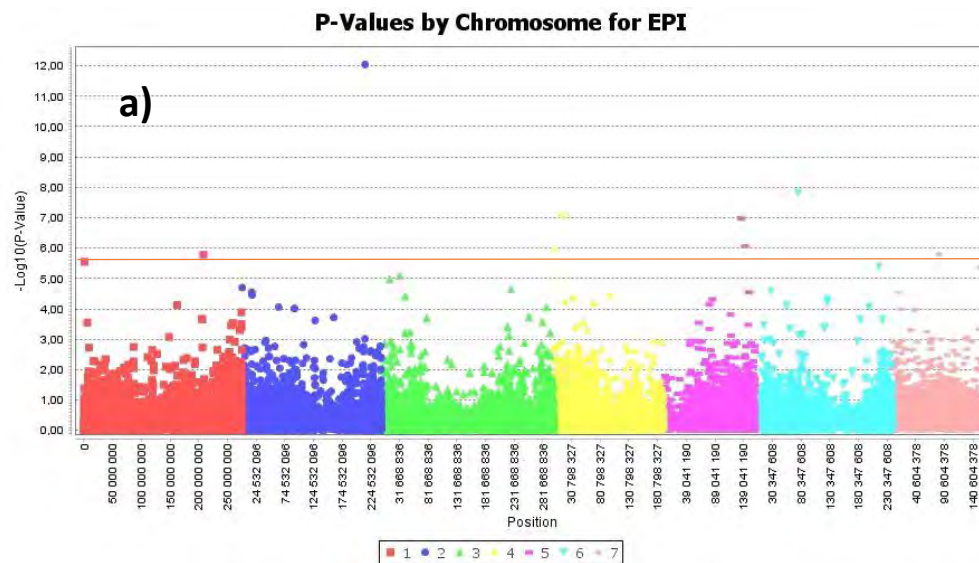


Figure 45 : Manhattan plot par rapport à la date d'épiaison et la date de floraison et QQ plot par rapport aux mêmes caractères, respectivement.

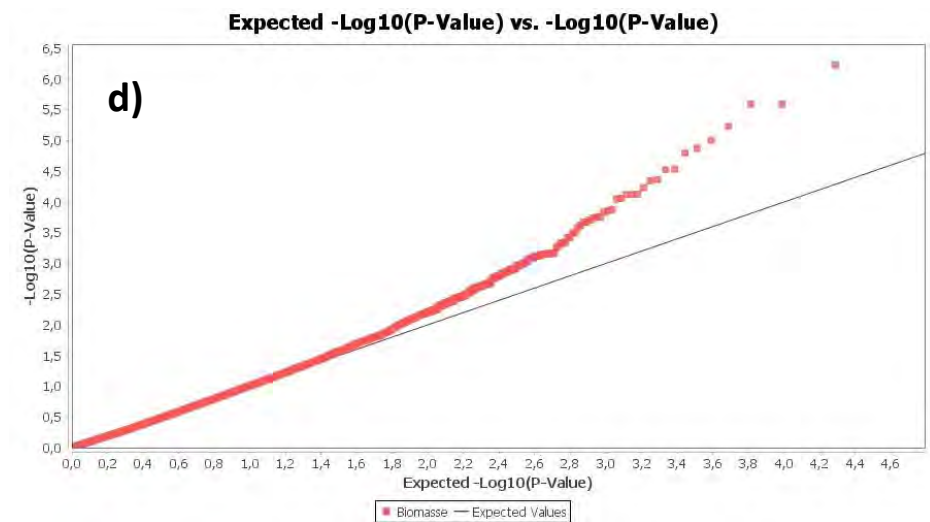
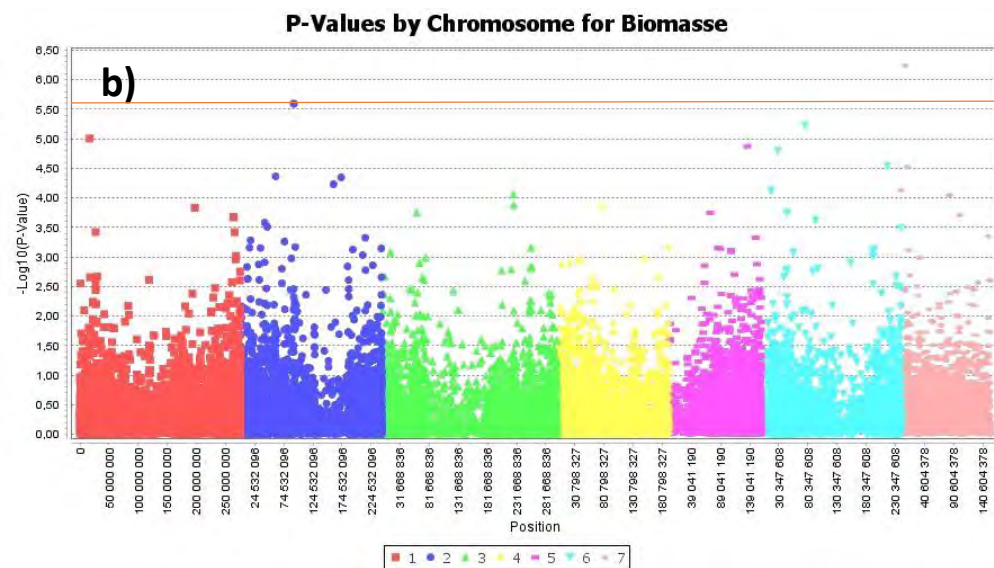
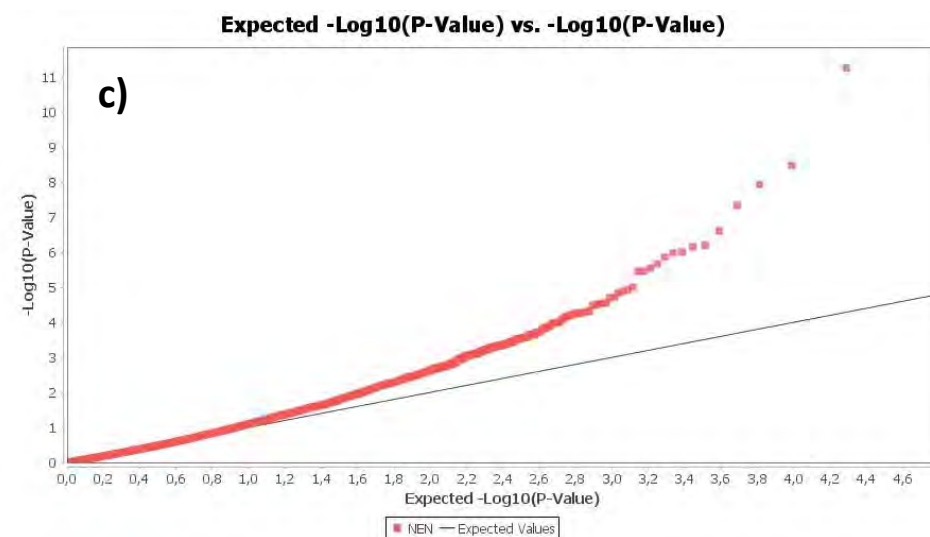
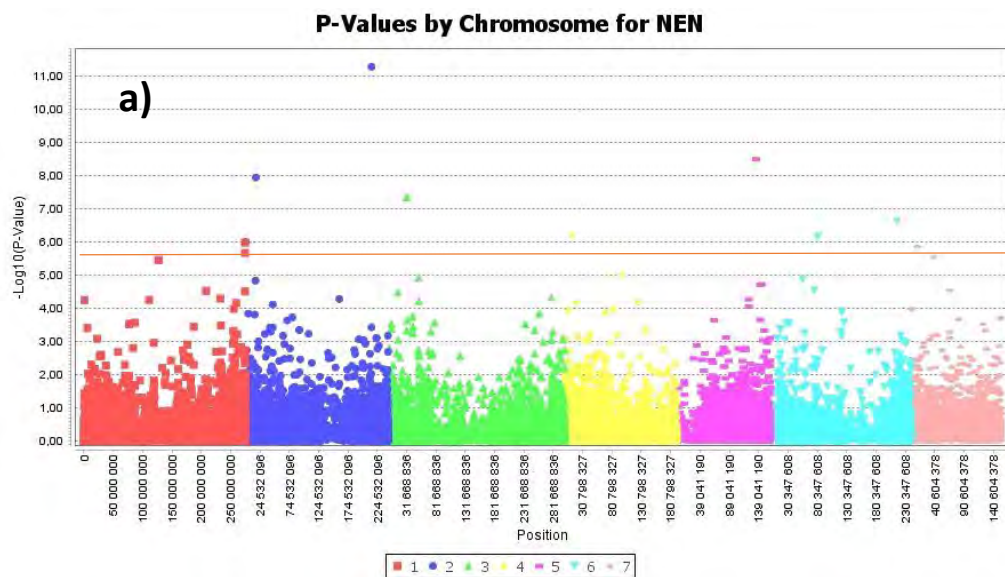


Figure 46 : Manhattan plot par rapport au nombre d'entrecœuds et la biomasse et QQ plot par rapport aux mêmes caractères respectivement.

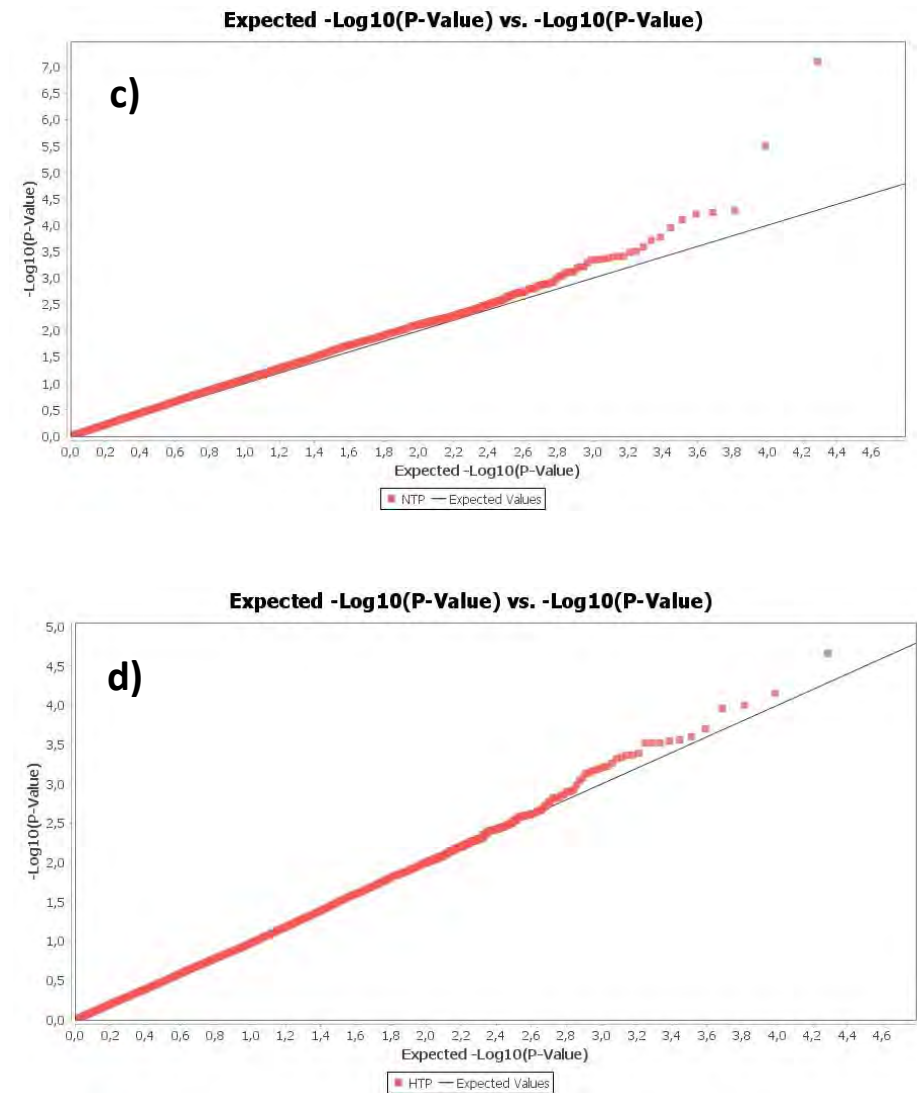
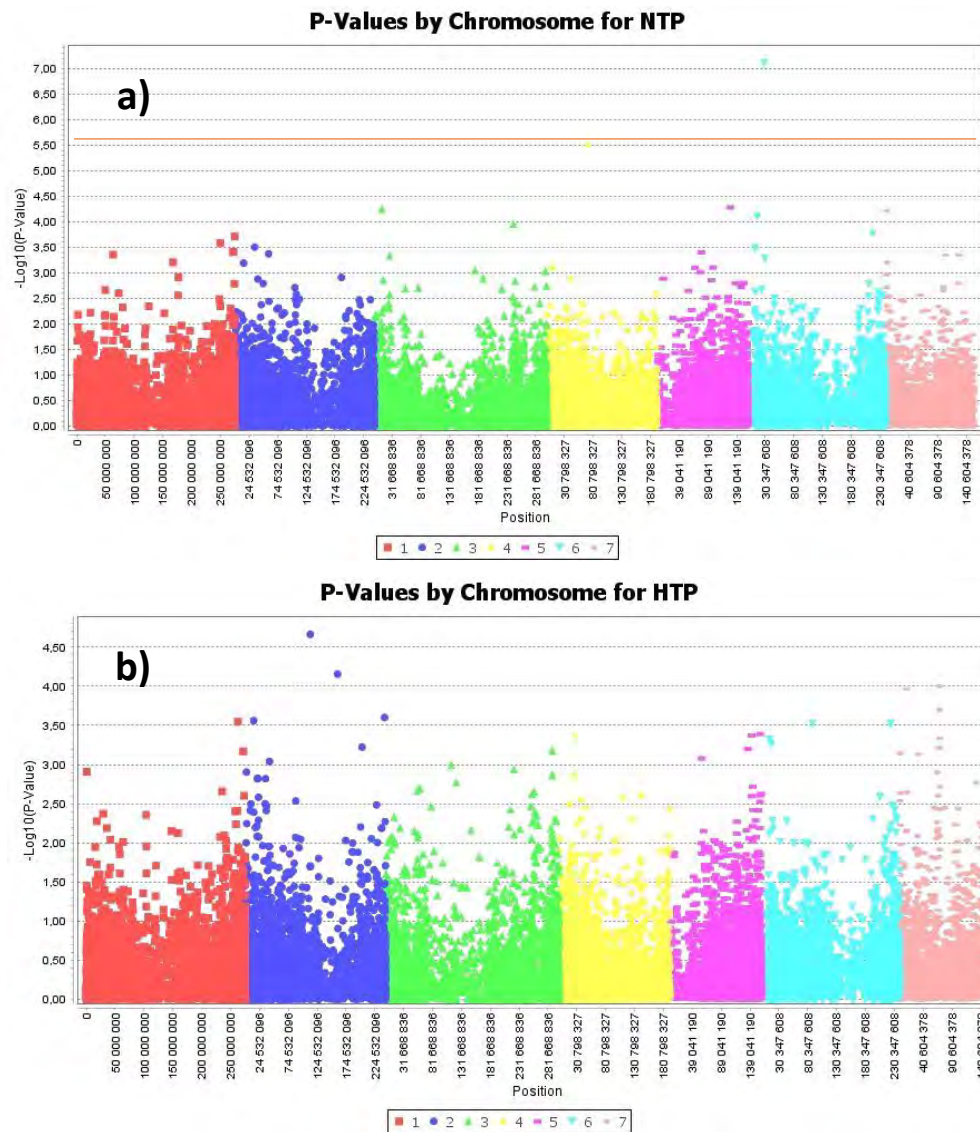


Figure 47 : Manhattan plot par rapport au nombre de talles productives et la hauteur de la tige principale et QQ plot par rapport aux mêmes caractères respectivement

L'analyse a détecté, au seuil de $\alpha = 2,56E-06$ (correction de Bonferroni), 21 SNPs associés aux caractères de différenciation entre Souna et Sanio, sauf à la hauteur de la tige principale pour qui aucun SNP n'a été significativement associé. 3 SNPs communs ont été détectés pour la date d'épiaison, de floraison et le nombre d'entrenœuds. Le caractère de nombre d'entrenœuds a compté 8 SNPs significatifs spécifiquement associés à ce trait, tandis que le nombre de talles productives a enregistré un seul SNP spécifique contre 3 pour la biomasse. Parmi tous les SNPs, le marqueur rs2_214640382, situé à la 214640382^{ème} paire de base du chromosome 2 du mil, s'est particulièrement démarqué avec un taux de significativité très élevé, sur chaque caractère de différenciation entre souna et Sanio auquel il est associé (à savoir la date d'épiaison, la date de floraison et le nombre d'entrenœuds). Le tableau 14 résume les différents marqueurs significatifs et leurs positions dans le génome.

Tableau 14 : Résumé des statistiques de l'analyse d'associations entre marqueurs SNPs et traits phénotypiques de différenciation entre Souna et Sanio

Trait phénotypique	SNP	Chromosome	Position	p-values	FDR
Date d'épiaison	rs2_214640382	2	214640382	8,99E-13	1,00E-08
	rs6_75259866	6	75259866	1,52E-08	1,00E-04
	rs4_14047671	4	14047671	8,59E-08	5,00E-04
	rs5_134902680	5	134902680	1,06E-07	5,00E-04
	rs5_141923948	5	141923948	8,48E-07	3,00E-03
	rs4_2024746	4	2024746	1,19E-06	3,00E-03
	rs7_81434382	7	81434382	1,58E-06	4,00E-03
	rs1_208305247	1	208305247	1,67E-06	4,00E-03
Date de floraison	rs2_214640382	2	214640382	2,30E-13	4,00E-09
	rs6_75259866	6	75259866	7,86E-09	7,00E-05
	rs4_14047671	4	14047671	2,19E-08	1,00E-04
	rs5_134902680	5	134902680	8,48E-08	4,00E-04
	rs5_141923948	5	141923948	2,39E-07	9,00E-04
	rs1_809090	1	809090	7,03E-07	2,00E-03
	rs1_208305247	1	208305247	8,22E-07	2,00E-03
Nombre d'entrenœuds	rs2_214640382	2	214640382	5,41E-12	1,00E-07
	rs5_134902680	5	134902680	3,30E-09	3,00E-05
	rs2_17691169	2	17691169	1,16E-08	7,00E-05
	rs3_31892095	3	31892095	4,55E-08	2,00E-04
	rs6_216688877	6	216688877	2,48E-07	9,00E-04
	rs4_14047671	4	14047671	6,34E-07	1,00E-03
	rs6_81072492	6	81072492	7,02E-07	1,00E-03
	rs2_505979	2	505979	9,95E-07	2,00E-03
	rs1_274445029	1	274445029	1,06E-06	2,00E-03
	rs7_10827013	7	10827013	1,39E-06	2,00E-03
	rs1_274583477	1	274583477	2,16E-06	3,00E-03
Nombre de talles productives	rs6_28155063	6	28155063	7,80E-08	1,00E-03
Biomasse	rs7_6854339	7	6854339	5,83E-07	1,00E-02
	rs2_91628557	2	91628557	2,55E-06	1,00E-02
	rs2_91628558	2	91628558	2,55E-06	1,00E-02

A l'issu de cette étude d'associations génotypes-phénotypes, 21 marqueurs ont été détectés comme significativement associés aux traits de différenciation des accessions de la core collection. Parmi ces 21 marqueurs, 11 SNPs sont en association indirecte, car étant en déséquilibre de liaison avec les loci causatifs. 10 SNPs sont en association directe avec les traits respectifs. Ces SNPs ont permis d'identifier 10 gènes causatifs liés à ces caractères dont quelques-uns codent pour des protéines connues. Parmi ces protéines figurent des facteurs d'épissage de transcription, des protéines intervenant dans la réception de signaux transmembranaires, etc...

Tableau 15 : Liste des marqueurs SNPs et informations connexes

Marqueur	Allèles	Chromosome	Position	Rapport de causalité	Région	Identifiant gène	Nom gène	Correspondance espèce	Score	E-Value	Identité
rs1_208305247	G/C	1	208305247	Indirect	Inter-génique						
rs1_274445029	G/A	1	274445029	Direct	CDS	Pgl_GLEAN_10004427	Non caractérisé	<i>Setaria italica</i>	1323	0.0	92.94%
rs1_274583477	C/G	1	274583477	Direct	CDS	Pgl_GLEAN_10004445	NAC1	<i>Setaria italica</i>	325	4,00E-111	83.98%
rs1_809090	G/A	1	809090	Indirect	Inter-génique						
rs2_17691169	A/T	2	17691169	Direct	CDS	Pgl_GLEAN_10002066	Non caractérisé	<i>Setaria italica</i>	1226	0.0	88.33%
rs2_214640382	C/T	2	214640382	Indirect	Inter-génique						
rs2_505979	T/C	2	505979	Indirect	Inter-génique						
rs2_91628557	C/T	2	91628557	Indirect	Inter-génique						
rs2_91628558	C/T	2	91628558	Indirect	Inter-génique						
rs3_31892095	T/G	3	31892095	Indirect	Inter-génique						
rs4_14047671	G/A	4	14047671	Direct	CDS	Pgl_GLEAN_10007732	LRR-RLKs	<i>Panicum miliaceum</i>	1546	0.0	78.53%
rs4_2024746	G/C	4	2024746	Indirect	Inter-génique						
rs5_134902680	C/T	5	134902680	Direct	CDS	Pgl_GLEAN_10029328	SRRM1	<i>Setaria italica</i>	584	0.0	78.63%
rs5_141923948	T/C	5	141923948	Direct	CDS	Pgl_GLEAN_10009101	RS-rich 19	<i>Setaria italica</i>	701	0.0	87.64%
rs6_216688877	T/C	6	216688877	Direct	CDS	Pgl_GLEAN_10030550	Non caractérisé	<i>Dichanthelium oligosanthos</i>	51.6	6,00E-05	55.36%
rs6_28155063	C/T	6	28155063	Direct	CDS	Pgl_GLEAN_10017678	Non caractérisé	<i>Setaria italica</i>	636	0.0	88.86%
rs6_75259866	A/T	6	75259866	Indirect	Inter-génique						
rs6_81072492	G/A	6	81072492	Direct	CDS	Pgl_GLEAN_10004087	DRP 1	<i>Setaria italica</i>	1020	0.0	83.60%
rs7_10827013	T/C	7	10827013	Direct	CDS	Pgl_GLEAN_10023229	Non caractérisé	<i>Setaria italica</i>	1330	0.0	87.99%
rs7_6854339	T/G	7	6854339	Indirect	Inter-génique						
rs7_81434382	A/C	7	81434382	Indirect	Inter-génique						

DISCUSSION

L'analyse de diversité génétique réalisée sur les marqueurs SNPs obtenus à travers cette étude a confirmé les résultats obtenus dans les chapitres précédents, avec différents types de marqueurs (SSRs). Ces observations confortent davantage la qualité des différents types de marqueurs utilisés dans cette thèse et aussi la qualité des différentes analyses réalisées.

Dans cette étude, nous avons utilisé un modèle linéaire mixte (MLM), pour analyser les associations possibles entre marqueurs génétiques et traits phénotypiques sur la core collection de mil du Sénégal. Le modèle MLM est connu comme étant un modèle puissant pour réduire les faux positifs. Les diagrammes quantile-quantile obtenus dans cette étude illustre, la capacité de ce modèle à prendre en compte les effets de structure et de relations cryptiques entre individus.

Les meilleures correspondances obtenues lors du blast des séquences de gènes, l'ont été avec l'espèce *Setaria italica*, avec 8 correspondances sur les 10 séquences blastées. Ces résultats confirment davantage la proximité déjà établie entre le mil et la setaie.

La moitié des gènes qui sont associés à des caractères de différenciation entre Souna et Sanio dans notre étude, est constituée de gènes qui codent pour des protéines non encore identifiées. Certains des gènes documentés dont les fonctions sont connues, qui sont associés à des caractères de différenciation entre Souna et Sanio, sont des gènes codant pour des facteurs de transcription (NAC) ou des facteurs d'épissage (RS-rich 19 et SRRM1). D'autres gènes codent pour des protéines qui interviennent dans la réception de signaux transmembranaires (LRR-RLK) ou des protéines cytoplasmiques (DRP1).

Les gènes RS-rich 19 et SRRM1 codent pour des protéines appartenant à une famille (protéines RS) qui joue un rôle clé dans la régulation de l'expression des gènes (Shepard and Hertel, 2009) et serait aussi impliquées dans la réponse des plantes aux stress abiotiques (Duque, 2011).

Le gène DRP1 code pour une famille de protéines cytoplasmiques appelées dynamine qui sont impliquées dans l'endocytose cellulaire (Bednarek and Backues, 2010). Vítámvás et al., (2015) auraient identifié, à travers une analyse protéomique, un différentiel d'expression de la dynamine chez des plantes d'orge sous différentes conditions hydriques. Cependant, aucune étude précédente n'aurait reporté des observations similaires chez les céréales (Vítámvás et al., 2015).

Les LRR-RLKs sont des récepteurs à activité tyrosine kinase qui joueraient un rôle dans un grand nombre de voies de transduction de signaux liées aux réponses hormonales et à des stress abiotiques (Colette et al., 2011). En effet certaines protéines appartenant à cette famille de gènes comme ERECTA ont été associées à l'adaptation à la sécheresse chez le haricot (Blair et al., 2016) et chez *Arabidopsis thaliana* (Masle et al., 2005).

Le gène NAC code pour des protéines qui constituent l'une des plus grandes familles de facteurs de transcription des plantes (Jensen et al., 2010a). Ce sont des régulateurs clé intervenant dans la perception du stress et dans le développement de la plante (Jensen et al., 2010b). En effet des études ont mis en évidence leur implication dans la résistance au stress abiotiques comme le stress salin chez la tomate (Chen et al., 2009). Elles seraient aussi impliquées dans la résistance à la sécheresse chez le haricot (Wu et al., 2016), le soja (Hussain et al., 2017) et chez le riz (Chung et al., 2018). Chez *Arabidopsis thaliana* et chez le riz, des études utilisant une approche biotechnologique, ayant produit des plantes transgéniques par la surexpression du facteur NAC ont permis d'améliorer leur tolérance à la sécheresse (Nakashima et al., 2012). D'autres études d'expressions de ce gène, menées sur le Peuplier de l'Euphrate ont suggéré que la capacité de cette plante à s'adapter à des environnements de haute salinité et de sécheresse pourrait être liée aux profils d'expressions différentiels de certaines protéines de ce gène NAC.

Il semble donc que la plupart des gènes associés aux caractères de différenciation entre Souna et Sanio, soient impliqués dans la résistance aux stress abiotiques et à l'adaptation à la sécheresse. Les Souna et Sanio du Sénégal présentent une disposition géographique qui suit un gradient climatique de pluviométrie. Il n'est donc absolument pas surprenant que la plupart des gènes associés à la différenciation de ces deux morphotypes soient des gènes jouant un rôle la résistance aux stress abiotiques (surtout la sécheresse) et dans l'adaptation des plantes.

CONCLUSION

Cette étude a permis d'identifier 21 marqueurs SNPs et 10 de gènes candidats associés à des traits qui interviennent dans la différenciation des morphotypes du mil au Sénégal et des traits phrénologiques. Elle a permis de disposer ainsi de nouveaux outils génétiques et génomiques du mil.

Grâce à cette étude, des milliers de marqueurs SNPs sont à présent disponibles dans la core collection de mil du Sénégal. Des bases génétiques des traits de différenciation entre morphotypes Souna et Sanio du mil sont identifiées et mieux connues. Les résultats ainsi obtenus pourront servir aux sélectionneurs dans les programmes de sélections assistées par marqueurs. Ils pourraient aussi être utilisés pour la biologie évolutive afin de comprendre l'histoire évolutive du mil qui lui a permis de passer du mil sauvage aux deux morphotypes observés actuellement au Sénégal (Souna et Sanio).

En perspectives, il serait important d'adjoindre à cette étude d'association, des études génétiques fonctionnelles : (1) afin d'étudier la régulation de l'expression des gènes chez le mil, car même si certains de leurs orthologues sont documentés chez d'autres espèces, d'autres sont à l'état hypothétique et peu sont documentés chez le mil. En effet, il peut exister des différences fonctionnelles entre gènes orthologues d'une espèce à l'autre comme c'est le cas du gène *Hdl* chez le riz comparé à son orthologue chez *Arabidopsis* (Tsuji et al., 2011). (2) Afin d'approfondir les connaissances en génétique fonctionnelle sur ces gènes comme par exemple, l'identification de différentiels d'expressions de ces gènes en fonction des phases de développement du mil.

CHAPITRE V : CONCLUSION GENERALE ET PERSPECTIVES

CONCLUSION

A l'entame de ces travaux de thèse, l'état des ressources phytogénétiques de mil du Sénégal n'était pas bien connu. En effet, la conservation qui en était faite n'était pas représentative de la diversité génétique des accessions rencontrées au Sénégal. En plus, leur connaissance surtout sur le plan génétique faisait défaut.

Ce travail a permis de réaliser une collection constituée de quatre cent soixante-dix-sept (477) accessions en couvrant tout le territoire national. Ces accessions ont été évaluées sur le plan génétique (neutre et adaptative) et sur le plan agro-morphologique. Ces études morphologiques et génétiques ont permis d'en sélectionner 91 accessions ayant capturer au mieux la diversité de cette collection. Cette diversité comparée à la diversité de mils de la sous-région a montré que les mils du Sénégal ont une large variabilité. Ces études ont aussi permis de distinguer clairement Souna et Sanio aussi bien au niveau génétiques, par une analyse bayésienne, que morphologiques. Cette étude poussée sur le plan morphologique au sein de ces différents pools génétiques a révélé que trois sous-groupes pour le Souna se distinguent en fonction du rendement en grains et en épi, de la longueur de l'épi, de la circonférence de l'épi, la hauteur de la plante, le nombre d'entrenœuds, la longueur et largeur de la feuille paniculaire et de l'incidence du mildiou et trois sous-groupes pour le Sanio se distinguent en fonction du rendement en épi, la date de floraison et la longueur de la feuille paniculaire. Pour adresser cette différenciation entre Souna et Sanio, de manière fine, nous avons utilisé les nouvelles technologies de séquençage dont le GBS et avons identifié six caractères morphologiques de différenciation de ces derniers à savoir la date d'épiaison et de floraison, la biomasse, la hauteur de la plante, le nombre moyen d'entrenœuds et le nombre de talles productives. Vingt-et-un (21) SNPs associés à ces caractères ont été détectés, grâce à une étude d'associations génotypes-phénotypes (GWAS), dont onze (11) de ces SNPs se situent dans des régions inter-géniques du génome du mil et dix (10) dans des régions géniques. Certains de ces gènes associés à ces caractères de différenciation entre Souna et Sanio sont impliqués dans l'adaptation et la résistance au stress environnementaux. Enfin, des associations supplémentaires de marqueurs des gènes adaptatifs *PgPHYC* et *PgMADS11* par rapport à cette différenciation ont été obtenus. Tous les SNPs des gènes *PgPHYC*, *PgMADS11* et des gènes issus du GWAS peuvent être validés et utilisés dans les programmes de sélection.

PERSPECTIVES

Par préoccupation du devenir cette diversité du mil du Sénégal en réponse aux prévisions des changements climatiques, une modélisation de l'adaptation du mil suivant les scénarii futurs prédits par le GIEC (Groupe d'experts intergouvernemental sur l'évolution du climat) a été initiée. Les résultats préliminaires montrent que dans les 50 ans à venir, le mil sera préservé au Sénégal mais par contre, les zones de cultures vont changer.

Dans le souci d'avoir un suivi dynamique de la diversité des ressources de mils cultivés du Sénégal, il serait aussi intéressant de procéder à des échantillonnages temporels de l'ordre d'une à deux décennies. Ce qui permettrait de savoir vers quel scénario futur évolueraient ces mils et de pouvoir anticiper dans leur gestion et leur valorisation.

Beaucoup de programmes de sélection ont été réalisés pour la précocité de floraison (Souna) laissant de côté les Sanio. Il serait intéressant de mieux utiliser cette diversité du mil en intégrant les Sanio qui n'ont pas encore été exploités et qui présentent pourtant des traits d'intérêts à prendre en compte dans les schémas de sélection considérant l'aspect cultures à double usage. Ces schémas pourraient permettre de produire, par exemple, des hybrides pour les caractères de rendement en grain et en paille et de précocité, en croisant Souna et Sanio.

REFERENCES BIBLIOGRAPHIQUES

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., et al. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* *46*, W537–W544.

Akanvou, L., Akanvou, R., Kouakou, C.K., N'Da, H.A., and Koffi, K.G.C. (2012). Evaluation de la diversité agro morphologique des accessions de mil [*Pennisetum glaucum* (L.) R. Br.] collectées en Côte d'Ivoire. *J. Appl. Biosci.* *50*, 3468– 3477.

Allouis, S., Qi, X., Lindup, S., Gale, M.D., and Devos, K.M. (2001). Construction of a BAC library of pearl millet, *Pennisetum glaucum*. *Theor. Appl. Genet.* *102*, 1200–1205.

Angé, A. (1991). La fertilité des sols et les stratégies paysannes de mise en valeur des ressources naturelles. Le mil dans les systèmes de culture du Sud du bassin arachidier sénégalais.

Balfourier, F., Charmet, G., Prosperi, J.-M., Goulard, M., and Monestiez, P. (1998). Comparison of different spatial strategies for sampling a core collection of natural populations of fodder crops. *Genet. Sel. Evol.* *30*, S215.

Bataillon, T.M., David, J.L., and Schoen, D.J. (1996). Neutral genetic markers and conservation genetics: simulated germplasm collections. *Genetics* *144*, 409–417.

Bednarek, S.Y., and Backues, S.K. (2010). Plant dynamin-related protein families DRP1 and DRP2 in plant development.

Bezançon, G., and Pham, J.-L. (2004). Ressources génétiques des mils en Afrique de l'Ouest : Diversité, conservation et valorisation (IRD Éditions).

Bezançon, G., Renno, J.-F., and Kumar, K.A. (1997a). Le mil.

Bezançon, G., Renno, J.-F., and Anand Kumar, K. (1997b). Le mil. In *L'amélioration des plantes tropicales*, A. Charrier, M. Jacquot, S. Hamon, and D. Nicolas, eds. (Montpellier: CIRAD), pp. 457–482.

Bezançon, G., Pham, J.-L., Deu, M., Vigouroux, Y., Sagnard, F., Mariac, C., Kapran, I., Mamadou, A., Gérard, B., Ndjeunga, J., et al. (2008). Changes in the diversity and geographic distribution of cultivated millet (*Pennisetum glaucum* (L.) R. Br.) and sorghum (*Sorghum*

bicolor (L.) Moench) varieties in Niger between 1976 and 2003. *Genet. Resour. Crop Evol.* 56, 223–236.

Bhattacharjee, R., Khairwal, I.S., Bramel, P.J., and Reddy, K.N. (2007). Establishment of a pearl millet [*Pennisetum glaucum* (L.) R. Br.] core collection based on geographical distribution and quantitative traits. *Euphytica* 155, 35–45.

Bilquez, A., and Lecomte, J. (1969). Relations entre mils sauvages et mils cultivés: Etude de l'hybride. *Agron Trop.*

Bisht, I.S., Mahajan, R.K., Loknathan, T.R., and Agrawal, R.C. (1998). Diversity in Indian sesame collection and stratification of germplasm accessions in different diversity groups. *Genet. Resour. Crop Evol.* 45, 325–335.

Blair, M.W., Cortés, A.J., and This, D. (2016). Identification of an ERECTA gene and its drought adaptation associations with wild and cultivated common bean. *Plant Sci.* 242, 250–259.

Bodian, A. (2014). Caractérisation de la variabilité temporelle récente des précipitations annuelles au Sénégal (Afrique de l'Ouest). *Physio-Géo* 297–312.

Brown, A.H.D. (1989a). Core collections: a practical approach to genetic resources management. *Genome* 31, 818–824.

Brown, A.H.D. (1989b). The case for core collections. 136–156.

Budak, H., Pedraza, F., Cregan, P.B., Baenziger, P.S., and Dweikat, I. (2003). Development and Utilization of SSRs to Estimate the Degree of Genetic Relationships in a Collection of Pearl Millet Germplasm. *Crop Sci.* 43, 2284.

Burgarella, C., Chantret, N., Gay, L., Prosperi, J.-M., Bonhomme, M., Tiffin, P., Young, N.D., and Ronfort, J. (2016). Adaptation to climate through flowering phenology: a case study in *Medicago truncatula*. *Mol. Ecol.* n/a-n/a.

Chandrashekhar, H., Jayaram, G., and Jayashree, U. (2012). Formation of core set in Indian and African finger millet [*Eleusine coracana* (L.) Gaertn] germplasm accessions. *Indian J. Genet. Plant Breed.* 72, 358–363.

- Chanterreau, J., Deu, M., Pham, J.L., Kapran, I., Vigouroux, Y., and Bezancon, G. (2010). Evolution des diversites phenotypique et genetique des sorghos et mils cultives au Niger de 1976 a 2003. *Sel. Francais* 33–45.
- Charmet, G., and Balfourier, F. (1995). The use of geostatistics for sampling a core collection of perennial ryegrass populations. *Genet. Resour. Crop Evol.* 42, 303–309.
- Chen, S., Gollop, N., and Heuer, B. (2009). Proteomic analysis of salt-stressed tomato (*Solanum lycopersicum*) seedlings: effect of genotype and exogenous application of glycinebetaine. *J. Exp. Bot.* 60, 2005–2019.
- Chowdari, K.V., Davierwala, A.P., Gupta, V.S., Ranjekar, P.K., and Govila, O.P. (1998). Genotype identification and assessment of genetic relationships in pearl millet [*Pennisetum glaucum* (L.) R. Br.] using microsatellites and RAPDs. *Theor. Appl. Genet.* 97, 154–162.
- Chung, P.J., Jung, H., Choi, Y.D., and Kim, J.-K. (2018). Genome-wide analyses of direct target genes of four rice NAC-domain transcription factors involved in drought tolerance. *BMC Genomics* 19.
- Clotault, J., Thuillet, A.-C., Buiron, M., De Mita, S., Couderc, M., Haussmann, B.I.G., 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.
- Colette, A., Bochdanovits, Z., Jansweijer, V.M., Koning, F.G., Berke, L., Sanchez-Perez, G.F., Scheres, B., and Heidstra, R. (2011). Probing the roles of LRR RLK genes in *Arabidopsis thaliana* roots using a custom T-DNA insertion set. *Plant Mol. Biol.* 76, 69–83.
- Cronquist, A. (1981). *An Integrated System of Classification of Flowering Plants* (Columbia University Press).
- Dancette, C. (1980). Water requirements and adaptations to the rainy season of millet in Senegal. *Int. Crops Res. Inst. Semi-Arid Trop. Proc. Int. Workshop Agroclimatol. Res. Needs Semi-Arid Trop.* Hyderabad India 22–24 Novemb. 1978 106–120.
- Dancette, C. (1983a). Besoins en eau du mil au Senegal: adaptation en zone semi-aride tropicale. *Agron. Trop.* 38, 267–280.

Dancette, C. (1983b). Besoins en eau du mil au Sénégal: adaptation en zone semi-aride tropicale. *Agron. Trop.* 38, 267–280.

Debieu, M., Kanfany, G., and Laplaze, L. (2017). Pearl Millet Genome: Lessons from a Tough Crop. *Trends Plant Sci.*

Dembélé, N.N. (2001). Sécurité alimentaire en Afrique Sub-saharienne: Quelle Stratégie de Réalisation? *Rapp. Dévaluation Mich. State Univ. Agric. Mali Bamako.*

Diack, O., Kane, N.A., Berthouly-Salazar, C., Gueye, M.C., Diop, B.M., Fofana, A., Sy, O., Tall, H., Zekraoui, L., Piquet, M., et al. (2017). New Genetic Insights into Pearl Millet Diversity As Revealed by Characterization of Early- and Late-Flowering Landraces from Senegal. *Front. Plant Sci.* 8.

Diakhaté, S., Villenave, C., Diallo, N.H., Ba, A.O., Djigal, D., Masse, D., Sembène, P.M., and Chapuis-Lardy, L. (2013). The influence of a shrub-based intercropping system on the soil nematofauna when growing millet in Senegal. *Eur. J. Soil Biol.* 57, 35–41.

Diangar, S., Fofana, A., Diagne, M., Yamoah, C.F., and Dick, R.P. (2004). Pearl millet-based intercropping systems in the semiarid areas of Senegal. *Afr. Crop Sci. J.* 12, 133-139–139.

Diop, M. (1999). Caractérisation du facteur hydrique en agriculture pluviale au Sénégal : le cas de l'arachide et du mil (Dijon).

Diop, M., Reyniers, F.-N., and Sarr, B. (2005). Apport du photopériodisme à l'adaptation du mil à la sécheresse en milieu soudano-sahélien. *Sci. Chang. Planétaires Sécher.* 16, 35–40.

Diouf, M. (1991). Diagnostic agronomique en parcelles paysannes. Une méthode d'amélioration des systèmes de culture.

Doyle, J.L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.

Duque, P. (2011). A role for SR proteins in plant stress responses. *Plant Signal. Behav.* 6, 49–54.

Dussert, Y., Snirc, A., and Robert, T. (2015a). Inference of domestication history and differentiation between early- and late-flowering varieties in pearl millet. *Mol. Ecol.* *24*, 1387–1402.

Dussert, Y., Snirc, A., and Robert, T. (2015b). Inference of domestication history and differentiation between early- and late-flowering varieties in pearl millet. *Mol. Ecol.* *24*, 1387–1402.

Dutta, M., Phogat, B.S., Kumar, S., Kumar, N., Kumari, J., Pandey, A.C., Singh, T.P., Tyagi, R.K., Jacob, S.R., Srinivasan, K., et al. (2015). Development of Core Set of Wheat (*Triticum* spp.) Germplasm Conserved in the National Genebank in India. In *Advances in Wheat Genetics: From Genome to Field*, Y. Ogihara, S. Takumi, and H. Handa, eds. (Springer Japan), pp. 33–45.

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* *14*, 2611–2620.

Falush, D., Stephens, M., and Pritchard, J.K. (2003). Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics* *164*, 1567–1587.

FAO (2010). *The Second Report on the State of the World's Plant Genetic Resources*. (Rome, Italy).

FAOSTAT (2017). FAOSTAT World Total Population.

Faye, I., Diouf, O., Guissé, A., Sène, M., and Diallo, N. (2006). Characterizing Root Responses to Low Phosphorus in Pearl Millet [*Pennisetum glaucum* (L.) R. Br.]. *Agron. J.* *98*, 1187–1194.

Field, D., Tiwari, B., Booth, T., Houten, S., Swan, D., Bertrand, N., and Thurston, M. (2006). Open software for biologists: from famine to feast. *Nat. Biotechnol.* *24*, 801–803.

Frankel, O.H., and Brown, A.H.D. (1984). Current plant genetic resources--a critical appraisal. *Genet. New Front. Proc. XV Int. Congr. Genet.* Ed. VL Chopra Al.

Frary, A., Tekin, P., Celik, I., Furat, S., Uzun, B., and Doganlar, S. (2015). Morphological and molecular diversity in Turkish sesame germplasm and core set selection. *Crop Sci.* *55*, 702–711.

- Gahukari, R.T. (1990). Reaction of Locally Improved Pearl Millets to Three Insect Pests and Two Diseases in Senegal. *J. Econ. Entomol.* 83, 2102–2106.
- Ganry, F., Guiraud, G., and Dommergues, Y. (1978). Effect of straw incorporation on the yield and nitrogen balance in the sandy soil-pearl millet cropping system of Senegal. *Plant Soil* 50, 647–662.
- Gepts, P., and Clegg, M.T. (1989). Genetic Diversity in Pearl Millet (*Pennisetum glaucum* [L.] R. Br.) at the DNA Sequence Level. *J. Hered.* 80, 203–208.
- Glaszmann, J., Kilian, B., Upadhyaya, H., and Varshney, R. (2010). Accessing genetic diversity for crop improvement. *Curr. Opin. Plant Biol.* 13, 167–173.
- Gouesnard, B., Bataillon, T.M., Decoux, G., Rozale, C., Schoen, D.J., and David, J.L. (2001). MSTRAT: An Algorithm for Building Germ Plasm Core Collections by Maximizing Allelic or Phenotypic Richness. *J. Hered.* 92, 93–94.
- Govindaraj, M., Selvi, B., Prabhu, D.A., and Rajarathinam, S. (2009). Genetic diversity analysis of pearl millet (*Pennisetum glaucum* [L.] R. Br.) accessions using molecular markers. *Afr. J. Biotechnol.* 8, 6046–6052.
- Gowda, J., Krishanappa, M., Pathak, N., and et al (2013). Use of Heuristic Approach for the Development of a Core Set from Large Germplasm Collection of Foxtail Millet (*Setaria italica* L.). *Indian J. Plant Genet. Resour.* 26, 13–18.
- Hammer, K. (2003). A paradigm shift in the discipline of plant genetic resources. *Genet. Resour. Crop Evol.* 50, 3–10.
- Hammer, K., Arrowsmith, N., and Gladis, T. (2003). Agrobiodiversity with emphasis on plant genetic resources. *Naturwissenschaften* 90, 241–250.
- Hamon, S., Hodgkin, T., Dussert, S., Anthony, F., Noirot, M., and Eucarpia, Clermont-Ferrand (FRA), 1994/03/15-18 (1994). Core collection : theoretical and applied aspects. In *Evaluation and Exploitation of Genetic Resources Pre-Breeding : Proceedings of the Genetic Resources Section*, (s.l.: s.n.), pp. 73–82.

- Hausmann, B.I.G., Boubacar, A., Boureima, S.S., and Vigouroux, Y. (2006). Multiplication and preliminary characterization of west and central African pearl millet landraces. *Int. Sorghum Millets Newsl.* 47, 110–112.
- Hausmann, B.I.G., Boureima, S.S., Kassari, I.A., Moumouni, K.H., and Boubacar, A. (2007). Mechanisms of adaptation to climate variability in West African pearl millet landraces – a preliminary. *J. SAT Agric. Res.* 3, 1–3.
- Hintum, T.J.L. van, Brown, A.H.D., Spillane, C., and Hodgkin, T. (2000). Core collections of plant genetic resources (Rom: International Plant Genetic Resources Institute).
- Hu, J., Zhu, J., and Xu, H.M. (2000). Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. *Theor. Appl. Genet.* 101, 264–268.
- Hu, Z., Mbacké, B., Perumal, R., Guèye, M.C., Sy, O., Bouchet, S., Prasad, P.V.V., and Morris, G.P. (2015). Population genomics of pearl millet (*Pennisetum glaucum* (L.) R. Br.): Comparative analysis of global accessions and Senegalese landraces. *BMC Genomics* 16, 1048.
- Hurlbert, S.H. (1971). The Nonconcept of Species Diversity: A Critique and Alternative Parameters. *Ecology* 52, 577–586.
- Hussain, R.M., Ali, M., Feng, X., and Li, X. (2017). The essence of NAC gene family to the cultivation of drought-resistant soybean (*Glycine max* L. Merr.) cultivars. *BMC Plant Biol.* 17.
- IPGRI, and ICRISAT (1993). Descripteurs du mil penicillaire [*Pennisetum glaucum* (L.) R. Br.] (Conseil international des ressources phytogénétiques, Rome, Italie: Institut International de Recherche sur les Cultures des Zones Tropicales Semi-arides, Patancheru, Inde).
- Jensen, M.K., Kjaersgaard, T., Petersen, K., and Skriver, K. (2010a). NAC genes: time-specific regulators of hormonal signaling in Arabidopsis. *Plant Signal. Behav.* 5, 907–910.
- Jensen, M.K., Kjaersgaard, T., Nielsen, M.M., Galberg, P., Petersen, K., O'shea, C., and Skriver, K. (2010b). The Arabidopsis thaliana NAC transcription factor family: structure–function relationships and determinants of ANAC019 stress signalling. *Biochem. J.* 426, 183–196.

- Johnson, J.B., Peat, S.M., and Adams, B.J. (2009). Where's the ecology in molecular ecology? *Oikos* 118, 1601–1609.
- Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405.
- Kane, N.A. (2014). Untapped genetic diversity of pearl millet landraces from Senegal as reveals by agro-morphological traits and SSRs markers (San Diego, USA).
- Kapila, R.K., Yadav, R.S., Plaha, P., Rai, K.N., Yadav, O.P., Hash, C.T., and Howarth, C.J. (2008). Genetic diversity among pearl millet maintainers using microsatellite markers. *Plant Breed.* 127, 33–37.
- Kim, K.-W., Chung, H.-K., Cho, G.-T., Ma, K.-H., Chandrabalan, D., Gwag, J.-G., Kim, T.-S., Cho, E.-G., and Park, Y.-J. (2007). PowerCore: a program applying the advanced M strategy with a heuristic search for establishing core sets. *Bioinformatics* 23, 2155–2162.
- Kimura, M., and Crow, J.F. (1964). The Number of Alleles That Can Be Maintained in a Finite Population. *Genetics* 49, 725–738.
- Kouakou, P.K., Muller, B., Guissé, A., Yao, R.N., Fofana, A., and Cissé, N. (2013). Étude et prise en compte en modélisation de l'effet de la latitude sur la réponse à la photopériode chez divers génotypes de mil (*Pennisetum glaucum* (L.) R. Br.) du Sénégal. *J. Appl. Biosci.* 67, 5289–5301.
- Kouressy, M., Vaksman, M., Niangado, O., and Sanogo, M.D. (2004). Valorisation et préservation de la diversité génétique du mil au Mali (IRD Éditions).
- Lakis, G., Navascués, M., Rekima, S., Simon, M., Remigereau, M.-S., Leveugle, M., Takvorian, N., Lamy, F., Depaulis, F., and Robert, T. (2012a). Evolution of Neutral and Flowering Genes along Pearl Millet (*Pennisetum glaucum*) Domestication. *PLoS ONE* 7, e36642.
- Lakis, G., Ousmane, A.M., Sanoussi, D., Habibou, A., Badamassi, M., Lamy, F., Jika, N., Sidikou, R., Adam, T., Sarr, A., et al. (2012b). Evolutionary dynamics of cycle length in pearl millet: the role of farmer's practices and gene flow. *Genetica* 139, 1367–1380.

- Lewis, L.R. (2010). Biogeography and Genetic Diversity of Pearl Millet (*Pennisetum glaucum*) from Sahelian Africa. *Prof. Geogr.* 62, 377–394.
- Maiti, R., and Bidinger, F.R. (1981). Growth and Development of the Pearl Millet Plant. Research Bulletin no. 6.
- Manning, K., Pelling, R., Higham, T., Schwenniger, J.-L., and Fuller, D.Q. (2011). 4500-Year old domesticated pearl millet (*Pennisetum glaucum*) from the Tilemsi Valley, Mali: new insights into an alternative cereal domestication pathway. *J. Archaeol. Sci.* 38, 312–322.
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27, 209–220.
- Marchais, L. (1982). Diversité phénotypique des mils pénicillines cultivés au Sénégal et au Mali.
- Marchais, L., Tostain, S., and Amoukou, I. (1993). Signification taxonomique et évolutive de la structure génétique des mils penicillaires.
- Mariac, C., Luong, V., Kapran, I., Mamadou, A., Sagnard, F., Deu, M., Chantereau, J., Gerard, B., Ndjeunga, J., Bezançon, G., et al. (2006a). Diversity of wild and cultivated pearl millet accessions (*Pennisetum glaucum* [L.] R. Br.) in Niger assessed by microsatellite markers. *TAG Theor. Appl. Genet. Theor. Angew. Genet.* 114, 49–58.
- Mariac, C., Robert, T., Allinne, C., Remigereau, M.S., Luxereau, A., Tidjani, M., Seyni, O., Bezançon, G., Pham, J.L., and Sarr, A. (2006b). Genetic diversity and gene flow among pearl millet crop/weed complex: a case study. *Theor. Appl. Genet.* 113, 1003–1014.
- Mariac, C., Jehin, L., SaïDou, A.-A., Thuillet, A.-C., Couderc, M., Sire, P., Jugdé, H., Adam, H., Bezançon, G., Pham, J.-L., et al. (2011a). Genetic basis of pearl millet adaptation along an environmental gradient investigated by a combination of genome scan and association mapping: ADAPTATION TO CLIMATIC VARIATION. *Mol. Ecol.* 20, 80–91.
- Mariac, C., Jehin, L., SaïDou, A.-A., Thuillet, A.-C., Couderc, M., Sire, P., Jugdé, H., Adam, H., Bezançon, G., Pham, J.-L., et al. (2011b). Genetic basis of pearl millet adaptation along an environmental gradient investigated by a combination of genome scan and association mapping: ADAPTATION TO CLIMATIC VARIATION. *Mol. Ecol.* 20, 80–91.

Masle, J., Gilmore, S.R., and Farquhar, G.D. (2005). The ERECTA gene regulates plant transpiration efficiency in Arabidopsis. *Nature* 436, 866.

Mbaye, D.F. (1994). Etude du pathosysteme pennisetum glaucum-sclerospora graminicola. Application a la gestion du mildiou du mil au senegal. thesis. École nationale supérieure agronomique (Montpellier).

Moe, K.T., Gwag, J.-G., and Park, Y.-J. (2012). Efficiency of PowerCore in core set development using amplified fragment length polymorphic markers in mungbean. *Plant Breed.* 131, 110–117.

Nakashima, K., Takasaki, H., Mizoi, J., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2012). NAC transcription factors in plant abiotic stress responses. *Biochim. Biophys. Acta BBA-Gene Regul. Mech.* 1819, 97–103.

Negre, T., Imbernon, J., Guinot, J.P., Seguin, B., Berges, J.C., and Guillot, B. (1988). Estimation et suivi de la pluviometrie au Senegal par satellite Meteosat. *Agron. Trop.* 1975 43, 279–288.

Noba, K., Ba, A., Caussanel, J., Mbaye, M., and Barralis, G. (2004). Flore adventice des cultures vivrières dans le sud du Bassin arachidier (Sénégal). *Webbia* 59, 293–308.

Noirot, M., Hamon, S., and Anthony, F. (1996). The principal component scoring: A new method of constituting a core collection using quantitative data. *Genet. Resour. Crop Evol.* 43, 1–6.

Olsen, K.M., and Wendel, J.F. (2013). Crop plants as models for understanding plant adaptation and diversification. *Front. Plant Sci.* 4, 290.

Oumar, I., Mariac, C., Pham, J.-L., and Vigouroux, Y. (2008). Phylogeny and origin of pearl millet (*Pennisetum glaucum* [L.] R. Br) as revealed by microsatellite loci. *TAG Theor. Appl. Genet. Theor. Angew. Genet.* 117, 489–497.

Peakall, R., and Smouse, P.E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28, 2537–2539.

Perez-de-Castro, A.M., Vilanova, S., Canizares, J., Pascual, L., Blanca, J.M., Diez, M.J., Prohens, J., and Pico, B. (2012). Application of Genomic Tools in Plant Breeding. *Curr. Genomics* 13, 179–195.

Perrier, X., Flori, A., and Bonnot, F. (1999). Les methodes d'analyse des donnees. In *Diversite Genetique Des Plantes Tropicales Cultivees / Hamon Perla (Ed.), Seguin Marc (Ed.), Perrier Xavier (Ed.), Glaszmann Jean-Christophe (Ed.)*, (Montpellier: CIRAD), pp. 43–87.

Poncet, V., Lamy, F., Enjalbert, J., Joly, H., Sarr, A., and Robert, T. (1998). Genetic analysis of the domestication syndrome in pearl millet (*Pennisetum glaucum* L., Poaceae): inheritance of the major characters. *Heredity* 81, 648–658.

Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155, 945–959.

Pucher, A., Sy, O., Angarawai, I.I., Gondah, J., Zangre, R., Ouedraogo, M., Sanogo, M.D., Boureima, S., Hash, C.T., and Haussmann, B.I.G. (2015). Agro-morphological Characterization of West and Central African Pearl Millet Accessions. *Crop Sci.* 55, 737–748.

QGIS Development Team (2012). QGIS Geographic Information System. Open Source Geospatial Foundation Project.

Qi, X., Lindup, S., Pittaway, T.S., Allouis, S., Gale, M.D., and Devos, K.M. (2001). Development of simple sequence repeat markers from bacterial artificial chromosomes without subcloning. *BioTechniques* 31, 355–362.

R Development Core Team (2008). R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria).

Rao, S.A., Mengesha, M.H., and Reddy, C.R. (1986). Pearl millet (*Pennisetum americanum*) germplasm. *Indian J. Genet. Plant Breed.* 46, 413–422.

Romagosa, I., Fox, P.N., Moral, L.F.G. del, Ramos, J.M., Moral, B.G. del, Togores, F.R. de, and Molina-Cano, J.L. (1993). Integration of statistical and physiological analyses of adaptation of near-isogenic barley lines. *Theor. Appl. Genet.* 86, 822–826.

Sabaghnia, N., Sabaghpour, S. ?H., and Dehghani, H. (2008). The use of an AMMI model and its parameters to analyse yield stability in multi-environment trials. *J. Agric. Sci.* 146.

- Saïdou, A.-A., Mariac, C., Luong, V., Pham, J.-L., Bezançon, G., and Vigouroux, Y. (2009). Association studies identify natural variation at PHYC linked to flowering time and morphological variation in pearl millet. *Genetics* 182, 899–910.
- Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Sandmeier, M. (1993). Selfing rates of pearl millet (*Pennisetum typhoides* Stapf and Hubb.) under natural conditions. *Theor. Appl. Genet.* 86, 513–517.
- Schoen, D.J., and Brown, A.H. (1993). Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. *Proc. Natl. Acad. Sci.* 90, 10623–10627.
- Shepard, P.J., and Hertel, K.J. (2009). The SR protein family. *Genome Biol.* 10, 242.
- Smouse, P.E., and Long, J.C. (1992). Matrix correlation analysis in anthropology and genetics. *Am. J. Phys. Anthropol.* 35(S15), 187–213.
- Smouse, P.E., Long, J.C., and Sokal, R.R. (1986). Multiple Regression and Correlation Extensions of the Mantel Test of Matrix Correspondence. *Syst. Zool.* 35, 627.
- Stich, B., Haussmann, B.I., Pasam, R., Bhosale, S., Hash, C.T., Melchinger, A.E., and Parzies, H.K. (2010). Patterns of molecular and phenotypic diversity in pearl millet [*Pennisetum glaucum* (L.) R. Br.] from West and Central Africa and their relation to geographical and environmental parameters. *BMC Plant Biol.* 10, 216.
- Sultan, B., Roudier, P., Quirion, P., Alhassane, A., Muller, B., Dingkuhn, M., Ciais, P., Guimberteau, M., Traore, S., and Baron, C. (2013). Assessing climate change impacts on sorghum and millet yields in the Sudanian and Sahelian savannas of West Africa. *Environ. Res. Lett.* 8, 014040.
- Sy, O., Fofana, A., Cisse, N., Noba, K., Diouf, D., Ndoye, I., Sane, D., Kane, A., Kane, N., Hash, T., et al. (2015). Étude de la variabilité agromorphologique de la collection nationale de mils locaux du Sénégal. *J. Appl. Biosci.* 87, 8030.
- Szpiech, Z.A., Jakobsson, M., and Rosenberg, N.A. (2008). ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* 24, 2498–2504.

Thioulouse, J., Chessel, D., Dole'dec, S., and Olivier, J.-M. (1997). ADE-4: a multivariate analysis and graphical display software. *Stat. Comput.* 7, 75–83.

Tiwari, K.K., Singh, A., Pattnaik, S., Sandhu, M., Kaur, S., Jain, S., Tiwari, S., Mehrotra, S., Anumalla, M., Samal, R., et al. (2015). Identification of a diverse mini-core panel of Indian rice germplasm based on genotyping using microsatellite markers. *Plant Breed.* 134, 164–171.

Tostain, S. (1992). Enzyme diversity in pearl millet (*Pennisetum glaucum* L.). *Theor. Appl. Genet.* 83, 733–742.

Tostain, S. (1998). Le mil, une longue histoire : Hypothèses sur sa domestication et ses migrations.

Tostain, S., and Marchais, L. (1993). Evaluation de la diversité génétique des mils (*Pennisetum glaucum* (L.) R. Br.) au moyen de marqueurs enzymatiques et relations entre formes sauvages et cultivées.

Tsuji, H., Taoka, K., and Shimamoto, K. (2011). Regulation of flowering in rice: two florigen genes, a complex gene network, and natural variation. *Curr. Opin. Plant Biol.* 14, 45–52.

Upadhyaya, H.D., Gowda, C.L.L., Reddy, K.N., and Singh, S. (2009). Augmenting the Pearl Millet Core Collection for Enhancing Germplasm Utilization in Crop Improvement. *Crop Sci.* 49, 573.

Upadhyaya, H.D., Yadav, D., Reddy, K.N., Gowda, C.L.L., and Singh, S. (2011). Development of Pearl Millet Minicore Collection for Enhanced Utilization of Germplasm. *Crop Sci.* 51, 217.

Varshney, R.K., Nayak, S.N., May, G.D., and Jackson, S.A. (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol.* 27, 522–530.

Varshney, R.K., Glaszmann, J.-C., Leung, H., and Ribaut, J.-M. (2010). More genomic resources for less-studied crops. *Trends Biotechnol.* 28, 452–460.

Varshney, R.K., Shi, C., Thudi, M., Mariac, C., Wallace, J., Qi, P., Zhang, H., Zhao, Y., Wang, X., Rathore, A., et al. (2017). Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nat. Biotechnol.*

Vigouroux, Y., Mariac, C., De Mita, S., Pham, J.-L., Gérard, B., Kapran, I., Sagnard, F., Deu, M., Chantereau, J., Ali, A., et al. (2011a). Selection for Earlier Flowering Crop Associated with Climatic Variations in the Sahel. *PLoS ONE* 6, e19563.

Vigouroux, Y., Mariac, C., De Mita, S., Pham, J.-L., Gérard, B., Kapran, I., Sagnard, F., Deu, M., Chantereau, J., Ali, A., et al. (2011b). Selection for Earlier Flowering Crop Associated with Climatic Variations in the Sahel. *PLoS ONE* 6, e19563.

Vítámvás, P., Urban, M.O., Škodáček, Z., Kosová, K., Pitelková, I., Vítámvás, J., Renaut, J., and Prášil, I.T. (2015). Quantitative analysis of proteome extracted from barley crowns grown under different drought conditions. *Front. Plant Sci.* 6, 479.

Wagner, H.W., and Sefc, K.M. (1999). *IDENTITY 1.0*.

Westengen, O.T., Jeppson, S., and Guarino, L. (2013). Global Ex-Situ Crop Diversity Conservation and the Svalbard Global Seed Vault: Assessing the Current Status. *PLoS ONE* 8, e64146.

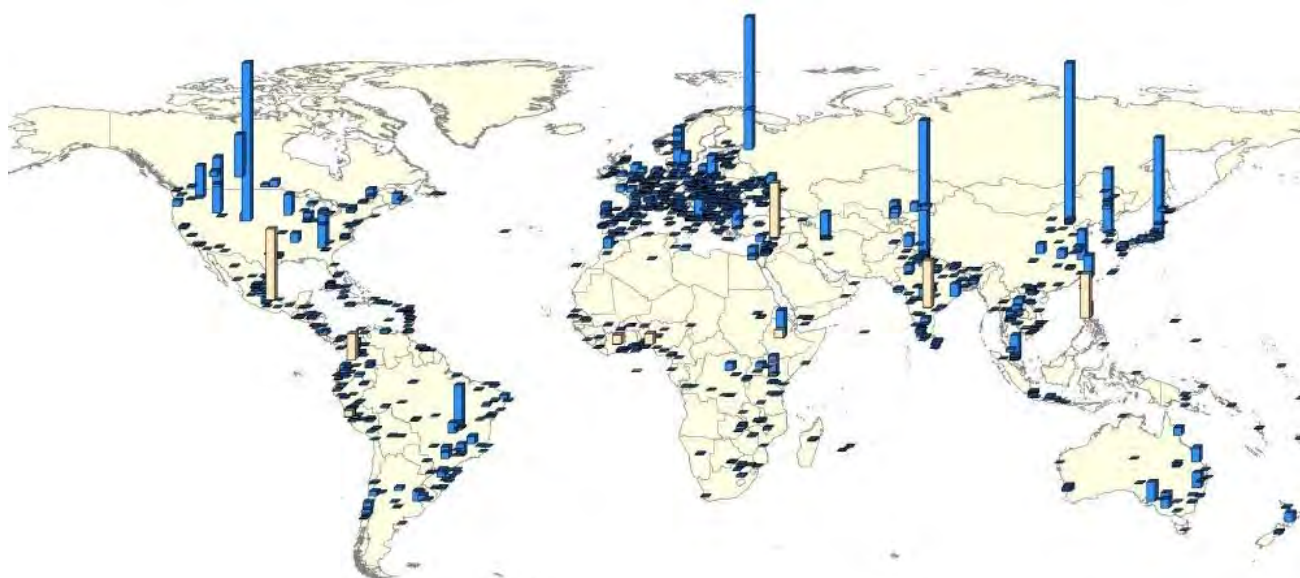
Wu, J., Wang, L., and Wang, S. (2016). Comprehensive analysis and discovery of drought-related NAC transcription factors in common bean. *BMC Plant Biol.* 16.

Zon, V.D. (1992). *Graminees du Cameroun*.

(2016). *agricolae: Statistical Procedures for Agricultural Research*. R package version 1.2-4.

ANNEXES

Annexe 1 : Les banques de gènes à travers le monde. La hauteur de la barre traduit le nombre d'entrées conservées. Les barres en beige sont les banques de gènes internationales et les barres en bleu sont des banques de gènes nationales.



Annexe 2

Tableau 1 : Accessions de la collection et données passeports

Echantillon	Type varietal	Latitude	Longitude	Date_de_collecte	Region
N_01-14	Souna	15,3985	-16,63315	01/09/2013	Louga
N_02-14	Souna	15,3725667	-16,45835	01/09/2013	Louga
N_03-14	Souna	16,3533833	-16,2745167	01/09/2013	Louga
N_05-14	Souna	15,2446333	-15,9704333	01/09/2013	Louga
N_07-14	Souna	15,8697833	-15,5883667	01/09/2013	Louga
N_08-14	Souna	16,5199167	-15,3347167	02/09/2013	ST-Louis
N_09-14	Souna	16,1805167	-14,0003167	02/09/2013	ST-Louis
N_10-14	Souna	16,0680333	-13,91145	02/09/2013	ST-Louis
N_11-14	Souna	16,0527333	-13,7529667	02/09/2013	ST-Louis
N_12-14	Souna	15,9504333	-13,57325	03/09/2013	Matam
N_13-14	Souna	15,9504333	-13,57325	03/09/2013	Matam
N_14-14	Souna	15,67245	-13,3986667	03/09/2013	Matam
N_16-14	Souna	15,4550333	-13,1847167	03/09/2013	Matam
N_176-14	Souna	12,9923167	-15,7579333	17/12/2013	Sédhiou
N_177-14	Souna	12,8368833	-15,18055	19/12/2013	Sédhiou
N_178-14	Souna	12,83605	-15,18145	19/12/2013	Sédhiou
N_179-14	Souna	12,8372	-15,1804667	19/12/2013	Sédhiou
N_180-14	Souna	13,7883667	-14,8532167	11/02/2014	Kaffrine
N_181-14	Souna	13,7876667	-14,8563167	11/02/2014	Kaffrine
N_18-14	Souna	15,0847667	-12,9015333	03/09/2013	Matam
N_182-14	Souna	13,7902	-14,8529	11/02/2014	Kaffrine
N_183-14	Souna	13,7879333	-14,8526667	11/02/2014	Kaffrine
N_185-14	Souna	13,7793333	-14,8340667	11/02/2014	Kaffrine
N_187-14	Souna	13,7790333	-14,8351667	11/02/2014	Kaffrine
N_188-14	Souna	13,8554167	-14,9027	12/02/2014	Kaffrine
N_189-14	Souna	13,854	-14,9033	12/02/2014	Kaffrine
N_190-14	Souna	13,8536167	-14,9030833	12/02/2014	Kaffrine
N_191-14	Souna	13,8536167	-14,9030833	12/02/2014	Kaffrine
N_19-14	Souna	14,9976167	-12,63385	03/09/2013	Matam
N_192-14	Souna	13,8545333	-14,90305	12/02/2014	Kaffrine
N_194-14	Souna	13,8488667	-14,8528667	12/02/2014	Kaffrine
N_196-14	Souna	13,848	-14,8519667	12/02/2014	Kaffrine
N_198-14	Souna	13,8473667	-14,8530667	12/02/2014	Kaffrine
N_200-14	Souna	13,8468167	-14,85185	12/02/2014	Kaffrine
N_201-14	Souna	13,9053667	-14,8672167	13/02/2014	Kaffrine
N_20-14	Souna	14,8756333	-12,497	03/09/2013	Tamba
N_202-14	Souna	13,9058	-14,8661833	13/02/2014	Kaffrine
N_203-14	Souna	13,9055167	-14,8693667	13/02/2014	Kaffrine
N_204-14	Souna	13,90805	-14,8663667	13/02/2014	Kaffrine
N_205-14	Souna	13,90805	-14,8663667	13/02/2014	Kaffrine
N_206-14	Souna	13,90805	-14,8663667	13/02/2014	Kaffrine

Echantillon	Type varietal	Latitude	Longitude	Date_de_collecte	Region
N_208-14	Souna	13,9047333	-14,8681667	13/02/2014	Kaffrine
N_209-14	Souna	13,90715	-14,87005	13/02/2014	Kaffrine
N_210-14	Souna	13,9052833	-14,8694833	13/02/2014	Kaffrine
N_211-14	Souna	13,8752833	-14,9212333	13/02/2014	Kaffrine
N_21-14	Souna	13,775	-13,7513	04/09/2013	Tamba
N_213-14	Souna	13,8749	-14,9213833	13/02/2014	Kaffrine
N_22-14	Souna	13,9167167	-14,2136	04/09/2013	Tamba
N_23-14	Souna	14,1039667	-15,4665333	04/09/2013	Kaffrine
N_24-14	Souna	14,0845667	-15,6105333	04/09/2013	Kaffrine
N_25-14	Souna	15,65725	-16,3975333	14/11/2013	Louga
N_26-14	Souna	15,65515	-16,325	14/11/2013	Louga
N_29-14	Souna	15,96935	-15,9673	15/11/2013	Louga
N_30-14	Souna	15,7943333	-15,9374833	15/11/2013	Louga
N_31-14	Souna	15,7277	-16,06485	15/11/2013	Louga
N_32-14	Souna	15,6118167	-16,0275333	15/11/2013	Louga
N_33-14	Souna	15,521	-13,4082667	16/11/2013	Matam
N_34-14	Souna	15,3936833	-13,5365	16/11/2013	Matam
N_35-14	Souna	15,2525833	-14,0531667	16/11/2013	Matam
N_36-14	Souna	15,2199667	-14,33165	16/11/2013	Matam
N_38-14	Souna	15,4030167	-15,1578333	17/11/2013	Louga
N_39-14	Souna	15,3264	-14,8721833	17/11/2013	Louga
N_40-14	Souna	15,1395	-14,9518167	17/11/2013	Louga
N_41-14	Souna	14,9661333	-15,0417833	17/11/2013	Louga
N_42-14	Souna	15,4018333	-15,13495	17/11/2013	Louga
N_43-14	Souna	14,8240333	-15,3079833	17/11/2013	Louga
N_44-14	Souna	13,8065833	-13,5814667	08/04/2014	Tamba
N_47-14	Souna	13,5659667	-13,5659667	08/04/2014	Matam
N_48-14	Souna	14,4895	-13,5328333	09/04/2014	Matam
N_49-14	Souna	14,37465	-13,5308	09/04/2014	Tamba
N_50-14	Souna	14,2719333	-13,5897667	09/04/2014	Tamba
N_51-14	Souna	14,1518	-13,6285333	09/04/2014	Tamba
N_52-14	Souna	14,0617333	-13,6718167	09/04/2014	Tamba
N_53-14	Souna	13,02195	-13,6584	09/04/2014	Tamba
N_54-14	Souna	13,8818167	-13,64275	09/04/2014	Tamba
N_56-14	Souna	13,9579	-13,3404333	09/04/2014	Tamba
N_57-14	Souna	13,88645	-13,4702167	09/04/2014	Tamba
N_58-14	Souna	15,2789833	-13,58135	09/04/2014	Matam
N_59-14	Souna	13,9431	-14,6115667	08/01/2014	Tamba
N_60-14	Souna	13,8158167	-14,4367	08/01/2014	Tamba
N_61-14	Souna	13,9413333	-14,2969833	08/01/2014	Tamba
N_62-14	Souna	14,0495333	-14,3091333	08/01/2014	Tamba
N_63-14	Souna	13,7974333	-14,1404167	08/01/2014	Tamba
N_64-14	Souna	13,7374	-14,2404667	08/01/2014	Tamba
N_65-14	Souna	13,8424167	-14,1097	08/01/2014	Tamba
N_66-14	Souna	13,7897167	-13,7758833	09/01/2014	Tamba

Echantillon	Type varietal	Latitude	Longitude	Date_de_collecte	Region
N_67-14	Souna	13,8892833	-13,68905	09/01/2014	Tamba
N_68-14	Souna	13,9100167	-13,4265333	10/01/2014	Tamba
N_69-14	Souna	14,0506	-13,4041333	10/01/2014	Tamba
N_70-14	Souna	13,29605	-13,3517667	11/01/2014	Tamba
N_71-14	Souna	13,6623333	-13,67045	11/01/2014	Tamba
N_72-14	Souna	12,9783333	-14,7392667	14/01/2014	Kolda
N_73-14	Souna	13,2173167	-14,6180333	14/01/2014	Kolda
N_75-14	Souna	12,7745333	-15,0942667	15/01/2014	Kolda
N_76-14	Souna	12,8485167	-15,0345333	15/01/2014	Kolda
N_77-14	Souna	12,73825	-15,1627833	12/02/2014	Sédhiou
N_78-14	Souna	12,8047	-15,1346667	13/02/2014	Sédhiou
N_79-14	Souna	12,8446167	-15,43995	13/02/2014	Sédhiou
N_80-14	Souna	12,9725333	-15,6103	15/02/2014	Sédhiou
N_81-14	Souna	13,1173333	-15,59185	15/02/2014	Sédhiou
N_82-14	Souna	13,1157333	-15,5931167	15/02/2014	Sédhiou
N_83-14	Souna	13,2411333	-15,5559333	15/02/2014	Sédhiou
PLS_01	Souna	14,5927389	-15,6397	1992	Fatick
PLS_02	Souna	14,483	-15,5	1992	Kaffrine
PLS_03	Souna	14,4166667	-15,5833333	1992	Kaffrine
PLS_05	Souna	14,3650611	-15,7497889	1992	Kaffrine
PLS_10	Souna	14,1333333	-15,8833333	1992	Kaffrine
PLS_100	Souna	NA	NA	1993	Diourbel
PLS_101	Souna	14,7399028	-16,1095806	1993	Diourbel
PLS_103	Souna	14,754799	-16,3499821	1993	Diourbel
PLS_104	Souna	14,8333333	-16,15	1993	Diourbel
PLS_106	Souna	14,8943722	-16,2036778	1993	Diourbel
PLS_107	Souna	14,6666667	-16,3333333	1993	Diourbel
PLS_108	Souna	NA	NA	1993	Diourbel
PLS_11	Souna	14,1333333	-15,75	1992	Kaffrine
PLS_111	Souna	14,7666667	-16,3	1993	Diourbel
PLS_112	Souna	14,8833333	-16,2166667	1993	Diourbel
PLS_113	Souna	14,8564659	-16,1572591	1993	Diourbel
PLS_114	Souna	14,9600008	-16,3308025	1993	Diourbel
PLS_115	Souna	14,9533414	-16,2989732	1993	Diourbel
PLS_116	Souna	14,9333333	-16,2	1993	Diourbel
PLS_117	Souna	NA	NA	1993	Diourbel
PLS_118	Souna	14,8666667	-15,9833333	1993	Diourbel
PLS_12	Souna	13,8195611	-15,4367417	1992	Kaffrine
PLS_120	Souna	14,8	-16,0166667	1993	Diourbel
PLS_121	Souna	14,8964472	-15,9541139	1993	Diourbel
PLS_123	Souna	14,95	-15,8833333	1993	Diourbel
PLS_124	Souna	15,6502778	-16,0477778	1993	Louga
PLS_125	Souna	14,9666667	-15,7833333	1993	Louga
PLS_127	Souna	14,7227944	-15,8300472	1993	Diourbel
PLS_129	Souna	14,5666667	-15,95	1993	Diourbel

Echantillon	Type varietal	Latitude	Longitude	Date_de_collecte	Region
PLS_13	Souna	14,1833333	-15,4333333	1992	Kaffrine
PLS_130	Souna	14,7166667	-15,8333333	1993	Diourbel
PLS_131	Souna	14,7333333	-16	1993	Diourbel
PLS_137	Souna	15,0666667	-16,3333333	1993	Thies
PLS_138	Souna	15,1769444	-16,5697222	1993	Thies
PLS_14	Souna	14,2833333	-15,3	1992	Kaffrine
PLS_140	Souna	NA	NA	1993	Thies
PLS_147	Souna	14,9394361	-16,4190528	1993	Diourbel
PLS_15	Souna	14,3666667	-15,2	1992	Kaffrine
PLS_150	Souna	14,7833333	-16,4166667	1993	Diourbel
PLS_151	Souna	15	-16,4166667	1993	Thies
PLS_153	Souna	14,8166667	-16,7666667	1993	Thies
PLS_154	Souna	NA	NA	1993	Thies
PLS_155	Souna	14,6391667	-17,0877778	1993	Thies
PLS_157	Souna	NA	NA	1993	Thies
PLS_158	Souna	14,7	-16,95	1993	Thies
PLS_159	Souna	14,6833333	-16,85	1993	Thies
PLS_16	Souna	NA	NA	1992	Tamba
PLS_160	Souna	14,6833333	-16,7166667	1993	Thies
PLS_161	Souna	14,6833333	-16,65	1993	Thies
PLS_162	Souna	14,5666667	-16,6166667	1993	Thies
PLS_163	Souna	NA	NA	1993	Thies
PLS_164	Souna	NA	NA	1993	Thies
PLS_166	Souna	14,389505	-16,7072358	1993	Fatick
PLS_167	Souna	14,302254	-16,7329627	1993	Thies
PLS_170	Souna	14,3	-16,8666667	1993	Thies
PLS_171	Souna	14,35	-16,8166667	1993	Thies
PLS_173	Souna	14,6286623	-16,7617644	1993	Thies
PLS_174	Souna	14,4833333	-16,95	1993	Thies
PLS_175	Souna	14,45	-16,9333333	1993	Thies
PLS_176	Souna	14,4572222	-17,01	1993	Thies
PLS_177	Souna	14,5402778	-17,0636111	1993	Thies
PLS_178	Souna	14,5577778	-17,0655556	1993	Thies
PLS_179	Souna	14,5575	-17,1138889	1993	Thies
PLS_18	Souna	13,8250611	-15,1375889	1992	Kaffrine
PLS_180	Souna	14,6275	-17,1058333	1993	Thies
PLS_181	Souna	14,95	-16,9833333	1993	Thies
PLS_182	Souna	14,9053549	-16,861726	1993	Thies
PLS_183	Souna	14,9666667	-16,9	1993	Thies
PLS_184	Souna	14,9166667	-16,9333333	1993	Thies
PLS_185	Souna	14,9333333	-16,9833333	1993	Thies
PLS_186	Souna	15,25	-16,5166667	1993	Thies
PLS_187	Souna	15,3667722	-16,5428	1993	Louga
PLS_188	Souna	15,8333333	-16,3	1993	Louga
PLS_189	Souna	15,1666667	-16,1166667	1993	Louga

Echantillon	Type varietal	Latitude	Longitude	Date_de_collecte	Region
PLS_19	Souna	13,8275306	-15,1830944	1992	Kaffrine
PLS_190	Souna	15,5666667	-16,5333333	1993	Louga
PLS_191	Souna	15,3	-16,6833333	1993	Louga
PLS_192	Souna	15,2333333	-16,6666667	1993	Thies
PLS_193	Souna	NA	NA	1993	Thies
PLS_194	Souna	NA	NA	1993	Thies
PLS_195	Souna	15,1769444	-16,7425	1993	Thies
PLS_196	Souna	15,1175	-16,68	1993	Thies
PLS_197	Souna	14,565	-17,1255556	1993	Thies
PLS_198	Souna	14,9333333	-16,3666667	1993	Diourbel
PLS_199	Souna	14,7727167	-16,5334667	1993	Diourbel
PLS_20	Souna	13,9833333	-14,8	1992	Kaffrine
PLS_200	Souna	14,8021248	-16,5528833	1993	Diourbel
PLS_201	Souna	14,7333333	-16,4166667	1993	Diourbel
PLS_202	Souna	15,1097222	-16,6216667	1993	Diourbel
PLS_203	Souna	14,7727167	-16,5334667	1993	Diourbel
PLS_204	Souna	14,9008778	-16,4450861	1993	Diourbel
PLS_206	Souna	14,7727167	-16,5334667	1993	Diourbel
PLS_207	Souna	14,771725	-16,5331047	1993	Diourbel
PLS_208	Souna	15,1287972	-16,6568056	1993	Diourbel
PLS_209	Souna	14,8056833	-16,57075	1993	Diourbel
PLS_210	Souna	NA	NA	1993	Diourbel
PLS_211	Souna	14,82355	-16,4164	1993	Diourbel
PLS_212	Souna	14,7981	-16,5469833	1993	Diourbel
PLS_213	Souna	14,7974194	-16,5469693	1993	Diourbel
PLS_214	Souna	14,7974194	-16,5469693	1993	Diourbel
PLS_216	Souna	14,9971074	-16,7584134	1993	Diourbel
PLS_217	Souna	14,819375	-16,4457722	1993	Diourbel
PLS_22	Souna	14,04475	-15,161725	1992	Kaffrine
PLS_220	Souna	14,8021248	-16,5528833	1993	Diourbel
PLS_221	Souna	14,9380417	-16,3741	1993	Diourbel
PLS_222	Souna	14,9971074	-16,7584134	1993	Diourbel
PLS_223	Souna	14,9110528	-16,2700694	1993	Diourbel
PLS_224	Souna	14,4333333	-16,8	1993	Thies
PLS_225	Souna	NA	NA	1993	Diourbel
PLS_226	Souna	15,5333333	-16,2166667	1993	Louga
PLS_227	Souna	NA	NA	1993	Thies
PLS_228	Souna	NA	NA	1993	Thies
PLS_23	Souna	14,0354139	-15,6454861	1992	Kaffrine
PLS_230	Souna	15,0376971	-16,2503158	1993	Thies
PLS_231	Souna	NA	NA	1993	Thies
PLS_232	Souna	NA	NA	1993	Thies
PLS_235	Souna	15,0376971	-16,2503158	1993	Thies
PLS_24	Souna	13,9137639	-15,6676472	1992	Kaffrine
PLS_25	Souna	13,8333333	-15,4333333	1992	Kaffrine

Echantillon	Type varietal	Latitude	Longitude	Date_de_collecte	Region
PLS_26	Souna	13,8333333	-15,4166667	1992	Kaolack
PLS_27	Souna	13,6101833	-16,03855	1992	Kaolack
PLS_29	Souna	13,6163667	-15,6044528	1992	Kaolack
PLS_30	Souna	13,6271222	-15,6427194	1992	Kaolack
PLS_301	Souna	14,0833333	-16,1333333	2010	Kaolack
PLS_302	Souna	13,61	-16,3833333	2010	Fatick
PLS_303	Souna	14,0572194	-16,0493222	2010	Kaolack
PLS_304	Souna	13,6833333	-16,1166667	2010	Kaolack
PLS_305	Souna	13,7333333	-16	2010	Kaolack
PLS_307	Souna	13,7	-15,8333333	2010	Kaolack
PLS_308	Souna	13,65	-15,6333333	2010	Kaolack
PLS_309	Souna	13,7712611	-15,5870111	2010	Kaolack
PLS_31	Souna	13,7166667	-15,4333333	1992	Kaolack
PLS_310	Souna	13,65	-15,5833333	2010	Kaolack
PLS_312	Souna	13,7833333	-15,75	2010	Kaolack
PLS_313	Souna	13,8666667	-15,6333333	2010	Kaffrine
PLS_314	Souna	14,0692806	-15,9332694	2010	Kaffrine
PLS_315	Souna	14,1333333	-15,55	2010	Kaffrine
PLS_316	Souna	14,1180348	-15,7362926	2010	Kaffrine
PLS_317	Souna	14,0187757	-15,4781766	2010	Kaffrine
PLS_318	Souna	14,0187757	-15,4781766	2010	Kaffrine
PLS_319	Souna	13,8279611	-15,4053444	2010	Kaffrine
PLS_32	Souna	13,6906139	-15,8300556	1992	Kaolack
PLS_320	Souna	13,8100139	-15,4020417	2010	Kaffrine
PLS_321	Souna	14,145301	-14,9898724	2010	Kaffrine
PLS_322	Souna	14,0166667	-15,3	2010	Kaffrine
PLS_325	Souna	13,9749165	-14,9407421	2010	Kaffrine
PLS_326	Souna	14	-14,7666667	2010	Kaffrine
PLS_327	Souna	NA	NA	2010	Matam
PLS_328	Souna	NA	NA	2010	Matam
PLS_329	Souna	NA	NA	2010	Matam
PLS_33	Souna	13,93205	-16,1872222	1992	Kaolack
PLS_330	Souna	NA	NA	2010	Matam
PLS_331	Souna	NA	NA	2010	Matam
PLS_35	Souna	13,9391	-15,9456917	1992	Kaolack
PLS_36	Souna	13,9629556	-15,9501944	1992	Kaolack
PLS_37	Souna	14,0166667	-16,0333333	1992	Kaolack
PLS_38	Souna	13,767475	-15,9983472	1992	Kaolack
PLS_39	Souna	13,8246013	-16,266913	1992	Fatick
PLS_40	Souna	13,8009361	-16,3272472	1992	Fatick
PLS_41	Souna	NA	NA	1992	Fatick
PLS_42	Souna	14,1	-15,6166667	1992	Kaolack
PLS_43	Souna	14,0889639	-16,1672278	1992	Kaolack
PLS_44	Souna	14,0858917	-16,0683861	1992	Kaolack
PLS_45	Souna	14,2166667	-16,2333333	1992	Kaolack

Echantillon	Type varietal	Latitude	Longitude	Date_de_collecte	Region
PLS_46	Souna	14,9666667	-16,8666667	1992	Thies
PLS_47	Souna	14,2833306	-16,75	1992	Fatick
PLS_48	Souna	14,2271944	-16,5775917	1992	Fatick
PLS_49	Souna	14,2265278	-16,6050944	1992	Fatick
PLS_50	Souna	14,3906214	-16,6609008	1992	Thies
PLS_52	Souna	14,4250056	-16,6070278	1992	Fatick
PLS_54	Souna	14,0391658	-16,3180106	1992	Fatick
PLS_56	Souna	14,3469806	-16,7404556	1992	Fatick
PLS_57	Souna	14,3610444	-16,7076972	1992	Fatick
PLS_61	Souna	14,5333333	-16,4333333	1992	Fatick
PLS_64	Souna	NA	NA	1992	Fatick
PLS_66	Souna	14,4166667	-16,2166667	1992	Fatick
PLS_67	Souna	14,5422417	-16,2559861	1992	Fatick
PLS_68	Souna	14,71015	-16,5211833	1993	Diourbel
PLS_69	Souna	14,6666667	-16,5333333	1993	Diourbel
PLS_70	Souna	14,6403833	-16,5052667	1993	Diourbel
PLS_71	Souna	14,7666667	-16,6333333	1993	Diourbel
PLS_72	Souna	14,7211167	-16,5672833	1993	Diourbel
PLS_73	Souna	14,7609667	-16,5764167	1993	Diourbel
PLS_74	Souna	14,7922833	-16,6055667	1993	Diourbel
PLS_75	Souna	14,8023333	-16,5363111	1993	Diourbel
PLS_77	Souna	14,7219833	-16,5631333	1993	Diourbel
PLS_79	Souna	14,7638833	-16,7972333	1993	Diourbel
PLS_80	Souna	14,9720333	-16,46045	1993	Diourbel
PLS_81	Souna	NA	NA	1993	Diourbel
PLS_82	Souna	14,7666667	-16,45	1993	Diourbel
PLS_83	Souna	14,8144917	-16,4113611	1993	Diourbel
PLS_84	Souna	14,6623333	-16,3363167	1993	Diourbel
PLS_85	Souna	14,75	-16,3833333	1993	Diourbel
PLS_87	Souna	14,7333333	-16,4	1993	Diourbel
PLS_88	Souna	15,0010361	-16,4048139	1993	Diourbel
PLS_89	Souna	14,6749667	-16,3530667	1993	Diourbel
PLS_90	Souna	14,6261167	-16,3946333	1993	Diourbel
PLS_91	Souna	14,6217833	-16,4608167	1993	Diourbel
PLS_92	Souna	14,6869	-16,47535	1993	Diourbel
PLS_93	Souna	14,6453144	-16,3234063	1993	Diourbel
PLS_94	Souna	14,6417722	-16,4346972	1993	Diourbel
PLS_95	Souna	14,6150667	-16,152225	1993	Diourbel
PLS_97	Souna	14,6	-16,15	1993	Diourbel
PLS_98	Souna	14,6027949	-16,2080732	1993	Diourbel
THIALACK_LOCAL	Souna	NA	NA	NA	NA
N_119-14	Sanio à épis court	14,6696	-12,8493667	13/01/2014	Kolda
N_100-14	Sanio	14,0506	-13,4041333	10/01/2014	Tamba
N_101-14	Sanio	13,9881833	-13,0948833	10/01/2014	Tamba

Echantillon	Type varietal	Latitude	Longitude	Date_de_collecte	Region
N_103-14	Sanio	13,71625	-13,6443333	11/01/2014	Tamba
N_105-14	Sanio	13,5654667	-13,56065	11/01/2014	Tamba
N_106-14	Sanio	13,44585	-13,4326833	11/01/2014	Tamba
N_107-14	Sanio	13,2926167	-13,3553167	11/01/2014	Tamba
N_108-14	Sanio	13,3087333	-13,2479667	11/01/2014	Tamba
N_109-14	Sanio	13,2133833	-13,11025	11/01/2014	Tamba
N_110-14	Sanio	13,6700667	-13,6625833	11/01/2014	Tamba
N_111-14	Sanio	13,4453667	-13,75645	12/01/2014	Kolda
N_112-14	Sanio	13,2571333	-13,9076167	12/01/2014	Kolda
N_113-14	Sanio	13,1718333	-14,03855	12/01/2014	Kolda
N_115-14	Sanio	13,01735	-14,0125833	12/01/2014	Kolda
N_116-14	Sanio	12,88605	-14,1242667	12/01/2014	Kolda
N_117-14	Sanio	12,7459667	-14,1638667	12/01/2014	Kolda
N_118-14	Sanio	12,8481667	-14,8305833	13/01/2014	Kolda
N_121-14	Sanio	12,87985	-14,53065	13/01/2014	Kolda
N_123-14	Sanio	12,9229833	-14,9146333	14/01/2014	Kolda
N_126-14	Sanio	13,2180667	-14,6179	14/01/2014	Kolda
N_127-14	Sanio	13,4002333	-14,8208667	14/01/2014	Kolda
N_128-14	Sanio	13,40425	-14,9494333	15/01/2014	Kolda
N_130-14	Sanio	12,694	-14,8721	15/01/2014	Kolda
N_131-14	Sanio	12,67875	-15,3436	12/02/2014	Sédhiou
N_132-14	Sanio	12,6026	-15,4985167	12/02/2014	Sédhiou
N_133-14	Sanio	12,4983667	-15,6091667	12/02/2014	Sédhiou
N_134-14	Sanio	12,5753	-15,9319833	13/02/2014	Sédhiou
N_136-14	Sanio	12,7318	-15,5294667	13/02/2014	Sédhiou
N_138-14	Sanio	12,6233667	-15,6239333	14/02/2014	Sédhiou
N_139-14	Sanio	12,6206	-15,7859333	14/02/2014	Sédhiou
N_140-14	Sanio	12,6729167	-15,9446667	14/02/2014	Sédhiou
N_141-14	Sanio	12,7943167	-15,9693333	14/02/2014	Sédhiou
N_142-14	Sanio	12,8639667	-15,8705167	14/02/2014	Sédhiou
N_144-14	Sanio	13,0265833	-15,7334333	15/02/2014	Sédhiou
N_145-14	Sanio	12,9760833	-15,9164	15/02/2014	Sédhiou
N_146-14	Sanio	12,8855167	-16,06785	15/02/2014	Ziguinchor
N_147-14	Sanio	12,6735167	-16,2098833	16/02/2014	Ziguinchor
N_148-14	Sanio	12,7092167	-16,0564833	16/02/2014	Ziguinchor
N_149-14	Sanio	12,8274667	-16,2525333	16/02/2014	Ziguinchor
N_151-14	Sanio	12,82285	-16,52385	16/02/2014	Ziguinchor
N_152-14	Sanio	12,8461167	-16,4167167	16/02/2014	Ziguinchor
N_153-14	Sanio	12,7592167	-16,23765	17/02/2014	Ziguinchor
N_154-14	Sanio	12,9078167	-16,4024	17/02/2014	Ziguinchor
N_155-14	Sanio	12,9983	-16,5512667	17/02/2014	Ziguinchor
N_156-14	Sanio	12,9961667	-16,5511833	17/02/2014	Ziguinchor
N_157-14	Sanio	13,0027	-16,69555	17/02/2014	Ziguinchor
N_158-14	Sanio	12,56755	-16,1673	18/02/2014	Ziguinchor
N_159-14	Sanio	12,62965	-16,0158667	18/02/2014	Ziguinchor

Echantillon	Type varietal	Latitude	Longitude	Date_de_collecte	Region
N_160-14	Sanio	12,4179	-16,7521667	19/02/2014	Ziguinchor
N_161-14	Sanio	12,4570167	-16,6366667	19/02/2014	Ziguinchor
N_162-14	Sanio	12,6527333	-16,3693833	20/02/2014	Ziguinchor
N_164-14	Sanio	14,5166667	-16,4833333	NA	Fatick
N_165-14	Sanio	14,4	-16,4	NA	Fatick
N_166-14	Sanio	14,4833333	-16,4833333	NA	Fatick
N_167-14	Sanio	14,4833333	-16,4833333	NA	Fatick
N_168-14	Sanio	14,5	-16,5	NA	Fatick
N_169-14	Sanio	14,5	-16,5	NA	Fatick
N_170-14	Sanio	14,5833333	-16,5166667	NA	Fatick
N_171-14	Sanio	14,5666667	-16,5166667	NA	Fatick
N_172-14	Sanio	14,4166667	-16,4	NA	Fatick
N_173-14	Sanio	14,4397167	-16,3900167	NA	Fatick
N_175-14	Sanio	14,5	-16,3166667	NA	Fatick
N_214-14	Sanio	12,4641333	-15,5968833	22/12/2013	Sédhiou
N_215-14	Sanio	12,4674333	-15,5989833	22/12/2013	Sédhiou
N_216-14	Sanio	12,7388167	-14,7792667	27/02/2014	Kolda
N_217-14	Sanio	12,7388833	-14,7789667	27/02/2014	Kolda
N_218-14	Sanio	12,72225	-14,4535	28/02/2014	Kolda
N_219-14	Sanio	12,7225167	-14,4551667	28/02/2014	Kolda
N_220-14	Sanio	12,7612333	-13,9615667	04/03/2014	Kolda
N_221-14	Sanio	12,76345	-13,96435	04/03/2014	Kolda
N_222-14	Sanio	12,7643667	-13,9657667	04/03/2014	Kolda
N_223-14	Sanio	13,3290333	-13,8385667	05/03/2014	Kolda
N_224-14	Sanio	13,3323667	-13,8367667	05/03/2014	Kolda
N_225-14	Sanio	13,2245833	-14,0613667	06/03/2014	Kolda
N_226-14	Sanio	13,2268	-14,06175	06/03/2014	Kolda
N_88-14	Sanio	13,8806667	-14,0999333	08/01/2014	Tamba
N_89-14	Sanio	13,7960833	-14,1413667	08/01/2014	Tamba
N_90-14	Sanio	13,2371833	-14,2391667	08/01/2014	Tamba
N_91-14	Sanio	13,8422333	-14,1102	08/01/2014	Tamba
N_93-14	Sanio	13,89	-13,6886667	09/01/2014	Tamba
N_94-14	Sanio	13,9181	-13,85885	09/01/2014	Tamba
N_95-14	Sanio	13,7002	-13,8155167	09/01/2014	Tamba
N_96-14	Sanio	13,5895	-13,8654333	09/01/2014	Tamba
N_97-14	Sanio	13,5906333	-13,8657	09/01/2014	Tamba
N_98-14	Sanio	13,8097833	-13,5798167	10/01/2014	Tamba
N_99-14	Sanio	13,9099667	-13,42695	10/01/2014	Tamba
GAWANE	Ameliorée				
IBMV_8402	Ameliorée				
ICMV_89305	Ameliorée				
ICMV_IS_92222	Ameliorée				
ICMV_IS_99001	Ameliorée				
ISMI_9301	Ameliorée				
PEO_2830	Ameliorée				

Echantillon	Type varietal	Latitude	Longitude	Date_de_collecte	Region
PEO_8030	Ameliorée				
SOSATC_88	Ameliorée				
SOUNA_3	Ameliorée				
THIALACK_2	Ameliorée				

Tableau 2 : Liste des marqueurs microsatellites (SSR) qui sont utilisés et de leurs caractéristiques moléculaires.

Nom	Taille	Nb allèles	Dye	Multiplex	Motif	Tm°C
<i>PgIRD46*</i>	80-107	7	Ned	A	(CTC) ₆	55
<i>PsmP2227</i> ϕ	192-246	7	Ned	A	(GT) ₇	55
<i>PsmP2249</i> ϕ	130-155	3	Fam	A	(GT) ₇	55
<i>PsmP2247</i> ϕ	199-206	4	Fam	A	(TG) ₆	55
<i>PsmP2201</i> $\diamond$	310-375	5	Red	A	(GT) ₆	55
<i>PsmP2237</i> ϕ	233-265	9	Red	A	(GT) ₈	55
<i>CTM8</i> $\square$	225-275	8	Vic	A	(CT) ₈	55
<i>PgIRD12*</i>	121-147	13	Vic	A	(TA) ₈	55
<i>PgIRD25*</i>	160-184	11	Red	B	(GA) ₉	55
<i>PsmP2214</i> $\diamond$	234-244	10	Red	B	(GT) ₉	55
<i>PsmP2219</i> $\diamond$	227-291	12	Ned	B	(GT) ₇	55
<i>PsmP2220</i> $\diamond$	118-149	14	Ned	B	(GT) ₁₁	55
<i>PsmP2206</i> $\diamond$	197-213	9	Fam	B	(GT) ₁₁	55
<i>PsmP2202</i> $\diamond$	146-163	2	Vic	B	(GT) ₈	55
<i>PsmP2231</i> ϕ	212-262	21	Vic	B	(TG) ₁₂ GG(TA) ₄	55

*= (Mariac et al., 2006a), $\diamond$ = (Qi et al., 2001), $\square$ = (Budak et al., 2003), ϕ = (Allouis et al., 2001). T°C = température d'hybridation

Annexe 3

PROTOCOL D'EXTRACTION D'ADN

Solutions :

Tampon Lyse: qsp d'eau milliQ

Tris 100 mM

NaCl 1.25 mol/l

EDTA 0.02 mol/l

Ajuster à pH 8 avec du HCl et stocker à température ambiante. (ou mettre le tris et l'EDTA à pH 8 avant). *à rajouter extemporanément dans la quantité de tampon juste nécessaire*

2 % de MATAB (0.019 mol/L) ou CTAB

2 % de β -mercaptoethanol ou DTT (0.01M)

Solution de chloroforme/isoamylique 24:1 :

48 mL de chloroforme

2 mL d'alcool isoamylique

Acétate de sodium 3 M pH 5 : qsp 250 mL

20412 g d'acétate de sodium (3 mol/L)

qsp 500 mL d'eau milliQ

Ajuster à pH 5 avec de l'acide acétique glacial

Tampon TE 1X : qsp 500 mL d'eau milliQ

06055 g de Tris (0.01 mol/L)

01861 g d'EDTA (0.001 mol/L)

Ajuster à pH 8 avec de l'acide chlorhydrique HCl 5 N et stocker à température ambiante.

PROTOCOL PCR MULTIPLEX KIT QIAGEN

Le PCR multiplex kit a été mis au point pour les PCR en multiplex.

Préparer un volume de réaction avec 10% de plus que nécessaire pour les réactions.

Concentrations finales du mix :

1x de Master mix (solution 2x)	5 μ l
0.2 μ M de chaque primer	0.2 μ l (solution à 10 μ M)
1x de solution Q (solution à 5x)	2 μ l
Volume de l'échantillon d'ADN	1 μ l (concentration à 25 ng/ μ l)

Qsp eau Rnase free.

Distribuer le mix dans les tubes. 9 μ l

Distribuer ensuite les échantillons d'ADN. 1 μ l

Annexe 4 :

Tableau 1 : Statistiques sur la qualité des reads séquencés par échantillon

Echantillon	Nombre de séquences	Séquences de mauvaise qualité	Longueur des séquences	%G C	% de duplicatas	longueur moyenne des séquences
N_09-14	3126386	0	35	45	14.44	35
N_119-14	221519	0	35	45	47.78	35
N_131-14	2503283	0	35	45	18.26	35
N_132-14	193170	0	35	42	42.21	35
N_180-14	299971	0	35	44	39.88	35
N_182-14	2576835	0	35	44	18.32	35
N_198-14	1604469	0	35	45	25.25	35
N_203-14	92753	0	35	45	59.8	35
PLS_57	3415576	0	35	45	15.03	35
PLS_94	1233101	0	35	45	22.22	35
N_101-14	2698275	0	36	44	16.12	36
N_103-14	2239123	0	36	44	20.16	36
N_115-14	2919586	0	36	46	17.75	36
N_121-14	1153871	0	36	45	27.62	36
N_138-14	172552	0	36	44	50.21	36
N_146-14	1004742	0	36	46	28.48	36
N_153-14	1673672	0	36	44	21.36	36
N_175-14	431176	0	36	45	39.01	36
N_192-14	974025	0	36	43	23.59	36
N_31-14	4306487	0	36	46	17.64	36
N_34-14	1199031	0	36	43	20.82	36
N_42-14	1655806	0	36	46	23.32	36
N_56-14	2409016	0	36	44	16.11	36
N_98-14	1520339	0	36	45	20.85	36

PLS_01	7191363	0	36	45	10.53	36
PLS_23	2174741	0	36	44	17.84	36
PLS_31	1755406	0	36	43	17.26	36
PLS_49	7660495	0	36	47	16.32	36
PLS_61	3088900	0	36	46	20.37	36
N_113-14	2420778	0	37	45	16.67	37
N_140-14	6947471	0	37	46	14.35	37
N_171-14	2135590	0	37	46	20.53	37
N_191-14	1352439	0	37	44	21.33	37
N_33-14	371492	0	37	43	39.62	37
N_44-14	1545865	0	37	45	20.31	37
N_71-14	5922012	0	37	45	12.9	37
N_81-14	18556286	0	37	45	6.79	37
N_88-14	732733	0	37	46	36.58	37
PLS_115	3447983	0	37	47	28.01	37
PLS_151	4175818	0	37	46	14.17	37
PLS_160	4992543	0	37	45	13.66	37
PLS_228	660168	0	37	46	36.66	37
PLS_326	798203	0	37	42	23.92	37
PLS_54	2319121	0	37	45	20.04	37
PLS_56	4439595	0	37	45	16.1	37
N_03-14	1480476	0	38	45	21.3	38
N_106-14	1746697	0	38	46	23.07	38
N_11-14	1428698	0	38	45	25.42	38
N_139-14	224719	0	38	45	46.38	38
N_155-14	1889944	0	38	44	22.37	38
N_156-14	334414	0	38	42	36.29	38
N_187-14	1482113	0	38	44	21.02	38
N_211-14	1804097	0	38	45	23.46	38
N_69-14	790378	0	38	44	29.33	38

N_83-14	2460833	0	38	44	17.32	38
N_99-14	289846	0	38	44	36.26	38
PLS_163	4514199	0	38	45	15.85	38
PLS_204	2126066	0	38	44	18.85	38
PLS_224	1041019	0	38	43	22.75	38
PLS_312	1192012	0	38	45	24.59	38
PLS_319	1171015	0	38	44	24.61	38
PLS_38	673993	0	38	44	29.31	38
PLS_95	4056778	0	38	45	15.95	38
N_117-14	2038821	0	39	45	20.72	39
N_118-14	1375622	0	39	44	20.8	39
N_170-14	2744476	0	39	45	19.14	39
N_177-14	1590324	0	39	46	24.13	39
N_196-14	6338105	0	39	46	23.48	39
N_91-14	1636719	0	39	45	24.75	39
N_95-14	2551287	0	39	44	16.64	39
PLS_100	1861765	0	39	43	15.67	39
PLS_104	2422398	0	39	45	17.25	39
PLS_112	2270193	0	39	44	18.56	39
PLS_121	2483112	0	39	45	21.3	39
PLS_138	4041337	0	39	46	19.24	39
PLS_175	2948571	0	39	44	16.12	39
PLS_318	4889475	0	39	44	11.57	39
PLS_48	1666892	0	39	45	26.73	39
PLS_50	1895173	0	39	45	24.66	39
N_07-14	3859778	0	40	45	15.41	40
N_160-14	1065427	0	40	44	27.76	40
N_167-14	3655866	0	40	45	15.28	40
N_214-14	2146027	0	40	46	26.64	40
N_223-14	2762819	0	40	45	19.15	40

N_52-14	2125194	0	40	43	15.7	40
N_58-14	1797679	0	40	43	17.9	40
N_66-14	3822733	0	40	45	15.63	40
N_96-14	1406467	0	40	45	26.76	40
PLS_05	4226932	0	40	45	16.86	40
PLS_158	1359543	0	40	42	19.5	40
PLS_209	4584969	0	40	46	18.14	40

Annexe 5 : Articles publiés ou en préparation de soumission, dans le cadre de cette thèse

Article 1 : New Genetic Insights into Pearl Millet Diversity As Revealed by Characterization of Early- and Late-Flowering Landraces from Senegal. *Frontiers in Plant Science* 8 (2017). <https://doi.org/10.3389/fpls.2017.00818>.

Article 2 : Establishment of Senegalese pearl millet (*Pennisetum glaucum*) core collection for breeding high yield, flowering earliness and drought tolerance
(*En préparation de soumission*)

Article 3 : Genome-wide association study reveals candidate genes involved in adaptation and resistance to abiotic stress that differentiate Senegalese early and late-flowering pearl millet
(*En préparation de soumission*)



New Genetic Insights into Pearl Millet Diversity As Revealed by Characterization of Early- and Late-Flowering Landraces from Senegal

Oumar Diack^{1,2}, Ndjido A. Kane^{2,3*}, Cecile Berthouly-Salazar^{2,4}, Mame C. Gueye⁵, Baye M. Diop⁶, Amadou Fofana⁶, Ousmane Sy⁶, Hamidou Tall⁷, Leila Zekraoui^{2,4}, Marie Piquet^{2,4}, Marie Couderc^{2,4}, Yves Vigouroux^{2,4}, Diaga Diouf^{1,2} and Adeline Barnaud^{2,4}

¹ Laboratoire Campus de Biotechnologies Végétales, Faculté des Sciences et Techniques, Université Cheikh Anta Diop de Dakar, Dakar, Senegal, ² Laboratoire Mixte International Adaptation des Plantes et Microorganismes Associés aux Stress Environnementaux, Centre de Recherche de Bel Air, Dakar, Senegal, ³ Laboratoire National de Recherches sur les Productions Végétales, Institut Sénégalais de Recherches Agricoles, Centre de Recherche de Bel Air, Dakar, Senegal, ⁴ Unité Mixte de Recherche Diversité et Adaptation des Espèces, Institut de Recherche pour le Développement, Montpellier, France, ⁵ Centre d'Etude Régional pour l'Amélioration de l'Adaptation à la Sécheresse, Institut Sénégalais de Recherches Agricoles, Thiès, Senegal, ⁶ Centre National de Recherches Agronomiques de Bambey, Institut Sénégalais de Recherches Agricoles, Bambey, Senegal, ⁷ Centre de Recherches Zootechniques de Kolda, Institut Sénégalais de Recherches Agricoles, Kolda, Senegal

OPEN ACCESS

Edited by:

Khaled Masmoudi,
Center of Biotechnology, Stax, Tunisia

Reviewed by:

Sureshkumar Balasubramanian,
Monash University, Australia
Liwang Liu,
Nanjing Agricultural University, China

*Correspondence:

Ndjido A. Kane
ndjido.kane@isra.sn

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 10 April 2016

Accepted: 01 May 2017

Published: 17 May 2017

Citation:

Diack O, Kane NA, Berthouly-Salazar C, Gueye MC, Diop BM, Fofana A, Sy O, Tall H, Zekraoui L, Piquet M, Couderc M, Vigouroux Y, Diouf D and Barnaud A (2017) New Genetic Insights into Pearl Millet Diversity As Revealed by Characterization of Early- and Late-Flowering Landraces from Senegal. *Front. Plant Sci.* 8:818. doi: 10.3389/fpls.2017.00818

Pearl millet (*Pennisetum glaucum* (L.) R. Br.) is a staple food and a drought-tolerant cereal well adapted to Sub-Saharan Africa agro-ecosystems. An important diversity of pearl millet landraces has been widely conserved by farmers and therefore could help coping with climate changes and contribute to future food security. Hence, characterizing its genetic diversity and population structure can contribute to better assist breeding programs for a sustainable agricultural productivity enhancement. Toward this goal, a comprehensive panel of 404 accessions were used that correspond to 12 improved varieties, 306 early flowering and 86 late-flowering cultivated landraces from Senegal. Twelve highly polymorphic SSR markers were used to study diversity and population structure. Two genes, *PgMADS11* and *PgPHYC*, were genotyped to assess their association to flowering phenotypic difference in landraces. Results indicate a large diversity and untapped potential of Senegalese pearl millet germplasm as well as a genetic differentiation between early- and late-flowering landraces. Further, a fine-scale genetic difference of *PgPHYC* and *PgMADS11* (SNP and indel, respectively) and co-variation of their alleles with flowering time were found among landraces. These findings highlight new genetic insights of pearl millet useful to define heterotic populations for breeding, genomic association panel, or crosses for trait-specific mapping.

Keywords: millet, *Pennisetum glaucum*, flowering, genetic diversity, SSR, *PgMADS11*, *PgPHYC*

INTRODUCTION

Low agricultural productivity causes food insecurity and malnutrition in Sub-Saharan Africa (SSA). This is due to climate variability, increased growth and needs of worldwide population, water demand, intensive exploitation of natural resources and environmental degradation. Therefore, better uses of natural resources could help overcoming some of these constraints and greatly contribute to improving productivity.

Pearl millet (*Pennisetum glaucum* (L.) R. Br.) is a major staple food for many SSA and Asian countries. Pearl millet is highly allogamous and mainly a rainfed crop, covering a wide range of different ecological zones and production systems. However, yield is low and variable, rarely reaching 1000 kg/ha. This is mainly explained by the limited exploitation of genetic resources and availability of improved varieties, but also by low soil fertility, drought, heat and highly variable rainfall (Waddington et al., 2010). Moreover, production is threatened by downy mildew disease, striga parasitic weed and predation by insects (Waddington et al., 2010). Characterization of crops genetic resources is a prerequisite to build up a breeding program for sustainable productivity enhancement.

Senegal is one of the top 10 pearl millet producers in the world (FAOSTAT, 2013), where farmers distinguish two main types of cultivars based on growth duration. Cultivars, called *Souna*, are sensitive to photoperiod, with a short cycle between 65 and 90 days and adapted to low (350–600 mm) rainfall regions. Cultivars, called *Sanio*, are a less photoperiod-sensitive type than *Souna*, with a long cycle between 120 and 150 days, and are adapted to high (900–1200 mm) rainfall regions. Nationwide, *Souna* type occupies nearly half the area sown to cereals (51%) while *Sanio* type, mainly cultivated in the South, represents about 15% of total millet production (ANSD, 2014).

Investigating genetic diversity and patterns of early- and late-flowering landraces is very important since flowering cycles and photoperiod sensitivity play a crucial role in the adaptation to climatic conditions. It is assumed a direct effect of selection for earliness associated with climate variations (Vigouroux et al., 2011). Genome scans and genetic association mapping have identified several genes tightly linked to adaptive traits of pearl millet in semi-arids areas. A SNP in the *Phytochrome C* locus (*PgPHYC*) and an indel variation in *PgMADS11* gene were associated with flowering time variation, annual rainfall and spike length of pearl millet (Saïdou et al., 2009; Mariac et al., 2011; Vigouroux et al., 2011). From farmers point of view, distinction between early- and late-flowering landraces is very clear but it is not always associated with clear genetic differentiation (Dussert et al., 2015). Flowering time in pearl millet is derived from a common domestication event and a strong gene flow between early- and late-flowering landraces is observed (Dussert et al., 2015). The level of gene flow would depend on cycle overlapping, agricultural practices and spatial distribution (Mariac et al., 2006b; Allinne et al., 2007; Lakis et al., 2012).

Because there is a critical need for adapting local agriculture to harsher future conditions, landraces and improved varieties adaptation will mostly rely on standing genetic variation available within the cultivated compartment. Recently, phenotyping

(Sy et al., 2015) and genotyping by sequencing (Hu et al., 2015) studies were carried out on a set of Senegalese pearl millet landraces. Sampling was restricted to only one agro-ecological area of Senegal, the Groundnut Basin, and included only nine so-called intermediate-flowering landraces (flower between 75 to 100 days after sowing). Based on phenology, head architecture and grain color, these accessions were classified into three cultivar groups, indicating a morphological diversity between early flowering landraces (Sy et al., 2015). Using 83,875 single nucleotide polymorphisms (SNPs) on the same set of accessions in addition to 252 global accessions, a higher genetic diversity was observed in Senegal accessions compared to millet accessions in India, South and Western Africa (Hu et al., 2015). Any local structure was evidenced, therefore studies of loci that control the cycle length would be necessary to assess more accurately the evolution of cultivated millet varieties (Dussert et al., 2015).

Here, a fine scale sampling strategy and genetic characterization are described that differentiate early- and late-flowering landraces of Senegalese pearl millet. Using highly polymorphic SSRs markers, genetic diversity and population structure of the landraces was assessed. Allelic diversity of *PgPHYC* and *PgMADS11* genes, both linked with flowering time variation and rainfall, was further investigated (Saïdou et al., 2009; Mariac et al., 2011; Vigouroux et al., 2011).

MATERIALS AND METHODS

Plant Materials

Collects were done in 1992 and 1994 in the main areas of millet production in the Groundnut Basin as previously described (Sy et al., 2015). Geographical coordinates of these accessions were partially retrieved (88%) by using village names. Additional collects were done in 2010 and 2014 to cover pearl millet production areas, except the city of Dakar and the eastern-south area where the Niokolo-Koba Wildlife Park is located. Geographical coordinates of these new accessions were recorded using a GPS. As our focus is on local landraces, villages near major roads or markets were avoided. In total, 392 accessions were collected from 316 villages, i.e., 1.24 accessions per village on average. A panel of 404 accessions was analyzed including 12 improved varieties bred locally and widely used by farmers, 306 early flowering landraces (252 villages) and 86 late-flowering landraces (74 villages). Among the 316 villages, 10 villages were sampled with both early- and late- flowering landraces (Supplementary Table S1). Cycles were recorded following farmer interviews.

DNA Extraction and SSR Genotyping

Five seeds per accession were grown in the greenhouse 3–4 weeks according to sampling date. About 200 mg of leaf sample from one individual per accession were collected and DNA extraction was carried out using the previously described protocol (Mariac et al., 2006a). Twelve highly polymorphic microsatellites distributed throughout the pearl millet genome were used (Supplementary Table S2). These markers have been previously described (Allouis et al., 2001; Qi et al., 2001;

Budak et al., 2003; Mariac et al., 2006a). PCR reactions were performed using the Multiplex PCR Kit (Qiagen, Inc) following the recommended protocol. PCR were conducted using a thermal cycler TC-Plus (TECHNE): pre-denaturation of 95°C for 15 min then 35 cycles consisting of a denaturation step at 94°C for 30 s, annealing at 55°C for 90 s, elongation at 72°C for 60 s and a final extension at 60°C for 30 min). Four positive and four negative controls were repeated on each PCR plate. Samples were genotyped on an ABI 3130 Prism® (Applied Biosystems®) and read with Genemapper™ software (version 3.7; Applied Biosystems®).

Genetic Diversity and Population Structure Analyses

Genetic diversity, heterozygosities (expected and observed) and F-statistics were calculated using Genalex 6.5 (Peakall and Smouse, 2012). For genetic structure, a principal component analysis (PCA) using the package ade4 (Thioulouse et al., 1997) implemented in R software (R Development Core Team, 2008) was first performed. Then, population structure was investigated using STRUCTURE software 2.3.3 (Pritchard et al., 2000). Analysis was performed with the admixture model (Falush et al., 2003) with K ancestral populations ranging from 1 to 6. We used 500,000 iterations and a burn-in period of 100,000, 10 runs for each K-value were performed. The values for the number of clusters (K) were assessed according to Evanno et al. (2005) by the ($D, \Delta K$) criterion and the log-likelihood ($\ln P(D|K)$) plot. Individuals were assigned to a cluster if their ancestry was higher than 70%, $q \geq 0.7$.

For spatial analysis of genetic variability, a total of 367 geo-referenced accessions including 281 early- and 86 late-flowering landraces were used. Spatial principal component analysis (sPCA) was performed using the *ade4* package (Jombart, 2008) with R software. The spatial genetic structure was assessed using spatial autocorrelation analyses of kinship coefficients between individuals (Loiselle et al., 1995) following the standard procedure (Vekemans and Hardy, 2004) implemented in SPAGeDi version 1.2 (Hardy and Vekemans, 2002). Mean multilocus kinship coefficient values, F_{ij} , i.e., genetic similarity between individuals i and j relative to the mean genetic similarity between random individuals in the sample, were regressed on both the linear (d_{ij}) and the logarithmic ($\ln(d_{ij})$) spatial distance between individuals. This distance was calculated as the Euclidian distances using spatial coordinates. The regression slopes b_d and $b_{\ln d}$ were jointly assessed. Standard errors for the kinship coefficients were estimated using a jackknife procedure over all loci. We tested the significance of the kinship coefficients and the regression slopes b_d and $b_{\ln d}$ estimates by comparing the observed values to those obtained after 10,000 random permutations.

PgPHYC and PgMADS11 Genotyping

Polymorphisms in both genes were associated with flowering time variation (Saidou et al., 2009; Mariac et al., 2011; Vigouroux et al., 2011). The panel of accessions was genotyped with *PgPHYC* (Acc numbers FN376885–FN377564) (Vigouroux et al., 2011)

and *PgMADS11* (Acc numbers FN552468–FN552522) (Mariac et al., 2011) to test the allelic differences in genotype frequencies between early- and late-flowering landraces.

For *PgPHYC*, a polymorphism at the 5' of the gene was assessed. A C/G SNP at that position is cleaved by PvuII restriction enzyme and therefore accessions scored as C/C, G/G and C/G according to their digestion pattern (Saidou et al., 2009).

For *PgMADS11*, an indel polymorphism of 24 bp was assessed as previously reported (Mariac et al., 2011).

Logistic regressions between genotypes, genetic cluster, latitude and longitude for each gene were further performed.

RESULTS

Genetic Diversity of Senegalese Germplasm

The germplasm collected through this study is the most comprehensive sampling to date of landraces from Senegal (Figure 1). Genotyping data revealed a total of 101 alleles with an average of 8.4 alleles per locus (Supplementary Table S2 and Figures S1–S3). The final data set contained only 1.8% of missing data. High levels of genetic diversity characterize both groups (Table 1). Early flowering landraces presented the highest level of genetic diversity as measured by observed $H_{Obs} = 0.481$ or expected heterozygosity, $H_{Exp} = 0.567$. In contrast, improved varieties showed the lowest, with $H_{Obs} = 0.391$. For early flowering landraces, only one locus (PSMP2249) was found not to be at Hardy-Weinberg equilibrium (HWE). Two loci were not at HWE (PSMP2249 and PSMP2246) in late-flowering landraces. Groups showed low levels of inbreeding with F_{IS} values of 0.160, 0.128, and 0.258 for early and late landraces and improved varieties, respectively. Furthermore, early-, late-flowering landraces and improved varieties present 22, 8 and 1 private alleles, respectively. Low F_{ST} differentiation was found between improved varieties and early flowering landraces (0.004), between early- and late-flowering landraces (0.052) and between late-flowering landraces and improved varieties (0.063).

Population Genetic Structure

Bayesian clustering analyses showed a clear structure between early- and late-flowering landraces (Figure 2). The value of the Evanno criterion (ΔK) was the highest for $K = 2$ (Supplementary Figure S1), supporting the evidence of two major clusters. A total of 89% of the early flowering landraces were assigned to a single cluster, and 90% of the late-flowering landraces were assigned to the other cluster. Thirty-four accessions of early- and nine of late-flowering landraces were misassigned ($q < 0.7$), but showed intermediate ancestries. For the 10 villages where both early- and late-flowering landraces were sampled, all accessions of similar phenotype were assigned to their respective cluster.

Similar results were found with PCA showing a clear distinction between early- and late-flowering landraces (Figure 3). The two first principal components explained, respectively, 5.2 and 2.8% of the inertia. PSMP2247 and PSMP2202 showed the highest contribution to the PC1. Two alleles, PSMP2247-199 and PSMP2202-146, showed

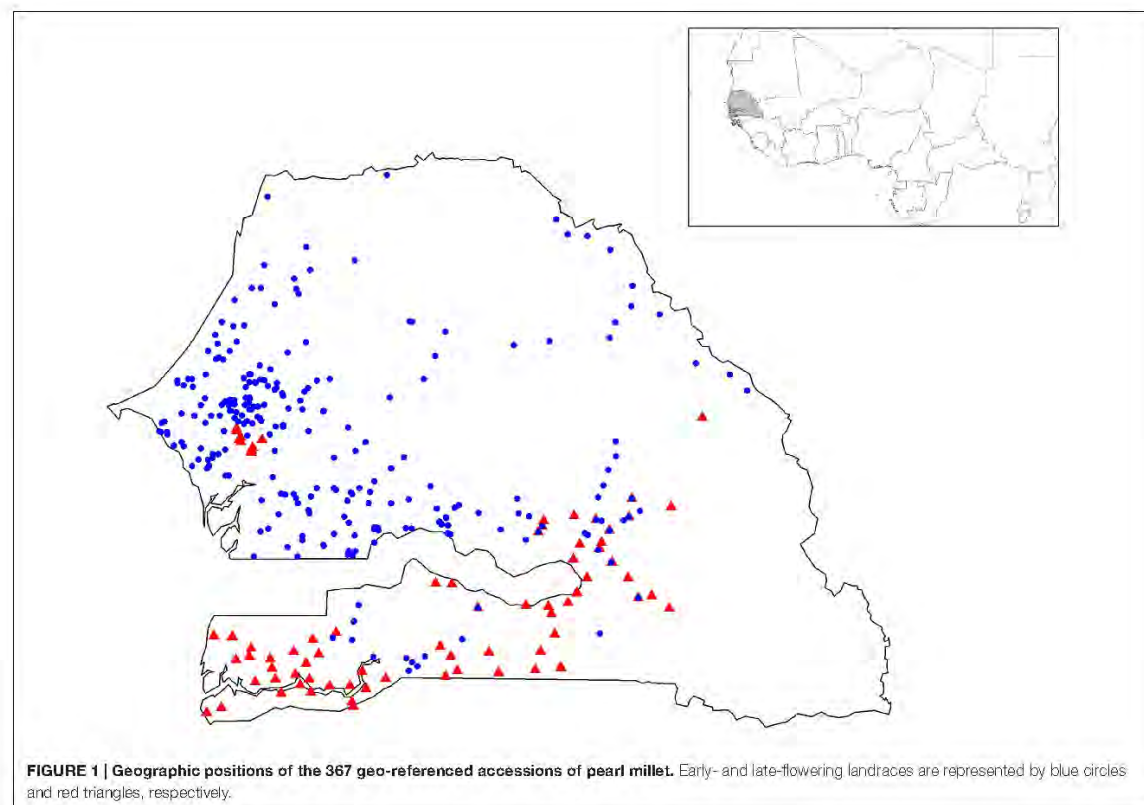


TABLE 1 | Genetic diversity statistics for improved varieties, early- and late-flowering landraces.

	Nb	Na	H _{Obs}	H _{Exp}	F _{IS}
Improved varieties	12	4.4	0.350	0.513	0.258
Early flowering landraces	306	7.7	0.446	0.542	0.180
Late-flowering landraces	86	5.8	0.434	0.494	0.128

Nb, number of accessions; Na, mean number of alleles; H_{Obs}, observed heterozygosity; H_{Exp}, expected heterozygosity; F_{IS}, inbreeding coefficient.

high frequencies in late-flowering landraces (0.92 and 0.83, respectively). However, removing these two loci did not affect PCA results. Finally, any genetic differentiation was found between improved and early accessions.

Spatial Analysis of Genetic Variability

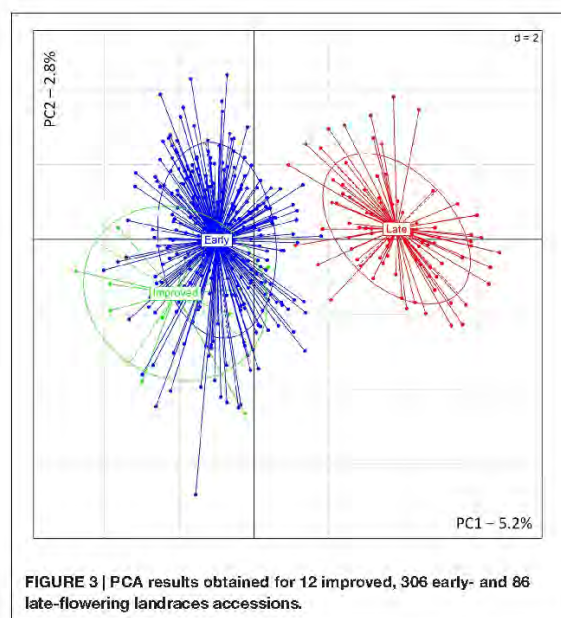
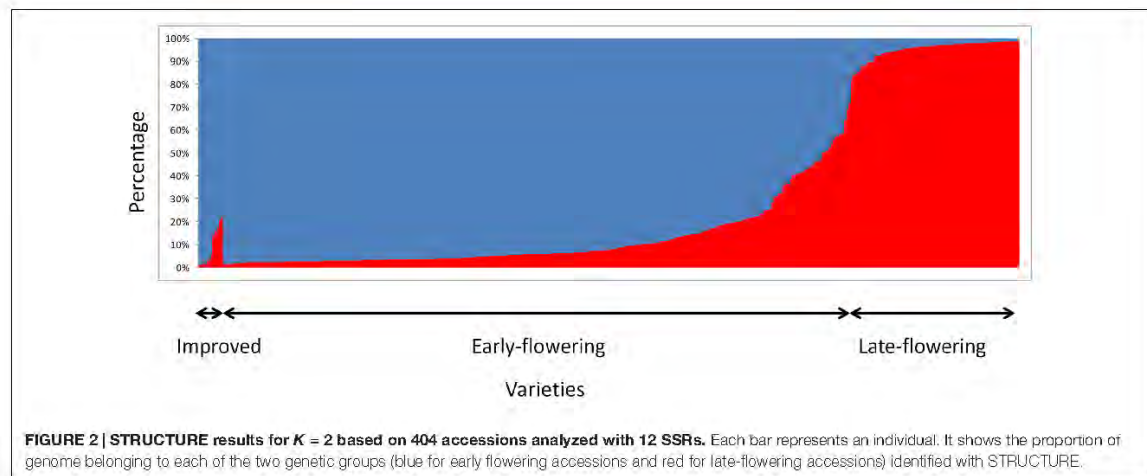
Spatial principal component analysis revealed a more cryptic genetic structure (Figure 4). Global structures, i.e., large spatial scale, were significant (p -value = 0.0001) with the first principal component, showing a high autocorrelation (Moran's I = 0.50). In contrast, local structures were not significant (p -value = 0.77). The first axis of the sPCA identified two clusters. Considering spatial information, early flowering landraces from Southern Senegal showed a more admixed pattern than observed with

STRUCTURE, while no differences were observed considering late-flowering landraces. Bayesian and multivariate approaches confirm genetic assignments for late-flowering landraces from Central Senegal.

The pattern of isolation by distance (IBD) for early- and late-flowering landraces was investigated. Low IBD slopes were obtained for early flowering (b_E = $-2.41E-05$) and late-flowering (b_L = $-9.89E-05$) landraces (Figure 5). Similar results were found for logarithmic distances. This suggests no significant pattern of isolation-by-distance.

PgPHYC and PgMADS11 Allele Diversity

Landraces were genotyped for both *PgPHYC* and *PgMADS11* alleles to assess genetic diversity in relation to flowering time (Table 2 and Supplementary Figures S4, S5). Only early landraces carried early flowering allele (G) at the *PgPHYC* locus leading to a significant difference in genotypes frequencies (F -test, p = 0.006). For *PgMADS11*, a significant difference in genotypes frequencies (F -test, p = 0.004) was observed with early flowering landraces having higher frequency of the allele (363 bp fragment). A significant correlation between allele G from *PgPHYC* with latitude (p -value = 0.0348) was observed. However, significance disappeared when taking into consideration genetic clustering ($q \leq 0.7$). For



PgMADS11, data reveal a significant correlation with longitude (p -value < 0.001) even when considering genetic clustering in the mode.

DISCUSSION

Large and Untapped Diversity in Senegalese Germplasm

In this study, a relatively high genetic diversity (H_{Exp} : 0.516) of pearl millet germplasm of Senegal is reported. Previous

studies have reported a higher (H_{Exp} : 0.69) (Stich et al., 2010) and a lower (H_{Exp} = 0.49) genetic diversity in Niger pearl millet (Mariac et al., 2006a). These differences are due to the number of SSRs used. In the meanwhile, a similar number of alleles per locus (6 alleles per locus for Senegal vs. 6.2 alleles per locus for Niger) was observed. Comparing heterozygosity levels, observation showed a higher coefficient (F_{IS} = 0.30) in Niger germplasm than in Senegal germplasm (F_{IS} = 0.18). However, these results are consistent with data from others studies carried out on accessions across SSA regions and India (Oumar et al., 2008; Dussert et al., 2015; Hu et al., 2015). Pearl millet shows a high genetic diversity that can be explained by its strong outcrossing rate (75%) and the still on-going gene flow with its wild relative (Mariac et al., 2006a; Lewis, 2010). This high genetic diversity is in line with its high phenotyping diversity observed in Senegal (Sy et al., 2015) as in Western Africa (Pucher et al., 2015). Together, these findings highlight the untapped potential of Senegalese pearl millet germplasm for breeding.

Genetic proximity between early flowering varieties and improved varieties highlight a history of breeding programs and agricultural practices. In Senegal, few breeding programs have been undertaken on pearl millet but all focused on reducing the flowering cycle and were “population” varieties whose parental seeds were collected from local landraces. In addition, farmers still grow improved varieties jointly with landraces in their field, increasing gene flow and thus genetic proximity.

Genetic Structure of Early- and Late-Flowering Landraces

Bayesian results clearly help identifying genetic structure associated with early- and late-flowering landraces. This genetic structure was partly explained by the geographic distribution of landraces as shown through the sPCA analysis, making it difficult to entirely disentangle spatial and genetic structure. More

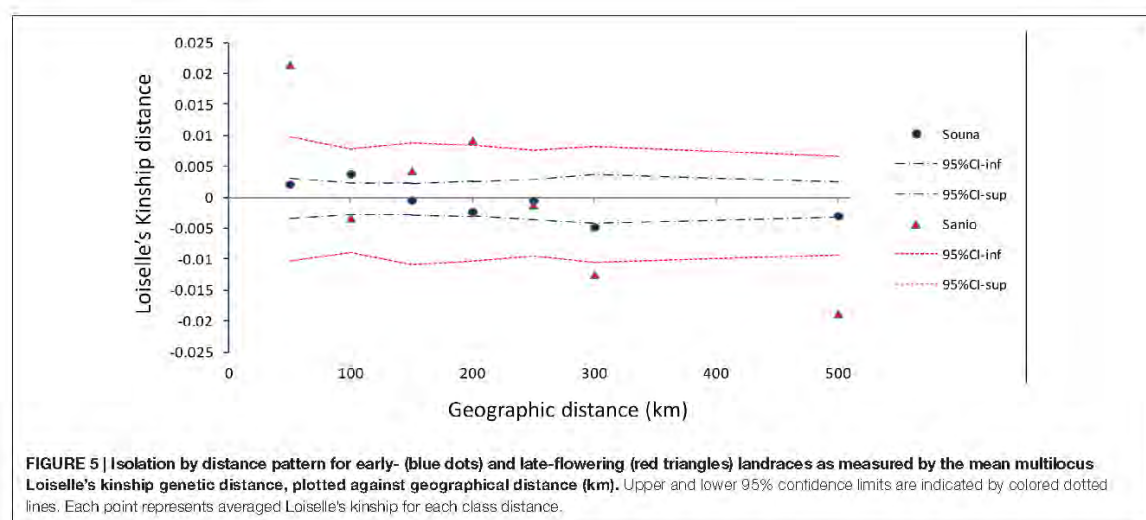
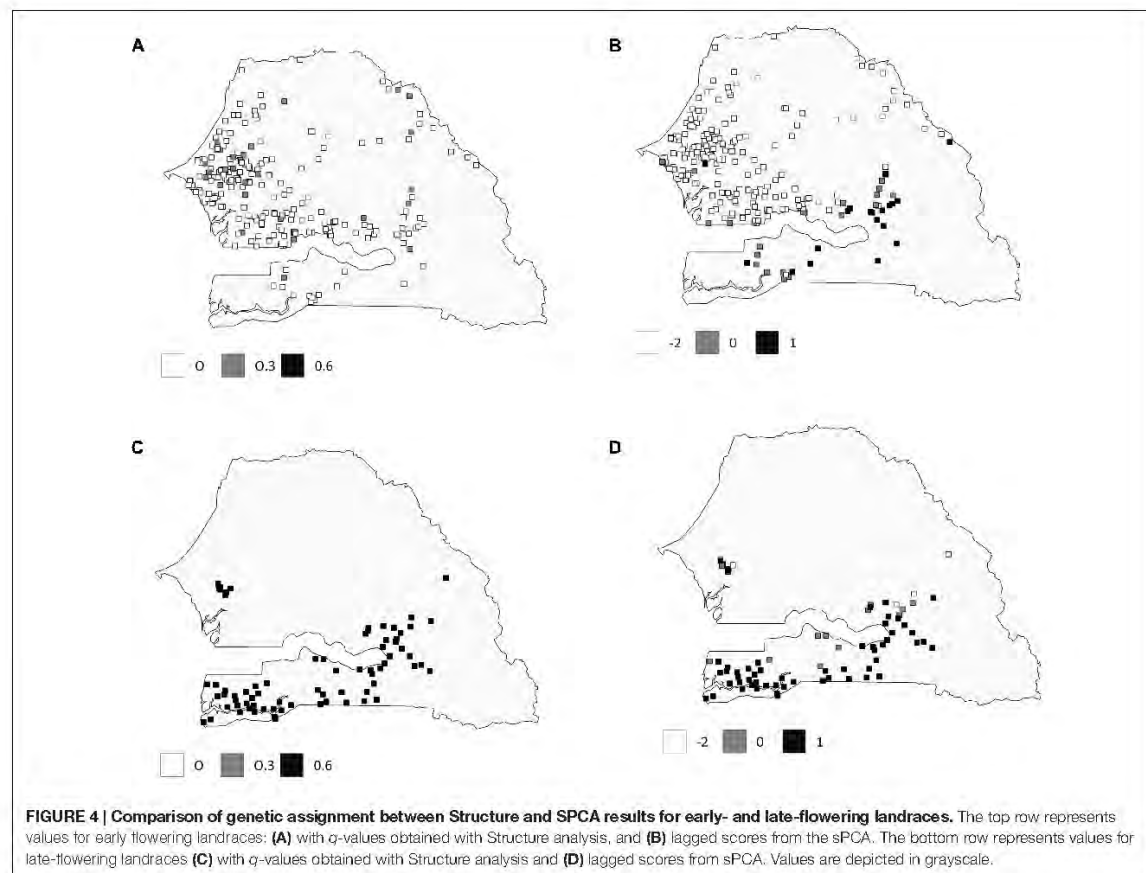


TABLE 2 | Genotypes counts for improved varieties, early- and late-flowering landraces with q -values ≤ 0.7 .

	<i>PgPHYC</i> alleles			<i>PgMADS11</i> alleles		
	C/C	C/G	G/G	363/363	363/387	387/387
Improved varieties	8	3	1	1	2	9
Early flowering landraces	239	21	0	18	65	185
Late-flowering landraces	79	0	0	2	16	62

contrast in admixture patterns might have been obtained with a higher number of SSR markers. However, this was compensated by our sampling effort and we were able to reveal a clear structure in Senegal landraces where other studies failed with a higher number of markers (Hu et al., 2015).

At a local scale, more contrasted admixture patterns were observed. In Central Senegal, clear differentiation between early- and late-flowering landraces was found. In contrast, more admixed patterns were observed in the Southern Senegal. The results might be related to the agricultural practices of the farmers, such as spatial cropping and seed circulation (Mariac et al., 2006b; Lakis et al., 2012; Kouakou et al., 2013).

Our data indicated two distinct genetic clusters in Senegalese pearl millet germplasm. In addition, comparing early- and late-flowering landraces, we found a slight genetic differentiation between the two groups of 0.052, which is of the same magnitude (0.053) found in Niger (Lakis et al., 2012). A single domestication event led to early- and late-flowering landraces, partly explaining the limited genetic differentiation observed at a regional scale (Dussert et al., 2015).

Flowering Traits Diversity

Photoperiod-sensitivity of pearl millet landraces and thus variation of flowering cycle constitute a key response for adaptation to future climate conditions (Sultan et al., 2013). Indeed, pearl millet landraces from Niger show reduced flowering cycle associated with drought episodes from 1976 to 2003 (Vigouroux et al., 2011). This reduction was correlated with changes in allele frequencies for *PgPHYC*. Correlation between allele frequencies and rainfall were also found with *PgMADS11* (Saidou et al., 2009; Mariac et al., 2011). Our data showed significant genetic differences for both genes with early flowering landraces enriched in precocity alleles.

Flowering time has been correlated with latitude, early flowering landraces being grown in northern latitude where environmental conditions are more arid (Haussmann et al., 2006; Pucher et al., 2015). In our study, the latitude effect was confounded with spatial genetic structure for *PgPHYC*. On the other hand, a correlation with longitude was found for *PgMADS11*. Further investigation would be needed to fully address this correlation. In any case, the use of flowering genes *PgPHYC* and *PgMADS11* in marker assisted-selection programs presents some interest.

Challenges for Adaptation of Pearl Millet and Breeding Strategies

Sub-Saharan Africa recorded long dry spells in the 1970s and 1980s that led to breeding for short cycles improved varieties. Indeed, pearl millet breeding programs were predominantly built on restricted genetic resources of early flowering landraces. Strong differentiation between early- and late-flowering landraces from Senegal suggests the existence of an important gene pool that has not been exploited yet. The high genetic diversity could explain the wider range of pearl millet adaptation to dry areas and this potential may further contribute to breeding programs in response to the specific needs or target areas (Kouressy et al., 2004). For instance, a key strategy to cope against climate changes within SSA agrosystems is to tap into diversity of flowering time (Haussmann et al., 2012; Sultan et al., 2013) and resilience (Prieto et al., 2015).

Analyses of allelic variation of *PgPHYC* and *PgMADS11* indicate fine-scale genetic difference (SNP and indel, respectively) among individuals and/or genotypes. Knowing that responses to photoperiod and rainfall were genetically associated with both genes, implication could be their use to detect/track climate adaptive changes to environment variations. For example, earliness of flowering and latitude correlation observed in early landraces support assumption that a direct effect of selection for that trait which is associated with climate variations such as photoperiod and rainfall. Both traits are key targets in selection for millet genotypes to be cultivated in rainfed areas.

CONCLUSION

The genetic diversity and population structure of Senegalese pearl millet landraces were assessed using a large panel of accessions and a limited number of SSRs markers. Results highlight a high genetic diversity and an untapped potential of the germplasm. However, two clusters were clearly distinguished as revealed by differentiation between early- and late-flowering landraces. Further, genetic difference and allelic co-variation in flowering genes *PgPHYC* and *PgMADS11* were found among individuals. These findings give new insights into Senegalese pearl millet germplasm and are promising for developing new cultivars and heterotic groups that can be used to breed synthetic and hybrid varieties with higher degrees of heterozygosity in order to intensify yield production under harsh semiarid environments.

AUTHOR CONTRIBUTIONS

NK, AF, DD, MG, YV, and AB designed the study. OD, AF, MG, MP, HT, BD, and OS collected samples. OD, MC, MP, and LZ performed DNA extraction, PCR and sequencing. AB, CB-S, and OD performed the genetic analyses. OD, AB, CB-S, DD, and NK drafted the manuscript. All authors contributed to the final version.

FUNDING

We are thankfully to the West Africa Agricultural Productivity Program (WAAPP/PPAAO 2A) for funding research activities (Grant No 51350 SE to NK) and OD thesis.

ACKNOWLEDGMENTS

We also thank farmers and local authorities for their consent and collaboration for seeds sharing. We acknowledge

Claire Billot and Monique Deu (CIRAD) for sampling collection and Cédric Mariac (IRD) for scientific and technical assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00818/full#supplementary-material>

REFERENCES

- Allinne, C., Mariac, C., Vigouroux, Y., Bezançon, G., Couturon, E., Moussa, D., et al. (2007). Role of seed flow on the pattern and dynamics of pearl millet (*Pennisetum glaucum* [L.] R. Br.) genetic diversity assessed by AFLP markers: a study in south-western Niger. *Genetica* 133, 167–178. doi: 10.1007/s10709-007-9197-7
- Allouis, S., Qi, X., Lindup, S., Gale, M. D., and Devos, K. M. (2001). Construction of a BAC library of pearl millet, *Pennisetum glaucum*. *Theor. Appl. Genet.* 102, 1200–1205. doi: 10.1007/s001220100559
- ANSD (2014). *Bulletin Mensuel des Statistiques Économiques de Mars 2014*. Dakar: Agence Nationale de la Statistique et de la Démographie.
- Budak, H., Pedraza, F., Cregan, P. B., Baenziger, P. S., and Dweikat, I. (2003). Development and utilization of SSRs to estimate the degree of genetic relationships in a collection of pearl millet germplasm. *Crop Sci.* 43, 2284–2290. doi: 10.2135/cropsci2003.2284
- Dussert, Y., Snirc, A., and Robert, T. (2015). Inference of domestication history and differentiation between early- and late-flowering varieties in pearl millet. *Mol. Ecol.* 24, 1387–1402. doi: 10.1111/mec.13119
- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x
- Falush, D., Stephens, M., and Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164, 1567–1587.
- FAOSTAT (2013). *Millets Production in de World*. Rome: FAOSTAT.
- Hardy, O. J., and Vekemans, X. (2002). spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes* 2, 618–620. doi: 10.1046/j.1471-8286.2002.00305.x
- Hausmann, B. I. G., Boubacar, A., Boureima, S. S., and Vigouroux, Y. (2006). Multiplication and preliminary characterization of west and central African pearl millet landraces. *Int. Sorghum Millets Newsl.* 47, 110–112.
- Hausmann, B. I. G., Fred Rattunde, H., Weltzien-Rattunde, E., Traoré, P. S. C., vom Brocke, K., and Parzies, H. K. (2012). Breeding strategies for adaptation of pearl millet and sorghum to climate variability and change in West Africa. *J. Agron. Crop Sci.* 198, 327–339. doi: 10.1111/j.1439-037X.2012.00526.x
- Hu, Z., Mbacké, B., Perumal, R., Guèye, M. C., Sy, O., Bouchet, S., et al. (2015). Population genomics of pearl millet (*Pennisetum glaucum* (L.) R. Br.): Comparative analysis of global accessions and Senegalese landraces. *BMC Genomics* 16:1048. doi: 10.1186/s12864-015-2255-0
- Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405. doi: 10.1093/bioinformatics/btn129
- Kouakou, P. K., Muller, B., Guissé, A., Yao, R. N., Fofana, A., and Cissé, N. (2013). Étude et prise en compte en modélisation de l'effet de la latitude sur la réponse à la photopériode chez divers génotypes de mil (*Pennisetum glaucum* (L.) R. Br.) du Sénégal. *J. Appl. Biosci.* 67, 5289–5301. doi: 10.4314/jab.v67i0.95051
- Kouressy, M., Vaksman, M., Niangado, O., and Sanogo, M. D. (2004). *Valorisation et Préservation de la Diversité Génétique du Mil au Mali*. Paris: IRD Éditions. doi: 10.4000/books.irdeditions.539
- Lakis, G., Ousmane, A. M., Sanoussi, D., Habibou, A., Badamassi, M., Lamy, F., et al. (2012). Evolutionary dynamics of cycle length in pearl millet: the role of farmer's practices and gene flow. *Genetica* 139, 1367–1380. doi: 10.1007/s10709-012-9633-1
- Lewis, L. R. (2010). Biogeography and genetic diversity of pearl millet (*Pennisetum glaucum*) from Sahelian Africa. *Prof. Geogr.* 62, 377–394. doi: 10.1080/00330124.2010.483640
- Loiselle, B. A., Sork, V. L., Nason, J., and Graham, C. (1995). Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *Am. J. Bot.* 82, 1420–1425. doi: 10.2307/2445869
- Mariac, C., Jehin, L., Saïdou, A.-A., Thuillet, A.-C., Couderc, M., Sire, P., et al. (2011). Genetic basis of pearl millet adaptation along an environmental gradient investigated by a combination of genome scan and association mapping: adaptation to climatic variation. *Mol. Ecol.* 20, 80–91. doi: 10.1111/j.1365-294X.2010.04893.x
- Mariac, C., Luong, V., Kapran, I., Mamadou, A., Sagnard, F., Deu, M., et al. (2006a). Diversity of wild and cultivated pearl millet accessions (*Pennisetum glaucum* [L.] R. Br.) in Niger assessed by microsatellite markers. *Theor. Appl. Genet.* 114, 49–58.
- Mariac, C., Robert, T., Allinne, C., Remigereau, M. S., Luxereau, A., Tidjani, M., et al. (2006b). Genetic diversity and gene flow among pearl millet crop/weed complex: a case study. *Theor. Appl. Genet.* 113, 1003–1014.
- Oumar, I., Mariac, C., Pham, J.-L., and Vigouroux, Y. (2008). Phylogeny and origin of pearl millet (*Pennisetum glaucum* [L.] R. Br.) as revealed by microsatellite loci. *Theor. Appl. Genet.* 117, 489–497. doi: 10.1007/s00122-008-0793-4
- Peakall, R., and Smouse, P. E. (2012). GenALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28, 2537–2539. doi: 10.1093/bioinformatics/bts460
- Prieto, I., Vielle, C., Barre, P., Durand, J.-L., Ghesquiere, M., and Litrico, I. (2015). Complementary effects of species and genetic diversity on productivity and stability of sown grasslands. *Nat. Plants* 1:15033. doi: 10.1038/nplants.2015.33
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Pucher, A., Sy, O., Angarawai, I. I., Gondah, J., Zangre, R., Ouedraogo, M., et al. (2015). Agro-morphological characterization of west and central african pearl millet accessions. *Crop Sci.* 55, 737–748. doi: 10.2135/cropsci2014.06.0450
- Qi, X., Lindup, S., Pittaway, T. S., Allouis, S., Gale, M. D., and Devos, K. M. (2001). Development of simple sequence repeat markers from bacterial artificial chromosomes without subcloning. *Biotechniques* 31, 355–362.
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Saïdou, A.-A., Mariac, C., Luong, V., Pham, J.-L., Bezançon, G., and Vigouroux, Y. (2009). Association studies identify natural variation at PHYC linked to flowering time and morphological variation in pearl millet. *Genetics* 182, 899–910. doi: 10.1534/genetics.109.102756
- Stich, B., Hausmann, B. I., Pasam, R., Bhosale, S., Hash, C. T., Melchinger, A. E., et al. (2010). Patterns of molecular and phenotypic diversity in pearl millet [*Pennisetum glaucum* (L.) R. Br.] from West and Central Africa and their relation to geographical and environmental parameters. *BMC Plant Biol.* 10:216. doi: 10.1186/1471-2229-10-216
- Sultan, B., Roudier, P., Quirion, P., Alhassane, A., Muller, B., Dingkuhn, M., et al. (2013). Assessing climate change impacts on sorghum and millet yields in the Sudanian and Sahelian savannas of West Africa. *Environ. Res. Lett.* 8, 014040. doi: 10.1088/1748-9326/8/1/014040

- Sy, O., Fofana, A., Cisse, N., Noba, K., Diouf, D., Ndoeye, I., et al. (2015). Étude de la variabilité agromorphologique de la collection nationale de mils locaux du Sénégal. *J. Appl. Biosci.* 87, 8030–8046. doi: 10.4314/jab.v87i1.1
- Thioulouse, J., Chessel, D., Dolédec, S., and Olivier, J.-M. (1997). ADE-4: a multivariate analysis and graphical display software. *Stat. Comput.* 7, 75–83. doi: 10.1023/A:1018513530268
- Vekemans, X., and Hardy, O. J. (2004). New insights from fine-scale spatial genetic structure analyses in plant populations. *Mol. Ecol.* 13, 921–935. doi: 10.1046/j.1365-294X.2004.02076.x
- Vigouroux, Y., Mariac, C., De Mita, S., Pham, J.-L., Gérard, B., Kapran, I., et al. (2011). Selection for earlier flowering crop associated with climatic variations in the sahel. *PLoS ONE* 6:e19563. doi: 10.1371/journal.pone.0019563
- Waddington, S. R., Li, X., Dixon, J., Hyman, G., and de Vicente, M. C. (2010). Getting the focus right: production constraints for six major food crops in Asian and African farming systems. *Food Secur.* 2, 27–48. doi: 10.1007/s12571-010-0053-8
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2017 Diack, Kane, Berthouly-Salazar, Gueye, Diop, Fofana, Sy, Tall, Zekraoui, Piquet, Couderc, Vigouroux, Diouf and Barnaud. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Annexe 6 : Participations aux congrès/conférences nationaux et internationaux :

- Oumar Diack*, Ghislain Kanfany, Ousmane Sy, Amadou Fofana, Mame Codou Guèye Cissé, Diaga Diouf, Ndjido Ardo Kane. **Mise en place d'une core collection de mil [Pennisetum glaucum ssp. glaucum (L.) R. Br.] du Sénégal, par approche heuristique.** (2017). Communication orale, *2ème édition des journées scientifique du campus de Bel-Air, 4-5 Octobre 2017, Pole de recherche de Hann, Institut Sénégalais de Recherches Agricoles, Dakar (Sénégal).*

- Oumar Diack*, Adeline Barnaud, Leila Zekraoui, Jésaëlle Piquet, Marie Couderc, Hamidou Tall, Ousmane Sy, Amadou Fofana, Mame Codou Guèye Cissé, Cecile Berthouly-Salazar, Yves Vigouroux, Diaga Diouf, Ndjido Ardo Kane. **Caractérisation génétique neutre et adaptative des mils cultivés [Pennisetum glaucum ssp. glaucum (L.) R. Br.] du Sénégal.** Communication orale, *6ème édition des doctoriales de l'Ecole Doctorale des Sciences de la vie, de la Santé et de l'Environnement , 8-9 Decembre 2015, Campus UCAD II, Université Cheikh Anta Diop, Dakar (Sénégal).*

- Oumar Diack*, Adeline Barnaud, Leila Zekraoui, Jésaëlle Piquet, Marie Couderc, Hamidou Tall, Ousmane Sy, Amadou Fofana, Mame Codou Guèye Cissé, Cecile Berthouly-Salazar, Yves Vigouroux, Diaga Diouf, Ndjido Ardo Kane. **Neutral genetic characterization of Senegalese cultivated pearl millet germplasm [Pennisetum glaucum ssp. glaucum (L.) R. Br.]** (2015), Poster. *Annual meeting of the Society for Molecular Biology and Evolution, July 12th to the 16th, 2015, Hofburg Palace, Vienne (Autriche).*

- Oumar, Diack*. Adeline, Barnaud. Diaga, Diouf. Ndjido, Kane. (2014). **Etude de la diversité génétique de la collection de mil [Pennisetum glaucum ssp. glaucum (L.) R. Br.].** Communication orale. *2èmes journées du Laboratoire mixte international Adaptation des Plantes et microorganismes associés aux Stress Environnementaux (LAPSE). 18-19 Février 2014, Campus UCAD II, Université Cheikh Anta Diop, Dakar (Sénégal).*

Neutral genetic characterization of Senegalese pearl millet germplasm

[*Pennisetum glaucum* (L.) R. Br.]

Oumar Diack^{1,2}, Adeline Barnaud^{3,4}, Leila Zekraoui^{3,4}, J  s  lle Piquet⁴, Marie Couderc⁴, Hamidou Tall⁵, Ousmane Sy⁵, Amadou Fofana⁵, Mame Codou Gu  ye Ciss  ^{1,4}, Cecile Berthouly-Salazar^{3,4}, Yves Vigouroux^{3,4}, Diaga Dion⁶, Ndjido Ardo Kane^{1,3}

E-mail: oumar.diack@ucad.edu.sn, oumar.diack@ird.fr

¹ National Research Laboratory for Plant Production, Senegalese Institute of Agricultural Research (ISRA), Dakar, Senegal,

² Plant Biotechnology Laboratory, Department of Plant Biology, Faculty of Science and Technology, University Cheikh Anta Diop (UCAD), Dakar, Senegal,

³ Laboratory on the adaptation plant and associated microorganisms to environmental stresses (LMI LAPSE), Dakar, Senegal,

⁴ UMR DIADE, Institute of Research for Development (IRD), Montpellier, France,

⁵ National Center for Agricultural Research (CNRA), Senegalese Institute of Agricultural Research (ISRA), Bambey, Senegal,

⁶ Regional Study Centre for the Improvement and Adaptation to Drought (CERAAS), Senegalese Institute of Agricultural Research (ISRA), Thi  s, Senegal



INTRODUCTION

Pearl millet is widely cultivated in Sub Saharian Africa, therefore could contribute to fight malnutrition and food insecurity. It is a *Poaceae*, found under wild (*P. monodii*), cultivated (*P. glaucum*) and hybrid (*P. stenotachyum*) subspecies (Fig 1). Its domestication occurs over 4,800 years in Africa (1). In cultivated subspecies, it is distinguished early flowering and late flowering varieties (Fig 2).



Figure 1: Illustration of wild, cultivated and hybrid pearl millet subspecies

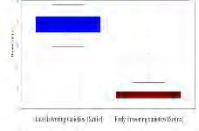


Figure 2: Flowering dates of pearl millet in Senegal

OBJECTIVES

The objective was to assess the genetic diversity of Senegalese pearl millet germplasm using microsatellite markers.

Specifically, we studied :

- The structure of genetic diversity of nationwide cultivated varieties
- The genetic variation of Senegalese germplasm in comparison with a worldwide collection.
- The introgressions between wild and cultivated subspecies.

METHODOLOGY

We collected 349 landraces (Fig 3) consisting of 253 early flowering (Souna) and 96 late flowering (Sanio) types. We also used 7 improved varieties homologated in Senegal.

The level of diversity was determined by assessing the diversity index (Ho, Hex, and Fis).

The structure of diversity was assessed using Bayesian analysis, isolation by distance, Principal Components Analysis as previously described (2).

Bayesian approach was used to estimate introgression between wild and cultivated subspecies as previously described (3).

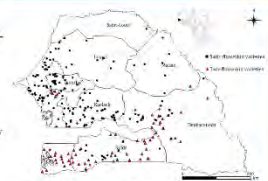


Figure 3: Geographical repartition pearl millet germplasm used in this study

GENETIC CHARACTERIZATION OF SENEGALESE CULTIVATED PEARL MILLET

LEVEL OF NEUTRAL DIVERSITY OF CULTIVATED PEARL MILLETS

Analyses of diversity indexes (Fig 4) reveals that :

- the allelic contribution of the early flowering varieties is higher than that of late flowering varieties
- the genetic diversity of the Senegalese germplasm is 0.54.

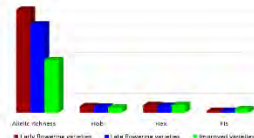


Figure 4: Contribution of early, late and improved varieties in genetic diversity in the Senegalese germplasm

ISOLATION BY DISTANCE (IBD)

Since we observed a structuration between Souna and Sanio, we estimate IBD within early and late flowering varieties using geographic and genetic distance matrices (Mantel test). Data (Fig 11) revealed no sub-structuration in Senegalese pearl millet.

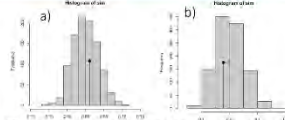


Figure 8: Histogram of mantel test simulation for (a) early varieties and (b) late varieties

BAYESIAN ANALYSIS OF CULTIVATED PEARL MILLETS

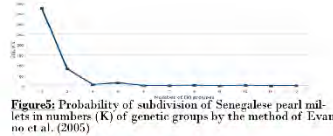


Figure 5: Probability of subdivision of Senegalese pearl millets in numbers (K) of genetic groups by the method of Evanno et al. (2005)

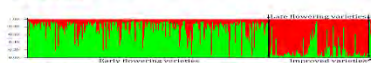


Figure 6: Graphical representation of individuals assignment test, in genetic groups

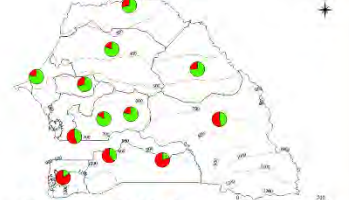


Figure 7: Geographical distribution of genetic groups according to rainfall

K means and STRUCTURE analyses revealed two distinct genetic groups based on flowering cycle (Fig 5 and 6). The genetic background of each group is geographically distributed based on rainfall: early flowering Souna genetic are more cultivated up north (300mm) while late flowering Sanio are cultivated in the South (1200mm) and Est of Senegal (700mm) (Fig 7).

INTROGRESSIONS BETWEEN WILD AND CULTIVATED

Bayesian approach analysis using wild (51 individus/population from West Africa) and cultivated (357 accessions) revealed that some cultivated varieties incorporate up to 40% of the wild genomes, indicating that gene flow is detected in Senegalese germplasm (Fig.10).

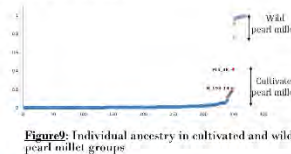


Figure 9: Individual ancestry in cultivated and wild pearl millet groups

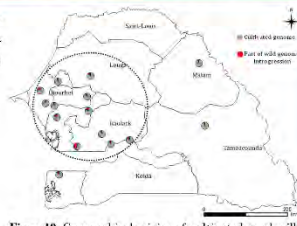


Figure 10: Geographical origins of cultivated pearl millet that have at least 5% of the wild genome introgressions

CONCLUSION

- The genetic diversity of Senegalese germplasm ($H_{ex}=0.54$) contribute the west Africa and worldwide millets diversity. There are two groups of pearl millet in Senegal according to their genetic background as well as their geographical repartition. This could help better management and uses of the germplasm in pearl millet selection/improvement.
- There is an existence of introgression between wild and cultivated millet most abundant in the West Centre of Senegal.
- This study provided the first assessment of the genetic diversity of millet in Senegal.

PERSPECTIVE

In prospect, we aim to assess the importance of wild-to-crop introgressions as potential source of adaptations in cultivated pearl millet.

We assume that wild genes conferring adaptations to harsher conditions have introgress into cultivated populations in regard to ongoing climate changes and/or anthropic selection.

References:

1. Cloutault, J., et al. (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.
2. Oumar, I., et al. (2008). Phylogeny and origin of pearl millet (*Pennisetum glaucum* [L.] R. Br.) as revealed by microsatellite loci. *TAG Theor. Appl. Genet. Theor. Angew. Genet.* 117, 489–497.
3. Mariac, C., et al. (2006). Diversity of wild and cultivated pearl millet accessions (*Pennisetum glaucum* [L.] R. Br.) in Niger assessed by microsatellite markers. *TAG Theor. Appl. Genet. Theor. Angew. Genet.* 114, 49–58.
4. Dussert, Y., et al. (2015). Inference of domestication history and differentiation between early- and late-flowering varieties in pearl millet. *Mol. Ecol.* 24, 1387–1402.



Acknowledgements: This study is supported by

-WAAPP mil projet funding under the PPAO/WAAPP programme in Senegal

-Agropolis Fondation under the reference ID 1403-057 through the «Investissement d'Avenir» Programme (Labex Agro:ANR-10-LABX-0001-01)

