

Mémoire de fin d'études

Présenté pour l'obtention du diplôme « **Master 2** »

Mention : Sciences et technologies de l'agriculture, de l'alimentation et de l'environnement

Spécialité : Sélection et Evolution des Plantes Méditerranéennes et Tropicales

Caractérisation et évaluation des prédictions génomiques des lignées de sorgho développées par le Chibas en Haïti.



Par Jean-Rigaud CHARLES

Année de soutenance : 2017

Organisme d'accueil : Université Quisqueya et Chibas

Mémoire de fin d'études

Présenté pour l'obtention du diplôme « **Master 2** »

Mention : Sciences et technologies de l'agriculture, de l'alimentation et de l'environnement

Spécialité : Sélection et Evolution des Plantes Méditerranéennes et Tropicales

Caractérisation et évaluation des prédictions génomiques des lignées de sorgho développées par le Chibas en Haïti.



Par Jean Rigaud CHARLES

Année de soutenance : 2017

Mémoire préparé sous la direction
de : Dominique THIS

Présenté le : 14/09/2017

Devant le jury :

Jean-Luc REGNARD

Muriel TAVAUD

David POT

Organisme d'accueil : Université
Quisqueya et Chibas

Maître de stage : Gaël PRESSOIR

RESUME

Ce travail de recherche s'inscrit dans le cadre d'un partenariat entre le Chibas et les Universités de Quisqueya, du Kansas et de Cornell, qui vise à mettre en place un programme de sélection génomique récurrente en Haïti. Le Chibas a démarré son programme d'amélioration du sorgho sucré en 2010 avec une méthode de sélection récurrente. Dans un premier cycle, les parents sont croisés avec une population mâle stérile et les meilleures familles ont été sélectionnées pour être inter-croisées dans un deuxième cycle. Ainsi, les lignées sous étude sont le résultat de ce programme de sélection récurrente. L'objectif de cette étude est d'analyser la variabilité génétique des lignées de sorgho développées par le Chibas pour voir s'il y a suffisamment de diversité pour faire de la sélection et si elles peuvent être utilisées comme population d'entraînement dans un programme de sélection génomique suite au goulot d'étranglement provoqué par un nouveau biotype de *Melanaphis sacchari* identifié en Haïti en 2015. Dans cette étude, 1660 marqueurs SNPs sont utilisés à partir du génotypage de 1283 lignées de sorgho suivant la technique de génotypage par séquençage. Pour les analyses de structuration, les méthodes suivantes sont utilisées : logiciel Structure, AFC, DAPC. Le package hierFstat est utilisé pour les calculs basiques en génétique. Les méthodes Fst outliers et outFlank sont utilisées pour la détection des traces de sélection. Et, pour la prédiction génomique, le modèle GBLUP est utilisé. Les résultats de l'AFC montrent que les lignées du Chibas sont moins éloignées du groupe Caudatum de l'Afrique de l'Est que tous les autres groupes du panel de diversité du sorgho. Ce résultat prouve qu'il y a une forte contribution des Caudatum de l'Afrique de l'Est à ce programme. Quant aux résultats des calculs de diversité, la population du Chibas a un indice de Nei égal à 0.11 qui est inférieur aux indices du groupe Caudatum de l'Afrique de l'Est (0.14) et du panel de diversité (0.25). Des traces de sélection ont été identifiées sur les chromosomes 6 et 10 au sein des lignées de sorgho du Chibas. Cette région du chromosome 6 a déjà été identifiée comme étant une région conférant la résistance aux pucerons du sorgho (*Melanaphis sacchari*). Pour la prédiction génomique, les niveaux de précision pour les caractères hauteur, degré brix et rendement grain sont respectivement de 50 %, 60 % et 5 %. La sélection génomique est réalisable avec les lignées de sorgho développées par le Chibas car la diversité de ces lignées n'est pas si faible et on obtient de bonnes prédictions génomiques pour les caractères héréditaires.

Mots clés : *Melanaphis sacchari*, GBLUP, Sorgho, prédiction génomique, diversité génétique, sélection génomique récurrente, programme de sélection.

ABSTRACT

This research is part of a partnership between Chibas and the universities of Quisqueya, Kansas and Cornell, which aims to establish a program of recurrent genomic selection in Haiti. Chibas started its sweet sorghum improvement program in 2010 with a recurrent selection method. In the first cycle, the parents were crossed with a male sterile population and the best families were selected to be intercrossed in a second cycle. The inbred lines used are the result of this recurrent selection program. The aim of this work is to analyze the genetic variability of the lines developed by Chibas to see if there is enough diversity to make the selection and if they can be used as training population for a selection program genomic after the bottleneck provokes by a new biotype of *Melanaphis sacchari* detected in 2015 in Haiti. In this study, 1660 SNPs are used by genotyping 1283 lines of sorghum with genotyping by sequencing methodology. For population structure analyze, the following methods are used: Structure software, AFC and DAPC. HierFstat package was used for basic genetic calculations. Fst outliers and outFlank methods were used to detect selection signature. For genomic prediction, GBLUP model was used. AFC results show that Chibas lines are less distant to the East African Caudatum group than to any other groups in the sorghum diversity panel. This result shows that the parents of Chibas' breeding program come from the East African Caudatum group. Chibas lines have a Nei index equal to **0.11** which is lower than the East African Caudatum (**0.14**) and that of the diversity panel (**0.25**). Areas of genus under selection analysis identified traces of selection on chromosomes 6 and 10 within the Chibas sorghum lines. This region of chromosome 6 has already been identified as a region conferring resistance to *Melanaphis sacchari*. For genomic prediction with GBLUP model, brix, height and grain yield are used. With GBLUP, precision levels for height, degree of brix and grain yield are 50 %, 60 % and 5 %, respectively. This low level of prediction for grain yield is explained by the low heritability of this trait. Genomic selection is feasible with sorghum lines of Chibas because the diversity of these lines is not so low and good genomic predictions are obtained for heritable traits.

Keywords: *Melanaphis sacchari*, GBLUP, Sorghum, genomic prediction, genetic diversity, genomic recurrent selection, selection program.

REMERCIEMENTS

Je tiens à remercier Gaël PRESSOIR, directeur du Chibas et qui est aussi mon maître de stage, pour son support et ses conseils dans la réalisation de ce travail.

Je tiens également à remercier madame Dominique THIS, ma tutrice et professeure à Supagro Montpellier, qui a assuré le bon déroulement du stage et qui m'a orienté sur certains points.

Un grand remerciement aux employés du Chibas, Docteur Raphael Wesly COLBERT, Raoul AMISIAL, Dorval Marie Darline, Joseph Chrisnel Alciné, Joseph Diana, Jean Wisguen Inozile, Durosiers Renaud, pour leur collaboration à la mise en place des essais expérimentaux et aussi à la collecte des données de phénotypage.

Un remerciement très spécial aux staffs des Universités du Kansas et de Cornell pour leurs accompagnements au cours de ce stage et pour le génotypage des individus. Geoffrey Morris, Sarah Jensen, ED Buckler et aux techniciens à Kansas State qui ont réalisé le génotypage par séquençage. Merci à Kebede Muleta pour le travail préliminaire de bio-informatique et pour avoir extrait les marqueurs pour cette étude

Je tiens aussi à remercier Ingrid VILMUS, future professeure à Supagro Montpellier, pour m'avoir éclairci sur certaines parties de ce travail.

Je tiens à remercier les projets suivants : le projet AKOSAA réalisé en partenariat avec l'Université Laval qui a financé le développement de lignées résistantes à *Melanaphis sacchari*; le projet S3F réalisé en partenariat avec le CIRAD qui a permis le démarrage du programme d'amélioration au Chibas; et enfin le projet USAID-SMIL en partenariat avec l'Université du Kansas et Cornell qui finance le programme de sélection génomique à Chibas.

Je remercie mes parents pour leurs supports et leurs encadrements tout au long de mon parcours quotidien. Un grand remerciement à mes amis, mes collègues et toutes les personnes qui ont contribué à la réussite de ce travail.

Je remercie ma chérie, Samantha PIERRE, et ma belle-mère, Claire Altagrâce PIERRE pour leurs supports et leurs encouragements.

Un remerciement de façon très spéciale au staff de la formation APIMET-SEPMET de Supagro Montpellier pour cette formation de haute qualité offerte aux jeunes.

This study is made possible by the support of the American People provided to the Feed the Future Innovation Lab for Collaborative Research on Sorghum and Millet through the United States Agency for International Development (USAID). The contents are the sole responsibility of the authors and do not necessarily reflect the views of USAID or the United States Government. Program activities are funded by the United States Agency for International Development (USAID) under Cooperative Agreement No. AID-OAA-LA-16-00003.

DEDICACE

Je dédie ce travail à une personne qui m'est très chère, **Lili VALÈRE**, ma maman qui est décédée durant ma première année en Master.

SOMMAIRE

RESUME	i
ABSTRACT	ii
REMERCIEMENTS	iii
DEDICACE	iv
SOMMAIRE	v
GLOSSAIRE	viii
LISTE DES TABLEAUX	ix
LISTE DES FIGURES	x
1. INTRODUCTION	1
1.1. Contexte et problématique	1
1.2. Objectif principal	2
1.3. Questions de recherche	2
2. REVUE DE LITTERATURE	3
2.1. Importance du sorgho dans le monde	3
2.2. Génétique et génomique du sorgho (<i>Sorghum bicolor</i>)	3
2.2.1. Le déséquilibre de liaison (DL) chez le sorgho	4
2.2.2. Déterminisme génétique du rendement grain chez le sorgho	5
2.2.3. Déterminisme génétique de la hauteur chez le sorgho	5
2.2.4. Déterminisme génétique du pourcentage de brix chez le sorgho	6
2.2.5. Déterminisme génétique de la résistance au puceron (<i>Melanaphis sacchari</i>).....	6
2.3. Diversité génétique du sorgho dans le monde	7
2.4. Structuration génétique du sorgho	8
2.5. Amélioration du sorgho et les critères de sélection	9
2.5.1. La sélection généalogique	9
2.5.2. Le rétrocroisement.....	10
2.5.3. La sélection récurrente	10
2.5.4. La sélection assistée par marqueurs	11
2.5.5. La sélection génomique	11
2.6. Progrès génétique et sélection génomique	12

2.7.	Diversité génétique et sélection génomique	12
2.8.	Les limites de la sélection génomique	13
2.8.1.	Héritabilité du phénotype	13
2.8.2.	Composition de la population de référence	13
2.8.3.	Densité de marquage	14
2.8.4.	Interaction Génotype x Environnement (G*E).....	14
2.8.5.	Performance du modèle de prédiction en sélection génomique	14
2.9.	BLUP génomique	15
2.10.	Objectifs spécifiques de ce travail de recherche.....	17
3.	MATERIELS ET METHODES.....	18
3.1.	Matériel biologique (espèce Sorghum bicolor).....	18
3.2.	Origine des lignées de sorgho du Chibas	18
3.3.	Dispositif expérimental	19
3.4.	Itinéraires techniques adoptés.....	19
3.5.	Extraction d'ADN	20
3.6.	Technique de séquençage.....	20
3.7.	Méthode 1 : Structuration et diversité génétique chez les populations de sorgho ...	20
3.7.1.	Analyse Factorielle des Correspondances.....	20
3.7.2.	Logiciel Structure de Pritchard	21
3.7.3.	Calcul des statistiques de base.....	21
3.7.3.1.	Hétérozygotie observée (Ho)	21
3.7.3.2.	Indice de diversité génétique de Nei (Hs)	21
3.7.3.3.	Indice de fixation (F_{IS})	21
3.7.3.4.	Différenciation entre les groupes (F_{ST}).....	21
3.8.	Méthode 2 : Détection de la signature de la sélection	22
3.9.	Méthode 3 : Protocole d'utilisation des modèles de prédiction	22
3.9.1.	Population de référence.....	23
3.9.2.	Analyse discriminante en composantes principales	23
3.9.3.	Matrice d'apparentement.....	24
3.9.4.	Modèle de prédiction	24
3.9.5.	Densité de marquage	24

3.10.	Paramètres phénotypiques évalués.....	24
3.10.1.	Détermination du rendement grain.....	24
3.10.2.	Détermination du degré brix.....	25
3.10.3.	Évaluation de la hauteur des lignées.....	25
4.	RESULTATS.....	26
4.1.	Caractérisation des lignées de sorgho développées par le Chibas.....	26
4.1.1.	Analyse Factorielle des Correspondances.....	26
4.1.2.	Structuration avec le logiciel Structure de Pritchard.....	27
4.2.	Paramètres de diversité génétique.....	28
4.3.	Détection des traces de sélection.....	28
4.4.	Prédiction génomique.....	30
4.4.1.	Présentation des traits phénotypiques utilisés dans la prédiction.....	30
4.4.2.	Corrélation entre les traits phénotypiques utilisés dans la prédiction.....	30
4.4.3.	Distribution des lignées du Chibas en fonction de leur apparentement.....	31
4.4.4.	Identification des individus très apparentés.....	31
4.4.5.	Modèle de prédiction GBLUP.....	32
5.	DISCUSSIONS.....	33
5.1.	Structuration des génotypes de sorgho.....	33
5.2.	Paramètre de diversité.....	33
5.3.	Détection des traces de sélection.....	34
5.4.	Distribution des individus de la population d'entraînement en fonction des relations d'apparentement.....	34
5.5.	Modèle GBLUP.....	35
5.6.	Densité de marquage.....	36
6.	CONCLUSION ET PERSPECTIVES.....	37
	BIBLIOGRAPHIE.....	39
	ANNEXE.....	i

GLOSSAIRE

BLUP : Best Linear Unbiased Prediction - Meilleur prédicteur linéaire non biaisé

GBLUP: Genomic Best Linear Unbiased Prediction - Meilleur prédicteur linéaire non biaisé basé sur les marqueurs moléculaires.

SG : Sélection génomique

DL : Déséquilibre de liaison

AFC : Analyse Factorielle des Correspondances

CAE : Caudatum de l'Afrique de l'Est

DAPC : Analyse discriminante en composantes principales

SAM: Sélection assistée par marqueurs

CIRAD: Centre de coopération internationale en recherche agronomique pour le développement

PG: Prédiction génomique

SNP: Marqueur « *Single-nucleotide polymorphism* »

ABLUP: Meilleur prédicteur linéaire non biaisé basé sur le pedigree.

GBS: Genotyping By Sequencing - Génotypage par séquençage

ICRISAT: International Crops Research Institute for the Semi-Arid Tropics - Institut international de recherche sur les cultures des zones tropicales semi-arides

Rdt : rendement

G * E : Interaction génotype environnement

LISTE DES TABLEAUX

TABLEAU 1: LES CINQ CEREALES LES PLUS IMPORTANTES EN TERMES DE PRODUCTION ET DE SUPERFICIE	3
TABLEAU 2 : VARIATION DU DL (R^2) A TRAVERS LES 10 CHROMOSOMES DU SORGHO SUIVANT LA VARIATION DU NOMBRE DE PAIRES DE BASE (Y.-H. WANG ET AL., 2013).	4
TABLEAU 3: VARIATION GENETIQUE AU SEIN DES RACES DE SORGHO (DEU ET AL., 2006)	8
TABLEAU 4: VARIETES EXOTIQUES RESISTANTES A <i>MELANAPHIS</i> FAISANT PARTIE DE LA COLLECTION DU CHIBAS	19
TABLEAU 5: MESURES DE DIVERSITE POUR LES LIGNEES DU CHIBAS, DU GROUPE CAE ET DU PANEL DE DIVERSITE MONDIALE DU SORGHO	28
TABLEAU 6 : VALEURS DE F_{ST} PAR PAIRES OBTENUES ENTRE LES TROIS GROUPES	28
TABLEAU 7 : LOCUS SOUS SELECTION ENTRE LIGNEES DU CHIBAS/CAE ET LIGNEES DU CHIBAS/PANEL DIVERSITE	29
TABLEAU 8 : CORRELATION ENTRE LES TRAIT PHENOTYPIQUES	30
TABLEAU 9: COMPARAISON DES NIVEAUX DE PRECISION (R) DE DEUX STRATEGIES D'ECHANTILLONNAGE SUR GBLUP	32
TABLEAU 10: LES TRAIT EVALUES PENDANT LE PHENOTYPAGE	I
TABLEAU 11: PANEL DE DIVERSITE DE SORGHO UTILISE DANS CETTE ETUDE.....	IV

LISTE DES FIGURES

FIGURE 1: DIMINUTION DU DL EN FONCTION DE LA DISTANCE (KBP) (MACE ET AL., 2013).	5
FIGURE 2: EVOLUTION DES APHIDES SUR UNE VARIETE DE SORGHO RESISTANTE (HN16) ET UNE VARIETE DE SORGHO SENSIBLE (BTx623) ET POSITION DU GENE RMSE1 SUR LE CHROMOSOME 6 (F. WANG ET AL., 2013).....	7
FIGURE 3 : ONZE SOUS POPULATIONS DE SORGHO DETECTEES PAR UNE ANALYSE DE STRUCTURE (Y.-H. WANG ET AL., 2013).	9
FIGURE 4 : SELECTION RECURRENTTE SUR DESCENDANCE (GALLAIS, 2011)	10
FIGURE 5 : DIAGRAMME DE LA SELECTION GENOMIQUE (HEFFNER ET AL., 2009).	12
FIGURE 6: DIFFERENCE ENTRE LES MODELES BAYES B ET GBLUP POUR DIFFERENTS NIVEAUX D'HERITABILITE ET POUR UNE VARIATION DU NOMBRE DE QTL (DAETWYLER ET AL., 2010).	15
FIGURE 7: PRESENTATION DE LA MATRICE A ET DE LA MATRICE G (CLARK ET VAN DER WERF, 2013).	16
FIGURE 8 : STRUCTURATION DES DIFFERENTS GROUPES DE SORGHO VIA L'ANALYSE FACTORIELLE DES CORRESPONDANCES ...	26
FIGURE 9 : RESULTATS DE L'ANALYSE STRUCTURE DE PRITCHARD (LENGHT OF BURNING EGAL A 5000 ET NOMBRE DE REPETITIONS EGAL A 5000).	27
FIGURE 10 : (A) LOCUS SOUS SELECTION AU SEIN DES LIGNEES DU CHIBAS ET LE GROUPE CAUDATUM DE L'AFRIQUE DE L'EST. (B) LOCUS SOUS SELECTION AU SEIN DES LIGNEES DU CHIBAS ET LE PANEL DE DIVERSITE.	29
FIGURE 11 : DISTRIBUTION DES INDIVIDUS EN FONCTION DES TRAITS PHENOTYPIQUES	30
FIGURE 12 : DISTRIBUTION DES LIGNEES DE SORGHO DU CHIBAS EN FONCTION DE LEUR APPARENTEMENT	31
FIGURE 13 : STRUCTURATION DES INDIVIDUS DE LA POPULATION D'ENTRAINEMENT AVEC DAPC	31
FIGURE 14: CORRELATION ENTRE RENDEMENT GRAIN OBSERVE ET RENDEMENT GRAIN PREDIT	II
FIGURE 15 : CORRELATION ENTRE BRIX OBSERVE ET BRIX PREDIT	II
FIGURE 16 : CORRELATION ENTRE HAUTEUR OBSERVEE ET HAUTEUR PREDITE	II
FIGURE 17 : REPARTITION DES ZONES DE CULTURE DU SORGHO EN HAÏTI.....	II
FIGURE 18: PLAN DU DISPOSITIF EXPERIMENTAL.....	III

1. INTRODUCTION

1.1. Contexte et problématique

Le sorgho est la cinquième culture céréalière la plus cultivée dans le monde derrière le maïs, le riz, le blé et l'orge avec une production de 69 millions de tonnes. En Haïti, le sorgho est la troisième céréale la plus cultivée, derrière le maïs (320 747 tonnes) et le riz paddy (143 650 tonnes) avec une production estimée à 122 500 tonnes cultivée sur une superficie de 137 872 hectares (FAOSTAT, 2014). Le sorgho est la principale céréale des zones sèches d'Haïti (Figure 17). Il est cultivé soit en monoculture soit en association avec le maïs ou/et avec le pois d'Angole. C'est une céréale qui permet aux agriculteurs de faire la jonction entre la récolte de Décembre-Janvier au milieu de la saison sèche (période de récolte du sorgho et du pois d'Angole) et les premières récoltes de printemps en Mai. La période comprise entre ces deux récoltes est la période de soudure où les chefs de ménages n'ont que les grains de sorgho pour nourrir leurs foyers et les pailles de sorgho pour nourrir les animaux. De ce fait, on peut constater que le sorgho est une culture d'importance en termes de sécurité alimentaire en Haïti. Cependant, la culture de sorgho fait face à de nombreuses contraintes qui sont les conditions edapho climatiques, les variations climatiques et, depuis l'an 2015, les attaques d'une nouvelle peste (*Melanaphis sacchari*).

Melanaphis sacchari est un puceron qui appartient à la famille des aphides. Un nouveau biotype de *Melanaphis* a été constaté en Amérique du Nord à partir de l'année 2013 et a occasionné beaucoup de dégâts à la culture de sorgho (Bowling et al., 2016). Les premières infestations de *Melanaphis* ont été observées dans les champs de sorgho d'Haïti durant l'année 2015. En 2016, la production du sorgho a accusé une perte de 70 % à cause des attaques de *Melanaphis sacchari*. Cette baisse de production affecte non seulement les ménages mais aussi les entreprises n'avaient pas suffisamment d'intrants qui sont les grains de sorgho pour la production, ce qui les obligeaient à importer les grains de sorgho. Pendant, cette catastrophe en 2016, le sorgho a failli disparaître en Haïti. Heureusement, un centre de recherche agricole connu sous le nom de Chibas a su répondre en développant des lignées de sorgho résistantes à ce puceron. La majorité de ces lignées de sorgho a été créée par un schéma de sélection récurrente qui a débuté en 2012. Dans ce schéma de sélection, deux cycles sont effectués. Un premier cycle où les parents sont croisés avec une population ségréguée pour la stérilité mâle et ensuite les meilleures familles S1 ont été sélectionnées. Dans un deuxième cycle, ces meilleures familles ont été inter croisées pour créer des familles de pleins frères. Ce schéma de sélection récurrente est en cours d'évolution vers un schéma de sélection génomique récurrente dans le but de réduire l'intervalle entre les

générations et d'avoir une population efficace de taille plus importante afin d'augmenter les gains génétiques attendus.

Dans le but de caractériser la population de lignées développées par le Chibas avant le démarrage du programme de sélection génomique, nous avons souhaité (1) de caractériser la diversité et la structuration de cette diversité au sein de la population du Chibas par rapport à un panel représentatif de la diversité du sorgho ; (2) de tester la présence de locus sous balayage sélectif par rapport au panel de diversité et aux populations les plus proches des lignées du Chibas ; et enfin (3) de valider les lignées développées jusqu'à date par le Chibas comme population d'entraînement pour le programme de sélection génomique en évaluant la précision des prédictions génomiques réalisées. Ce travail est réalisé pour trois caractères au déterminisme génétique contrasté qui sont le brix ou teneur en sucre des tiges, la hauteur des plantes et le rendement grain. Ce travail de recherche s'inscrit dans le cadre d'un partenariat entre le Chibas et les Universités de Quisqueya, du Kansas et de Cornell, qui vise à mettre en place un programme de sélection génomique récurrente en Haïti.

1.2. Objectif principal

- ✓ Caractériser les lignées de sorgho développées par le Chibas pour l'établissement d'un programme de sélection génomique récurrente.

1.3. Questions de recherche

- ✓ Existe-t-il suffisamment de diversité génétique au sein des lignées de sorgho du Chibas ?
- ✓ Est-ce qu'il y a des locus sous sélection pour la résistance à *Melanaphis sacchari* ?
- ✓ Les lignées de sorgho développées par le Chibas permettent-elles d'avoir de bonnes prédictions génomiques ?

2. REVUE DE LITTÉRATURE

La revue littérature se focalise sur l'importance du sorgho dans le monde, sa diversité, les déterminismes génétiques de quelques caractères, les critères et les méthodes de sélection du sorgho notamment la sélection génomique.

2.1. Importance du sorgho dans le monde

Le sorgho est la cinquième plus importante céréale dans le monde, qu'il s'agisse de production ou de superficie, avec une production de 69 millions de tonnes par an cultivée sur une superficie estimée à environ 45 millions d'hectares (FAOSTAT, 2014). Les USA, le Mexique, l'Argentine et l'Australie sont reconnus comme étant des pays qui utilisent des hybrides de sorgho dans des conditions favorables et dont le rendement grain à l'hectare est très élevé (Rooney et al., 2007). En termes de superficies, les trois grands pays producteurs sont le Soudan avec une superficie de 7.1 Mha, l'Inde avec une superficie de 6.2 Mha et le Nigeria avec une superficie de 5.5 Mha. Par rapport au volume de production, les cinq plus grands producteurs sont les USA avec une production de 11 Mt, le Mexique avec une production de 8.4 Mt, le Nigeria avec une production de 6.7 Mt, le Soudan avec une production de 6.3 Mt et l'Inde avec une production de 5.4 Mt (FAOSTAT, 2014).

Tableau 1: Les cinq céréales les plus importantes en termes de production et de superficie

	Production mondiale (Mt)	Superficie mondiale (Mha)
Maïs	1038	185
Riz Paddy	741	163
Blé	729	220
Orge	145	49
Sorgho	69	45

Source: FAOSTAT 2014

2.2. Génétique et génomique du sorgho (*Sorghum bicolor*)

Le sorgho est une espèce diploïde avec 10 chromosomes ($2n = 20$) dont la taille du génome est d'environ 730 méga bases. Environ 34 496 gènes sont identifiés chez cette espèce qui est une espèce modèle en génomique fonctionnelle pour étudier les plantes fourragères C4 (Paterson, 2009). Le sorgho est une espèce monoïque qui est autogame dont le taux d'allogamie est d'environ 6 % (Doggett, 1988). Cependant, certaines variétés totalement cléistogames ont un taux d'allogamie de 0 % et d'autres ont des taux d'allogamie de plus de 30 % (Charrier et al., 1997).

Le génome du sorgho a été reséquencé avec une couverture de 16-45 x par Mace et al., (2013) en utilisant des marqueurs SNPs à partir de 44 lignées qui représentent le groupe génétique primaire du sorgho.

Les résultats de l'étude de Mace (2013) et d'autres études précédentes ont montré qu'au niveau des régions codantes du sorgho le nombre de mutation non synonyme (112 255) est plus faible que le nombre de mutation synonyme (112 108). Ainsi, le ratio DN/DS, étant un critère pour détecter des traces de sélection, est sensiblement égal à 1 au niveau des régions codantes du sorgho. Ce qui prouve que des zones au niveau du génome du sorgho sont soumises à la sélection puisque la valeur obtenue est supérieure à 1 (McDonald and Kreitman, 1991). Sur l'ensemble du génome du sorgho, le nombre de SNPs est plus élevé chez les espèces sauvages de sorgho (3 465 947) que chez les variétés traditionnelles (3 092 165) et les variétés améliorées (2 284 285).

2.2.1. Le déséquilibre de liaison (DL) chez le sorgho

Le déséquilibre de liaison chez le sorgho a été étudié par Wang et al., en 2013 en utilisant 13 390 marqueurs SNPs après avoir génotypé une collection de 242 variétés locales de sorgho de l'ICRISAT. Cette collection contient les cinq principales races de sorgho. Ainsi, les résultats de cette étude ont montré que la valeur moyenne de déséquilibre de liaison varie de 0.04 à 0.06 pour les dix chromosomes du sorgho. Sur le chromosome 2, on observe une augmentation du DL entre 20-30 kb et sur le chromosome 4 une augmentation du DL entre 30-40 kb. Cependant, on observe une augmentation de r^2 sur le chromosome 6 au-delà de 10 Mb (Tableau 2). Les résultats de l'étude de Mace et al., en 2013 ont montré qu'il y a une augmentation du DL moyen chez les variétés améliorées par rapport aux espèces sauvages et aux variétés traditionnelles. De plus, ils ont montré qu'au-delà de 19.7 kbp et 10.3 kbp, la valeur du DL diminue de moitié respectivement pour des variétés améliorées et des variétés locales. Suite à ces résultats, pour une couverture complète du génome du sorgho, on estime qu'il faut au minimum 37 055 marqueurs SNPs pour des variétés améliorées.

Tableau 2 : variation du DL (r^2) à travers les 10 chromosomes du sorgho suivant la variation du nombre de paires de base (Y.-H. Wang et al., 2013).

	SBI-01	SBI-02	SBI-03	SBI-04	SBI-05	SBI-06	SBI-07	SBI-08	SBI-09	SBI-10
<1 kb	0.51	0.37	0.5	0.53	0.43	0.49	0.5	0.44	0.54	0.51
1–10 kb	0.15	0.11	0.12	0.14	0.15	0.14	0.14	0.16	0.22	0.11
10–20 kb	0.1	0.1	0.07	0.11	0.08	0.11	0.11	0.05	0.11	0.1
20–30 kb	0.08	0.15	0.08	0.09	0.07	0.09	0.08	0.1	0.08	0.08
30–50 kb	0.07	0.09	0.08	0.14	0.04	0.1	0.08	0.09	0.08	0.07
50–100 kb	0.07	0.08	0.06	0.07	0.06	0.07	0.06	0.05	0.05	0.05
100 kb–1 Mb	0.04	0.04	0.03	0.04	0.03	0.05	0.04	0.03	0.04	0.03
1–10 Mb	0.03	0.03	0.03	0.03	0.02	0.03	0.03	0.03	0.03	0.03
>10 Mb	0.03	0.03	0.02	0.05	0.04	0.08	0.03	0.06	0.03	0.05
Average	0.05	0.05	0.04	0.05	0.04	0.06	0.05	0.04	0.05	0.05

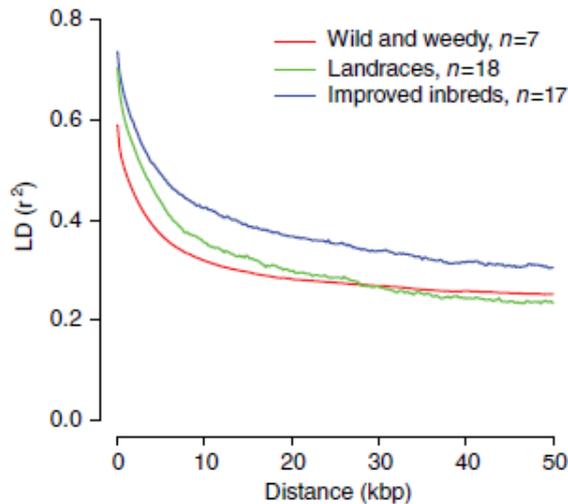


Figure 1: diminution du DL en fonction de la distance (kbp) (Mace et al., 2013).

2.2.2. Déterminisme génétique du rendement grain chez le sorgho

Le rendement grain est un caractère complexe polygénique qui est sous le contrôle de plusieurs gènes. Pour illustration, le travail de Ritter et al., 2008 a identifié un grand nombre de QTLs qui affectent le rendement grain. L'héritabilité observée dans une population sous sélection peut être faible et varier considérablement d'un environnement à un autre. Par exemple, Naim et al., (2012) ont mesuré une héritabilité au sens large pour le rendement grain de 3 % pendant une première année et de 8 % l'année suivante (El Naim et al., 2012). Cette faible héritabilité du rendement grain est expliquée par le fait que c'est un caractère complexe ayant plusieurs composantes qui sont : le nombre de grains par panicule, le rendement grain par plante, le poids des panicules, le poids de 1000 grains. En effet, chacune de ces composantes peut être sous le contrôle de nombreux gènes mais également sous l'influence de nombreux facteurs environnementaux. Dans l'équation du modèle génétique : $Phénotype = Génotype + Environnement$, la variance phénotypique dépend beaucoup plus de la variance environnementale que de la variance génétique pour le rendement grain. Ce qui se traduit, en effet, par une faible héritabilité de ce caractère.

2.2.3. Déterminisme génétique de la hauteur chez le sorgho

La hauteur du sorgho est influencée par 4 gènes majeurs (dw1, dw2, dw3 et dw4) qui ont un effet brachytique sur la plante en réduisant la longueur des entre-nœuds (ICRISAT, 1995). Quinby et Karper (1954) ont montré que des variétés de sorgho dont le génotype est récessif pour un gène (par ex. Dw1Dw2Dw3dw4) ont une taille plus importante que celles dont le génotype est récessif pour 2 à 3 gènes (par ex. Dw1Dw2dw3dw4, dw1dw2Dw3dw4). Il existe d'autres gènes qui agissent non seulement sur la longueur des tiges mais également sur la longueur du pédoncule, la

longueur de la panicule et le nombre de feuilles. Le gène Dw1 est localisé sur le chromosome 9 à 57 Mbp, le gène Dw2 est localisé sur le chromosome 6 à 42 Mbp, le gène Dw3 est localisé sur le chromosome 9. La position du gène Dw4 sur le génome du sorgho n'est pas bien définie. En effet, il est localisé soit au niveau du chromosome 6 à 6 Mbp soit au niveau du chromosome 4 à 67 Mbp (Hilley, 2017). Le gène Dw3 a un effet pléiotropique en agissant sur le nombre de grains par panicule, le poids des panicules et la taille des panicules (Cassady, 1965). Cependant, le gène Dw2 semble avoir un effet sur la longueur de la panicule, le poids des graines, la surface des feuilles (Graham et Lessman, 1966). En revanche, des effets pléiotropiques n'ont pas été détectés jusqu'à présent pour les gènes Dw1 et Dw4. D'après Rami et al., 1998) et Ritter et al., 2007), la hauteur est un caractère phénotypique chez le sorgho qui est très héritable (95 %).

Pour illustrer l'influence d'autres régions du génome c'est-à-dire des QTLs en plus des gènes majeurs décrits plus haut, on peut citer les résultats du travail de Ritter et al., 2007, qui à partir de 228 marqueurs microsatellites et AFLP, ont identifié quatre régions au niveau du génome du sorgho qui affectent la hauteur. Ces quatre régions se trouvent sur les chromosomes : 1, 3, 5 et 6. Les régions du chromosome 5 et 6 viennent d'un parent nain et les régions du chromosome 1 et 3 viennent d'un parent élite pour le rendement grain (Ritter et al., 2007).

2.2.4. Déterminisme génétique du pourcentage de brix chez le sorgho

Le brix est un caractère très héritable ($h^2 = 0.84$) qui est sous le contrôle de plusieurs gènes (Ritter et al., 2007). Les résultats de l'étude de Ritter et al., en 2007 ont identifiés 7 marqueurs qui ont un impact significatif sur la teneur en brix. Parmi ces marqueurs, deux marqueurs dans le groupe de liaison 5 au niveau du chromosome 5 qui sont localisés à 0 centimorgan et à 149.2 centimorgan et 5 marqueurs dans le groupe de liaison 6 sur le chromosome 6 qui sont localisés à 0, 7.4, 14.4, 14.9 et 15.4 centimorgan.

2.2.5. Déterminisme génétique de la résistance au puceron (*Melanaphis sacchari*)

Il existe trois types de résistance au puceron qui sont : l'antixenose, l'antibiose et la tolérance (Fartek, 2011). L'antixenose est un type de résistance par lequel la plante met en place des mécanismes qui empêchent la colonisation de l'insecte. L'antibiose est un autre type de résistance par lequel la plante secrète des substances biochimiques qui affectent de façon négative la survie, la croissance et la reproduction de l'insecte. Le dernier type de résistance est la tolérance où la plante n'est pas affectée par le développement et la colonisation de l'insecte. Le criblage de 462 lignées de sorgho a été effectué dans une étude publiée par Teetes et al., en 2017 dans le but d'identifier des lignées qui sont résistantes à *Melanaphis sacchari*. Parmi ces 462 lignées, 32 ont montré une forte résistance à ce puceron et les mécanismes de cette résistance

pour ces lignées résistantes se sont révélés être à la fois antixenose et antibiose. Les résultats de l'étude de Wang et al., (2013) ont pu montrer à partir des marqueurs moléculaires microsatellites, l'existence du gène dominant RMES1 au niveau du petit bras du chromosome 6 qui confère une résistance à *Melanaphis sacchari* qui n'est pas le même biotype que celui présent en Amérique du Nord et dans la Caraïbe. Dans cette étude, les descendants du croisement d'un parent sensible et d'un parent résistant ont été utilisés pour identifier les régions du génome du sorgho conférant cette résistance. Les descendants pour lesquels l'allèle favorable du gène RMES1 est absent sont sensibles et ceux pour lesquels l'allèle favorable est présent sont résistants.

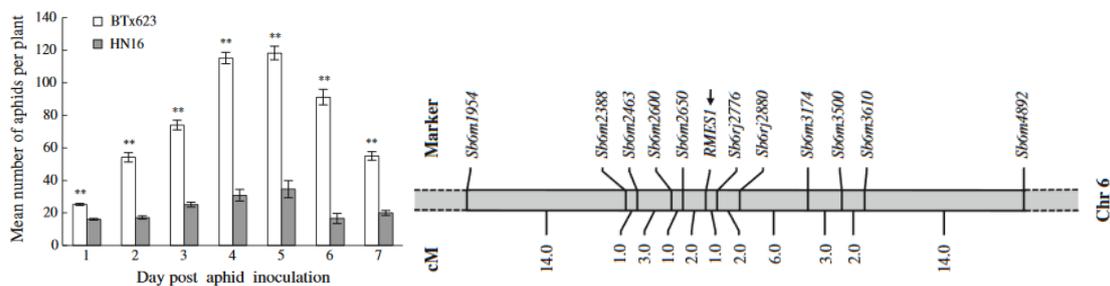


Figure 2: évolution des aphides sur une variété de sorgho résistante (HN16) et une variété de sorgho sensible (BTx623) et position du gène RMSE1 sur le chromosome 6 (F. Wang et al., 2013).

2.3. Diversité génétique du sorgho dans le monde

Cinq races de sorgho sont répertoriées dans le monde ; la race Bicolor, la race Kafir, la race Caudatum, la race Durra et la race Guinea (Harlan et de Wet, 1972). A partir de ces cinq races principales, des combinaisons deux à deux permettent d'obtenir dix races intermédiaires lesquelles ayant des caractères empruntés à plusieurs races principales. Zera-zera est un groupe de sorgho d'origine Éthiopienne et Soudanienne appartenant à la race Caudatum et à la race intermédiaire Caudatum-Guinea. Ce groupe est très utilisé en amélioration des plantes en raison de son rendement grain élevé, de la qualité des grains, de la résistance aux maladies et à la sécheresse (Prasada Rao et Mengesha, 1981). Toutes les lignées parentales de la population du Chibas résistantes à *Melanaphis sacchari* sont de type Zera-zera. Elles avaient été croisées à la population initiale en raison de leur niveau élevé de brix et de la qualité du jus. Un jus qui a une couleur jaune et donne un bon sirop et un alcool savoureux (G Pressoir, communication personnelle).

Il existe environ deux cent mille accessions de sorgho qui sont collectées par différentes institutions dans le monde (Chantereau et al., 2013). En fait, cette grande variabilité génétique du sorgho permet aux sélectionneurs d'exploiter cette gamme de diversité pour aboutir à de meilleurs résultats dans leurs programmes de développement de nouveaux génotypes qui offrent

des réponses aux contraintes environnementales. Les résultats d'une étude de diversité génétique du sorgho réalisée à partir de 45 marqueurs microsatellites (EST-SSR) ont montré que la race bicolor contient le maximum d'allèles, soit 58.89 % de l'ensemble des allèles détectés chez le sorgho, suivi de la race Caudatum (57.78 %) (Ramu et al., 2013). En revanche, L'étude réalisée par (Deu et al., 2006) sur 210 accessions, représentant les cinq races de sorgho en se basant sur des marqueurs RFLP, a montré que c'est la race Guinea qui a le plus grand nombre d'allèles (176) suivie de la race bicolor (175). Cette même étude a montré que la race Caudatum a le pourcentage de locus polymorphe le plus élevé, les races Guinea et bicolor ont le plus grand nombre d'allèles rares et la race bicolor a l'indice de Nei le plus élevé.

Dans un programme de sélection qui a pour objectif de produire des lignées avec de hauts rendements grains, il est intéressant de choisir des parents de la race Caudatum car cette race possède des caractéristiques génétiques reconnus comme favorables à l'augmentation du rendement grain (Mace et al., 2008).

Tableau 3: variation génétique au sein des races de sorgho (Deu et al., 2006)

<i>(a) Variation within races.</i>					
Race (no. of accessions)	% polymorphic loci ^a	Total no. of alleles	No. of rare alleles ^b	Mean no. of alleles/ polymorphic loci ^c	H_{exp} ^d
Bicolor (22)	0.92	175	11	2.38	0.37
Caudatum (43)	0.93	172	4	2.34	0.34
Durra (29)	0.91	169	6	2.30	0.33
Guinea (62)	0.86	176	11	2.38	0.35
Kafir (18)	0.45	110	1	1.49	0.14
Total (205)	0.99	219	31	3.00	0.40

2.4. Structuration génétique du sorgho

Les résultats de l'étude de Wang et al., 2013 ont montré qu'il existe onze sous populations de sorgho à partir de 242 lignées de sorgho de la collection de l'ICRISAT. Des marqueurs SNPs (13 390) ont été sélectionnés pour étudier cette structuration. Cette étude a montré une bonne structuration en fonction des cinq principales races de sorgho et en fonction des pays d'origine. Par exemple, les groupes SG5, SG7, SG8, et SG9 (Figure 3) représentent respectivement les races Caudatum, Guinea, Kafir et Durra. Cependant, une même race peut être structurée en plusieurs groupes suivant ses origines. En effet, les groupes SG9 et SG11 sont constitués par la race Durra provenant de deux régions différentes qui sont les Durra de l'Afrique et les Durra de L'Asie. Les résultats de cette étude ont aussi montré que la race Bicolor est la moins structurée des cinq races de sorgho.

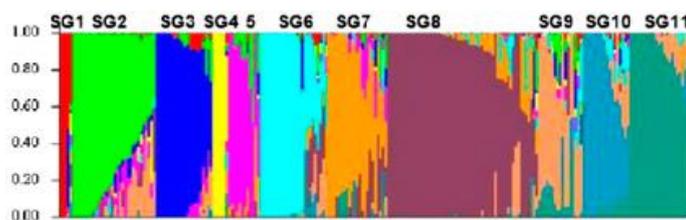


Figure 3 : Onze sous populations de sorgho détectées par une analyse de Structure (Y.-H. Wang et al., 2013).

2.5. Amélioration du sorgho et les critères de sélection

Le sorgho est amélioré en fonction du type de besoins des consommateurs. Par exemple, les critères de sélection pour un sorgho de type biomasse diffèrent de ceux d'un sorgho de type grain ou sucrier. Par exemple, en Haïti, ils se portent principalement sur les sorghos de double ou triple usage avec une sélection simultanée pour le rendement grain et le rendement sucre. A noter que le rendement grain a un poids prépondérant dans l'index de sélection. Le sorgho est surtout amélioré pour le rendement et la qualité des grains, la résistance aux maladies, la hauteur, la précocité, la photo périodicité. Pour l'amélioration du sorgho, plusieurs méthodes de sélection sont adoptées par les sélectionneurs et les agriculteurs dont les plus utilisées sont : La sélection généalogique, la sélection récurrente, la sélection par rétrocroisement (back-cross) et la sélection assistée par marqueur (SAM). Cependant, la sélection génomique (SG) est de plus en plus utilisée dans les pays comme les USA, la France et le Canada car elle permet d'avoir recours au génotypage dont le coût est moins élevé que le coût du phénotypage qui est souvent couteux à cause de la main d'œuvre et de l'utilisation des terres. En fait, le choix de la méthode dépend de l'héritabilité du caractère et de l'objectif du programme de sélection.

2.5.1. La sélection généalogique

La sélection généalogique est une technique basée sur la culture des panicules-lignes. Une fois que le produit du croisement des deux parents est effectué, les graines sont semées pour avoir des hybrides F1. Ensuite, les panicules les plus intéressantes de ces F1 sont choisies et semées séparément sur des lignes. Le processus se répète jusqu'à ce qu'on aboutisse à une homogénéité raisonnable. Par la suite, les lignées produites vont être comparées à des variétés déjà existantes dans le but de vérifier si elles sont supérieures (House, 1987). Certaines lignées développées par le Chibas sont issues d'un croisement entre la meilleure lignée présente en Haïti avant l'infestation par *Melanaphis sacchari* (Papèsèk) et la lignée de type Zera-zera dénommée Dekabès (IS 23572) sélectionnée pour la qualité de son jus (excellent rhum), un rendement grain convenable (2,5 tonne/ha) et un brix élevé (G Pressoir, communication personnelle).

2.5.2. Le rétrocroisement

Le rétrocroisement, backcross en anglais, permet de transférer un caractère monogénique ou Oligo-génique intéressant d'une variété dans une autre variété dans laquelle le caractère représente un défaut (Chantereau et al., 2013). L'un des programmes de rétrocroisement importants pour le sorgho consiste à créer de nouvelles lignées mâles stériles utilisées comme parents femelles dans les hybridations (House, 1987). Cette méthode est utilisée dans le programme d'amélioration du Chibas pour transférer le caractère résistance à *Melanaphis* en croisant la variété améliorée locale qui est sensible (parent récurrent, Papèsèk) avec une variété résistante (Dekabès) afin d'introduire la résistance dominante dans la variété Papèsèk.

2.5.3. La sélection récurrente

La sélection récurrente permet d'avoir une augmentation de la fréquence des gènes désirables au sein d'une même population (Gallais, 2011). Chez les autogames comme le sorgho, le schéma général d'un cycle comprend 2 phases, une phase de sélection et une phase d'inter-croisement (Gallais, 1990). Généralement, la sélection se fait sur les meilleurs génotypes qui vont être intercroisés durant une prochaine campagne. C'est ce schéma de sélection (Figure 4) que le Chibas a adopté pour créer la majorité des lignées de sorgho. Depuis le démarrage du programme en 2012, le Chibas a réalisé 2 cycles complets de sélection. La plupart des lignées résistantes à *Melanaphis* (plus de 90%) développées par le Chibas sont issues de cette stratégie de sélection. Dans le programme du Chibas nous réalisons des familles de plein frère entre les meilleures familles S1 ; les plantes mères (des S1) sont alors sélectionnées au sein des familles de plein frère. Les familles S1 sélectionnées font également l'objet d'autofécondation pour développer les lignées.

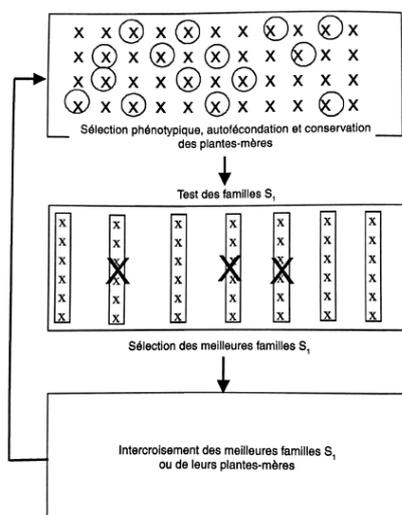


Figure 4 : Sélection récurrente sur descendance (Gallais, 2011)

2.5.4. La sélection assistée par marqueurs

L'utilisation des marqueurs dans les programmes d'amélioration permet d'avoir beaucoup plus de précision dans les programmes d'amélioration variétale en retenant seulement la portion du génome du parent non récurrent qui est intéressant. D'après A. Gallais (2007), la sélection assistée par marqueur est efficace quand il existe un déséquilibre de liaison très important entre le marqueur et le gène d'intérêt en question. Les marqueurs moléculaires peuvent être utilisés pour identifier les allèles favorables et la proportion du génome du parent non récurrent (Najimi et al., 2003). L'identification de marqueurs par GWAS et leur utilisation comme prédicteur est également envisagé dans le schéma du Chibas surtout pour les gènes à effets majeurs.

2.5.5. La sélection génomique

La sélection génomique (SG) est une forme de sélection sur marqueurs seuls utilisée en amélioration des plantes qui vise à utiliser la variation génétique de tous les QTL, et même ceux qui n'ont pas pu être détectés en raison de leur faible contribution à la variance génétique totale (Hayes et al., 2001) ; Bernardo and Yu, 2007). A travers cette méthode de sélection, une relation statistique est établie entre le phénotype et le génotype à partir des marqueurs du génome dans le but de prédire le phénotype à partir des valeurs génétiques. La SG semble plus efficace que les méthodes classiques et notamment la SAM quand la sélection porte sur les traits polygéniques (Endelman, 2011). D'une manière générale, la SG part d'une population d'entraînement qui est phénotypée pour des caractères intéressants et génotypée pour un grand nombre de marqueurs. Puis, un modèle de prédiction est conçu à partir de ces individus qui sont génotypés et phénotypés. Par la suite, ce modèle de prédiction est utilisé pour estimer les valeurs propres des individus d'une population de validation génotypée pour les mêmes marqueurs. En fait, le modèle de prédiction est établi dans le but de capturer la totalité de la variance génétique additive afin d'estimer la valeur propre des individus en sommant les effets de tous les marqueurs. Au final, les meilleurs individus sont choisis sur la base des valeurs propres estimées (Figure 5).

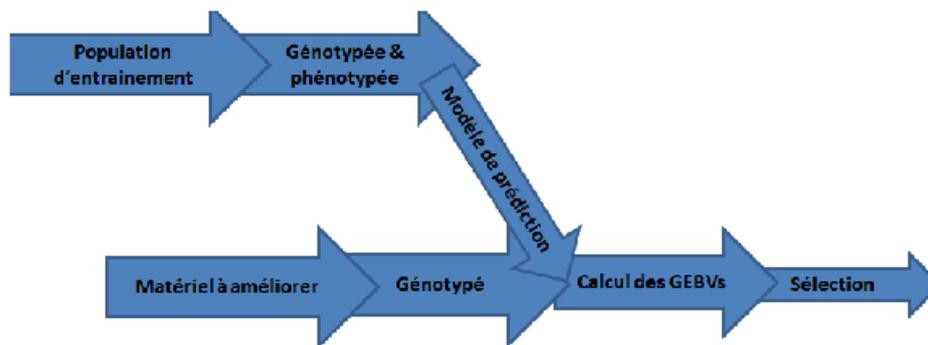


Figure 5 : diagramme de la sélection génomique (Heffner et al., 2009).

Cette méthode est en cours d'implémentation au Chibas. L'idée est de réduire le nombre de génération par cycle de 3 (voir figure 4) à 1.

2.6. Progrès génétique et sélection génomique

L'équation du progrès génétique s'écrit de la manière suivante (Falconer, 1989) : Progrès génétique = intensité de sélection * précision de la sélection * variabilité génétique / intervalle de génération. Les quatre paramètres de cette équation font varier le progrès génétique. En effet, la SG pourrait augmenter l'intensité de sélection en permettant d'évaluer beaucoup plus d'individus que si on voulait évaluer ces individus au champ. Le niveau de précision est potentiellement plus élevé avec la SG quand les valeurs de prédiction génomiques sont élevées. De plus, la SG permet d'avoir une diminution de l'intervalle de génération car les évaluations ne se réalisent pas à partir des descendants. La sélection génomique récurrente peut donc permettre de réduire chaque cycle à une seule génération. Si on arrive à maintenir une bonne précision avec le phénotypage alors on peut espérer considérablement augmenter les gains avec une intensité de sélection élevée et une réduction de l'intervalle entre les générations.

2.7. Diversité génétique et sélection génomique

La sélection génomique est susceptible d'entraîner une perte de la diversité génétique par la fixation rapide des locus à gros effet et par l'apparementement entre les individus de la population d'entraînement (De Beukelaer et al., 2017). La consanguinité provoque une réduction de l'hétérozygotie au sein des populations et impacte négativement la fitness (Clark et al., 2013). En revanche, il n'est pas possible de faire la sélection génomique en utilisant des individus non apparentés entre la population d'entraînement et la population de validation. En sélectionnant des individus sur la base de leurs valeurs propres, la proportion de sélectionner des individus qui sont très proches génétiquement augmente, ce qui pourrait être amené à une réduction de la diversité génétique. Il est recommandé ainsi de choisir au plus un individu par famille de telle

sorte qu'une telle situation soit évitée (Clark et al., 2013). Une autre solution consiste à veiller à ne pas faire chuter le niveau d'hétérozygotie (diversité) au sein de la population d'une génération à l'autre (De Beukelaer et al., 2017).

2.8. Les limites de la sélection génomique

La performance du modèle de prédiction, l'héritabilité du caractère, l'apparentement entre les individus, la densité de marquage, la composition de la population de référence et l'interaction G*E sont les principaux facteurs qui affectent la précision de la prédiction en sélection génomique (Burgueño et al., 2012; Crossa et al., 2010). Cette précision se définit comme étant la corrélation entre les valeurs propres estimées obtenues à partir des effets additifs et le phénotype observé (BLUE) (Ly et al., 2013).

2.8.1. Héritabilité du phénotype

L'héritabilité d'un caractère est la proportion de la variance qui est expliquée par la variance génotypique. Ainsi, un caractère est considéré comme étant très héritable quand il y a une grande proportion de la variance qui est expliquée par la variation génotypique. On distingue deux types d'héritabilité qui sont l'héritabilité au sens large et l'héritabilité au sens strict. L'héritabilité au sens strict ne tient compte que des effets additifs. D'une manière générale, plus un caractère est héritable plus la précision du modèle de prédiction est élevée. Ainsi, une étude réalisée par Hayes et al., (2009) a montré que la précision de la prédiction est de 0.6 et 0.8 pour deux caractères ayant des héritabilités de 0.2 et 0.7 respectivement. Néanmoins, les résultats de l'étude de Kumar et al., (2012) sur la pomme ont trouvé des niveaux de précision de 0.68 et de 0.89 respectivement pour les caractères de l'astringence ayant une héritabilité au sens strict de 0.26 et de la teneur en matières solubles ayant une héritabilité au sens strict de 0.19.

2.8.2. Composition de la population de référence

La prédiction génomique n'est pas possible s'il n'y a aucun lien d'apparentement entre les individus de la population d'entraînement et ceux de la population de validation. Ainsi, Beaulieu et al., (2014) ont montré que la précision des valeurs propres estimées est deux fois moins grande pour des individus non apparentés que pour des individus qui sont des demi-frères. Cependant, quand il y a trop d'individus très apparentés dans la population d'entraînement cela entraîne une surestimation du niveau de prédiction. Une telle situation doit être évitée en évaluant la population d'entraînement de telle sorte qu'on ait pas beaucoup d'individus trop apparentés dans la population d'entraînement (Ly et al., 2013).

2.8.3. Densité de marquage

La densité de marquage est un paramètre très important à prendre en compte en sélection génomique. Plusieurs études ont montré que la densité de marquage et la précision du modèle de prédiction sont corrélées positivement (Calus et al., 2008; Calus et Veerkamp, 2007; Solberg et al., 2008). Cependant, il existe un plateau quand on fait augmenter la densité de marquage (Solberg et al., 2008). Il est à préciser que plus l'apparentement est faible entre les individus, moindre est le déséquilibre de liaison entre les marqueurs et les QTL et plus la densité de marquage doit être importante pour ne pas louper des marqueurs associés aux QTL.

2.8.4. Interaction Génotype x Environnement (G*E)

L'interaction G*E est un paramètre à prendre en compte dans les modèles de prédiction, surtout pour les caractères faiblement héritables. En fait, un modèle de prédiction qui a été conçu à partir des individus d'une population d'entraînement au cours d'une année « t » doit pouvoir prédire des individus d'une population de validation au cours d'une année « t + 1 ». La présence de cette interaction affecte négativement la capacité prédictive du modèle de prédiction quand il n'y a pas une évaluation des environnements dans lesquels les génotypes seront évolués (Ly et al., 2013). Pour bien maîtriser l'interaction G*E dans un modèle de prédiction, les individus de la population d'entraînement et ceux de la population de validation doivent représenter des environnements différents.

2.8.5. Performance du modèle de prédiction en sélection génomique

Le choix du modèle de prédiction, considéré comme étant l'une des principales limites de la sélection génomique, doit se faire de façon très rationnelle pour avoir un niveau de précision élevé. La précision des modèles de prédiction dépend des hypothèses et des effets des marqueurs. Par exemple, le modèle RR-BLUP suppose que tous les marqueurs aient une même variance avec des effets aléatoires qui sont faibles et non nuls (Hayes et al., 2001; Heffner et al., 2011). Ce modèle marche mieux pour le modèle infinitésimal, c'est-à-dire quand le caractère est sous le contrôle de plusieurs gènes et que chacun de ces gènes a un effet très minime dans la variation du caractère. Certains auteurs ont montré que le modèle RR-BLUP est l'équivalent du modèle de prédiction GBLUP qui est conçu à partir des relations d'apparentement entre les individus (Habier et al., 2008; VanRaden, 2008). A l'instar du modèle RRBLUP, le modèle GBLUP ne prend pas en compte les effets de SNP de grandes tailles (Robert-Granié et al., 2011).

Plusieurs types d'approches bayésiennes sont utilisés en sélection génomique à savoir Bayes A, Bayes B, Bayes C et BRR (Bayesians Ridge Regression). Ces méthodes bayésiennes ont été proposées par Meuwissen et al., en 2001. Cependant, les effets des marqueurs sont vus

différemment suivant la méthode bayésienne utilisée. Ainsi, la méthode « Bayes A » suppose que les effets des SNP suivent une distribution normale avec une variance spécifique qui est différente d'un SNP à l'autre. La méthode « Bayes B » suppose que la majorité des SNP ont une variance génétique nulle et que peu de SNP ont une variance génétique non nulle. Cette méthode semble être efficace pour des caractères qui sont sous le contrôle d'un nombre limité de QTL. La méthode BRR est la version bayésienne de RRBLUP qui utilise des informations sur les variances des effets des marqueurs et des effets des résiduels. La méthode « Bayes C » suit les mêmes principes que ceux de la méthode BRR mais avec la possibilité d'avoir des marqueurs avec des effets nuls (de los Campos et al., 2013).

Les méthodes GBLUP et « Bayes B » ont été comparées pour voir laquelle est meilleure suivant que le caractère est héritable ou pas et qui est sous le contrôle d'un grand nombre de QTL ou pas. Ainsi, la méthode « Bayes B » donne de meilleures précisions pour des caractères gouvernés par peu de QTL et qui sont hértables. En revanche, la méthode GBLUP est meilleure pour des caractères sous le contrôle d'un grand nombre de QTL et qui sont hértables (Daetwyler et al., 2010).

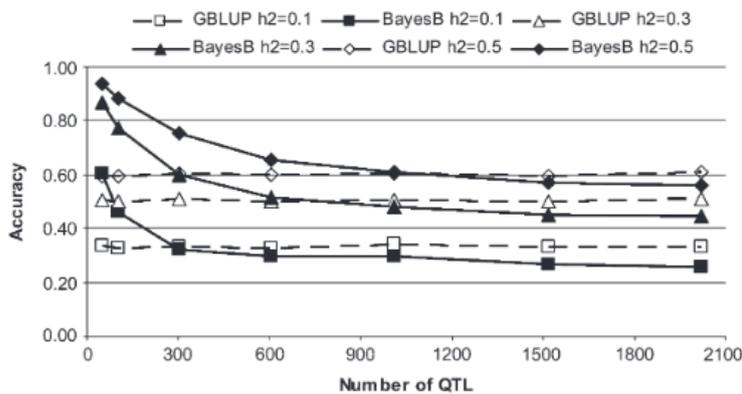


Figure 6: différence entre les modèles Bayes B et GBLUP pour différents niveaux d'héritabilité et pour une variation du nombre de QTL (Daetwyler et al., 2010).

2.9. BLUP génomique

C'est une méthode qui utilise les relations génomiques entre des individus dans le but d'estimer leurs valeurs propres. Pour ce faire, une matrice **G** de relation génomique est constituée à partir des informations moléculaires (MRG). Cette matrice de parenté génomique permet de mesurer la proportion d'allèles qui est partagée entre les individus et elle s'obtient à partir de la formule suivante : $G = ZZ' / 2 \sum p_i q_i$. Dans cette équation, p est la fréquence de l'allèle de référence, q la

fréquence de l'allèle alternatif et Z la matrice d'incidence qui relie les observations aux effets des marqueurs. La prédiction du phénotype en fonction du génotype se fait à partir de cette matrice.

Autrefois, c'était la méthode BLUP sur pedigree (ABLUP) qui était très utilisée pour la prédiction dans les programmes d'amélioration des animaux. Une matrice **A** constituée des informations de pedigree est utilisée pour la prédiction. La méthode de BLUP sur pedigree, est une méthode qui était introduite par Henderson, en 1973) dans la prédiction des valeurs propres chez les animaux. La matrice A utilisée dans le BLUP sur pedigree capture moins d'informations génétiques que la matrice G. Par exemple, les individus 1 et 4 dans la Figure 7 n'ont aucun lien d'apparentement d'après la matrice **A**. Tandis qu'avec la matrice **G**, ils ont un niveau d'apparentement de 0.02. Les individus 2 et 5 ont un coefficient d'apparentement de 0.25 dans la matrice **A**. Tandis qu'ils ont un coefficient de 0.20 avec la matrice G. On voit que la matrice **A** sous-estime et surestime les coefficients d'apparentement. Les marqueurs moléculaires permettent d'avoir beaucoup plus de précision dans le calcul de l'apparentement entre les individus.

$$\mathbf{A} = \begin{bmatrix} 1 & 0.5 & 0.5 & 0 & 0.5 \\ 0.5 & 1 & 0.25 & 0 & 0.25 \\ 0.5 & 0.25 & 1 & 0 & 0.25 \\ 0 & 0 & 0 & 1 & 0 \\ 0.5 & 0.25 & 0.25 & 0 & 1 \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} 1.0 & 0.50 & 0.50 & 0.02 & 0.50 \\ 0.50 & 1.0 & 0.20 & 0.015 & 0.20 \\ 0.5 & 0.20 & 1.0 & 0.025 & 0.30 \\ 0.02 & 0.015 & 0.025 & 1.0 & 0.20 \\ 0.5 & 0.20 & 0.30 & 0.02 & 1.0 \end{bmatrix}$$

Figure 7: Présentation de la matrice A et de la matrice G (Clark et van der Werf, 2013).

Le BLUP sur pedigree ne marche pas en amélioration des plantes à cause des problèmes de consanguinité chez les plantes. À travers plusieurs générations d'autofécondation les valeurs attendues d'apparentement en espérance peuvent diverger grandement des valeurs réelles. Cependant, Les résultats des travaux de Yu et Pressoir (2006) ont montré que des marqueurs moléculaires (IBS) pouvaient être utilisés dans la génétique d'association pour contrôler l'effet du partage de QTL à d'autres locus que celui testé. Les études de Bernado et Yu (2007) ont montré que les marqueurs moléculaires pouvaient être utilisés en sélection génomique.

Cependant, l'une des limites du GBLUP est qu'il ne prend pas en compte les effets de SNP de grandes tailles (Robert-Granié et al., 2011). Ce modèle marche mieux pour des caractères quantitatifs sous le contrôle de plusieurs gènes et que chacun de ces gènes a un effet très minime dans la variation du caractère.

2.10. Objectifs spécifiques de ce travail de recherche

- ✓ Etudier la structuration et la diversité génétique au sein des lignées de sorgho développées par le Chibas.
- ✓ Détecter des traces de sélection au sein des lignées de sorgho développées par le Chibas.
- ✓ Evaluer la qualité de la population d'entraînement pour la prédiction génomique.

3. MATERIELS ET METHODES

3.1. Matériel biologique (espèce *Sorghum bicolor*)

Dans cette étude, le jeu de données est constitué de 1283 génotypes de sorgho dont 296 lignées qui sont développées par le Chibas et 967 génotypes qui représentent la diversité mondiale du panel de diversité. Dans ce panel sont incluses les cinq principales races de sorgho, les dix races intermédiaires et les quatre races de sorgho sauvage (Tableau 11).

3.2. Origine des lignées de sorgho du Chibas

Les 296 lignées de sorgho développées par le Chibas sont issues essentiellement d'un programme de sélection récurrente mais également d'un croisement entre les variétés Papèsèk x Dekabès (IS 23572) et d'un programme de rétrocroisement pour introgresser la résistance à *Melanaphis* dans la variété Papèsèk. Le schéma de sélection récurrente ressemble à celui décrit par A. Gallais en 2011 (Figure 4). La population initiale a été constituée en croisant une population F2 (BC1 BF 95 11/110) fournie par Gilles TROUCHE, un sélectionneur du CIRAD, et croisée avec une lignée ICSV 25280 de l'ICRISAT. Le produit de ce croisement ségrégeait pour la stérilité mâle est autofécondé et par la suite la génération F2 est croisée avec plusieurs parents (NTJ2, Papèsèk, ICSV 25275, IS 23562 et IS23572). Puis, la deuxième génération de ces croisements est croisée avec des variétés de sorgho sucré avec un jus de bonne qualité (OO-SB-F5DT-427, BF 95-11/195, VARIETE-INCONNUE-1, WILEY, IS 23562, IS 23572, IS 23538, IS 23563). Cette population a fait l'objet de deux cycles de sélection récurrente ; une partie des lignées sélectionnées est issue par autofécondation des familles S1 sélectionnées du cycle 1 et l'autre partie des familles S1 sélectionnées du cycle 2.

Tableau 4: variétés exotiques résistantes à *Melanaphis* faisant partie de la collection du Chibas

Variétés exotiques/parents Chibas	
1	WILEY
2	SANTO 275
3	PCR-2>23C-1-M-1
4	IS 299 11
5	IS 23574-CS/09-10
6	IS 23 572
7	IS 23 563
8	IS 23 538
9	IS 23 519
10	VARIETE INCONNUE 1
11	CIR-2/7-3G-N.M (A27)
12	CIR-2/7-3G-16-M.M(A09)
13	CIR1/0G2-4G-1G-M-M
14	BF 95-11/195
15	BCI(CIR-2/7-3G-1G-MXIS37007)-P12-1-4-(A33)
16	00-SB-FSDT427

3.3. Dispositif expérimental

Un dispositif simple randomisé (un seul bloc, pas de répétition) est utilisé dans le cadre de cette expérience pour pouvoir phénotyper les 222 lignées de la population de référence pour la prédiction génomique (Figure 18). Ce dispositif est localisé en Haïti dans la commune de l'Archaie plus précisément à Cabaret. Ce dispositif a pour coordonnées géographiques : 18°43.466' de latitude et 072°25.064' de longitude. Il s'agissait de reproduire les semences de toutes les lignées du Chibas pour pouvoir réaliser les essais multi locaux. Les semis de ces essais multi locaux ayant plusieurs répétitions sont déjà réalisés. Chacune des parcelles variétales a une largeur de 3.5 mètres et une longueur de 2.8 mètres (4 rangées). En longueur, elles ne se sont pas séparées. Cependant, elles se sont séparées de 1 mètre en largeur (Figure 18).

3.4. Itinéraires techniques adoptés

Le semis des 222 lignées pour le phénotypage a été réalisé le 25 Mars 2017 sur une parcelle expérimentale dont le précédent cultural est le sorgho. Le sorgho est semé avec une distance de 25 centimètres entre les poquets d'une même rangée et 70 centimètres entre les poquets de deux rangées différentes (57744 plantes/ha). 15 jours après la germination, on a fait un dégarnissage pour laisser une seule plantule par poquet. Aucune fertilisation n'a eu lieu avant et pendant l'expérience. Les lignées de sorgho du Chibas ont été évaluées en saison pluvieuse. Cependant, s'il ne pleuvait pas pendant plus de huit jours on procédait à un arrosage pour éviter tout problème de stress hydrique.

3.5. Extraction d'ADN

Les 296 lignées de sorgho du Chibas et les 18 parents ont été génotypés. Pour ce faire, les échantillons de feuilles, prélevés sur les 5 premières plantules au sein de chaque parcelle variétale, sont collectés dans des plaques de 96 puits. Ensuite, une solution de lyse a été ajoutée aux échantillons préalablement lyophilisés pendant 2 jours. Cette solution de lyse, utilisée pour la libération de l'ADN, est obtenue à partir de RNase A, Buffer AP1 et Reagent DX. L'ADN obtenu a été expédié à l'Université du Kansas pour génotypage.

3.6. Technique de séquençage

La technique de génotypage par séquençage (GBS) est utilisée pour caractériser la variation génomique au sein des individus. La technologie *llumina Nextseq500 avec un multiplexage d'échantillons 384X et l'enzyme de restriction ApeKI* de l'Université Kansas est utilisée pour le génotypage des populations. Le génotypage et le traitement des données GBS ont été effectués à l'Université du Kansas aux Etats-Unis. Les contraintes imposées pour le traitement ont porté sur : la profondeur du séquençage (nombre de reads doit être supérieur à 2), les données manquantes et les fréquences alléliques.

3.7. Méthode 1 : Structuration et diversité génétique chez les populations de sorgho

Les méthodes d'analyse factorielle des correspondances et de Structure sont utilisées pour la structuration et le package hierFstat est utilisé pour le calcul des statistiques de base.

3.7.1. Analyse Factorielle des Correspondances

L'analyse factorielle des correspondances (AFC) est une méthode statistique qui a été créée par Benzécri en 1982. C'est une méthode qui est basée sur l'inertie et qui permet d'étudier l'association entre deux variables qualitatives à partir d'un tableau de contingence. Le but de l'AFC consiste à représenter un maximum de l'inertie totale sur le premier axe factoriel, un maximum de l'inertie résiduelle sur le second axe, et ainsi de suite jusqu'à la dernière dimension. La somme des inerties est égale à la statistique du χ^2 divisée par la fréquence totale qui est la somme des cellules du tableau de contingence. Le χ^2 permet de déterminer, sur la base de la statistique du χ^2 , si l'on doit rejeter l'hypothèse nulle selon laquelle les lignes et les colonnes du tableau sont indépendantes.

3.7.2. Logiciel Structure de Pritchard

Ce logiciel a été inventé par Pritchard et al., en 2000). Ce modèle permet de déterminer la structuration des individus à partir des données moléculaires obtenues pour plusieurs locus. Un nombre de populations ancestrales est identifié et les individus appartenant à une même population ancestrale se sont caractérisés à partir de leurs fréquences alléliques pour chaque locus. Cette méthode permet d'identifier des individus admixtes et des migrants au sein d'une population.

3.7.3. Calcul des statistiques de base

Les lignées du Chibas seront comparées avec le groupe le plus proche génétiquement en se basant sur des calculs statistiques comme : indice de fixation, hétérozygotie observée, Indice de Nei et F_{st} entre les groupes. Pour ces calculs, le package HierFstat du logiciel R est utilisé.

3.7.3.1. Hétérozygotie observée (H_o)

Le taux d'hétérozygotie observée est la moyenne des fréquences des hétérozygotes observées sur l'ensemble des locus. Il est calculé à partir de la formule suivante : $H_o = 1 - \sum_k \sum_i P_{kii} / np$ où k est l'échantillon, P est la proportion d'homozygotes et np le nombre d'échantillons.

3.7.3.2. Indice de diversité génétique de Nei (H_s)

L'indice de diversité génétique de Nei est un paramètre qui permet de mesurer le niveau de diversité génétique d'une population. Il est calculé à partir des fréquences alléliques et du niveau d'hétérozygotie observée au niveau des locus. Cet indice est calculé à partir de la formule suivante : $H_s = (n/n-1) * (1 - \sum p_i^2 - H_o/2n)$ (Nei and Chesser, 1983).

3.7.3.3. Indice de fixation (F_{IS})

Le F_{IS} est un paramètre qui permet de mesurer la différenciation des individus au sein d'une population donnée. Autrefois, ce paramètre est appelé coefficient de consanguinité. Il est calculé à partir de l'hétérozygotie observée (H_o) et de l'hétérozygotie attendue (H_s). Il est obtenu à partir de la formule suivante : $F_{IS} = 1 - (H_o/H_s)$.

3.7.3.4. Différenciation entre les groupes (F_{ST})

Le F_{st} est un indice de diversité génétique proposé par Wright (Wright, 1949) qui représente la différenciation des populations par rapport à la population globale pour des sites bi alléliques. Il est obtenu à partir de la formule suivante : $F_{st} = (H_t - H_s) / H_t$ où H_t est l'hétérozygotie attendue dans la population globale et H_s l'hétérozygotie moyenne attendue à l'intérieur des populations.

3.8. Méthode 2 : Détection de la signature de la sélection

La méthode d'outFlank et la méthode de F_{st} outliers se sont combinées pour la détection des traces de sélection. OutFlank est une nouvelle méthode développée par Michael C. Whitlock et Katie E. Lotterhos en 2015 pour détecter des locus qui sont sous sélection. Cette méthode recommande de supprimer tous les locus ayant des taux d'hétérozygotie inférieurs à 5 % avant de poursuivre avec le calcul de différenciation génétique pour chaque locus. Ensuite, un F_{st} moyen est calculé de telle sorte qu'on puisse prédire la distribution des F_{st} . Les locus avec de fortes valeurs de F_{st} sont ceux responsables d'une adaptation locale (sélection diversifiante) et ceux avec des valeurs de F_{st} faibles sont ceux affectés par une sélection balancée (Whitlock and Lotterhos, 2015). Avec cette méthode, dans un premier temps, les lignées du Chibas et du panel de diversité du sorgho sont fixées de telle sorte qu'on ait deux populations ancestrales. Et, dans un second temps, les lignées du Chibas et les lignées du groupe Caudatum de l'Afrique de l'Est sont fixées pour avoir aussi deux populations ancestrales c'est-à-dire K est égal à 2. Ensuite, les F_{st} sont calculés à partir de ces deux ensembles. Et par la suite, ces F_{st} sont utilisés par la méthode F_{st} outliers pour détecter les locus qui sont sous sélection. L'approche F_{st} outliers est basée sur la différenciation entre populations. Cette approche permet de mesurer la quantité de variabilité génétique qui est expliquée par la différenciation entre populations. Elle permet d'identifier des locus qui montrent une différenciation de fréquences alléliques entre des groupes significativement différents par rapport à la différenciation moyenne sur le génome. L'approche F_{st} outliers utilise des valeurs de F_{st} qui sont établies sur les fréquences alléliques des groupes ancestraux. Ensuite, on fait appel à une statistique, le z-score, qui mesure la distance de chaque valeur observée de la moyenne en unité de déviation standard. Le z-score permet de savoir si la différenciation pour un SNP est significativement plus grande que l'attendue. Les valeurs de z-score sont obtenues à partir des valeurs de F_{st} . En fait, pour chaque valeur de F_{st} , une p-value est calculée. Normalement, les valeurs de ces p-values doivent avoir une distribution uniforme avec un pic à proximité de zéro. Le critère de False Discovery Rate (FDR) est estimé pour chaque p-value pour sélectionner les locus les plus significatifs (Martins et al., 2016).

3.9. Méthode 3 : Protocole d'utilisation des modèles de prédiction

Le modèle de prédiction GBLUP est utilisé pour la prédiction génomique. La fonction `kin.blup` implémentée dans le logiciel R est utilisée. La fonction `kin.blup` utilise une matrice d'apparentement entre les individus pour pouvoir prédire les traits phénotypiques. Cette matrice d'apparentement a été obtenue via le logiciel TASSEL. Une analyse discriminante en composantes principales est faite pour regrouper les lignées qui sont très apparentées.

3.9.1. Population de référence

Deux stratégies d'échantillonnage sont utilisées pour la prédiction génomique. Pour la première stratégie d'échantillonnage, l'objectif est de ne pas avoir des lignées qui sont très apparentées dans la population de validation c'est-à-dire l'élimination des groupes de lignées sœurs très fortement apparentées. Ainsi, les phénotypes des lignées appartenant à un même groupe (groupe obtenu via DAPC) sont supprimés quand on prédit leurs phénotypes et ces phénotypes sont remplacés par leurs valeurs observées quand on prédit les lignées d'autres groupes. Par ce procédé, la population de validation est constituée des individus qui ne sont pas utilisés pour concevoir le modèle de prédiction car les phénotypes de ces individus ne sont pas utilisés dans la population d'entraînement. Dans la deuxième stratégie d'échantillonnage, la population de référence est subdivisée en une population d'entraînement et une population de validation pour la prédiction génomique. Ainsi, la population d'entraînement représente 75 % et la population de validation représente 25 % de la population de référence. Pour ce faire, un tirage aléatoire est réalisé pour constituer la population d'entraînement. Et le reste des individus fait partie de la population de validation.

3.9.2. Analyse discriminante en composantes principales

L'analyse discriminante en composantes principales (DAPC) est une nouvelle méthode multivariée, permettant d'extraire des informations à partir des données moléculaires. Elle a été développée par Jombart et al., en 2010 dans le but de regrouper des individus qui sont très apparentés. Cette méthode est une combinaison des méthodes de l'analyse en composantes principales (PCA) et de l'analyse discriminante (DA). PCA vise à résumer la variabilité globale chez les individus, c'est-à-dire la divergence entre les groupes qui se traduit par la structuration et la variation au sein d'un même groupe qui se traduit par la variabilité génétique intra population. En effet, PCA aboutit à un graphique qui montre une mauvaise représentation des informations (Jombart et al., 2010). Ces auteurs ont développé DAPC qui est une approche qui maximise la variation inter groupe et minimise la variation intra groupe.

Cette approche a été implémentée sur le logiciel R pour être utilisée plus facilement. Les fonctions `find.clusters` et `dapc` du logiciel R sont utilisées pour déterminer le nombre de groupes. La fonction « `find.clusters` » du logiciel permet d'identifier le nombre de groupes en utilisant l'algorithme « k-means ». Cet algorithme utilise les regroupements d'une analyse en composante principale qui est fait préalablement pour distinguer les groupes. La fonction « `dapc` » donne le coefficient d'appartenance des individus à chacun des groupes.

3.9.3. Matrice d'apparement

La matrice d'apparement est calculée pour les 222 lignées de sorgho qui sont phénotypées pour les caractères hauteur, degré brix et rendement grain. Cette matrice prend en compte l'apparement entre les individus mais également la structuration qui existe au sein de ces individus (Endelman and Jannink, 2012).

3.9.4. Modèle de prédiction

Pour la prédiction, un tirage avec remise est réalisé, à partir d'un processus itératif, en excluant à chaque fois le phénotype d'un groupe d'individus qui sont très apparementés. Un tel processus permet de prédire un groupe d'individus très apparementés sans tenir compte des phénotypes. Les individus ont deux valeurs de GEBV, une valeur de GEBV pour la population d'entraînement et une valeur de GEBV pour la population de validation. Pour calculer la précision du modèle de prédiction, seules les valeurs de GEBV pour la population de validation sont utilisées.

3.9.5. Densité de marquage

Pour l'ensemble des 1283 individus, on a utilisé 1660 marqueurs SNPs qui ont déjà été identifiés au niveau du génome du sorgho par le laboratoire Geoffrey Morris à l'Université du Kansas. On a utilisé que 1660 marqueurs et non pas un nombre plus élevé pour éviter de ne pas avoir trop de retard pour la soumission du travail.

3.10. Paramètres phénotypiques évalués

Le phénotypage des lignées du Chibas est réalisé sur les caractères suivants : le taux de germination, la vigueur et l'homogénéité à la levée, la date d'épiaison, la date de maturité, le degré brix, le volume jus, le rendement grain, la hauteur, la résistance à la verse et la présence/quantité de pucerons au sein des unités expérimentales. La manière de procéder pour le phénotypage est résumée dans un tableau en annexe (Tableau 10). La hauteur, le degré brix et le rendement grain sont trois caractères contrastés qui ont été choisis pour la prédiction génomique. La hauteur est un caractère polygénique héritable influencé par quelques gènes majeurs, le brix un caractère polygénique fortement héritable et le rendement grain est un caractère polygénique faiblement héritable.

3.10.1. Détermination du rendement grain

Avant de récolter les panicules pour chaque parcelle variétale, on tenait à ce que 80% des panicules aient été à maturité. Une fois mature, on a récolté les panicules sur l'ensemble des quatre rangées de chaque parcelle variétale. Ensuite, on a contrôlé le nombre de panicules

récoltées puis on a mesuré le poids des panicules récoltées pour chaque parcelle variétale. Le poids des grains représente environ 85% du poids des panicules.

3.10.2. Détermination du degré brix

Un réfractomètre est utilisé pour déterminer le degré brix du jus extrait de la canne du sorgho préalablement pressée. Pour être plus précis dans la mesure, on l'a prise trois fois avant de faire la notation.

3.10.3. Évaluation de la hauteur des lignées

La hauteur de chacune des lignées est évaluée en choisissant aléatoirement 15 plantes par parcelle variétale. Cette hauteur est mesurée à partir du collet jusqu'à la base du pédoncule en utilisant un PVC gradué.

4. RESULTATS

Dans cette partie sont présentés les différents résultats de ce travail de recherche.

4.1. Caractérisation des lignées de sorgho développées par le Chibas

Pour la caractérisation des lignées développées par le Chibas, une structuration de la diversité génétique de l'ensemble des lignées de sorgho du Chibas et de celles du panel de diversité du sorgho a été réalisée. Pour ce faire, une analyse factorielle des correspondances et une analyse de structuration (logiciel Structure de Pritchard) ont été réalisées à partir des données de génotypage de ces lignées.

4.1.1. Analyse Factorielle des Correspondances

Dans le graphique obtenu par l'AFC, les lignées développées par le Chibas (en rose) sont plus proches du groupe Caudatum de l'Afrique de l'Est (en bleu) que les autres groupes de sorgho. On remarque, aussi, qu'il y a une structuration au sein des lignées du Chibas. Ce graphique donne un aperçu sur la diversité génétique au sein des différents groupes. Il paraît que les lignées de sorgho développées par le Chibas sont moins diversifiées que le groupe Caudatum de l'Afrique de l'Est.

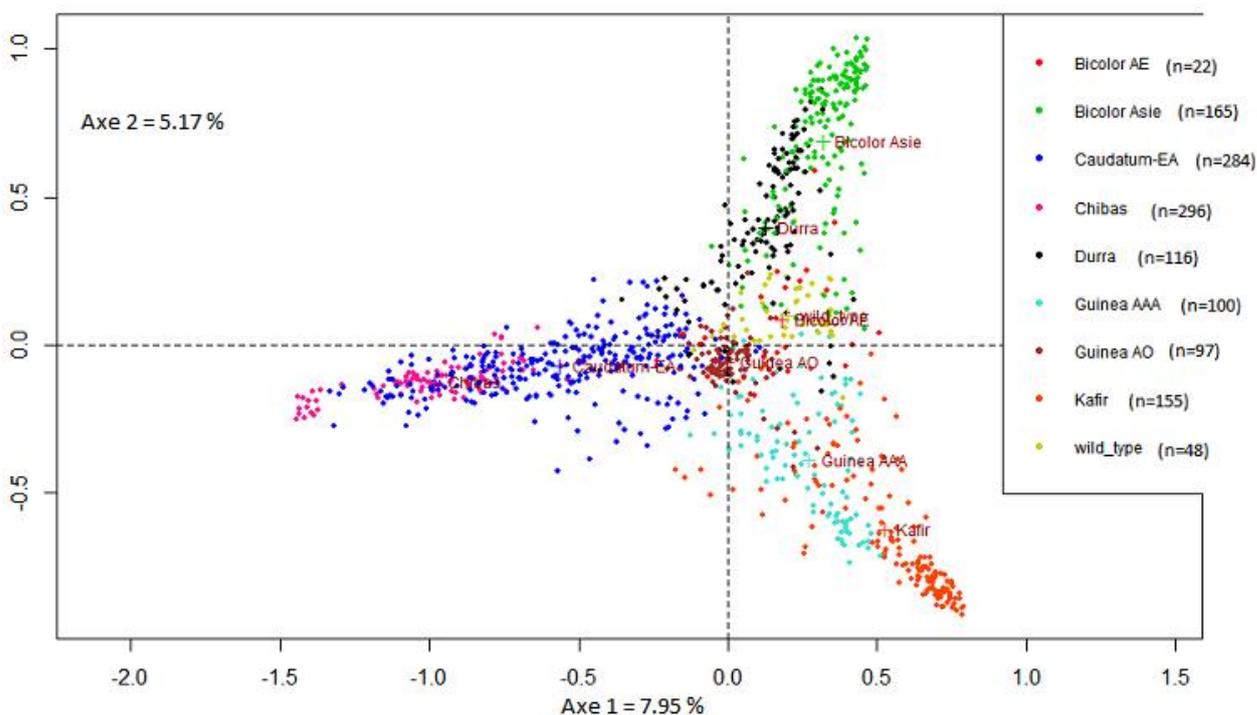


Figure 8 : Structuration des différents groupes de sorgho via l'Analyse Factorielle des Correspondances

Bicolor AE : Bicolor de l'Afrique de l'Est. **Caudatum EA** : Caudatum de l'Afrique de l'Est. **Guinea AAA** : Guinée de l'Asie et de l'Afrique Australe. **Guinea AO** : Guinée de l'Afrique de l'Ouest. **Wild-type** : sorgho sauvage. **Chibas** : les lignées de sorgho développées par le Chibas.

4.1.2. Structuration avec le logiciel Structure de Pritchard

Les 1283 génotypes de sorgho utilisés dans cette étude sont structurés en onze groupes (Figure 9). L'analyse de structure montre que les lignées du Chibas se divisent 3 groupes qui sont les groupes : 1, 5 et 7. Parmi ces trois groupes identifiés, le groupe 1 est un groupe mixte qui contient certaines lignées du Chibas et certaines lignées du groupe Caudatum de l'Afrique de l'Est notamment des lignées de type Zera-zera. Et, les groupes 5 et 7 ne contiennent que des lignées développées par le Chibas. On a observé des introgressions du groupe 11 dans le groupe 7 qui est constitué par des lignées caractérisées par des panicules très lâches à l'instar des sorghos sauvages (*verticilliflorum-aethiopicum*). Néanmoins, certaines variétés de ce groupe 7 font partie des lignées avec les rendements grains les plus élevés avec leurs panicules lâches.

Hormis la race bicolor, les principales races de sorgho sont bien structurées. Les résultats de cette analyse montrent que la structuration est non seulement en fonction des races mais aussi elle est fonction des origines.

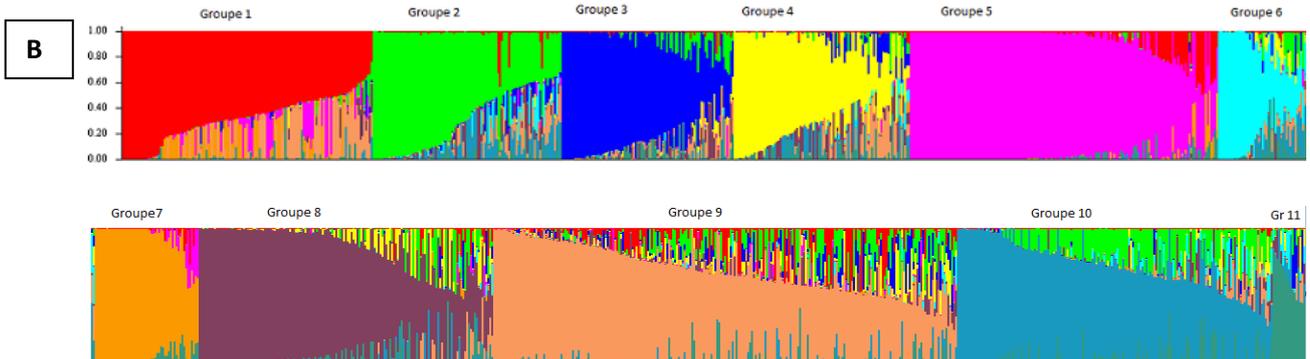
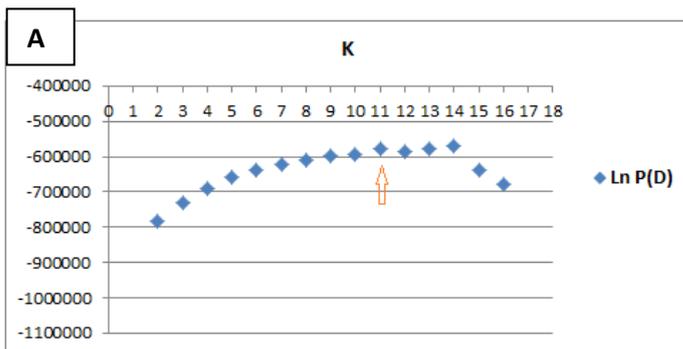


Figure 9 : Résultats de l'analyse structure de Pritchard (length of burning égal à 5000 et nombre de répétitions égal à 5000).

(A) nombre de populations ancestrales choisi en fonction de Ln P(D). La flèche sur ce graphique indique que la valeur de K choisi est 11. (B) Les groupes : 1, 5 et 7 sont constitués par les lignées de sorgho développées par le Chibas. Les Durra de l'Afrique de l'Est forment le groupe 2. Les Guinea de l'Afrique de l'Ouest forment le groupe 3. Les Guinea de l'Asie et de l'Afrique australe forment le groupe 4. Le groupe 6 est constitué des Guinea et certains sorghos sauvages. Les races Kafir de l'Afrique Australe et Caudatum forment respectivement les groupes 8 et 9. La race intermédiaire Bicolor-Caudatum forme le groupe 10. Et le groupe 11 est constitué des sorghos sauvages (*verticilliflorum-aethiopicum*).

4.2. Paramètres de diversité génétique

L'indice de fixation (F_{is}), l'hétérozygotie attendue, l'hétérozygotie observée, la richesse allélique et le F_{st} entre paires de population sont les principaux paramètres de diversité qui sont calculés pour comparer les lignées du Chibas des variétés du groupe Caudatum de l'Afrique de l'Est et des génotypes du panel de diversité mondiale du sorgho. En se basant sur l'hétérozygotie attendue, les lignées du Chibas sont les moins diversifiées des trois groupes avec 21 % diversité de moins que le groupe Caudatum de l'Afrique de l'Est et 56 % diversité de moins que le panel de diversité mondiale du sorgho. Toutefois, la diversité de la population du Chibas est comparable, voir plus élevée, à celle observée dans certaines races de sorgho. Les taux moyens d'allèles pour les lignées de sorgho du Chibas, le groupe Caudatum de l'Afrique de l'Est et le panel de diversité sont respectivement de 1.35, 1.62 et 1.94. L'indice de fixation est moins élevé dans les lignées du Chibas que les groupes Caudatum de l'Afrique de l'Est et le panel de diversité. Dans le Tableau 6, ayant les F_{st} par paires pour les trois groupes, le F_{st} le moins élevé (0.15) est obtenu pour les paires de lignées du Chibas et le groupe Caudatum de l'Afrique de l'Est.

Tableau 5: Mesures de diversité pour les lignées du Chibas, du groupe CAE et du panel de diversité mondiale du sorgho

	Fis	He	Ho	Richesse Allélique
Lignées Chibas	0.8	0.11	0.022	1.35
Caudatum AE	0.87	0.14	0.018	1.62
PD	0.88	0.25	0.03	1.94

Tableau 6 : Valeurs de F_{st} par paires obtenues entre les trois groupes

	Lignées Chibas	Caudaum AE	Panel diversité sans Caudatum AE
Lignées Chibas		0.1527	0.269
Caudaum AE	0.1527		0.1944
Panel diversité sans Caudatum AE	0.269	0.1944	

4.3. Détection des traces de sélection

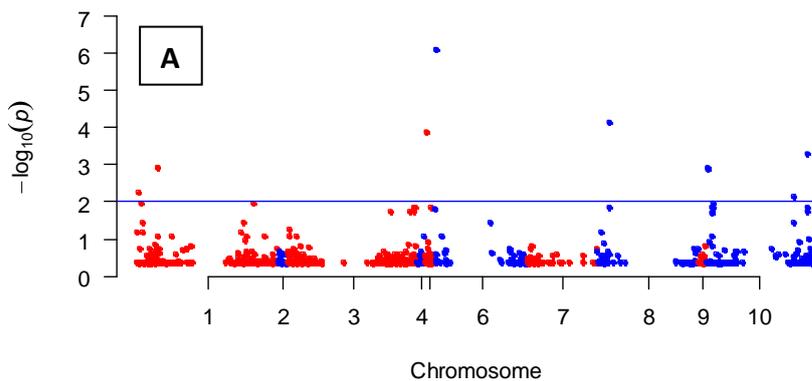
Des traces de sélection sont identifiées entre les lignées du Chibas et le groupe Caudatum de l'Afrique de l'Est et entre les lignées du Chibas et le Panel de diversité du sorgho (Figure 10). Pour un seuil de 0.01, seuil de probabilité de faux positif, neuf locus sont identifiés comme étant sous sélection entre les lignées du Chibas et le groupe Caudatum de l'Afrique de l'Est et 28 locus entre les lignées du Chibas et les génotypes du panel de diversité. Parmi ces locus, deux se trouvent sur le chromosome 6, chromosome sur lequel est localisé le gène RMES1 conférant une résistance à *Melanaphis sacchari*. Le locus « S6_1202180 » est proche d'une région du génome du sorgho qui a

été identifiée comme étant une zone qui confère la résistance à *Melanaphis sacchari* d'après l'étude de Wang et al., 2013. Les locus « S6_1202180 » et « S10_58437817 » sont détectés pour être à la fois sous sélection entre les lignées du Chibas / CAE et entre lignées du Chibas / panel de diversité.

Tableau 7 : Locus sous sélection entre lignées du Chibas/CAE et lignées du Chibas/panel diversité

Chibas vs caudatum AE		Chibas vs Panel diversité			
Chr	sites	Chr	sites	Chr	sites
1	S1_12227113	1	S1_66261086	6	S6_1202180
1	S1_12227115	2	S2_1217038	7	S7_39568768
5	S5_1456219	3	S3_1059859	8	S8_7567679
6	S6_1055730	3	S3_1072220	8	S8_7570897
6	S6_1202180	3	S3_4241910	10	S10_4117657
8	S8_7570897	3	S3_4270600	10	S10_4118102
10	S10_1514378	3	S3_12134519	10	S10_4258703
10	S10_1514379	3	S3_56662688	10	S10_52241870
10	S10_58437817	3	S3_58657895	10	S10_52241898
		3	S3_68840917	10	S10_52241907
		3	S3_73032285	10	S10_58276380
		4	S4_2283971	10	S10_58437817
		5	S5_3866567	10	S10_58553403
		6	S6_990451	10	S10_59697166

Locus sous sélection dans Chibas&caudatum



Locus sous sélection dans Chibas&PD

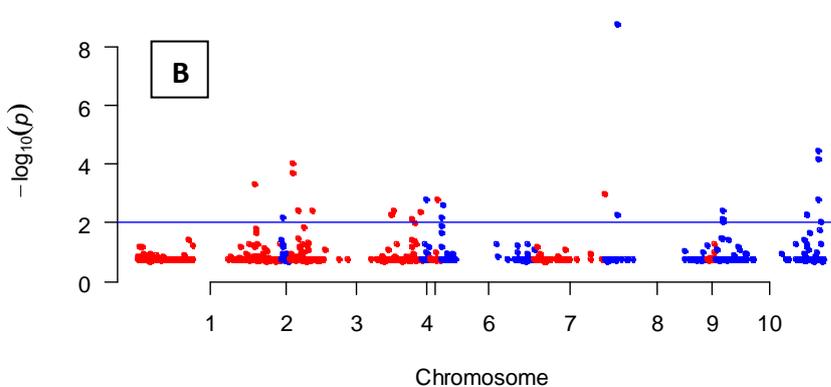


Figure 10 : (A) Locus sous sélection au sein des lignées du Chibas et le groupe Caudatum de l'Afrique de l'Est. (B) Locus sous sélection au sein des lignées du Chibas et le panel de diversité.

4.4. Prédiction génomique

Pour la prédiction génomique, dans un premier temps, une analyse discriminante en composantes principales est faite pour regrouper les lignées qui sont très apparentées. Ensuite, la prédiction génomique est réalisée en fonction de ces groupes.

4.4.1. Présentation des traits phénotypiques utilisés dans la prédiction

Le degré brix du jus extrait des tiges, la hauteur des tiges et le rendement grain sont les trois caractères qui sont utilisés pour l'évaluation de la qualité des prédictions génomiques. Le degré brix et la hauteur sont plus héréditaires que le rendement grain qui est un caractère complexe.

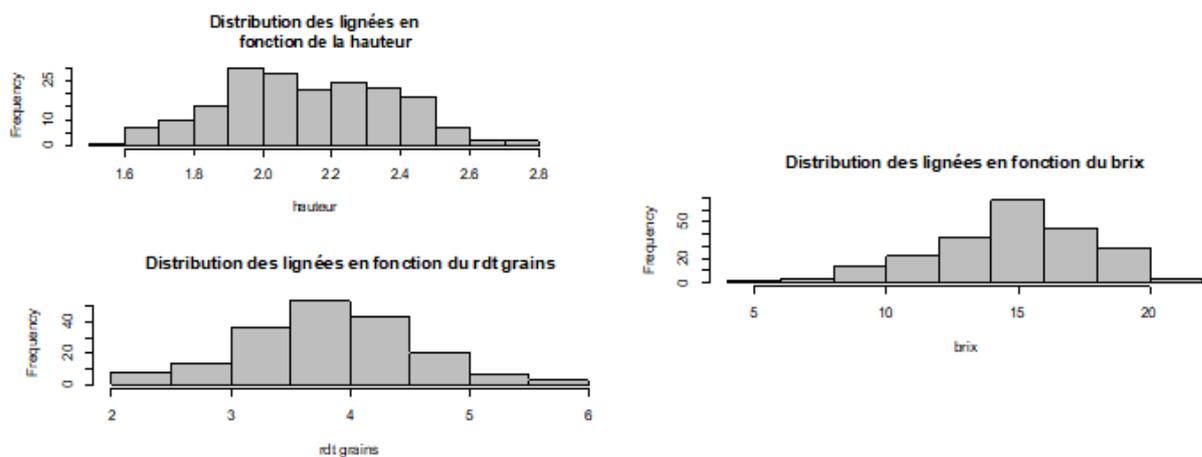


Figure 11 : distribution des individus en fonction des traits phénotypiques

4.4.2. Corrélation entre les traits phénotypiques utilisés dans la prédiction

Une corrélation positive est obtenue pour les caractères degré brix et hauteur pour une significativité très élevée (p -value = 0.000085). Les caractères hauteur et rendement grain sont aussi corrélés positivement pour une significativité moindre que les caractères degré brix et hauteur (p -value = 0.0027). Cependant, les caractères degré brix et rendement grain ne sont pas significativement corrélés.

Tableau 8 : Corrélation entre les traits phénotypiques

	degré brix	hauteur
hauteur	0.2824 ***	
rdt grain	-0.1048	0.2196 **

Niveau de significativité indiqué par *, ** et *** respectivement pour $P < 0.05$, $P < 0.01$ et pour $P < 0.001$.

4.4.3. Distribution des lignées du Chibas en fonction de leur apparentement

Le niveau d'apparentement des lignées de sorgho développées par le Chibas paraît très favorable pour faire de la prédiction génomique. Ainsi, la majorité des lignées du Chibas est relativement apparentée, une partie, très apparentée, qui est représentée par lignées sœurs et qui ne sont différentes que par un nombre limité de locus et une autre partie qui est faiblement ou moyennement apparentée.

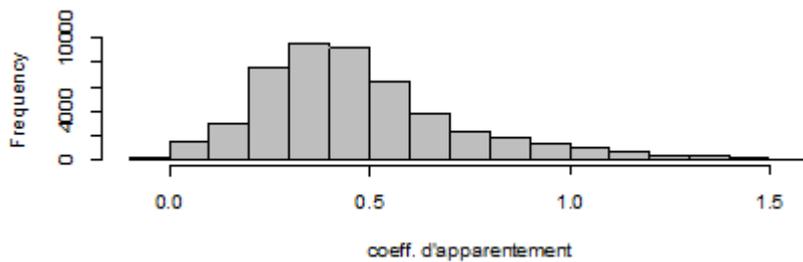


Figure 12 : Distribution des lignées de sorgho du Chibas en fonction de leur apparentement

4.4.4. Identification des individus très apparentés

La population de référence pour la prédiction génomique est constituée de 222 lignées de sorgho développées par le Chibas. Par une DAPC, on a identifié quatorze groupes au niveau de la population de référence. Les lignées faisant partie d'un même groupe sont plus apparentées entre elles que des lignées appartenant à un autre groupe. Ces groupes sont utilisés pour la prédiction génomique dans le but d'éliminer les lignées qui sont très apparentées dans la population de référence.

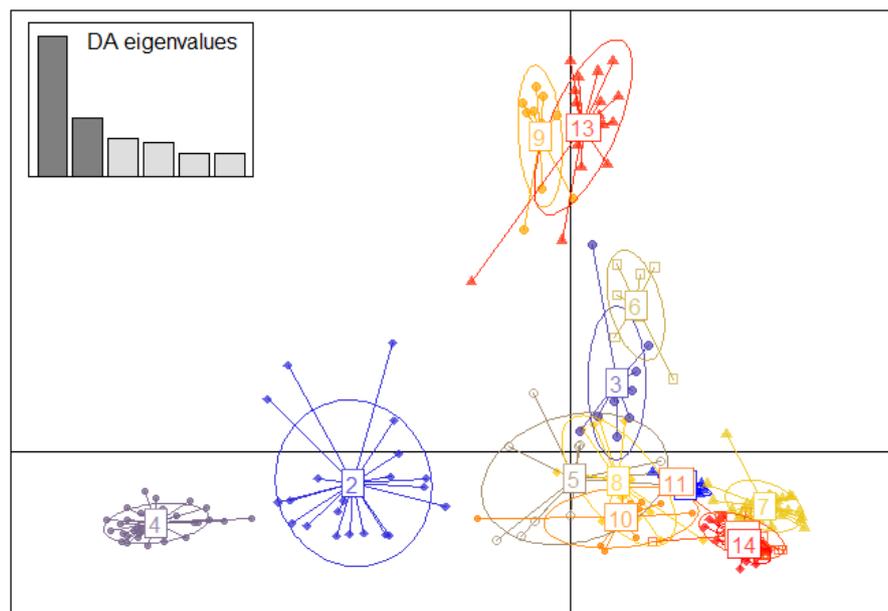


Figure 13 : Structuration des individus de la population d'entraînement avec DAPC

4.4.5. Modèle de prédiction GBLUP

Le modèle de prédiction GBLUP est utilisé dans ce travail. Deux stratégies d'échantillonnage sont mises en évidence. Les niveaux de précision pour les caractères rendement grain, degré brix et hauteur sont respectivement de **5 %**, **60 %** et **50 %** pour l'échantillonnage qui se base sur la non prise en compte des lignées très apparentées dans la population de validation (stratégie 1).

Etant donné que la stratégie d'échantillonnage aléatoire fait varier les niveaux de précision à chaque tirage, neuf tirages aléatoires sont effectués. Les résultats de l'échantillonnage aléatoire et de l'héritabilité génomique, calculée à partir de la variance environnementale et de la variance génétique obtenue des relations d'apparentement entre les individus, sont présentés dans le tableau ci-dessous.

Tableau 9: Comparaison des niveaux de précision (r) de deux stratégies d'échantillonnage sur GBLUP

	Précision de la prédiction (r) via la stratégie 1 (DAPC)	Précision de la prédiction (r) via la stratégie 2 (Echantillonnage aléatoire)	héritabilité génomique (h²g)
Hauteur	0.50	0.84 ± 0.03	0.61
Degré brix	0.60	0.86 ± 0.02	0.56
Rdt grain	0.05	0.68 ± 0.05	0.32

Stratégie 1 : regroupement des lignées très apparentées via DAPC pour calculer la précision de la prédiction. **Stratégie 2** : réalisation de 9 tirages aléatoires (intervalle de confiance : 95 %) pour constituer la population d'entraînement (75%) et la population de validation (25%).

5. DISCUSSIONS

Les principaux résultats de ce travail sont discutés dans cette partie du document.

5.1. Structuration des génotypes de sorgho

Les résultats de l'AFC montrent que les Lignées de sorgho développées par le Chibas sont proches du groupe Caudatum de l'Afrique de l'Est. De tels résultats prouvent que les génotypes qui ont été utilisés comme parents dans le schéma de sélection du Chibas viennent du groupe Caudatum de l'Afrique de l'Est. Certains parents de ce groupe Caudatum de l'Afrique de l'Est utilisés dans le programme d'amélioration du Chibas sont sensibles à *Melanaphis*. Trois des parents utilisés dans le programme d'amélioration du Chibas ont été perdus du fait qu'ils n'ont pas pu être maintenus et multipliés en 2015 et 2016 du fait de leur sensibilité à *Melanaphis*. Certains Caudatum d'Éthiopie connus sous le nom de Zera-zera sont très utilisés dans les programmes de sélection car ils sont résistants aux maladies et aux attaques des insectes (Prasada Rao, et Mengesha, 1981). Il est probable que la résistance à *Melanaphis* ait pour origine le groupe Zera-zera. En effet, les 8 lignées parentales du groupe Caudatum de l'Afrique de l'Est sont résistantes à *Melanaphis* et présentent des caractéristiques typiques du groupe Zera-zera.

Les lignées de sorgho du Chibas sont structurées en trois groupes (résultats de l'analyse de structure). L'un des trois groupes est constitué des lignées ayant des panicules lâches. Dans ce groupe, on observe des introgressions d'allèles du groupe qui comprend les sorghos sauvages de type verticilliflorum et aethiopicum. Il y avait possiblement des introgressions des sorghos sauvages dans certains parents Caudatum. La structuration au sein des lignées du Chibas pourrait être expliquée par l'utilisation de plusieurs parents qui présentent des caractéristiques différentes. Il serait aussi important de faire une caractérisation morphologique de ces trois groupes (groupe 7 avec panicules lâches) dans le but de mettre en évidence les différences.

5.2. Paramètre de diversité

En se basant sur l'indice de diversité génétique de Nei, les lignées de sorgho du Chibas ont une diversité moindre que celles du groupe Caudatum de l'Afrique de l'Est. Normalement, on devrait comparer la diversité génétique actuelle au Chibas avec la diversité qui existait avant le goulot d'étranglement pour savoir s'il y a une diminution de la diversité génétique au sein du panel de sorgho du Chibas. Malheureusement, une telle diversité n'est pas connue car les génotypes sensibles n'ont pas été génotypés. Cependant, la diversité des lignées de sorgho du Chibas est comparée avec celle du groupe Caudatum de l'Afrique de l'Est. Les résultats de ce travail de recherche montrent que le niveau de diversité génétique des lignées de sorgho du Chibas (**0.11**)

est moindre que celui du groupe Caudatum de l'Afrique de l'Est (**0.14**) soit 21 % de diversité génétique en moins. On s'attend à un niveau de diversité génétique plus faible au sein du panel du Chibas car il y avait une forte pression de sélection au *Melanaphis sacchari*. Il serait aussi intéressant de comparer la diversité génétique dans le programme d'amélioration du Chibas avec celle dans d'autres programmes d'amélioration du sorgho. Ce qui permettrait d'avoir une idée sur la diversité du Chibas. Mais, il est impossible de faire cette comparaison car les mesures de diversité ne sont pas calculées avec les mêmes marqueurs moléculaires (Deu et al., 2006 ; Ramu et al., 2013) ou elles sont mesurées suivant des paramètres de diversité différents (Mace et al., 2013). Le Chibas espère pouvoir prochainement comparer le niveau de diversité observé au sein de sa population avec la diversité des programmes d'amélioration variétale de l'ICRISAT ou de Texas. Toutefois, on peut déjà dire qu'il serait judicieux d'introduire de la variation en provenance d'autres races (Kafir, Dura, Guinea). En effet, la diversité présente dans la population du Chibas provient quasiment exclusivement des Caudatum et sorghos sauvages de l'Afrique de l'Est.

5.3. Détection des traces de sélection

Tout goulot d'étranglement aboutit toujours à une réduction de la diversité qui peut toucher un ou plusieurs locus dépendamment qu'il soit en DL avec un ou plusieurs autres locus. On peut dire qu'il y a une perte de diversité au sein du panel de sorgho du Chibas même lorsque la diversité avant le goulot n'est pas connue. La région du petit bras du chromosome 6 ayant un locus SNP sous sélection a déjà été identifiée par Wang et al., en 2013 comme étant une zone conférant la résistance à *Melanaphis*. Etant donné qu'on n'a pas suffisamment de marqueurs pour couvrir le génome entier du sorgho, ça pourrait arriver que le locus SNP le plus proche de cette zone ne soit pas détecté. L'utilisation d'un nombre de marqueurs plus élevé serait intéressante pour voir s'il n'y a pas d'autres marqueurs liés à la résistance au puceron et de mieux circonscrire la zone du petit bras du chromosome 6 soumise au balayage sélectif.

5.4. Distribution des individus de la population d'entraînement en fonction des relations d'apparentement

La majorité des lignées de sorgho du Chibas est relativement apparentée. Ainsi, une partie des lignées est très apparentée et une autre partie est faiblement ou moyennement apparentée. D'après Beaulieu et al., 2014, un niveau d'apparentement trop faible ne permet pas d'avoir un bon niveau de précision dans la prédiction génomique. Tandis que Clark et al., 2013) montrent qu'un niveau d'apparentement trop élevé dans la population de référence provoque une réduction de l'hétérozygotie qui impacte négativement la fitness des individus. Le niveau

d'apparement des lignées du Chibas semble favorable pour faire de la prédiction génomique et aboutir à de bons niveaux de précision. Le fait que la population d'entraînement est étroitement liée à la population de validation, cela permet d'obtenir une précision élevée de la prédiction (Wang et al., 2014) pour la stratégie d'échantillonnage aléatoire. Dans le cas où les individus étaient non-apparentés, on aboutirait à de faibles niveaux de précision pour la prédiction. La stratégie de regroupement des lignées très apparentées avec DAPC permet d'éviter toute surestimation de la précision en éliminant les individus qui sont très apparentés dans la population de référence.

5.5. Modèle GBLUP

On a abouti à des résultats différents pour les deux stratégies d'échantillonnage utilisées pour la prédiction génomique. Dans la première stratégie, les phénotypes des individus très apparentés n'ont pas été pris en compte dans la population d'entraînement. Cette stratégie est très rigide car elle fait diminuer le niveau d'apparement entre les individus de la population d'entraînement et ceux de la population de validation. La précision de la prédiction chute quand les populations ne sont pas étroitement liées. Pour les caractères hauteur et degré brix, la précision de la prédiction est élevée avec la première stratégie. Une telle précision est expliquée par le fait que ces caractères sont très héréditaires. L'hérédité génomique, calculée à partir de la matrice d'apparement, est de 61 %, 56 % et 32 % respectivement pour les caractères hauteur, degré brix et rendement grain. Nous constatons que pour les caractères à hérédité élevée (brix et hauteur), les niveaux de précision obtenus sont également élevés même en supprimant les individus très apparentés.

Il apparaît donc que la structure de la population et les relations d'apparement sont suffisantes dans la population du Chibas pour réaliser des prédictions génomiques et d'envisager un gain génétique important. A travers la stratégie DAPC, le niveau de précision est plus faible que celui qu'on aurait pu obtenir. Les niveaux de précision obtenus via la stratégie DAPC sont expliqués par le plus faible niveau d'apparement entre les individus (Wang et al., en 2014 ; Beaulieu et al., 2014). Cependant, avec la stratégie d'échantillonnage aléatoire, les niveaux de précision sont très élevés par rapport à ceux obtenus avec la stratégie DAPC et sans doute surestimés par rapport aux prédictions génomiques qui seront obtenus pour des individus issus de croisements entre des individus faiblement ou moyennement apparentés. La stratégie DAPC illustre elle sans doute justement ce pire cas de figure.

Pour le rendement grain, la précision de la prédiction obtenue avec la stratégie DAPC est très faible (5 %). Il serait intéressant d'augmenter le nombre de groupes DAPC qui se traduit par une augmentation du niveau d'apparentement entre les individus, de travailler sur plusieurs répétitions qui permet d'augmenter l'héritabilité au niveau des moyennes (Aaron et al., 2013 ; Hayes et al., 2009) et enfin d'augmenter le nombre d'individus de la population d'entraînement (VanRaden et al., 2010 ; Technow et al., 2013 ; Grattapaglia et al., 2014) pour vérifier les facteurs qui permettront d'avoir une meilleure précision sur un caractère tel que le rendement grain. La stratégie qui vise à augmenter la taille de la population d'entraînement prendra quelques années pour créer les nouvelles lignées. Les résultats de l'étude de Aaron et al., 2013 démontrent que la précision de la prédiction augmente avec le nombre de répétitions et elle augmente beaucoup plus avec la taille de la population d'entraînement. Cependant, l'augmentation du nombre de répétitions est plus facilement réalisable à court terme puisqu'il faut quelques années pour développer plus de lignées pour augmenter la taille de la population d'entraînement.

Dans ce travail de recherche, on n'a utilisé que le modèle de prédiction GBLUP pour la prédiction génomique. Le modèle de prédiction « Bayes B » semble être mieux adapté pour des caractères héréditaires qui sont sous le contrôle d'un nombre limité de QTL. De ce fait, on aurait pu avoir une meilleure précision pour le caractère hauteur avec le modèle « Bayes B » qu'avec le modèle GBLUP. Le modèle GBLUP marche mieux que « Bayes B » pour les caractères complexes et faiblement héréditaires comme le rendement grain.

5.6. Densité de marquage

Le nombre de marqueurs utilisé (1 660 SNPs) pour la détection des traces de sélection dans cette étude est insuffisant pour une couverture complète du génome du sorgho. Ainsi, pour couvrir complètement le génome du sorgho, il faut environ 37 000 marqueurs SNPs. Cependant, avec ces 1 660 marqueurs, on a pu identifier une région au niveau du petit bras du chromosome 6 qui est sous sélection.

6. CONCLUSION ET PERSPECTIVES

Dans tout programme de sélection variétale, avant même de se lancer, il est primordial de caractériser la population qui va être utilisée. Le Chibas n'a pas pu caractériser les génotypes à partir des marqueurs moléculaires au démarrage de son programme d'amélioration en raison des coûts de génotypage qui étaient trop élevés. Ainsi, le Chibas a profité du partenariat avec les Universités du Kansas et de Cornell pour évaluer les lignées de sorgho qui ont été développées et d'étudier la possibilité de démarrage d'un programme de sélection génomique récurrente à partir de ces lignées qui ont connu depuis ces deux dernières années une pression de sélection très importante pour la résistance à *Melanaphis sacchari*, un puceron qui est apparu en automne 2015 en Haïti. Les technologies développées dans le cadre du partenariat avec les Universités du Kansas et de Cornell devraient permettre de rendre la sélection génomique compétitive d'un point de vue économique pour les sélectionneurs ne disposant aucune plateforme de génotypage. A travers ce partenariat, le coût du génotypage par individu est inférieur à 10 USD.

Malgré, la sélection sévère pour *Melanaphis sacchari*, la diversité génétique qui se trouve au sein des lignées du Chibas n'est pas si faible. Même pour le chromosome 6 l'empreinte de la sélection est limitée à une petite section du bras court. Toutefois, les résultats de cette étude ont montré que la majorité des parents qui ont été sélectionnés pour la création de ces lignées étaient des variétés d'origine du groupe Caudatum de l'Afrique de l'Est. Il apparaît essentiel d'élargir la variabilité allélique de la population en intégrant de la variabilité allélique en provenance d'autres races. Cependant, il faudra le faire en essayant de limiter les effets négatifs sur le progrès génétique. En termes d'apparentement, la distribution des relations d'apparentement entre les individus est très favorable pour faire de la prédiction génomique. Ceci est illustré par les bons niveaux de prédictions sur les caractères fortement héritables.

Dans un premier temps, il serait donc utile d'augmenter le nombre de répétitions pour le rendement grain afin d'augmenter l'héritabilité au niveau des moyennes (Aaron et al, 2013). Dans un deuxième temps, le Chibas pourra augmenter la taille de sa population d'entraînement (Aaron et al, 2013). Une telle stratégie semble être la plus influente pour améliorer la précision des prédictions génomiques.

Aux termes de ce travail de recherche, on est arrivé à :

- Mettre en évidence la structuration au sein des lignées de sorgho développées par le Chibas.
- Etudier la diversité génétique au sein de ces lignées.

- Détecter des traces de sélection au sein des lignées de sorgho développées par le Chibas.
- Mettre en évidence les relations d'apparentement entre les lignées de sorgho.
- Evaluer la qualité de la population d'entraînement pour la prédiction génomique.

BIBLIOGRAPHIE

1. Beaulieu, J., Doerksen, T., Clément, S., MacKay, J., Bousquet, J., 2014. Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. *Heredity* 113, 343–352.
2. Bernardo, R., Yu, J., 2007. Prospects for Genomewide Selection for Quantitative Traits in Maize. *Crop Sci.* 47, 1082. doi:10.2135/cropsci2006.11.0690
3. Bowling, R.D., Brewer, M.J., Kerns, D.L., Gordy, J., Seiter, N., Elliott, N.E., Buntin, G.D., Way, M.O., Royer, T.A., Biles, S., Maxson, E., 2016. Sugarcane Aphid (Hemiptera: Aphididae): A New Pest on Sorghum in North America. *J. Integr. Pest Manag.* 7, 12. doi:10.1093/jipm/pmw011
4. Burgueño, J., de los Campos, G., Weigel, K., Crossa, J., 2012. Genomic Prediction of Breeding Values when Modeling Genotype \times Environment Interaction using Pedigree and Dense Molecular Markers. *Crop Sci.* 52, 707. doi:10.2135/cropsci2011.06.0299
5. Calus, M.P.L., Meuwissen, T.H.E., de Roos, A.P.W., Veerkamp, R.F., 2008. Accuracy of Genomic Selection Using Different Methods to Define Haplotypes. *Genetics* 178, 553–561. doi:10.1534/genetics.107.080838
6. Calus, M.P.L., Veerkamp, R.F., 2007. Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. *J. Anim. Breed. Genet.* 124, 362–368.
7. Cassady, A., 1965. Effect of a Single Height (Dw) Gene of Sorghum on Grain Yield, Grain Yield Components, and Test Weight. *Crop Sci.* 5: 385 – 388.
8. Chantreau, J., Cruz, J., Ratnadass, A., Trouche, G., 2013. *Le sorgho*, Quae. ed. Presses agronomiques de Gembloux.
9. Charrier, A., CIRAD (Organization), O.R.S.T.O.M. (Agency : France) (Eds.), 1997. *L'amélioration des plantes tropicales*, Repères. CIRAD ; ORSTOM, Montpellier, France] : [Bondy, France.
10. Clark, S.A., Kinghorn, B.P., Hickey, J.M., van der Werf, J.H., 2013. The effect of genomic information on optimal contribution selection in livestock breeding programs. *Genet. Sel. Evol.* 45, 44.
11. Crossa, J., Campos, G. d. l., Perez, P., Gianola, D., Burgueno, J., Araus, J.L., Makumbi, D., Singh, R.P., Dreisigacker, S., Yan, J., Arief, V., Banziger, M., Braun, H.-J., 2010. Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers. *Genetics* 186, 713–724. doi:10.1534/genetics.110.118521
12. Daetwyler, H.D., Pong-Wong, R., Villanueva, B., Woolliams, J.A., 2010. The Impact of Genetic Architecture on Genome-Wide Evaluation Methods. *Genetics* 185, 1021–1031.
13. De Beukelaer, H., Badke, Y., Fack, V., De Meyer, G., 2017. Moving Beyond Managing Realized Genomic Relationship in Long-Term Genomic Selection. *Genetics* genetics.116.194449.
14. De los Campos, G., Hickey, J.M., Pong-Wong, R., Daetwyler, H.D., Calus, M.P.L., 2013. Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. *Genetics* 193, 327–345. doi:10.1534/genetics.112.143313
15. Deu, M., Rattunde, F., Chantreau, J., 2006. A global view of genetic diversity in cultivated sorghums using a core collection. *Genome* 49, 168–180.
16. Doggett, H. (1988) *Sorghum*. 2nd Edition, Longmans Scientific and Technical, UK.
17. El Naim, A.M., Ibrahim, I.M., Rahman, M.E.A., Ibrahim, E.A., 2012. Evaluation of some local sorghum (*Sorghum bicolor* L. Moench) Genotypes in Rain-Fed. *Int. J. Plant Res.* 2, 15–20.
18. Endelman, J.B., 2011. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *Plant Genome J.* 4, 250. doi:10.3835/plantgenome2011.08.0024
19. Endelman, J.B., Jannink, J.-L., 2012. Shrinkage Estimation of the Realized Relationship Matrix. *G3amp58 GenesGenomesGenetics* 2, 1405–1413. doi:10.1534/g3.112.004259
20. Falconer, D. S. 1989. *Introduction to Quantitative Genetics*. 3rd ed. Longman Scientific and Technical, New York, NY

21. Fartek, B., 2011. Résistance de la canne à sucre à la maladie de la feuille jaune et à son vecteur *Melanaphis sacchari*. Université de la Réunion.
22. Graham, D., Lessman, K., 1966. Effect of Height on Yield and Yield Components of Two Isogenic Lines of *Sorghum vulgare* Pers [WWW Document]. URL (accessed 9.29.17).
23. Grattapaglia, D. (2014). Breeding forest trees by genomic selection: Current progress and the way forward. In *Genomics of Plant Genetic Resources* (pp. 651-682). Springer Netherlands.
24. Habier, D., Fernando, R.L., Dekkers, J.C.M., 2008. The impact of genetic relationship information on genome-assisted breeding values. *Genetics*. doi:10.1534/genetics.107.081190
25. Habier, D., Tetens, J., Seefried, F.-R., Lichtner, P., Thaller, G., 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genet. Sel. Evol.* 42, 5.
26. Harlan, J., de Wet, J., 1972. A Simplified Classification of Cultivated Sorghum. *Crop Sci.* doi:10.2135/cropsci1972.0011183X001200020005x
27. Hayes, B.J., Bowman, P.J., Chamberlain, A.J., Goddard, M.E., 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* 92, 433–443.
28. Hayes, B.J., Goddard, M.E., others, 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.
29. Heffner, E.L., Jannink, J.-L., Iwata, H., Souza, E., Sorrells, M.E., 2011. Genomic Selection Accuracy for Grain Quality Traits in Biparental Wheat Populations. *Crop Sci.* 51, 2597.
30. Henderson, C.R., 1973. Sire evaluation and genetic trends. *J. Anim. Sci.* 1973, 10–41.
31. Hilley, J., 2017. Taking Sorghum to New Heights: Identification of Genes Controlling Height Variation in Sorghum.
32. House, L.R., Institut international de recherches sur les cultures des zones tropicales semi-arides, 1987. Manuel pour la sélection du sorgho. ICRISAT, Andhra Pradesh, Inde.
33. Jannink, J.-L., Lorenz, A.J., Iwata, H., 2010. Genomic selection in plant breeding: from theory to practice. *Brief. Funct. Genomics* 9, 166–177.
34. Jean-Paul Benzécri, Histoire et préhistoire de l'analyse des données. Dunod, Paris, 1982, 159 p. (ISBN 2-04-015467-1)
35. Jombart, T., Devillard, S., Balloux, F., 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11, 94.
36. Kumar, S., Chagné, D., Bink, M.C.A.M., Volz, R.K., Whitworth, C., Carlisle, C., 2012. Genomic Selection for Fruit Quality Traits in Apple (*Malus domestica* Borkh.). *PLoS ONE* 7, e36674. doi:10.1371/journal.pone.0036674
37. Lorenz, A.J., 2013. Resource Allocation for Maximizing Prediction Accuracy and Genetic Gain of Genomic Selection in Plant Breeding: A Simulation Experiment. *G3amp58 GenesGenomesGenetics* 3, 481–491. doi:10.1534/g3.112.004911
38. Ly, D., Hamblin, M., Rabbi, I., Melaku, G., Bakare, M., Gauch, H.G., Okechukwu, R., Dixon, A.G.O., Kulakow, P., Jannink, J.-L., 2013. Relatedness and Genotype × Environment Interaction Affect Prediction Accuracies in Genomic Selection: A Study in Cassava. *Crop Sci.* 53, 1312. doi:10.2135/cropsci2012.11.0653
39. Mace, E.S., Tai, S., Gilding, E.K., Li, Y., Prentis, P.J., Bian, L., Campbell, B.C., Hu, W., Innes, D.J., Han, X., Cruickshank, A., Dai, C., Frère, C., Zhang, H., Hunt, C.H., Wang, X., Shatte, T., Wang, M., Su, Z., Li, J., Lin, X., Godwin, I.D., Jordan, D.R., Wang, J., 2013. Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nat. Commun.* 4. doi:10.1038/ncomms3320
40. Mace, E.S., Xia, L., Jordan, D.R., Halloran, K., Parh, D.K., Huttner, E., Wenzl, P., Kilian, A., 2008. DArT markers: diversity analyses and mapping in *Sorghum bicolor*. *BMC Genomics* 9, 26. doi:10.1186/1471-2164-9-26

41. Martins, H., Caye, K., Luu, K., Blum, M.G.B., François, O., 2016. Identifying outlier loci in admixed and in continuous populations using ancestral population differentiation statistics. *Mol. Ecol.* 25, 5029–5042. doi:10.1111/mec.13822
42. McDonald, J.H., Kreitman, M., 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351, 652–654. doi:10.1038/351652a0
43. Najimi, B., El Jaafari, S., Jlibène, M., Jacquemin, J.-M., 2003. Applications des marqueurs moléculaires dans l'amélioration du blé tendre pour la résistance aux maladies et aux insectes. *Biotechnol. Agron. Société Environ.* 7, 17–35.
44. Nei, M., Chesser, R.K., 1983. Estimation of fixation indices and gene diversities. *Ann. Hum. Genet.* 47, 253–259.
45. Paterson, A.H., 2009. The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457, 551–556. doi:10.1038/nature07723
46. Prasada Rao, K., Mengesha, M., 1981. A pointed collection of zera-zera sorghums in the Gambella area of Ethiopia.
47. Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
48. Quinby JR, Karper RE (1954) Inheritance of height in sorghum. *Agron J* 46:211–216
49. Rami, J.-F., Dufour, P., Trouche, G., Fliedel, G., Mestres, C., Davrieux, F., Blanchard, P., Hamon, P., 1998. Quantitative trait loci for grain quality, productivity, morphological and agronomical traits in sorghum (*Sorghum bicolor* L. Moench). *Theor. Appl. Genet.* 97, 605–616.
50. Ramu, P., Billot, C., Rami, J.-F., Senthilvel, S., Upadhyaya, H.D., Ananda Reddy, L., Hash, C.T., 2013. Assessment of genetic diversity in the sorghum reference set using EST-SSR markers. *Theor. Appl. Genet.* 126, 2051–2064. doi:10.1007/s00122-013-2117-6
51. Ritter, K.B., McIntyre, C.L., Godwin, I.D., Jordan, D.R., Chapman, S.C., 2007. An assessment of the genetic relationship between sweet and grain sorghums, within *Sorghum bicolor* ssp. *bicolor* (L.) Moench, using AFLP markers. *Euphytica* 157, 161–176. doi:10.1007/s10681-007-9408-4
52. Robert-Granié, C., Legarra, A., Ducrocq, V., 2011. Principes de base de la sélection génomique. *Prod. Anim.* 24, 331.
53. Rooney, W.L., Blumenthal, J., Bean, B., Mullet, J.E., 2007. Designing sorghum as a dedicated bioenergy feedstock. *Biofuels Bioprod. Biorefining* 1, 147–157.
54. Solberg, T.R., Sonesson, A.K., Woolliams, J.A., Meuwissen, T.H.E., 2008. Genomic selection using different marker types and densities. *J. Anim. Sci.* 86, 2447–2454. doi:10.2527/jas.2007-0010
55. Technow, F., Bürger, A., & Melchinger, A. E. (2013). Genomic prediction of northern corn leaf blight resistance in maize with combined or separated training sets for heterotic groups. *G3: Genes|Genomes| Genetics*, 3(2), 197-203.
56. VanRaden, P.M., 2008. Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* 91, 4414–4423. doi:10.3168/jds.2007-0980
57. VanRaden, P. M., & Sullivan, P. G. (2010). International genomic evaluation methods for dairy cattle. *Genetic, Selection and Evolution*, 42(7).
58. Wang, F., Zhao, S., Han, Y., Shao, Y., Dong, Z., Gao, Y., Zhang, K., Liu, X., Li, D., Chang, J., Wang, D., 2013. Efficient and fine mapping of RMES1 conferring resistance to sorghum aphid *Melanaphis sacchari*. *Mol. Breed.* 31, 777–784. doi:10.1007/s11032-012-9832-6
59. Wang, Y.-H., Upadhyaya, H.D., Burrell, A.M., Sahraeian, S.M.E., Klein, R.R., Klein, P.E., 2013. Genetic Structure and Linkage Disequilibrium in a Diverse, Representative Collection of the C4 Model Plant, *Sorghum bicolor*. *G3amp58 GenesGenomesGenetics* 3, 783–793. doi:10.1534/g3.112.004861
60. Wang, Y., Mette, M. F., Miedaner, T., Gottwald, M., Wilde, P., Reif, J. C., & Zhao, Y. (2014). The accuracy of prediction of genomic selection in elite hybrid rye populations

surpasses the accuracy of marker-assisted selection and is equally augmented by multiple field evaluation locations and test years. *BMC genomics*, 15(1), 556.

61. Whitlock, M.C., Lotterhos, K.E., 2015. Reliable Detection of Loci Responsible for Local Adaptation: Inference of a Null Model through Trimming the Distribution of F_{ST} . *Am. Nat.* 186, S24–S36. doi:10.1086/682949
62. Wright, S., 1949. The genetical structure of populations. *Ann. Hum. Genet.* 15, 323–354.

ANNEXE

Tableau 10: Les traits évalués pendant le phénotypage

Paramètres évalués au champ	Mode d'évaluation	Outils utilisés/évaluation
Taux de germination (%)	Le nombre de poquets ayant au moins une plantule levée divisé par le nombre total de poquets au sein de chaque parcelle variétale.	
Vigueur à la levée	La vigueur est évaluée sur une échelle de 1 à 9. La note 1 est donnée pour une très bonne vigueur et la note 9 est donnée pour une très mauvaise vigueur.	Visuelle
Homogénéité à la levée	L'homogénéité est évaluée sur une échelle de 1 à 9. La note 1 est donnée pour une parcelle variétale qui est très homogène et la note 9 pour une parcelle variétale qui très hétérogène.	Visuelle
Quantité de pucerons	La quantité de pucerons est évaluée avant et après épiaison. La note 0 est attribuée quand les pucerons ne sont pas détectés au niveau des parcelles variétales. La note 1 quand il y a très peu et la note 9 quand il y en a beaucoup.	Visuelle
Durée du cycle	Cette durée est obtenue à partir de la différence entre les dates de semis et de maturité.	
Diamètre tige (mm)	Le diamètre est mesuré sur 5 plantes au niveau du collet des plantes après l'épiaison. Cette mesure est prise sur 5 plantes dans chaque parcelle variétale.	Pied à coulisse
Hauteur	La hauteur est mesurée à partir du collet jusqu'à la feuille drapeau. Cette mesure est prise sur 15 plantes dans chaque parcelle variétale.	PVC gradué
degré brix	Le degré brix est mesuré au moment de la récolte.	Refractomètre
Jutosité	Le volume de jus est mesuré au moment de la récolte.	Pot gradué
Poids des tiges, feuilles et panicules	Les tiges, les feuilles et les panicules des deux rangées du milieu sont pesées.	Balance

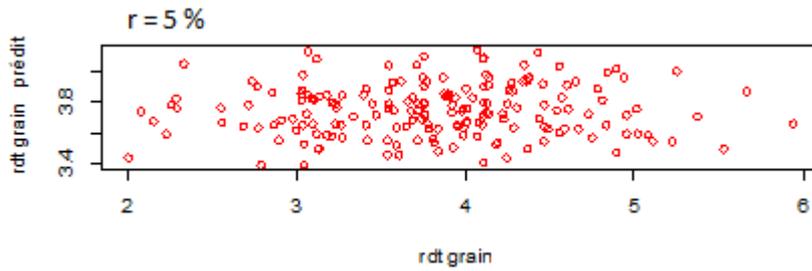


Figure 14: Corrélation entre rendement grain observé et rendement grain prédit

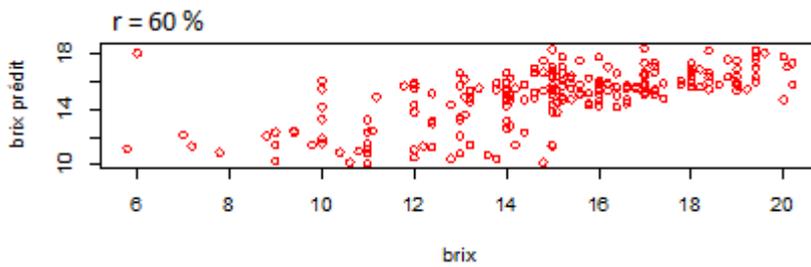


Figure 15 : Corrélation entre brix observé et brix prédit

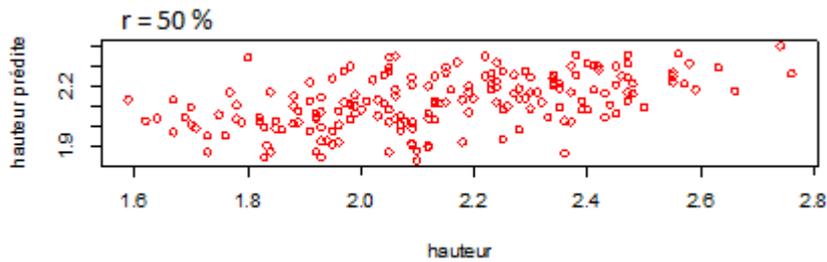


Figure 16 : Corrélation entre hauteur observée et hauteur prédite

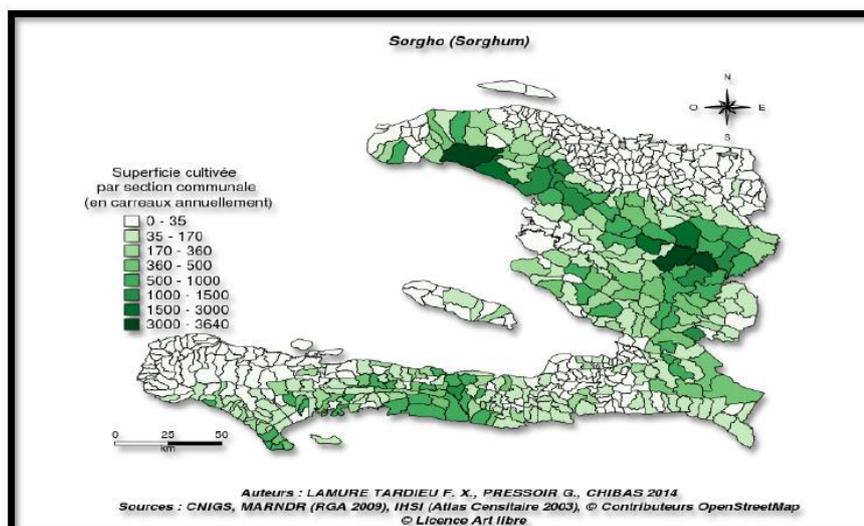


Figure 17 : Répartition des zones de culture du sorgho en Haïti

b o r d u r e	1	Bordure														
		DA-073		DB-077		DA-074		DA-081		DA-024		DB-039		DA-073		DB-039
		DA-103		DA-001		DB-021		DA-075		DB-011		DB-020		DA-103		DB-020
		C-23		DA-056		DB-090		DA-061		DA-104		C-18		C-23		C-18
		DB-005		DA-070		DA-092		DA-033		DB-106		DB-084		DB-005		DB-084
		DA-015		DB-003		C-22		DA-068		DB-037		DA-060		DA-015		DA-060
		DB-009		DA-038		DB-062		DB-052		DB-024		DB-064		DB-009		DB-064
		DA-002		DA-055		C-7		DB-046		DA-008		DB-023		DA-002		DB-023
		C-24		C-10		DB-105		DB-068		C-12		C-17		C-24		C-17
		DA-059		C-6		C-11		DB-004		DB-017		DA-020		DA-059		DA-020
		DA-095		DA-088		C-14		DA-046		DA-090		DA-048		DA-095		DA-048
		DA-047		DA-043		DA-054		C-19		DA-028		DB-010		DA-047		DB-010
		DB-012		DB-008		DA-080		DA-006		DB-025		C-16		DB-012		C-16
		DA-049		DA-101		C-4		DB-048		DB-002		DA-051		DA-049		DA-051
		DA-062		DB-056		DB-097		DB-027		DA-093		DB-098		DA-062		DB-098
		DB-092		DB-031		DA-005		DA-014		DB-080		DA-102		DB-092		DA-102
		DB-042		DA-057		DA-029		DB-038		DA-065		DB-066		DB-042		DB-066
		DA-022		DA-067		DA-089		DB-030		DA-018		DA-012		DA-022		DA-012
		C-9		DA-025		DB-075		DB-040		C-20		DB-035		C-9		DB-035
		DA-031		DB-079		DA-078		DA-091		DB-015		DB-029		DA-031		DB-029
		DB-054		DB-049		DB-060		DA-021		DA-058		DA-009		DB-054		DA-009
		DA-106		DB-047		DA-082		C-2		DB-050		C-21		DA-106		C-21
		DB-061		DB-081		C-15		DA-007		DA-053		C-3		DB-061		C-3
		DB-095		DB-018		DB-088		DB-063		C-8		DB-091		DB-095		DB-091
		DB-044		DA-034		DB-069		DB-006		DB-041		DA-017		DB-044		DA-017
		DA-011		D-2		DA-026		DA-010		DA-052		DB-036		DA-011		DB-036
		DA-044		DB-103		DA-094		D-1		DA-023		DA-050		DA-044		DA-050
		DB-071		C-13		DA-097		DA-087		DB-100		DB-070		DB-071		DB-070
		DA-086		DB-086		D-3		DB-016		DB-022		DB-074		DA-086		DB-074
		DA-039		DA-037		DB-078		DB-051		DB-014		DB-096		DA-039		DB-096
	DB-085		DB-001		DB-033		DB-082		DA-016		DB-028		DB-085		DB-028	
	C-1		DA-004		DA-096		DB-072		DB-073		DB-019		C-1		DB-019	
	DA-041		DB-099		DB-087		DA-040		DB-058		DA-036		DA-041		DA-036	
	DB-076		DA-035		DB-104		DA-077		C-5		DB-032		DB-076		DB-032	
	DA-013		DB-101		DA-085		DB-093		DB-007		DB-065		DA-013		DB-065	
	DA-064		DB-026		DB-045		DA-003		DA-030		DB-089		DA-064		DB-089	
	DB-013		DA-063		DB-055		DB-094		DA-083		DB-043		DB-013		DB-043	
	DA-099		DA-105		DB-083		DB-053		DA-098		DB-102		DA-099		DB-102	
	Bordure		Bordure		Bordure		Bordure		Bordure		Bordure		Bordure		Bordure	

Figure 18: Plan du dispositif expérimental

Tableau 11: Panel de diversité de sorgho utilisé dans cette étude.

Races de sorgho	Effectifs (n)
bicolor	57
bicolor-caudatum	56
bicolor-durra	32
bicolor-guinea	6
bicolor-kafir	5
caudatum	163
caudatum-durra	48
caudatum-guinea	67
caudatum-kafir	26
durra	81
durra-guinea	7
durra-kafir	5
guinea	96
guinea-kafir	6
kafir	47
NK	551
wild type	30

Participation à ce travail de recherche

Au cours de ce stage au Chibas, j'avais la responsabilité de gérer mon essai expérimental du semis à la récolte. Je gère une équipe qui assure l'entretien des expériences du Chibas au champ. La majorité des données de phénotypage est évaluée par moi. J'avais fait le prélèvement des échantillons de feuilles de sorgho pour l'extraction d'ADN. Et, les différentes analyses ont été réalisées par moi sous la vérification de mon maître de stage.

En fait, j'ai commencé à travailler au Chibas en 2012. J'ai été recruté à titre d'assistant à la recherche au Chibas. J'étais rentré au Chibas dans une période où il y avait beaucoup de croisements à réaliser. Ainsi, j'ai été formé par M. Pressoir pour réaliser ces croisements. J'ai été aussi sur toutes les expériences du Chibas. Et, par la suite, le Chibas m'avait donné l'opportunité de faire ce master pour que je puisse être utile au centre de recherche et améliorer les capacités de l'équipe d'enseignants-chercheurs du Chibas et de l'Université Quisqueya.