



**HAL**  
open science

**Analyse des déterminants génétiques contrôlant la  
production et la composition de la tige chez le sorgho  
(*Sorghum bicolor* [L.] Moench). Intégration des  
approches bi-et multi-parentales**

Korotimi Thera

► **To cite this version:**

Korotimi Thera. Analyse des déterminants génétiques contrôlant la production et la composition de la tige chez le sorgho (*Sorghum bicolor* [L.] Moench). Intégration des approches bi-et multi-parentales. Amélioration des plantes. Montpellier SupAgro, 2017. Français. NNT : 2017NSAM0030 . tel-04543717

**HAL Id: tel-04543717**

**<https://theses.hal.science/tel-04543717>**

Submitted on 12 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE MONTPELLIER SUPAGRO

En Sciences Agronomiques et Ecologie

École doctorale GAIA – Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau  
Portée par l'Université de Montpellier

Unité de recherche AGAP

## Analyse des déterminants génétiques contrôlant la production et la composition de la tige chez le sorgho (*Sorghum bicolor* [L.] Moench). Intégration des approches bi-et multi-parentales

Présentée par Mme Korotimi Théra  
Le 22 novembre 2017

Sous la direction de Dominique This

Devant le jury composé de

Gilles Charmet, Directeur de Recherches, INRA  
Charles-Eric Durel, Directeur de Recherches, INRA  
Mathilde Causse, Directrice de Recherches, INRA  
Matthieu Reymond, Chargé de Recherches, INRA  
Dominique This, Maître de conférences, SupAgro  
Niaba Tiémé, Chercheur, IER  
Michel Vaksman, Chargé de Recherches, CIRAD  
David Pot, Chargé de Recherches, CIRAD

Rapporteur  
Rapporteur  
Examinatrice  
Examineur  
Directrice de thèse  
Invité  
Invité  
Invité



UNIVERSITÉ  
DE MONTPELLIER





## Résumé

Le sorgho est une des espèces de céréales les plus importantes au niveau mondial en termes d'alimentation humaine et animale. En outre, de nouvelles voies de valorisations liées à son haut potentiel de production de biomasse émergent (énergie, biomatériaux) avec des possibilités d'applications aux niveaux des pays du Nord et du Sud. L'identification des facteurs génétiques contrôlant la variabilité de la production et de la qualité de la biomasse est un des vecteurs de développement de variétés adaptées aux différentes utilisations. Actuellement, seule une vision partielle de l'architecture génétique de la production et de la qualité de la biomasse est disponible, pénalisant l'efficacité des programmes de sélection. Les principales raisons qui peuvent être évoquées sont : l'utilisation quasi exclusive de populations à base génétique très étroite (population biparentales correspondant à des idéotypes peu divers) et un manque de compréhension des interactions existantes entre la qualité de la biomasse et la phénologie. Dans ce contexte, cette thèse s'est articulée autour de 3 objectifs : i) apporter une meilleure compréhension du déterminisme génétique de la durée du cycle « semis - floraison » et de la hauteur des plantes dans le contexte de variétés sensibles à la photopériode, lesquelles ont été relativement peu travaillées jusqu'à présent, ii) analyser les interactions entre la qualité de la biomasse, la croissance en hauteur, la durée du cycle et la sensibilité au photopériodisme et iii) affiner notre compréhension des mécanismes moléculaires impliqués dans la variation de la qualité de la biomasse. Pour mener à bien ces analyses, deux populations de cartographie biparentales, un panel de diversité correspondant aux 35 parents d'un schéma de croisement multiparental (Backcross Nested Association Mapping) ainsi que leurs descendance (29 populations totalisant environ 1200 BC1F4) ont été caractérisés en termes de croissance, de phénologie et de qualité de la biomasse. L'analyse du déterminisme génétique de la phénologie couplée à une approche de modélisation écophysiological a permis la mise en évidence d'une région chromosomique contrôlant une large part de la variabilité de la photopériode critique. Le gène ELF3 pour lequel plusieurs évidences fonctionnelles ont été mises en évidence chez d'autres espèces, est un candidat potentiel pour expliquer cet effet. Outre une large convergence entre les zones détectées pour la phénologie au sein de cette thèse et les études précédentes, une forte instabilité des zones détectées en fonction des conditions environnementales (dates de semis, année) a été mise en évidence. L'analyse des relations entre la qualité de la biomasse, la croissance en hauteur, la précocité et la sensibilité à la photopériode a mis en évidence des effets importants du nanisme et de la date de semis sur la composition des tiges et des relations entre ses composantes. En outre, des différences de réactions à la variabilité des dates de semis ont été observées en fonction de la sensibilité à la photopériode des génotypes considérés. Enfin, la caractérisation de la composition biochimique des tiges (teneur en cellulose, en hémicellulose et en lignine des parois) au sein des 29 populations BC1F4, a mis en évidence de nombreuses zones chromosomiques pour lesquelles des gènes potentiellement causatifs ont été proposés. Ces analyses ont permis d'alimenter un répertoire de régions chromosomiques contrôlant les caractères cibles de la sélection. Les potentialités de ces résultats et des populations développées (bi- et multiparentales) dans le contexte de la sélection ainsi que de la validation des gènes sous-jacents à ces zones sont discutées.

## Abstract

Sorghum ranks fifth in terms of grain production at the worldwide level. It contributes to food security in the developing countries and provides a significant share of the animal caloric intake. In addition, sorghum is expected to contribute significantly to the emerging energy and biomaterial value chains. In this context, identification of the genetic factors controlling biomass production and quality is of crucial interest to develop varieties fitting the expectations of the different uses. Currently only a partial vision of the genetic architecture of biomass production and quality is available, hampering optimal breeding efficiencies. This can be explained by a limited use of the genetic diversity pool to dissect the genetic control of the traits of interest (mainly restricted to non-photoperiod sensitive varieties until now) and a limited understanding of the relationships between biomass production (largely driven by phenology variability) and quality. In this context, this PhD aimed at i) providing a better understanding of the genetic bases of flowering and height variabilities using photoperiod sensitive genetic material ii) disentangling the interactions between biomass quality, height, cycle duration and photoperiod sensitivity and iii) refining our understanding of the molecular mechanisms involved in the variation in biomass quality. To tackle these challenges, two biparental QTL mapping populations, a diversity panel including 35 parents of a Backcross Nested Association Mapping design and a subset of their progenies (29 populations encompassing a total of 1200 BC1F4) were characterized for their phenology, growth and biomass composition. The analysis of the genetic determinism of phenology coupled with an ecophysiological modeling approach, revealed a chromosomal region controlling a large part of the variability (60%) of critical photoperiod (a component of cycle duration). The gene ELF3, for which functional evidence of impact on photoperiod sensitivity is available in other species can be proposed as a functional driver of the variability induced by this region. Although a significant proportion of the genomic regions detected in this study have already been revealed in previous studies, several new regions have been discovered, underlying the interest of multiparental based designs. Phenology analyses over different sowing dates highlighted the complexity of the genetic determinism of this trait through the detection of genomic regions with variable effects. In depth analyses of biomass quality relationships with height, precocity and photoperiod sensitivity underlined strong dwarfism and sowing dates effects on the components of biomass quality and their relationships. In addition, different behaviors of photoperiod sensitive and insensitive genotypes to delayed sowing dates have been observed. Finally, characterization of stem biochemical composition (cellulose, hemicellulose and lignin content of the cell walls) on the 29 BC1F4 populations, revealed numerous chromosomal regions with variable effects. Putative candidate genes from these regions of interest have been proposed. These analyses enriched a library of regions impacting the breeding targets linked to biomass production and quality. The potentials of these results and of the populations developed towards optimization of breeding efficiency and gene function validation have been discussed.

## Remerciements

---

J'adresse mes vifs et sincères remerciements à l'institut d'Economie Rurale du Mali (IER) pour m'avoir accueilli dans le cadre de ce projet de thèse. Je remercie tout particulièrement les Drs Niaba Témé et Mamoutou Kouressy de l'Unité Agro-climatologie du programme sorgho.

Mes remerciements vont aussi au CIRAD et en particulier à tout le personnel de l'équipe DDSE dirigée par Claire Billot et de l'UMR AGAP pilotée par Patrick This pour m'avoir accueillie au sein de leur équipe lors des passages en France.

Je remercie également le Service de Coopération et d'Action Culturelle « SCAC » de l'Ambassade France Mali pour avoir financé une partie de mes séjours à Montpellier.

Je voudrai aussi remercier chaleureusement ma Directrice de thèse Mme This Dominique pour son soutien et ses conseils.

Un grand merci à tous les membres de mon comité de thèse : Yves Vigouroux, Michel Vaksman, Jean François Rami, Matthieu Reymond, Delphine Luquet, Baptiste Guitton, Gilles Trouche pour les précieux conseils.

Mes vifs remerciements s'adressent également à mon encadrant Dr. David Pot, pour son soutien, sa rigueur, ses encouragements et surtout pour sa capacité d'écoute.

Je tiens à remercier toutes les personnes qui ont bien voulu participer à mon Jury de thèse : Gilles Charmet, Charles-Eric Durel, Mathilde cause, Matthieu Reymond, Niaba Temé, Dominique This, Michel Vaksman et David Pot.

Et enfin un grand merci à toutes les personnes qui ont contribué de près ou de loin à la réalisation de cette thèse.

Je ne saurais tourner cette page sans adresser mes plus sincères remerciements et ma grande reconnaissance à toute ma famille : ma mère, mes frères et sœurs, mes cousins, cousines, nièces et neveux ainsi que mes ami(e)s pour leurs soutiens tout le long de cette thèse.

# Sommaire

---

<b>Remerciements .....</b>	<b>4</b>
<b>Introduction générale.....</b>	<b>9</b>
<b>Contexte de la thèse.....</b>	<b>11</b>
<b>Chapitre I : Synthèse bibliographique .....</b>	<b>13</b>
<b>I. I. Le sorgho, une espèce « riche en ressources » .....</b>	<b>13</b>
I.1- Une espèce d'intérêt au niveau mondial.....	13
I.2- Un portefeuille d'utilisations possibles .....	13
I.2.1- Alimentation humaine .....	14
I.2.2- Alimentation animale .....	14
I.2.3- Bioénergie .....	14
I.2.4- Biomatériaux.....	16
I.3- Besoins pédo-climatiques, écophysiologicals et itinéraires culturels.....	16
I.4- Un réservoir de diversité vecteur de progrès génétique.....	17
<b>II. II. La floraison, une étape clé en termes évolutifs et agronomiques.....</b>	<b>18</b>
<b>III. III. La qualité de la biomasse .....</b>	<b>19</b>
III.1- Les composantes de la qualité de la biomasse.....	19
III.2- Déterminisme génétique de la qualité de la biomasse .....	20
<b>IV. IV. La sélection du sorgho.....</b>	<b>21</b>
IV.1- Les cibles premières de la sélection du sorgho pour les environnements tempérés.....	21
IV.2- La sélection du sorgho en Afrique de l'Ouest .....	22
IV.2.1- Un début difficile : l'échec « relatif » du modèle de révolution verte.....	22
IV.2.2- Une approche plus proche des attentes des agriculteurs et des contextes locaux : la sélection participative .....	22
IV.2.3- Les objectifs actuels de la sélection en Afrique de l'Ouest.....	22
IV.2- Les méthodes de sélection du sorgho .....	23
IV.4- Optimiser l'efficacité de la sélection au travers de l'utilisation de l'information moléculaire .....	24
<b>V. Améliorer la compréhension du déterminisme génétique de la production et de la qualité de la biomasse : stratégie .....</b>	<b>26</b>
<b>Chapitre II : Etude du déterminisme génétique de la floraison et de la hauteur des plantes chez le sorgho .....</b>	<b>28</b>
<b>I. La date de floraison et la hauteur des plantes : des caractères clés de la sélection chez le sorgho .....</b>	<b>28</b>
I.1- La date de floraison : des stratégies de sélection mouvantes et une connaissance du déterminisme génétique encore fragmentaire .....	28
I.1.1 – La date de floraison : des stratégies de sélection « mouvantes ».....	28
I.1.2– Les composantes de la longueur du cycle chez le sorgho .....	30
I.1.3– Déterminisme génétique et moléculaire de la longueur du cycle chez le sorgho .....	31
I.2 – La taille des plantes : Sélection et déterminisme génétique .....	34
I.2.1- Multi-utilisations vs intensification.....	34

I.2.2- Déterminisme génétique de la hauteur chez le sorgho.....	35
I.3 – Comment affiner notre compréhension du déterminisme génétique de la longueur du cycle et de la hauteur chez le sorgho ?.....	37
<b>II. Integrating genetic analysis and crop modeling: a major QTL can finely adjust photoperiod-sensitive sorghum flowering.....</b>	<b>40</b>
II.1. Article soumis à Plant Cell and Environment.....	40
II-2. Principaux résultats et discussion de l'article « Integrating genetic analysis and crop modeling: a major QTL can finely adjust photoperiod-sensitive sorghum flowering ».....	66
II.2.1. Analyse de la population P118 : Intégration de l'analyse génétique et de la modélisation du développement.....	66
II-2.2. La modélisation permet de décomposer le caractère complexe de floraison.....	67
II-2.3. Un QTL majeur à effet faible permet d'envisager un réglage fin de la sensibilité à la photopériode. ....	67
<b>III. Analyse du déterminisme génétique de la floraison et de la croissance en hauteur au sein d'une population biparentale de type rétrocroisement avancé (BC1F4).....</b>	<b>69</b>
III.1. BC06 : Matériels et méthodes .....	70
III.1.1- Développement de la population BC06 : E36-1 (parent donneur) X Keninkeni (parent récurrent) .....	70
III.1.2. Site expérimental et conditions environnementales .....	71
III.1.3. Dispositif expérimental .....	71
III.1.3.1- Hivernage 2014 .....	71
III.1.3.2- Hivernage 2015 .....	72
III.1.3.3- Contre saisons 2013 et 2015 .....	72
III.1.4. Caractérisation phénotypique : Phénologie et croissance .....	72
III.1.5 Analyse des données phénotypiques.....	74
III.1.5.1- Vérification des données.....	74
III.1.5.2- Ajustements des données en prenant en compte la structure des dispositifs expérimentaux .....	74
III.1.5.3- Corrélations phénotypiques.....	75
III.1.5.4- Estimation des variabilités génétiques.....	75
III.1.5.5- Analyse des effets de date de semis et des interactions « Génotype x Date de semis ».....	76
III.1.6 – Génotypage de la population BC06.....	76
III.1.7. Stratégie de détection de QTL au sein de la population BC06.....	79
III.2. Résultats .....	81
III.2.1. Variabilité des caractères et héritabilités .....	81
III.2.2. Identification des QTL au sein de la population BC06.....	83
III.2.2.1- Des régions chromosomiques à effet fort associées à la variation de la durée semis floraison et à la sensibilité à la photopériode .....	83
III.2.2.2- Régions chromosomiques associées à la hauteur des tiges et à ses composantes .....	84
III.2.2.3- Stabilité des QTL en fonction de la date de semis .....	85
III.3. Discussion .....	87
III.3.1- Durée du cycle « Semis – Floraison » : De nombreux QTL validés entre études et un nouveau QTL majeur mis en évidence sur le chromosome 3 .....	87
III.3.2- La hauteur et ses composantes : des gènes majeurs confirmés et plusieurs nouvelles régions à effet plus faibles mises en évidence. ....	88
III.3.3- Co-localisations détectées entre les caractères.....	88
III.3.4- Stabilité des QTL détectés entre les dates de semis.....	89

<b>IV. Analyse du déterminisme génétique de la durée du cycle et de la croissance en hauteur sur la population « GCP- BCNAM »</b> .....	<b>90</b>
IV.1. Le dispositif « GCP-BCNAM » : des ambitions pour « l’analyse génétique » et la « sélection ».....	90
IV.2. Matériels et Méthodes.....	90
IV.2.1- Matériel végétal .....	90
IV.2.1.1- GCP-BCNAM : des ambitions et des contraintes.....	90
IV.2.1.2- Variétés parentales du dispositif BCNAM .....	91
IV.2.1.3- Développement des populations BCNAM .....	91
IV.2.2. Dispositifs expérimentaux .....	91
IV.2.2.1- Essai 2014 .....	91
IV.2.2.2- Essai 2015 .....	92
IV.2.3- Caractérisation phénotypique des 29 populations BCNAM : Phénologie et croissance .....	93
IV.2.4. Analyse des données phénotypiques .....	94
IV.2.5. Génotypage des populations BCNAM.....	95
IV.2.6. Identification des zones chromosomiques impliquées dans le contrôle de la durée du cycle et de la croissance aérienne.....	95
<b>IV. 3. Résultats</b> .....	<b>97</b>
IV. 3.1. Analyse phénotypique des 29 populations BCNAM .....	97
IV. 3.1.1- Variabilité inter et intra population des caractères mesurés et calcul d’héritabilité.....	97
IV.3.1.2- Corrélations phénotypiques entre les caractères de phénologie et de croissance des familles BCNAM en fonction des années .....	98
IV.3.2-Analyse d’association « phénotype-génotype ».....	99
IV.3.2.1- Des régions génomiques à effet majeur sont associées à la variation de la durée du cycle et de la hauteur des tiges et de ses composantes.....	99
IV.3.2.2- Les longueurs du pédoncule et de la panicule sont majoritairement contrôlées par des régions génomiques à effet faibles .....	101
IV.3.2.3- Stabilité des régions chromosomiques détectées en fonction des années .....	101
IV.3.2.4- Une même région génomique associée à la variation de la hauteur des tiges et de ses composantes .....	102
IV.4. Discussion : .....	102
IV.4.1- Comparaison des régions chromosomiques identifiées et la bibliographie .....	102
IV.4.2- Colocalisations entre QTL de caractères différents .....	104
IV.4.3. Stabilité des QTL détectés en fonction des années .....	104
<b>Chapitre III : Relations entre production et qualité de la tige : variabilité génétique et impact de la date de semis</b> .....	<b>105</b>
I- Variabilité génétique de la composition de la biomasse .....	105
II- Impacts des conditions de culture sur la composition de la biomasse.....	106
III- Vers une compréhension plus fine de la variabilité de la composition de la biomasse .....	106
IV- Article en cours de finalisation à soumettre à Field Crop Research .....	106
<b>V. Principaux résultats et discussion de l’article : « Genotype and sowing date effects on the production and composition of sorghum stems, disentangling the contributions of dwarfism, earliness and photoperiod sensitivity »</b> .....	<b>141</b>
V.1- Un réservoir de diversité génétique pour la composition de la biomasse : un atout vers le développement de variétés adaptées aux différentes voies de valorisation .....	141



## Introduction générale

---

L'identification des facteurs génétiques contrôlant la variabilité des caractères d'intérêts agronomiques et adaptatifs est un des pivots de la gestion durable et efficace des ressources génétiques. Cette gestion a une importance primordiale, non seulement en raison de son intérêt à court terme pour l'amélioration génétique mais aussi à plus long terme pour garantir la sécurisation du réservoir de diversité génétique qui permettra de faire face aux nouveaux challenges de l'agriculture.

Le sorgho est une des espèces de céréales les plus importantes en termes d'alimentation humaine et animale au niveau mondial (*FAOSTAT*, 2017). Dans de nombreux pays en voie de développement, il constitue un des pivots de la sécurité alimentaire (Frère *et al.*, 2011). Dans les pays développés, où il est principalement utilisé pour l'alimentation animale, de nouvelles utilisations émergent pour l'énergie et les biomatériaux (Mathur *et al.*, 2017; Vo *et al.*, 2017) en raison de caractéristiques agronomiques favorables telles que la tolérance à la sécheresse et le potentiel de rendement en biomasse. Outre ces différentes utilisations et son intérêt agronomique, sa proximité phylogénétique avec plusieurs autres espèces de graminées d'intérêt et la faible complexité de son génome contribuent à son intérêt à un niveau fondamental pour la compréhension de l'organisation du génome et de possibles transferts de connaissances vers d'autres espèces (Ming *et al.*, 1998, Paterson *et al.*, 2004, Paterson *et al.*, 2009, Vermerris, 2011, Mullet *et al.*, 2014, McCormick *et al.*, 2017). De nombreux programmes de sélection du sorgho existent dans le monde avec des objectifs dépendant des contextes régionaux (types d'utilisations, et conditions climatiques). Outre l'adaptation aux conditions locales de cultures (contextes pédoclimatiques et parasitaires), l'objectif majeur qui a été, et est toujours poursuivi concerne l'amélioration du rendement en grains. Néanmoins, la production de biomasse végétative qui a été assez peu ciblée dans les programmes de sélection (tant au Nord qu'au Sud) revient aujourd'hui sur le devant de la scène, en raison de son importance pour l'alimentation animale et dans les filières de valorisation de la biomasse.

L'optimisation de la production de biomasse végétative comprend deux défis, d'une part l'optimisation de son rendement (qui doit être couplée ou non à de la production de grains en fonction des contextes de valorisation ciblés) et d'autre part l'adaptation de sa composition aux différentes utilisations.

Dans ce contexte, trois types de caractères cibles principaux sont à considérer : la phénologie, la quantité de biomasse produite et sa composition. Les deux premiers vont contribuer au rendement en biomasse et le dernier orienter le type d'utilisation possible.

L'objectif de cette thèse est de contribuer à une meilleure compréhension du déterminisme génétique de ces caractères afin de guider le développement de variétés adaptées aux différentes voies de production et de valorisation. Pour répondre à cet objectif, une stratégie en 3 étapes a été mise en place. Dans un premier temps, le déterminisme génétique de la durée du cycle « semis - floraison » et de la hauteur des plantes a été analysé au sein de 3 dispositifs complémentaires (deux populations biparentales contrastantes et un dispositif multi-parental). Dans un second temps, la variabilité de la qualité de la biomasse a été analysée sur un panel de diversité pertinent en termes de sélection durant trois années. En outre, les interactions entre la qualité de la biomasse, la croissance en hauteur, la durée du cycle et la sensibilité à la photopériode ont été explorées dans une perspective de pilotage de la qualité de la biomasse dans différentes configurations (tailles des plantes, phénologie variable, etc.) Enfin dans une troisième partie, le déterminisme génétique de la qualité de la biomasse a été analysé au sein d'un dispositif de croisement multi-parental de type BackCross Nested Association Mapping (BCNAM) (Jordan *et al.*, 2011) intéressant non seulement en termes d'analyse génétique mais aussi de sélection dans le but de développer des outils génomiques d'appui à la sélection.

Cette thèse est structurée en quatre chapitres. Un premier chapitre présente une synthèse bibliographique visant à clarifier l'intérêt du sorgho et fournir un aperçu des connaissances disponibles et des manques sur les différents caractères ciblés. Trois chapitres présentent ensuite les trois étapes de la thèse précédemment exposées, pour lesquels une présentation plus précise des connaissances déjà disponibles vient compléter l'introduction générale, suivi d'un exposé des stratégies mises en œuvre et des résultats obtenus pour répondre spécifiquement aux objectifs de cette thèse.

## Contexte de la thèse

---

Cette thèse a été réalisée dans le cadre d'un partenariat entre l'Institut d'Economie Rurale (IER) du Mali et le Centre de coopération internationale en recherche agronomique pour le développement (CIRAD). L'IER est la principale institution de recherche agricole du Mali et elle a pour mandat l'amélioration de la production et de la productivité agricole, pastorale et aquacole pour la sécurité et la souveraineté alimentaire ainsi que pour la préservation de la santé humaine et de la biodiversité au Mali. L'IER s'appuie pour ce faire sur un réseau de stations couvrant l'ensemble des conditions agro-écologiques du pays. Le CIRAD est un organisme français ayant pour mission de contribuer au développement économique et social des régions chaudes. Cette thèse s'inscrit dans une collaboration étroite et ancienne entre ces deux instituts. Cette thèse a été financée par le projet BIOSORGH « Diversifying and optimizing biomass sorghum value chains for Mediterranean and Semi-arid environments: an integrative approach merging biological sciences, biomaterial development and socio economic assessment » (financé par les fondations Agropolis (France) et Cariplo (Italie)) dont les ambitions étaient i) de contribuer à l'acquisition d'une meilleure compréhension de la mise en place de la biomasse végétative tant au niveau moléculaire (réseaux de gènes) qu'au niveau de la plante entière (i.e. en considérant les échanges et les relations entre les différents organes), ii) d'affiner l'identification des facteurs génétiques impliqués dans la variabilité des caractères de production et de qualité de la biomasse végétative, iii) d'initier des actions de sélection pour des variétés adaptées au contexte Malien (variétés double usage « alimentation humaine et alimentation animale », iv) de développer de nouvelles valorisations de la biomasse végétative, et v) de prendre en compte les contextes locaux et régionaux tant au niveau des conditions pédoclimatiques, des systèmes de cultures que des filières de valorisation pour définir les idéotypes à cibler dans les programmes de sélection.

Le projet BIOSORGH (2014-2017) impliquait, outre le CIRAD et l'IER, 4 partenaires complémentaires. L'université de Milan et le PTP (Parco Tecnologico Padano) assuraient avec le CIRAD la coordination du projet et contribuaient aux analyses génétiques. Le Centre des Matériaux des Mines d'Alès (C2MA) était impliqué dans le développement de nouveaux débouchés (biocomposites) pour la biomasse de sorgho. Enfin l'équipe Biens et Services des

Ecosystèmes Forestiers tropicaux (BESEF) du CIRAD contribuait à la définition des filières de production les plus pertinentes en fonction des contextes locaux et régionaux.

Les objectifs de cette thèse, initiée fin février 2014, étaient d'affiner notre compréhension de la variabilité de la production et de la qualité de la biomasse pour améliorer l'efficacité des programmes de sélection visant la valorisation de la biomasse (dans des contextes de valorisations multiples (double usage « alimentation humaine et animale ») ou ciblés sur une valorisation unique (énergie par exemple)). Dans ce contexte, du matériel végétal développé dans le cadre de deux projets précédents a été mobilisé pour répondre aux différentes questions de recherches de cette thèse.

Dans le cadre du projet GCP « MARS » « Improving Sorghum productivity in semi-arid environments of Mali through integrated marker assisted selection (MARS) », deux grandes populations biparentales ont été développées à partir de variétés élites adaptées au contexte Malien. Une de ces populations (P118) a été utilisée dans le cadre de ce projet de thèse pour analyser les bases génétiques des différentes composantes de la durée du cycle « semis – floraison » en combinant une approche de cartographie génétique et de modélisation écophysiological.

En outre, les lignées parentales utilisées dans le cadre du projet GCP « Enhancing sorghum grain yield and quality for the Sudano-Sahelian zone of West Africa using the Backcross Nested Association Mapping (BCNAM) approach » pour développer un plan de croisement multiparental de type « Backcross Nested Association Mapping » ont été mobilisées pour analyser les relations entre la croissance des plantes, leur phénologie et la qualité de la biomasse végétative. Enfin une partie des populations développées dans ce projet ont été utilisées dans cette thèse pour affiner l'analyse de l'architecture génétique de la durée du cycle « semis-floraison » et de la qualité de la biomasse végétative.

La colonne vertébrale de cette thèse a été la caractérisation phénotypique de parents et descendances de plusieurs plans de croisements dans différents environnements. Dans ce contexte, cette thèse a impliqué une alternance de séjours entre l'IER (Mali) et le CIRAD (France) afin de permettre l'acquisition des données phénotypiques, génotypiques et leurs analyses. Des appuis financiers de l'ambassade de France au Mali ont contribué au bon déroulement de ces nombreux déplacements et séjours.

# Chapitre I : Synthèse bibliographique

---

## I. I. Le sorgho, une espèce « riche en ressources »

### I.1- Une espèce d'intérêt au niveau mondial

Le sorgho est actuellement cultivé dans une centaine de pays sur une surface totale de 44.9 millions d'hectares ((*FAOSTAT*, 2017), données de 2014). Les plus importantes surfaces sont situées en Afrique sub-saharienne (Soudan : 8.33 M ha, Nigéria : 5.44 M ha, Niger : 3.57 M ha) en Inde (5.82 M ha) et aux USA (2.59 M ha). En Europe, en intégrant les pays de l'ex union soviétique, le sorgho représente 390 000 ha. Sur la base de ses besoins climatiques, la zone de culture potentielle du sorgho est très large s'étendant des latitudes 38°Sud jusqu'à 52 °Nord. Cette zone comprend la zone de culture de la canne à sucre et la plupart des zones de production de betterave sucrière (qui remonte quant à elle jusqu'à 60°Nord). En termes de production de grains, qui est l'objectif principal de la production du sorgho, le sorgho occupe la cinquième place au niveau mondial (69 Millions de tonnes (Mt) en 2014) après le maïs (1037 Mt), le riz (741 Mt), le blé (729 Mt) et l'orge (144 Mt) (*FAOSTAT*, 2014). Pour les pays en voie de développement le sorgho correspond même à la troisième céréale après le riz et le maïs, ce qui peut s'expliquer par ses faibles besoins en intrants.

### I.2- Un portefeuille d'utilisations possibles

Si le grain du sorgho est la première cible de production, le sorgho se définit plus généralement comme une espèce multi-usages. En effet, elle peut être utilisée pour l'alimentation humaine et animale, les biomatériaux (papier, bioplastiques, ...), et les bioénergies. Ces différentes utilisations ne sont pas exclusives, pour une même culture les grains pouvant par exemple être utilisés pour l'alimentation humaine, les sucres solubles extraits du jus des tiges peuvent être utilisés pour produire du bioéthanol, et les résidus des tiges (bagasse) et les feuilles utilisées pour l'alimentation animale, pour la production d'énergie (co-génération, biocarburant de seconde génération) ou pour la fertilisation du sol. Par ailleurs, le rendement en bioéthanol du sorgho, en particulier le sorgho à tiges sucrés lequel est considéré comme l'une des cultures les plus prometteuses pour la production d'éthanol à faible coût, est estimé à 3000 L/ha comparativement à la canne à sucre où le rendement en bioéthanol varie entre 3000-5000 L/ha (Almodares and Hadi, 2009).

### ***1.2.1- Alimentation humaine***

Près de la moitié de la production mondiale de grain (41 % en 2003), est dédiée à l'alimentation humaine, ce débouché étant essentiellement localisé en Afrique et en Asie (95 % de l'utilisation alimentaire du sorgho). En Afrique, la consommation humaine représente près des trois quarts de l'utilisation totale et une fraction importante de la ration calorique dans la plupart des pays producteurs. Parmi les utilisations alimentaires du sorgho, on peut distinguer les utilisations traditionnelles impliquant des processus de transformation artisanaux et locaux (tô, galettes, bouillies, couscous, tortillas, boissons fermentées ou non fermentées) et les utilisations mettant en œuvre des processus agro-industriels : malts, bières industrielles, farine sans gluten, produits extrudés, etc.

### ***1.2.2- Alimentation animale***

Dans les pays développés, le sorgho (grains et pailles) est actuellement essentiellement utilisé pour l'alimentation animale (97 % de l'utilisation totale) sous forme de concentrés, fourrages et d'ensilage. Les Etats-Unis, le Mexique et le Japon sont les principaux pays utilisateurs de sorgho sous cette forme, suivis de l'Argentine, de l'ex-URSS et du Venezuela. Il est aussi important de mentionner ici, le rôle clé du sorgho pour l'alimentation animale en Afrique et en Asie, même si ce type de valorisation n'est pas forcément visible économiquement. En effet le sorgho contribue significativement à l'alimentation animale lors des périodes de sécheresse.

### ***1.2.3- Bioénergie***

Depuis quelques années, le sorgho s'affirme comme une espèce phare pour la production de bioénergie (Berenji and Dahlberg, 2004, Carpita and McCann, 2008, Mathur et al., 2017, Mullet et al., 2014, Olson et al., 2012, Rooney et al., 2007, Sarath et al., 2008, van der Weijde et al., 2013, Vermerris et al., 2007, Fernandez et al., 2009).

La pertinence du sorgho dans ce contexte se base sur plusieurs facteurs comprenant i) un potentiel de rendement élevé en biomasse et une composition biochimique adaptée aux différentes voies de production de bioénergie (conversion de l'amidon du grain vs utilisation des sucres solubles de la tige vs utilisation de la biomasse lignocellulosique), ii) une tolérance à la sécheresse et une efficacité d'utilisation de l'eau élevées, iii) une bonne connaissance des exigences de la culture au niveau agronomique et iv) un fort potentiel d'amélioration génétique (Rooney et al 2007) auxquels peut s'ajouter la flexibilité d'utilisation de cette culture permettant la mise en place de différents schémas d'exploitation autorisant des débouchés simultanés pour

l'alimentation humaine (« food »), la production de bioénergie (« fuel »), l'alimentation animale (« feed »), les biomatériaux (« food-feed-fuel », « food-fuel », « feed-fuel »...). Enfin, en raison de la large gamme de variabilité phénotypique disponible en ce qui concerne la durée du cycle, le sorgho constitue une plante facilement « modelable » en fonction des contraintes environnementales et industrielles (approvisionnement des sites industriels).

En termes de production de bioénergie, trois principales stratégies d'exploitation peuvent être envisagées :

i) Le sorgho peut être utilisé pour produire du bioéthanol à partir de l'amidon présent dans les grains (Bioéthanol de première génération). Aux Etats-Unis, où la très grande majorité du bioéthanol produit est basé sur la conversion de l'amidon, le sorgho est d'ailleurs la seconde source de bioéthanol (10 % de la production) après le maïs (90 %). L'appareil végétatif et les résidus de la production de l'éthanol sont utilisés pour l'alimentation animale ou la cogénération (production simultanée d'électricité et de chaleur).

ii) Les sucres simples (saccharose, glucose, fructose) présents dans les tiges peuvent aussi, après pressage, être utilisés pour la production de bioéthanol en utilisant les processus industriels développés pour la canne à sucre. La bagasse (résidus obtenus après extraction des sucres de la tige) est ensuite utilisée pour l'alimentation animale ou la cogénération. Les grains qui ne sont pas utilisés dans cette stratégie peuvent être utilisés en alimentation humaine ou animale ou pour produire aussi de l'éthanol après transformation de l'amidon. L'optimisation de l'utilisation des co-produits en conservant comme priorité la production de bioéthanol permettrait de produire le bioéthanol à un coût de 250 euros / m<sup>3</sup> en Europe centrale, de l'Est et du (Grassi, 2008), un coût inférieur à celui du bioéthanol produit à partir de maïs aux USA (320 Euros / m<sup>3</sup>).

iii) Enfin le sorgho constitue une culture de choix pour les biocarburants de seconde génération basés sur la biomasse lignocellulosique. A ce niveau différents schémas sont envisageables. D'une part en valorisant les résidus cellulosiques disponibles provenant des cultures de sorgho-grain ou sorgho sucrier et d'autre part en exploitant des sorghos spécifiques à haute productivité en biomasse (sorgho de type « Biomasse »), et ce même en conditions pédoclimatiques difficiles, qui en font une des plantes les plus intéressantes. Ce troisième scénario est encore dépendant de la mise en œuvre de processus industriels véritablement efficaces (Voie biochimique ou voie thermochimique).

### ***1.2.4- Biomatériaux***

Des tests de productions papetières à partir de sorgho ont été réalisés et ont révélé des potentiels de production en termes qualitatifs élevés. Ce type de débouchés est envisageable à partir de la bagasse produite suite à l'extraction des sucres des tiges de sorgho (Belayachi and Delmas, 1995, Jiménez et al., 1993).

La Kafirine qui est la prolamine du sorgho, la principale protéine de réserve des grains (12 % du poids du grain) est un biopolymère aux propriétés intéressantes pour les emballages alimentaires biodégradables (Emmambux et al., 2004). Cette protéine peut être extraite des résidus issus de la production de bioéthanol à partir de l'amidon des grains.

L'acide aconitique qui est l'acide organique en plus forte concentration chez le sorgho peut être utilisé comme plastifiant biologique pour le polychlorure de vinyle (PVC : poly vinyl chloride). Traditionnellement, l'acide phtalique qui a un coût plus faible était utilisé (marché annuel de 500 000 tonnes/an). Cet acide est obtenu par voie de synthèse à partir d'additifs pour polymères pour l'élaboration de matériaux innovants. Néanmoins une toxicité des plastifiants basés sur l'acide phtalique a récemment été mise en évidence et leurs utilisations ont été interdites en Europe et dans d'autres pays ouvrant donc des possibilités d'utilisation pour l'acide aconitique issu du sorgho.

Enfin récemment de nouveaux biocomposites à base de sorgho ont été développés (Vo et al., 2017). Ces applications si elles sont actuellement relativement confidentielles ouvrent de nouveaux marchés aux potentiels très élevés (construction automobile, « plastiques du quotidien » (verres en plastiques, stylos), emballages alimentaires...

### **I.3- Besoins pédo-climatiques, écophysiologiques et itinéraires cultureux**

Le sorgho est traditionnellement une plante de jour court. Il présente une tolérance intrinsèque élevée au stress hydrique (ses besoins en eau correspondent à 1/3 des besoins de la canne à sucre et 1/2 de ceux du maïs) et il présente une très bonne efficacité d'utilisation de l'eau, parmi les plus élevées des graminées en C4 (Mastrorilli et al., 1995, Farré and Faci, 2006, Garofalo and Rinaldi, 2013, Enciso et al., 2015), et des nutriments, lui permettant de croître dans des zones marginales (sols pauvres, salins...). C'est une plante exigeante en température, sa température minimale de germination et de croissance étant aux alentours de 12-13 °C avec une température optimale de 25-30 °C. En climat tempéré, la somme de température au-dessus de 10°C doit se situer aux alentours de 1400 (variétés précoces) à 1900°C (variétés tardives) entre le semis et la floraison (de 100 à 130 jours). Au niveau du sol, le sorgho ne tolère pas une



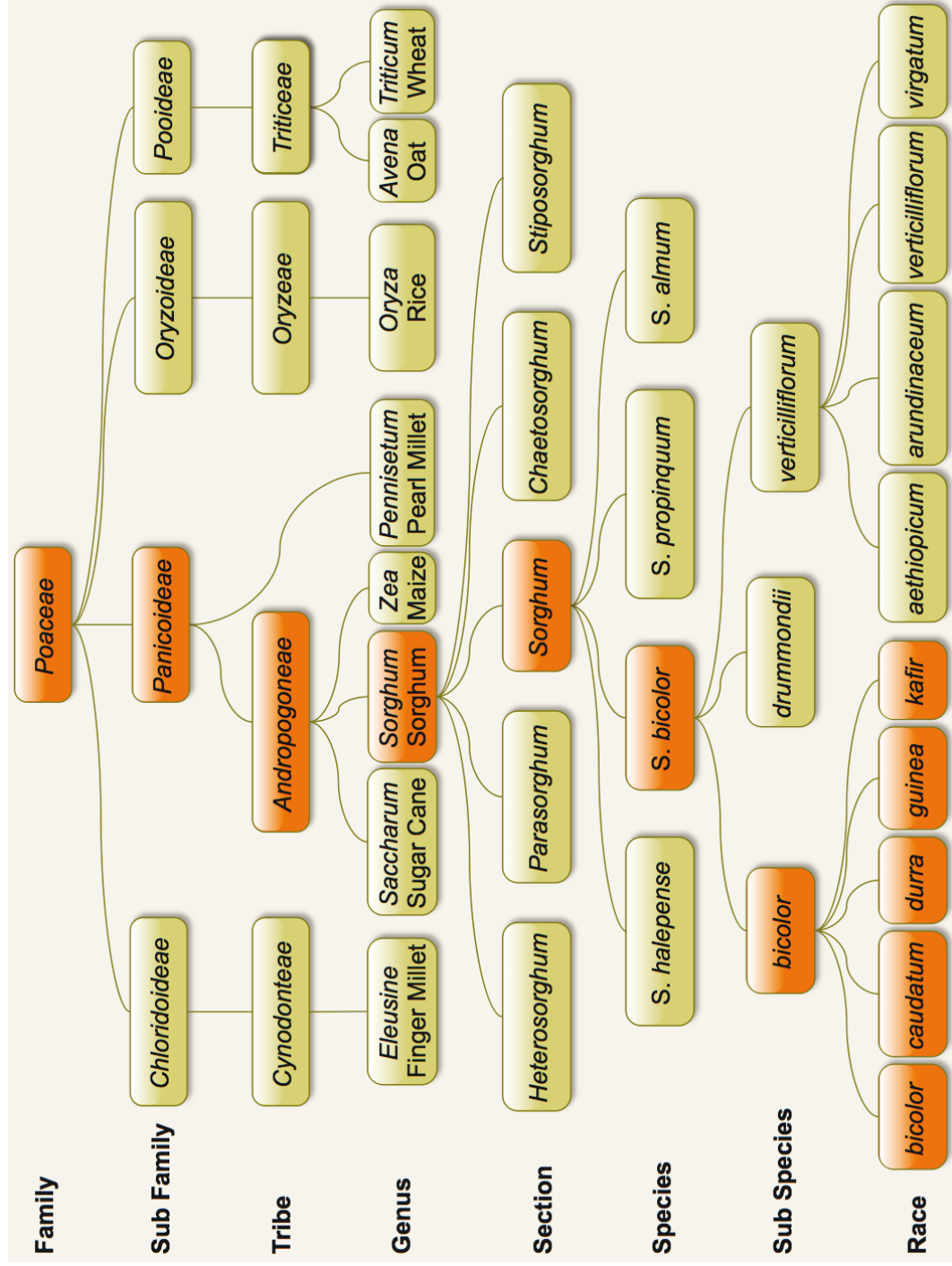


Figure I-1 : Représentation schématique de la classification du sorgho au sein de la famille des Poacées (d'après Bouchet (2008)).

température descendant en dessous de 15°C, et il s'adapte à une large gamme de sols allant des sols argileux jusqu'aux sols sablo- limoneux avec des PH compris entre 5 et 8.5. Même si plusieurs ravageurs et maladies peuvent affecter le sorgho, cette espèce peut tout de même être considérée comme une plante rustique relativement peu affectée par les stress biotiques. L'ensemble de ces caractéristiques contribue actuellement à l'intérêt qui est porté au sorgho.

#### **I.4- Un réservoir de diversité vecteur de progrès génétique**

Le genre *Sorghum* appartient à la famille des Poacées (Figure I-1), à la sous famille des Panicoïdées comme le mil et *Panicum virgatum* L. (Switchgrass) et à la tribu des Andropogonées au même titre que le maïs et la canne à sucre.

Le genre *Sorghum* est divisé en cinq sections (Doggett, 1988) parmi lesquelles se trouvent la section *Sorghum* composée de deux espèces sauvages et pérennes (*S. propinquum* et *S. halepense*) ainsi que d'une espèce annuelle (*S. bicolor*). *S. bicolor* et *S. propinquum* sont diploïdes ( $2n = 20$ ) et interfertiles malgré leur isolement géographique. *S. halepense* est une espèce tétraploïde ( $2n = 4x = 40$ ) qui peut s'hybrider avec les sorghos cultivés pour donner l'espèce *S. alnum* (Doggett, 1988, Grenier et al., 2000). Au sein de l'espèce *S. bicolor*, trois sous-espèces sont distinguées : la sous-espèce *S. bicolor* spp. *Bicolor* inclut les sorghos cultivés au sein desquels on rencontre les cinq races principales (*caudatum*, *durra*, *guinea*, *kafir* et *bicolor*); la sous-espèce *ssp. Arundinaceum*, inclut les sorghos sauvages qui comprennent quatre races (*aethiopicum*, *arundinaceum*, *verticilliflorum* et *virgatum*) avec une forte diversité morphologique et écologique ; la sous-espèce *ssp. drumondii* qui est issue de l'hybridation entre les deux précédentes sous-espèces. Les sorghos cultivés (*S. bicolor* spp *bicolor*) sont représentés par des types agronomiques très diversifiés comme les sorghos grain, les sorghos sucrés, l'herbe du soudan (fourrage) et les sorghos à balais (Berenji and Dahlberg, 2004). La grande diversité morphologique et génétique rencontrée chez les sorghos cultivés a permis de développer une classification simplifiée qui distingue cinq races botaniques principales et dix races intermédiaires qui représentent la combinaison des cinq races deux à deux (Snowdon, 1936, Harlan and De Wet, 1972).

L'origine, la systématique et l'évolution du sorgho ont été des sujets de discussion entre de nombreux auteurs (De Wet and Harlan, 1971). Certains considèrent que le sorgho provient d'Afrique de l'Ouest, d'autres de la bordure du Sahara dans la région nord-est de l'Afrique. Dans l'état actuel des connaissances, les recherches donnent raison à Dogget (Doggett, 1965) qui considère que le sorgho a été domestiqué il y a environ 5000 ans à partir des formes

sauvages en Ethiopie où se trouve la plus grande diversité du genre *Sorghum*. Il se propagea ensuite en Afrique de l'Ouest en passant par le Soudan, le Mali et la haute vallée du Niger qui sont devenus des centres de diversification secondaires du sorgho cultivé. Par voie maritime, le sorgho a atteint l'Asie avant l'ère chrétienne, l'Inde d'abord puis la Chine où l'on rencontre les sorghos kaoling. L'arrivée du sorgho en Europe date de l'époque romaine. Il s'est ensuite diffusé aux États-Unis vers la moitié du XIXe siècle par l'intermédiaire de la traite des noirs. Aujourd'hui, la plante originaire d'Afrique se trouve sur tous les continents.

## **II. La floraison, une étape clé en termes évolutifs et agronomiques**

Comme mentionné précédemment, la biomasse végétative peut être exploitée dans différentes filières de valorisation (alimentation animale, bioénergie, biomatériaux). L'ensemble de ces filières nécessite de hauts rendements de production de biomasse et l'adaptation de la qualité de la biomasse aux attentes des transformateurs et des producteurs. Au sein de cette partie, nous soulignerons l'importance de la phénologie dans le processus de production de la biomasse, puis nous décrirons les composantes de la qualité de la biomasse.

La floraison d'une plante marque le passage du développement végétatif au développement reproductif. En d'autres termes il marque la fin d'une phase d'accumulation de biomasse végétative et le passage vers une phase de développement des semences qui assureront la pérennité de l'espèce. Dans un contexte visant l'optimisation de la production de biomasse, la floraison peut donc être considérée comme un « coup d'arrêt » à une phase « favorable » d'accumulation. En outre, la transition entre ces deux phases de développement s'accompagne de modifications des allocations de ressources carbonées aux différents organes de la plante qui affectent donc en profondeur la taille et la composition des organes végétatifs.

Le pilotage de la date de floraison et la compréhension des impacts de cette transition « végétative – reproductrice » sont donc des aspects clés pour optimiser la production et la qualité de la biomasse. L'analyse du déterminisme génétique de la floraison fait l'objet d'études intensives chez le sorgho en mobilisant des approches de génétique moléculaire, d'analyse du transcriptome et de validation fonctionnelle (Tableau I-1). Néanmoins ces analyses ont été pour leur grande majorité focalisées sur des génotypes non sensibles à la photopériode ou sur des croisements entre génotypes très contrastants (non sensibles vs très sensibles) ne permettant pas une analyse fine des mécanismes impliqués dans le contrôle de la floraison au sein du pools de génotypes photopériodiques. Le déterminisme génétique de la floraison ainsi que les interactions entre le cycle « semis – floraison » et la qualité de la biomasse faisant l'objet du

deuxième et du troisième chapitre de ce document, une revue plus approfondie de l'état de l'art sur la floraison du sorgho est donc présentée en introduction du second chapitre.

### **III. La qualité de la biomasse**

Si le rendement en biomasse est le premier caractère à cibler pour des utilisations dans les filières de valorisation de la biomasse, sa composition et ses propriétés physiques qui définissent sa « qualité » sont aussi importantes pour optimiser sa transformation. En effet, les propriétés optimales de la biomasse dépendent largement des produits finis ciblés. La production d'ensilage ou de fourrages pour l'alimentation animale requiert par exemple une biomasse très digestible et riche en énergie (amidon). A l'opposé, la qualité des bio-composites produits à partir de biomasse végétale pâtira de l'utilisation de biomasses fortement digestibles et riches en sucres solubles et en amidon.

En amont de la sélection de variétés dédiées aux différents types d'utilisations, il est donc primordial de caractériser précisément les attentes des différentes filières au niveau de la composition et de la structure de la biomasse. Cette phase passe en premier lieu par une caractérisation précise de la composition et de la structure de la biomasse végétale.

#### **III.1- Les composantes de la qualité de la biomasse**

En fonction des contextes d'utilisation de la biomasse, ce sont soit des plantes entières (tiges, panicules et feuilles) soit des organes isolés (tiges uniquement) qui seront valorisés. Dans le premier cas, le facteur principal impactant la composition de la biomasse correspond aux proportions des différents organes composant la biomasse, les proportions de grains, tiges et feuilles pouvant en effet être très variables en fonction des types de plantes utilisées.

Dans un second temps, il est aussi intéressant de s'intéresser à la composition spécifique des différents organes. Pour les variétés utilisées dans le cadre des filières de valorisation de la biomasse, la tige est l'organe qui contribue le plus à la biomasse végétative. Au sein de la tige, ce sont les parois des cellules et notamment leurs parois secondaires qui contribuent le plus largement à la biomasse totale. Par ailleurs, la tige est composée d'une moelle riche en sucres solubles contenus dans les vacuoles des cellules et élaborés par la photosynthèse. Ils sont de deux types : des sucres réducteurs facilement métabolisés par la plante, le glucose et le fructose, et un sucre de réserve, le saccharose (disaccharide formé d'un glucose et d'un fructose relié par une liaison osidique) qui est la forme la plus fréquente du transport des glucides dans la sève

Tableau I-1 - Principales études portant sur l'analyse du déterminisme moléculaire et génétique des caractères impliqués dans la production et la qualité de la biomasse chez le sorgho.

	<b>Caractères liés à la production de biomasse (floraison, photopériode, hauteur, tallage)</b>	<b>Caractères liés aux grains</b>	<b>Métabolisme des sucres solubles</b>	<b>Composition de la biomasse et digestibilité</b>
Génomique fonctionnelle / Transcriptome	Ulanch <i>et al.</i> , 1996; Salas Fernandez <i>et al.</i> , 2009	Pedersen <i>et al.</i> , 2005; Gilding <i>et al.</i> , 2013; Shakoor <i>et al.</i> , 2014	S.-Y. Jiang <i>et al.</i> , 2013; Shakoor <i>et al.</i> , 2014; Bihmidine, Baker, <i>et al.</i> , 2015; Bihmidine, Julius, <i>et al.</i> , 2015	Saballos <i>et al.</i> , 2008, 2012; Sattler <i>et al.</i> , 2014; Rai <i>et al.</i> , 2016; Scully <i>et al.</i> , 2016
QTL biparental	Lin, Schertz and Paterson, 1995; Pereira and Lee, 1995; Hart <i>et al.</i> , 2001; Klein <i>et al.</i> , 2001; Feltus <i>et al.</i> , 2006; Srinivas <i>et al.</i> , 2009	Rami <i>et al.</i> , 1998; Murray, Sharma, <i>et al.</i> , 2008a; Srinivas <i>et al.</i> , 2009; Han <i>et al.</i> , 2015; Abdel-Ghany <i>et al.</i> , 2016	Murray, Sharma, <i>et al.</i> , 2008b; Ritter <i>et al.</i> , 2008; Shiringani, Frisch and Friedt, 2010; Shiringani and Friedt, 2011	Murray, Rooney, <i>et al.</i> , 2008; Abdel-Ghany <i>et al.</i> , 2016
QTL multi-parental	Mace, Hunt and Jordan, 2013; Higgins <i>et al.</i> , 2014; Bouchet <i>et al.</i> , 2017	Boyles <i>et al.</i> , 2016, 2016		
GWAS	Brown <i>et al.</i> , 2008; Murray <i>et al.</i> , 2009; Bangbol, 2013	Srinivas <i>et al.</i> , 2009; Figueiredo <i>et al.</i> , 2010; Sukumaran <i>et al.</i> , 2012; Boyles <i>et al.</i> , 2016	Murray <i>et al.</i> , 2009; Burks <i>et al.</i> , 2015	Brenton <i>et al.</i> , 2016

d'hémicellulose et de lignines (Hendriks and Zeeman, 2009). Les celluloses (40 à 60% de la paroi) et hémicelluloses (20 à 40% de la paroi) sont des chaînes de polysaccharides. La lignine est un polymère rigide inclus dans ce réseau et contribue à 10 à 25% de la biomasse des parois. Outre ces composés principaux les parois contiennent aussi des pectines, des protéines et d'autres composés mineurs.

Les proportions de ces différentes composantes et leurs compositions intrinsèques (proportion des monomères de la lignine, composition des hémicellulose, ...) ainsi que leurs imbrications contribuent à la définition de la qualité de la biomasse.

### **III.2- Déterminisme génétique de la qualité de la biomasse**

En termes de qualité de la biomasse, trois principaux groupes de caractères d'intérêt ont déjà été travaillés chez le sorgho. Il s'agit des caractères liés à la composition du grain (teneur en amidon et en protéines pour des utilisations en énergie de première génération), à la digestibilité des parois cellulaires et ses composantes (teneur en lignines, en cellulose, accessibilité à la cellulose...) et au métabolisme des sucres solubles de la tige.

Pour ces différents caractères, des analyses de génomique fonctionnelle basées sur des mutants ou au travers d'analyse du transcriptome, des analyses de QTL au sein de familles biparentales, de panels de génétique d'association ou de dispositifs multiparentaux ont été réalisées quasi systématiquement (Tableau I-1). Ces analyses ont permis l'identification d'un premier catalogue de gènes candidats et de régions chromosomiques potentiellement responsables de la variabilité naturelle des caractères impliqués dans la qualité de la biomasse. Néanmoins l'identification précise des gènes responsables de la variabilité naturelle des caractères d'intérêt reste encore à être complétée pour fournir des outils de sélection assistée par marqueurs véritablement pertinents.

En parallèle à l'exploitation de la variabilité naturelle pour l'identification des déterminants génétiques des caractères d'intérêt énergétiques, des banques de mutants EMS (éthyle méthyle sulfonate) ont été créés (Xin et al., 2008). Ces banques contiennent des mutants intéressants pour les aspects énergétiques (Sattler et al., 2014) et permettront sans aucun doute des avancées spectaculaires au niveau de l'identification des gènes candidats les plus intéressants et de la compréhension des mécanismes moléculaires impliqués dans les caractères cibles. En outre, des méthodes de transformation génétiques rapides et efficaces ont été développées chez le sorgho (Jiang et al., 2013, Liu et al., 2014, Liu and Godwin, 2012).

## **IV. La sélection du sorgho**

Le sorgho est une espèce diploïde ( $2n=10$ ), monoïque et préférentiellement autogame. La taille de son génome est de 1600 cM et 750 Mb (Paterson *et al.*, 2009). L'autogamie n'est pas stricte, 6% de pollinisation croisée est observée chez les formes à panicule compacte (Doggett, 1988) et jusqu'à 30% chez certains sorghos fourragers ou de race guinea à panicule lâche (Ollitrault, 1987, Chantereau *et al.*, 1994).

L'amélioration génétique du sorgho a commencé avec sa domestication. Les premiers acteurs de la sélection variétale ont été les agriculteurs qui ont toujours cherché à adapter les cultivars à leurs environnements et à leurs besoins alimentaires. Ainsi, les agriculteurs sont responsables en grande partie de la variabilité des caractères morphologiques et physiologiques que l'on rencontre de nos jours chez les sorghos cultivés (Chantereau *et al.*, 2013). Néanmoins, les variétés locales présentent une variabilité génétique d'une plante à l'autre en raison de l'allogamie résiduelle. Si cette variabilité offre une capacité d'adaptation à la variabilité des conditions environnementales elle réduit aussi l'homogénéité de la récolte et potentiellement les gains génétiques possibles. La sélection du sorgho a dans ce contexte visé à professionnaliser les actions de sélection pour augmenter les gains génétiques et répondre à la diversité des attentes des utilisateurs.

### **IV.1- Les cibles premières de la sélection du sorgho pour les environnements tempérés**

Les premiers travaux d'amélioration du sorgho pour les environnements tempérés ont été entrepris aux USA au début du XXe siècle à partir des sorghos introduits d'Afrique. Les objectifs principaux étaient l'augmentation de la productivité, de la précocité et la réduction de la taille des tiges. Dans ce contexte, un programme de conversion (Stephens *et al.*, 1967a) a été initié et a permis l'élimination du caractère de sensibilité à la photopériode et une réduction drastique de la taille des génotypes par l'introduction de gènes de nanisme permettant la mécanisation de la culture. Ce travail a permis de rendre accessible le germoplasme africain aux programmes de sélection américains. Suite à ces initiatives, les premiers hybrides (milo / kafir) ont fait leur apparition aboutissant à une avancée importante des rendements.

## **IV.2- La sélection du sorgho en Afrique de l'Ouest**

### ***IV.2.1- Un début difficile : l'échec « relatif » du modèle de révolution verte***

Jusqu'à très récemment, l'amélioration du sorgho en Afrique de l'Ouest a uniquement ciblé le développement de variétés de type « lignées pures ». Avant 1960, la sélection du sorgho en Afrique de l'Ouest consistait à augmenter la productivité du sorgho par l'amélioration du matériel local en utilisant la sélection massale afin de mettre rapidement à la disposition des paysans des variétés améliorées. Suite à la révolution verte des années 1960 et 1970 dans les pays développés, des variétés modernes de sorgho ont été développées pour les pays en voie de développement sur le même modèle (élimination du photopériodisme, taille courte...). Ces variétés de sorgho ont été peu adoptées (Dingkuhn et al., 2006, Chantereau et al., 2014), et plusieurs raisons peuvent être évoquées :

- Une faible intensification des méthodes de production alors que les variétés nécessitaient des apports significatifs d'intrants et un suivi des cultures précis
- Une faible plasticité des variétés face à la variabilité des conditions environnementales rencontrées en Afrique de l'Ouest
- Des qualités technologiques des grains peu adaptées aux utilisations traditionnelles
- Un transfert d'information trop limité vers les agriculteurs concernant les bonnes pratiques d'utilisation.

### ***IV.2.2- Une approche plus proche des attentes des agriculteurs et des contextes locaux : la sélection participative***

Dans le courant des années 1980, des approches de sélection participative prenant mieux en compte les contextes locaux de production et de valorisation ont été initiés. Les premiers programmes de sélection participative véritablement formalisés ont vu le jour à la fin des années 1990 (Chantereau et al., 2014).

### ***IV.2.3- Les objectifs actuels de la sélection en Afrique de l'Ouest***

Aujourd'hui, le principal objectif de sélection en Afrique de l'Ouest reste l'augmentation du rendement en grain afin de répondre aux enjeux de sécurité alimentaire. Cette amélioration du rendement doit cependant s'effectuer dans un contexte « en tension » :



- D'une part l'objectif est d'accompagner la transition d'un système extensif à un modèle plus intensif nécessitant un apport d'intrants plus important.
- D'autre part, malgré les soutiens des programmes agricoles, l'accès des agriculteurs aux intrants est encore limité
- La variabilité des conditions climatiques et pédologiques tant au niveau spatial qu'au niveau temporel est très large
- Enfin, l'amélioration doit prendre en compte la nécessité de transition de l'agriculture vers un modèle plus respectueux de l'environnement, limitant donc l'utilisation aux intrants et valorisant les interactions bénéfiques entre la variété et son milieu.

Outre la cible « rendement en grains », il est aussi clair que la qualité du grain et l'adaptation locale des variétés développées (résistance aux insectes, compétition avec les adventices, adaptation aux conditions climatiques (régime des pluies erratiques) sont des conditions sine qua non à l'adoption des variétés.

Enfin, dans le contexte d'Afrique de l'Ouest, où la biomasse végétative de sorgho est valorisée pour l'alimentation animale, les caractères de rendement en fourrage et sa qualité sont critiques pour le développement de nouvelles variétés.

#### **IV.2- Les méthodes de sélection du sorgho**

**Les méthodes de sélection généalogiques** sont les plus utilisées chez le sorgho pour développer des lignées pures. Le principe général de la sélection généalogique est de sélectionner au sein de populations diverses (F<sub>2</sub>, M<sub>2</sub> issu d'un traitement mutagène) les plantes les plus intéressantes. Les descendances de ces plantes obtenues par autofécondations sont suivies et sélectionnées en « panicule-ligne ».

**Les rétrocroisements** sont aussi souvent utilisés dans le cas de nécessité d'introgession de caractères oligogéniques dans des fonds génétique élités. Cette stratégie a été mobilisée à grande échelle dans le cadre du programme de conversion américain, où l'insensibilité à la photopériode et le caractère taille courte ont été introduits dans du matériel africains pour permettre son utilisation en conditions tempérées.

**La sélection récurrente** a aussi été utilisée dans de nombreux cas. Dans ce contexte, l'objectif est d'améliorer une population à base génétique large dont les géniteurs portent des caractères cibles de la sélection tout en conservant au mieux la diversité génétique disponible dans les différentes générations de sélection. Cette stratégie fait souvent appel à l'utilisation de gènes de

stérilité génique (ms3) qui sont introduits par rétrocroisements dans les lignées fondatrices de la population. Les populations en recombinaisons sont ensuite soumises à une pression de sélection qui va permettre d'augmenter la fréquence des allèles favorables.

Les trois méthodes présentées jusqu'ici visent le développement de lignées pures qui peuvent être valorisées directement ou qui peuvent ensuite être utilisées comme lignées parentales pour **la production d'hybrides**. Chez le sorgho, un fort hétérosis justifiant le développement d'hybrides F1 a été mis en évidence (sans mettre en avant l'intérêt économique pour les entreprises de sélection privées).

En raison du régime de reproduction principalement autogame du sorgho, la sélection pour des variétés hybrides implique l'utilisation d'un système de stérilité mâle cytoplasmique (CMS). Les lignées mâles doivent être « restaureuses de fertilité » (matériel R), tandis que les lignées femelles doivent être mainteneuses de la stérilité (matériel B). Les meilleures lignées femelles sont ensuite stérilisées par rétrocroisement pour donner des lignées mâles stériles (matériel A). La sélection se fait donc de façon indépendante dans les deux pools parentaux (R et B) et les idéotypes cibles de ces deux pools sont relativement différents. Les lignées A/B doivent être de bonnes productrices de semences, de tailles courtes et de floraison homogène. En complémentarité, les lignées mâles doivent outre leurs caractéristiques agronomiques classiques (rendement, qualité...) être de bonnes pollinisatrices.

#### **IV.4- Optimiser l'efficacité de la sélection au travers de l'utilisation de l'information moléculaire**

L'efficacité des méthodes de sélection présentées ci-dessus, définie ici comme le gain génétique par unité de temps pour un investissement donné, dépend de la capacité de l'équipe de sélection à :

- identifier au plus tôt les meilleurs génotypes
- définir le plus précisément possible la valeur des génotypes
- définir les croisements les plus pertinents
- piloter la diversité génétique disponible

Ces différentes actions peuvent être facilitées au travers de l'utilisation des informations moléculaires.

L'information moléculaire peut être, dans le contexte de la sélection, scindée en deux classes :

- L'information relative à des régions du génome ou des allèles impliqués dans le contrôle génétique des caractères cibles de la sélection,
- L'information correspondant aux zones du génome qui ne sont pas directement impliquées dans la manipulation des caractères cibles de la sélection.

En fonction des phases de sélection, l'objectif sera de maintenir ou de réduire la diversité de ces deux classes d'information moléculaire. L'accès à ces deux classes d'information nécessite une analyse du déterminisme génétique des caractères cibles, cette étape pouvant être réalisée en amont des actions de sélection ou au sein des actions de sélection.

L'analyse du déterminisme génétique des caractères d'intérêt fait appel à différentes méthodologies (détection de gènes majeurs, QTL, GWAS, recherche de signatures de sélection) et populations d'études (populations biparentales, populations peu ou pas apparentées, populations multiparentales).

Chez le sorgho, Mace et Jordan (Mace and Jordan, 2010) ont répertorié l'ensemble des gènes majeurs détectés à ce jour chez cette espèce. Cette information, cruciale pour piloter les caractères cibles de ces gènes, est aussi très utile dans le cadre de croisements biparentaux pour piloter la manipulation de QTL dont l'effet se situe entre celui de ces gènes majeurs en utilisant le phénotype de ces gènes majeurs comme des marqueurs génétiques. En 2011, les mêmes auteurs (Mace and Jordan, 2011) ont fait une synthèse des analyses QTL menées sur le sorgho entre 1995 et 2010. L'ensemble de ces études ont permis la mise en évidence de 771 QTL détectés pour 161 caractères différents. De nouvelles études QTL ont été réalisées depuis cette synthèse et ont permis d'enrichir la bibliothèque de régions chromosomiques d'intérêt. Néanmoins l'ensemble de ces QTL ont été assez peu utilisés jusqu'à présent dans le cadre de programmes de sélection publique. Plusieurs raisons peuvent être mises en avant : i) la relativement faible transférabilité des QTL entre populations, ii) le manque de diffusion des bonnes pratiques liées à l'intégration des marqueurs dans les programmes de sélection, iii) la relative difficulté d'accès aux méthodologies de génotypage...

Quelques cas d'implémentations de sélection assistée par marqueurs sont tout de même à noter. Des QTL de stay-green (effet forts) ont notamment été intégrés dans des variétés élites par rétrocroisement assistés par marqueurs (Harris et al., 2006, Hash et al., 2003, Kassahun et al., 2010). Un projet de sélection récurrente assistée par marqueurs visant l'amélioration conjointe de la productivité et de la qualité du grain pour son utilisation pour le Tô a notamment été mené à bien dans le cadre d'une collaboration CIRAD –IER (Guindo, 2016).

D'autres études de sélection récurrente assistée par marqueurs ont été également réalisées chez le sorgho (Ejeta and Knoll, 2007, Xu and Crouch, 2008, Mohamed et al., 2014). Il est important de noter que ces cas d'implémentation de sélection assistée par marqueurs ont été menés dans des populations biparentales et que leurs résultats sont difficilement extrapolables à d'autres populations.

Pour pallier à ces limites et étendre la connaissance du déterminisme génétique des caractères d'intérêt à l'ensemble de la diversité génétique susceptible d'être utilisée dans un programme de sélection, des stratégies complémentaires mettent en œuvre des panels de génotypes peu ou pas apparentés et présentant des histoires de recombinaisons plus complexes ont été développées. Les approches de génétique d'association permettent en outre d'améliorer la résolution de la cartographie des régions chromosomiques d'intérêt. Cette stratégie a été, et est encore, largement mobilisée chez le sorgho et contribue à enrichir la bibliothèque de régions chromosomiques d'intérêt. Néanmoins cette stratégie présente aussi des limites (déséquilibre des fréquences alléliques induisant une perte de puissance, difficulté de transfert des zones identifiées dans des fonds élités).

Aussi, des stratégies combinant les avantages des approches biparentales et de génétique d'association ont émergées grâce à l'utilisation de populations multiparentales (Nested Association Mapping (NAM) (Yu et al., 2008), Backcross Nested Association Mapping (BC-NAM) (Jordan et al., 2011), multi-parent advanced generation intercross (MAGIC) (Kover et al., 2009) combinées à des stratégies d'intégration des informations phénotypiques et moléculaires originales.

## **V. Améliorer la compréhension du déterminisme génétique de la production et de la qualité de la biomasse : stratégie**

Pour contribuer à la compréhension du déterminisme génétique de la production et de la qualité de la biomasse, une stratégie en 3 étapes a été développée au cours de ce travail de thèse et est présentée dans la suite de ce document.

Dans un premier temps, le déterminisme génétique de la durée du cycle « semis – floraison » a été analysé en combinant l'analyse de populations biparentales (P118, BC06) et multiparentales (BCNAM), et en combinant des approches de génétique et de modélisation écophysiological.

Dans un second temps, les relations entre la qualité de la biomasse, la croissance, la durée du cycle et la sensibilité à la photopériode ont été explorées sur les parents du dispositif BCNAM.



Dans un troisième temps, le déterminisme génétique de la qualité de la biomasse a été analysé sur le dispositif multiparental BCNAM.

Ces différentes étapes ont été ancrées sur un large nombre d'expérimentations de terrains qui ont été mises en place entre 2010 et 2015 au Mali (Tableau I.2). Ces essais au champ ont été basés sur l'utilisation de différentes populations :

La population biparentale P118 issue du projet GCP MARS

La population biparentale BC06 issue du projet GCP BCNAM

Un panel de lignées correspondant aux parents récurrents et donneurs du dispositif de croisement multiparental BCNAM développé dans le cadre du projet GCP BCNAM

Un ensemble de 29 populations (environ 1400 familles BC1F4) issu du dispositif de croisement multiparental BCNAM mentionné ci-dessus.

## Chapitre II : Etude du déterminisme génétique de la floraison et de la hauteur des plantes chez le sorgho

---

### I. La date de floraison et la hauteur des plantes : des caractères clés de la sélection chez le sorgho

#### I.1- La date de floraison : des stratégies de sélection mouvantes et une connaissance du déterminisme génétique encore fragmentaire

##### *I.1.1 – La date de floraison : des stratégies de sélection « mouvantes »*

La floraison, étape incontournable dans le cycle de vie d'une plante, est un caractère agronomique important qui contrôle l'adaptation des plantes à leur environnement. Le contrôle de la transition entre les phases végétative et reproductive est généralement sous la dépendance de deux signaux environnementaux : la durée du jour (ou photopériode) et la température (Quinby et al., 1973). A ces signaux exogènes, s'ajoutent des signaux endogènes principalement liés au développement de la plante (Amasino and Michaels, 2010). L'interaction de ces signaux permet de synchroniser la date de floraison avec les conditions environnementales favorables, permettant ainsi une adaptation optimale au milieu et une optimisation du succès reproducteur (Simpson et al., 1999).

On considère généralement trois types de comportements des plantes en fonction de leur réponse à la photopériode : les plantes de jours courts (nyctipériodiques), les plantes de jours longs (héméropériodiques) et les plantes indifférentes à la durée du jour (Thomas and Vince-Prue, 1996). Le sorgho, (*Sorghum bicolor ssp bicolor*) comme le mil (*Pennisetum glaucum*), le maïs (*Zea mays*), le soja (*Glycine max*) sont des plantes photopériodiques de jour court (Garner and Allard, 1923a), c'est-à-dire que l'induction florale se produit lorsque la durée du jour descend en dessous d'un seuil critique (photopériode critique) qui est caractéristique de la variété. De manière générale, les écotypes de sorgho Africain présentent différents niveaux de sensibilité par rapport à la variation de la durée du jour, contrairement aux régions tempérées où les sorghos ont été sélectionnés pour être très largement insensibles à la photopériode.

En règle générale, le climat tropical est caractérisé par des pluies d'été et les céréales doivent donc fleurir à l'automne afin de permettre une maturation des grains après les pluies dans des

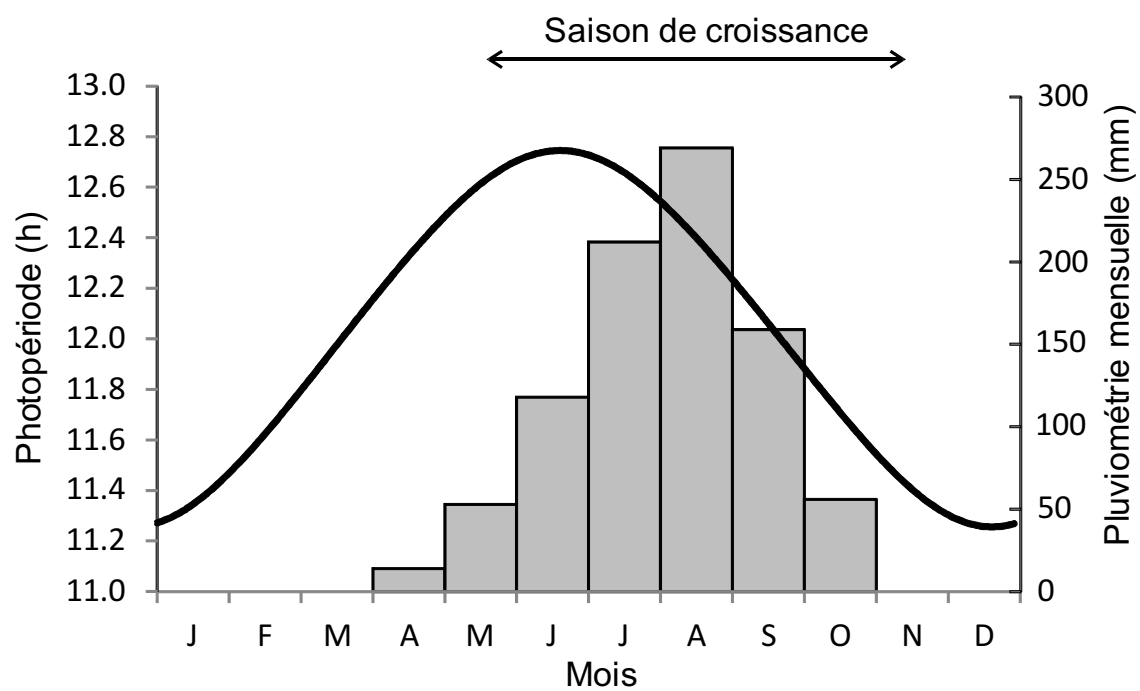


Figure II-1 : Evolution de la durée du jour au cours de l'année au niveau de Bamako, Mali (Latitude : 12°39' N)



conditions suffisamment sèches. Le photopériodisme de jours courts offre permet aux plantes cette capacité d'adaptation.

Cette adaptation est particulièrement importante dans les zones soudano-sahéliennes caractérisées par de faibles précipitations, où la saison de culture est monomodale (une seule saison par an) et se déroule en condition de jours décroissants c'est à dire de mai à août (Figure II-1).

Chez les variétés précoces non photopériodiques, la maturation des grains a lieu avant la fin de la saison des pluies, ceux-ci sont donc altérés par les moisissures, les attaques d'oiseaux ou d'insectes. Par contre les variétés photopériodiques à floraison tardive échappent à ce phénomène sauf en cas de sécheresse post florale pendant laquelle celles-ci épuisent les réserves en eau du sol avant la fin du remplissage des grains provoquant ainsi l'échaudage (Curtis, 1968).

La sensibilité à la photopériode (photopériodisme) des sorghos provoque alors un raccourcissement du cycle lorsque le semis est retardé favorisant ainsi une floraison groupée avec la fin de la saison des pluies (Garner and Allard, 1923a, Cocheme and Franquin, 1967, Curtis, 1968, Andrews, 1973, Vaksman et al., 1996, Kouressy Mamoutou, 2006). En fonction des écotypes de sorgho et notamment de leur sensibilité à la photopériode, la durée entre le semis et la maturité des grains peut varier de 90 à 190 jours (Vaksman et al., 1996).

Les efforts de sélection sur le sorgho en Afrique de l'Ouest, au cours des 40 dernières années, ont essentiellement porté sur la réduction de la taille des plantes et du nombre de talles, la précocité (élimination de la sensibilité au photopériodisme et raccourcissement des cycles de développement) et l'optimisation de l'indice de récolte.

Outre la faible adoption de ces variétés améliorées par les paysans, l'élimination de la sensibilité au photopériodisme « postulat fondamental de la révolution verte » (Swaminathan, 2006b) a induit à une diminution considérable du rendement en paille (biomasse végétative) en raison du raccourcissement du cycle et de la taille des plantes. Cette conséquence des stratégies de sélection peut, en partie, expliquer la faible adoption des variétés améliorées par les paysans (Yapi et al., 2000, Stoop et al., 1982) . La suppression de la sensibilité au photopériodisme a également entraîné une réduction du rendement en grains car les nouvelles variétés manquent de souplesse d'adaptation face aux aléas climatiques.

Une des raisons de la volonté d'élimination du caractère de sensibilité à la photopériode s'explique par le fait que les sélectionneurs pensaient que ce dernier était incompatible avec

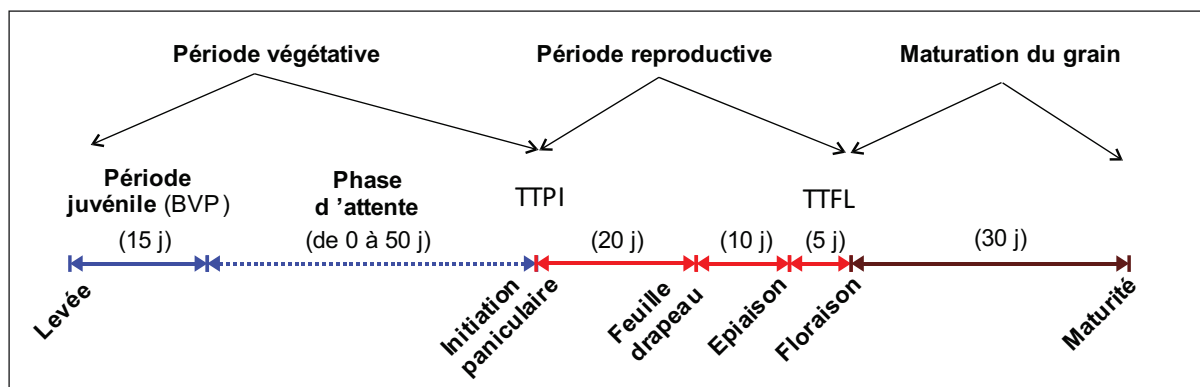


Figure II-2 : Cycle du sorgho montrant les différentes phases physiologiques. BVP : Base Vegetative Phase, FD : feuille drapeau, Ep : épiaison et Fl : floraison (d'après Belliard, 1982).

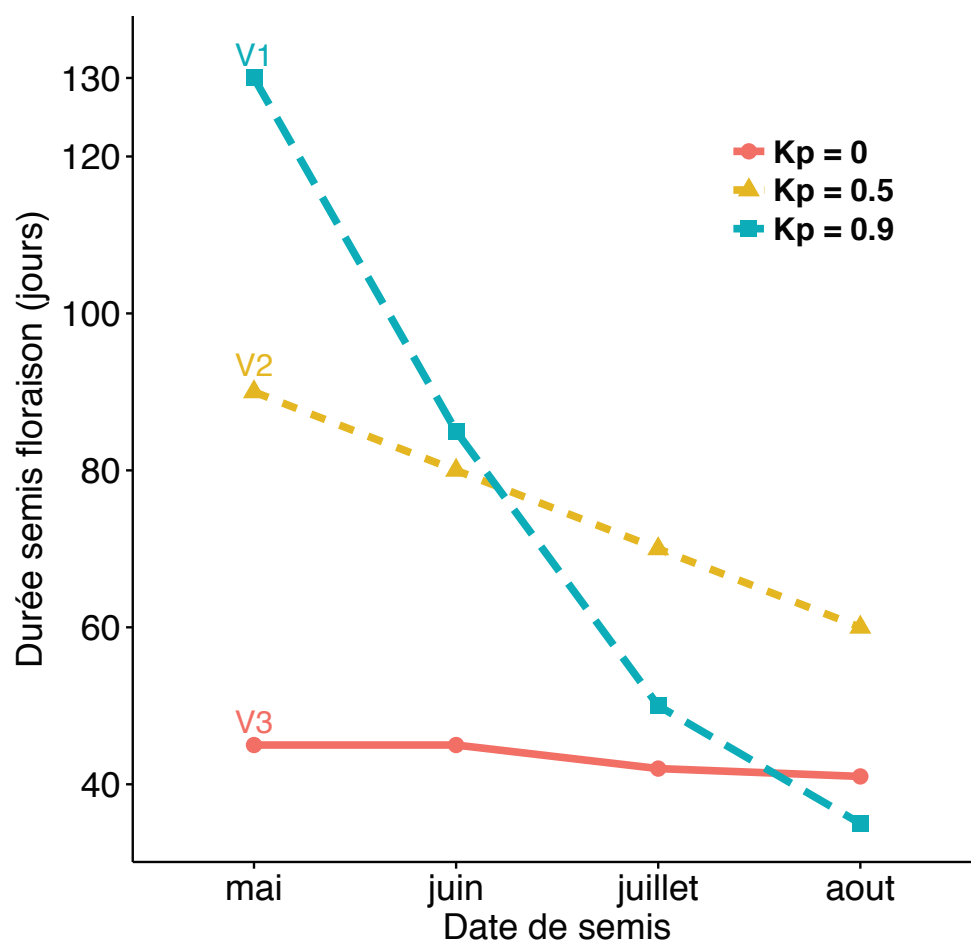


Figure II-3 : Relation entre la durée du cycle « semis floraison » et la date de semis chez 3 variétés hypothétiques de sorgho de sensibilité à la photopériode « Kp » très différentes : très photosensible (Kp=0.9), moyennement photosensible (Kp=0.5) et non photosensible (Kp=0).

l'augmentation de la productivité en raison notamment du grand nombre de feuilles développées (Sapin, 1983). En 1998, Kouressy et al, montrent néanmoins qu'il est possible de concilier les caractères de sensibilité à la photopériode et de productivité dans une même variété.

C'est à partir des années 2000, que le photopériodisme a été réintroduit comme critère de sélection dans les programmes de sélection du sorgho en Afrique. Depuis quelques années, les programmes de sélection en pays tempérés (USA, Europe) redécouvrent aussi l'intérêt de ce caractère qui, en retardant la floraison, permet d'améliorer la production de biomasse qui est un caractère clef pour les filières visant la valorisation de la biomasse (énergie, biomatériaux, bioraffinerie...) (Olson et al., 2012).

### ***1.1.2– Les composantes de la longueur du cycle chez le sorgho***

Comme mentionné auparavant, la durée du cycle « semis-floraison » du sorgho est un paramètre clé de son adaptation dans différentes zones agroclimatiques. Elle dépend de plusieurs facteurs qui peuvent être d'origine interne (précocité intrinsèque du génotype) ou des stimuli environnementaux, mais également de facteurs externes tels que la date de semis et la latitude du milieu. La plupart des variétés locales de sorgho rencontrées dans les zones soudano-sahélienne sont de grandes tailles et très photopériodiques.

Chez ces variétés, la durée de la période allant du semis à l'initiation paniculaire (Figure II-2) (Kouressy et al., 2008c), est fonction non seulement de la durée minimale de sa phase végétative mais également de la longueur du jour.

Pour une variété très sensible à la photopériode, la longueur du cycle est allongée lorsque le semis est réalisé en condition de jours longs. Ces variations de durée du cycle sont très marquées pour les variétés photopériodiques (allant jusqu'à 190 jours) tandis que pour les variétés non photosensibles le cycle reste inchangé (Figure II-3).

Par ailleurs, la durée du jour augmente avec la latitude. C'est pourquoi, une variété semée sous une basse latitude (durées du jour plus courtes) fleurira plus tôt que si elle était semée à une latitude plus élevée. Un autre facteur indirect, la date de semis, intervient également dans la variation de la longueur du cycle des variétés photosensibles. Chez ces variétés, un retard de semis entraîne une réduction de la durée du cycle, ce qui a pour conséquence d'induire une floraison groupée à la même période indépendamment de la date de semis.

Tableau II-1 : Liste des gènes majeurs de maturité chez le sorgho, leurs fonctions, positions, ainsi que leurs localisations chromosomiques.

<b>Locus de maturité</b>	<b>Identifiants des gènes</b>	<b>Fonctions</b>	<b>Chr</b>	<b>Position (CIM)</b>	<b>Références</b>	<b>Etude de cartographie</b>
Ma1	PRR37	Réprime la floraison en condition de jours longs	6	42.1-54	Quinby and Karper 1945	Lin et al. (1995), Murphy et al. (2011)
Ma2	?	Impliqué dans l'interaction photopériode x température	?	?	Quinby et Karper 1945	
Ma3	PhyB	Régule la date de floraison en réponse à la longueur du jour	1	115.5-125.7	Quinby et Karper 1945	Childs et al. 1997
Ma4	?	Influencé par la température	10	22.2-39.1	Quinby 1966	Hart et al. (2001)
Ma5	PhyC?	Contrôle la sensibilité à la photopériode	2	152.5-166	Rooney and Aydin 1999	Kim (2003)
Ma6	Ghd7	Contrôle la sensibilité à la photopériode	6	12cM	Rooney and Aydin 1999	Murphy et al. (2014), Yang et al. (2014a)

### ***1.1.3– Déterminisme génétique et moléculaire de la longueur du cycle chez le sorgho***

La compréhension des mécanismes moléculaires contrôlant la transition entre la phase de croissance végétative et la phase de reproduction a fait l'objet de nombreuses études chez certaines espèces comme le riz, le maïs, l'orge...

Au cours des dernières années, les approches de génétique moléculaire développées chez le sorgho (Cuevas et al., 2016, Wolabu et al., 2016, Murphy et al., 2014b, Yang et al., 2014b, Yang et al., 2014c, Bangbol Sangma, 2013, Murphy et al., 2011b, Childs et al., 1997a) ont facilité notre compréhension des mécanismes moléculaires contrôlant l'induction photopériodique de la floraison chez cette espèce.

Chez le sorgho, 6 loci influençant la maturité ont été identifiés (Quinby, 1967a, Quinby, 1966a) (Tableau II.1). Tous ces loci sont impliqués dans le contrôle de la date d'initiation florale et donc la durée de la croissance végétative du sorgho. Il existe de nombreuses interactions entre ces 6 loci de maturité, et les allèles dominant à chacun de ces loci permettent de retarder la floraison en condition de jours longs et allonger la durée de la phase végétative. Par contre, en présence de l'allèle récessif *ma1*, l'allèle dominant *Ma2* provoque une floraison précoce. Leurs expressions sont influencées à la fois par la photopériode mais également par la température (Quinby and Karper, 1945, Rooney and Aydin, 1999a, Major et al., 1990a, Quinby, 1966a).

Parmi ces 6 loci de maturité, l'allèle dominant *Ma1* cause le plus grand retard de floraison, et, lorsqu'il est en présence de *Ma2*, leurs actions provoquent une tardiveté extrême ; alors que les allèles dominant *Ma2* et *Ma4* sont plutôt sensibles à la température.

Comme mentionné précédemment, la réponse à la photopériode du sorgho dépend en plus des facteurs intrinsèques, de la photopériode critique ou seuil. Cette photopériode critique encore appelé P2O (Optimal photoperiod) du modèle CERES (Major, Rood and Miller, 1990; Ritchie and Alagarwamy, 1991) est définie comme étant la photopériode en dessous de laquelle la plante ne réagit pas à la variation de la durée du jour, la floraison restant donc inhibée.

Récemment, Bangbol (2013) a identifié chez le sorgho un gène (*SbSOC1*) dont l'expression est influencée par la température indépendamment de la durée du jour, bien que ce gène soit connu chez *Arabidopsis* pour son rôle de régulation de la floraison via les voies autonomes et hormonales (gibbérellines). Quant aux effets d'interactions entre la photopériode et la température, peu d'études ont été réalisées concernant la compréhension de ce mécanisme.

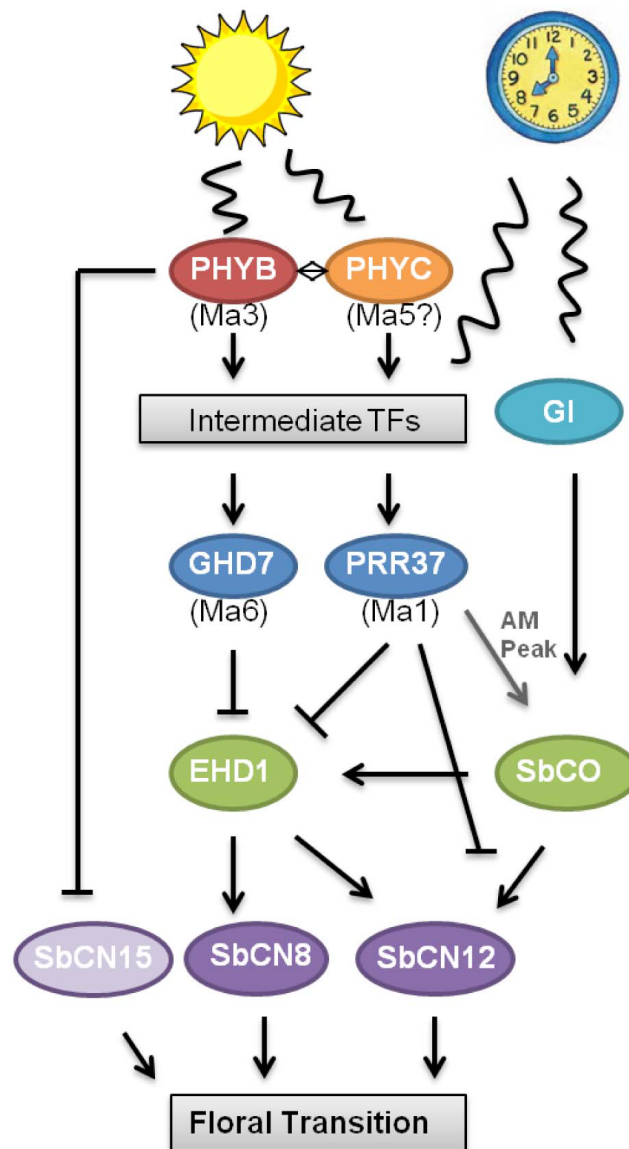


Figure II-4 : Modèle de contrôle moléculaire de la date de floraison impliquant les voies d'activation (les flèches pleines ) et de répression (les barres perpendiculaires) de la floraison chez le sorgho (d'après (Shanshan Yang, Rebecca L Murphy, et al., 2014).

Trouche et al. (1998) ont suggéré des interactions photopériode x température du gène majeur Ma2 dont la fonction demeure encore mal connue.

A ce jour, parmi ces 6 gènes de maturité décrits chez le sorgho, seulement 3 ont été clonés : Ma1, Ma3 et Ma6.

Le gène majeur Ma1 code pour une protéine régulatrice (*SbPRR37*) qui réprime la floraison en condition de jours longs (Murphy et al., 2011b), il augmente donc la sensibilité à la photopériode. Ce gène a été identifié comme le gène ayant le plus grand impact sur la date de floraison chez le sorgho (Quinby, 1974, Mace and Jordan, 2010).

Le second gène majeur Ma3 code pour un phytochrome B (*SbPhyB*) lequel réduit la sensibilité à la photopériode (Yang et al., 2014b, Childs et al., 1997a, Izawa et al., 2002) ; alors que le gène Ma6 a été identifié comme un fort répresseur de la floraison (augmente la sensibilité à la photopériode en condition de jours long). Ma6 code pour une protéine *SbGhd7* qui inhibe l'expression du gène de floraison « FT » retardant ainsi la floraison chez le sorgho (Murphy et al., 2014b, Yang et al., 2014b). Les deux gènes majeurs Ma1 et Ma6, en retardant la floraison favorisent l'accumulation accrue de biomasse végétative.

La figure II-4. développée par Yang et al. (2014b) chez le sorgho montre les différentes voies de signalisation qui interviennent dans l'induction de l'initiation florale « lesquels sont sous la dépendance de deux principaux facteurs environnementaux : la longueur du jour et la température ». En outre, les différents gènes sous-jacents sont également indiqués.

Le principe d'induction de la floraison est basé sur la coïncidence entre un signal interne propre à la plante correspondant à l'horloge circadienne (qui régule les rythmes circadiens sur une période d'environ 24h) et un signal externe exogène (photopériode et température). Ces signaux vont activer un réseau de plus de 80 gènes dont les expressions permettront de déclencher l'initiation de la floraison.

Le fonctionnement de cette cascade peut être résumé comme suit :

- la lumière est d'abord perçue par le photorécepteur phytochrome B (PHYB) au niveau des feuilles qui joue un rôle de médiation entre le signal lumineux et la plante pour moduler la date de floraison en réponse à la photopériode. En condition de jours longs, PHYB régule l'expression des gènes Ma1 et Ma6, qui sont deux gènes répresseurs de floraison, qui vont ensuite bloquer l'expression du gène *EHD1* retardant ainsi la floraison en condition de jours long. Le gène *PhyB* peut également interagir avec *PhyC*, un gène candidat pour le locus Ma5,

- en condition de jours courts, l'expression des gènes répresseurs de la floraison (Ma1 et Ma6) est réduite, ce qui entraîne une initiation de la floraison (lorsque les autres conditions nécessaires à la floraison sont remplies),
- en fonction de la photopériode, le PHYB peut directement réprimer l'expression du gène SbCN15, connu comme orthologue de « Hd3 » chez le riz et source potentielle de florigen chez le sorgho,
- les oscillateurs centraux de l'horloge circadienne c'est à dire l'alternance jour/nuit des gènes (TOC1, CCA1 et LHY) de la plante, module l'expression de GIGANTEA (Gi) un gène en sortie de l'horloge circadienne qui active à son tour l'expression de CONSTANS (CO) un gène clé du contrôle de l'induction florale par son action de régulation photopériodique de la date de floraison chez le sorgho Yang et al. (2014c). Le gène répresseur de la floraison Ma1 peut également agir sur CONSTANS « à l'aube » qui active à son tour l'expression des gènes d'induction florale (SbEHD1, SbCN8, SbCN12).

En plus de ces gènes majeurs, plusieurs études visant à identifier les régions chromosomiques impliquées dans la date de floraison via des approches d'analyse de liaison et de génétique d'association (GWAS : Genome Wide Association Study) ont été entreprises chez le sorgho. Ces études ont permis de mettre en évidence plus d'une centaine de QTL associés à la date de floraison, impliquant au moins 65 régions chromosomiques différentes (Lin et al., 1995a, Crasta et al., 1999, Hart et al., 2001, Kebede et al., 2001, Chanterreau et al., 2001a, Kim, 2004, Parh, 2005, Brown et al., 2006, Feltus et al., 2006, Murray et al., 2008a, Ritter et al., 2008, Srinivas et al., 2009, El Mannai et al., 2011, Mace et al., 2013a, Bangbol Sangma, 2013, Higgins et al., 2014). La plupart de ces études à l'exception de celles entreprises par Mace et al. (2013a) , Upadhyaya et al. (2013) et Higgins et al. (2014) ont utilisé des populations biparentales conventionnelles à faible résolution de cartographie (en raison de l'existence de déséquilibre de liaison à longue distance au sein de ce type de population).

Toutes ces études ont montré que le contrôle génétique de la date de floraison chez le sorgho est contrôlé par un nombre relativement important de locus avec des allèles multiples, dont la plupart présentent des effets faibles. La majorité des QTL liés à la date de floraison ont été identifiés chez les sorghos tempérés pour lesquels les gènes à effets majeurs pour la sensibilité à la photopériode ont été supprimés ; ou des populations issues de croisements entre lignées photopériodiques de grande taille et non photopériodique de taille courte ou naine.



## **I.2 – La taille des plantes : Sélection et déterminisme génétique**

### ***I.2.1- Multi-utilisations vs intensification***

Comparativement à beaucoup d'autres espèces (notamment les espèces céréalières), le sorgho est l'une des plantes présentant le plus grand éventail de variabilité génétique pour la date de floraison, la hauteur et le tallage.

La plupart des écotypes de sorgho d'Afrique de l'Ouest, sont de grandes tailles (hauteur pouvant atteindre 5m). Ces écotypes sont majoritairement composés de race Guinea : entre 70 et 93% respectivement pour les sorghos du Mali et du Burkina Faso. Ces variétés sont très sensibles à la photopériode et s'adaptent à la diversité des systèmes de cultures rencontrés dans ces zones soudano-sahélienne. Dans ces zones, le sorgho est principalement cultivé pour son grain pour l'alimentation humaine (préparation des plats traditionnels). Néanmoins, le fourrage produit à partir de la biomasse des tiges et des feuilles après la récolte des grains est intensivement utilisé pour l'alimentation animale ou comme matériaux pour la construction. Cette utilisation est très importante dans les systèmes de production d'Afrique de l'Ouest, où le fourrage de sorgho a une importance clé lors de la période de soudure (période de pénurie de fourrage).

Malgré ces atouts, la culture du sorgho dans ces zones d'Afrique de l'Ouest demeure encore extensive et se pratique principalement au niveau des systèmes de culture traditionnels avec utilisation des variétés à faible indice de récolte. En outre, la grande taille du sorgho peut constituer un handicap dans le contexte des systèmes intensifiés (Piroux et al., 1997), car les plantes sont exposées à la verse rendant ainsi la récolte des grains plus difficile (Campbell et al., 1975).

Dans les contextes Européen et Nord-Américain, où le sorgho est produit principalement pour l'alimentation animale à partir des grains et des plantes entières, les variétés présentent principalement des tailles courtes (<1.2m) et sont cultivées dans des systèmes de cultures intensifiés. D'autres variétés légèrement plus grandes (1.6 à 2.5m) sont rencontrées dans ces régions et sont destinées à une utilisation sous forme d'ensilage. Dans ces zones de culture, il existe également des variétés visant les marchés énergétiques, des biomatériaux et de la bioraffinerie. L'objectif principal dans ces filières étant principalement la production de biomasse, des variétés plus hautes ont été ainsi développées (>3m).

Cependant, que ce soit dans les contextes européens ou Africains, la valorisation de l'appareil végétatif nécessite un rendement en biomasse élevé (densité de peuplement, augmentation du

nombre et de la taille des entre-nœuds) ainsi qu'une composition et une structure de la biomasse adaptées aux différentes utilisations.

Dans ces différents contextes (systèmes de cultures et de valorisation d'Afrique de l'ouest et des pays tempérés), la compréhension fine du déterminisme génétique de la taille des plantes (indépendamment des aspects de sensibilité à la photopériode) constitue un enjeu clef pour le développement de variétés adaptées aux différentes utilisations.

### ***1.2.2- Déterminisme génétique de la hauteur chez le sorgho***

Comme mentionné préalablement, la hauteur de la plante est fortement associée à la date de floraison car elle dépend du nombre d'entre-nœuds allongés produits mais aussi de leur longueur moyenne. Si le nombre d'entre-nœuds allongés dépend principalement de la durée du cycle « semis- floraison », leur longueur moyenne est par contre indépendante des aspects de durée de cycle, elle dépend fortement du caractère de nanisme.

Le nanisme est contrôlé principalement par 4 gènes majeurs appelés : dwarf1 (dw1), 2, 3 et 4 (Quinby and Martin, 1954, Windscheffel et al., 1973). Les allèles récessifs de ces gènes provoquent un raccourcissement de la taille des entre-nœuds avec un effet cumulatif selon le nombre de gènes à l'état homozygote récessif (Casady, 1967, Campbell et al., 1975), sans affecter les autres caractères agronomiques tel que le nombre de feuilles développées (un proxy du nombre d'EN allongés), la date de floraison ou de maturité, et les longueurs des panicules et des pédoncules (Quinby, 1974).

Par ailleurs, le phyllochrone qui détermine le temps nécessaire à l'apparition de deux feuilles successives, a été démontré chez le sorgho, comme étant constant de la levée jusqu'à l'apparition de la feuille-drapeau (Alagarswamy et al., 1986, Muchow and Carberry, 1990, Craufurd and Peacock, 1993, Vaksman et al., 1998). Mais le phyllochrone peut être variable en fonction des génotypes au cours de la saison de culture (Goldsworthy, 1970, Mantineo et al., 1996).

A ce jour, sur les 4 gènes majeurs identifiés, seul dw1 et dw3 ont été clonés. Le gène dw3 est le premier gène de nanisme qui a été identifié et cloné chez le sorgho. Il code pour une glycoprotéine qui régule le transport de l'auxine polaire. dw3 est connu comme l'orthologue de brachytic2 (br2) qui est, chez le maïs, responsable du raccourcissement de la taille des entre-nœuds (Multani et al., 2003).

Le gène majeur *dw1* code pour une protéine membranaire (Yamaguchi et al., 2016, Hilley et al., 2016) qui a pour rôle de réguler de façon positive la signalisation d'une hormone stéroïde connue sous le nom de Brassinosteroid (BR) en inhibant la fonction d'une autre hormone Brassinosteroid-insensitive 2 (BIN2) (Hirano et al., 2017). Ces gènes de nanisme sont agronomiquement importants car ils sont impliqués dans la résistance à la verse, facilitant ainsi les travaux de mécanisation des récoltes dans les systèmes de culture à fortes intensification.

Une étude récente (Yamaguchi et al., 2016) a révélé des interactions significatives entre *dw1* et *dw3* (effet non additif) provoquant une réduction de la longueur des entre-nœuds situés au milieu de la tige principale.

En complément de l'identification des gènes majeurs de nanisme, de nombreuses études visant à identifier les autres régions chromosomiques impliquées dans la variation de la hauteur ont été effectuées. Ces études ont principalement portées sur des populations biparentales issues de croisements entre des génotypes très contrastants pour leur hauteur (Lin et al., 1995a, Pereira and Lee, 1995, Madhusudhana and Patil, 2013, Hart et al., 2001, Rami et al., 1998, Srinivas et al., 2009). Ces études ont permis l'identification de plusieurs régions génomiques associées à la hauteur, qui sont complémentaires des gènes majeurs. Elles ont également mises en évidence des co-localisations entre QTL de hauteur et de maturité/ sensibilité à la photopériode (Lin et al., 1995a, Upadhyaya et al., 2012, Brown et al., 2008, Higgins et al., 2014).

En plus de ces études d'analyse de liaison, des études récentes de cartographie d'association basée sur des panels d'accessions avec introgression du gène de nanisme (Brown et al., 2008), constitués principalement de sorghos sucrés (Murray et al., 2009) et ou des panels provenant des collections mondiales (Morris et al., 2013) ont été entreprises chez le sorgho.

En complément de ces études d'associations, des approches visant à combiner la méthode traditionnelle d'analyse de liaison versus analyse d'association ont récemment été développées chez le sorgho (Mace et al., 2013a, Higgins et al., 2014). Cette dernière approche a l'avantage de fournir plus de précisions sur la position des QTL identifiés c'est à dire, une grande résolution et bonne puissance de détection.

### **I.3 – Comment affiner notre compréhension du déterminisme génétique de la longueur du cycle et de la hauteur chez le sorgho ?**

La plupart des études visant l'identification des régions chromosomiques impliquées dans la variation de la phénologie et de la hauteur des plantes ont été basées sur des populations biparentales de type F2 ou RILS. Cette méthodologie profite de fréquences alléliques

Tableau II-2 : Comparaison des différents schémas de croisement utilisés pour la détection de QTL (adapté et complété à partir de (Yu *et al.*, 2008)).

Dispositifs ---Propriétés	Population biparentales		Cartographie d'Admixture*		BCNAM	MAGIC	Description des propriétés
	GWAS	NAM	NAM	MAGIC			
Richesse allélique	Faible	Elevée	Elevée	Moyen	Moyen	Moyen	Nombre d'allèles qui sera analysé
Nombre de marqueurs nécessaire pour l'analyse complète du génome	Faible	Elevé	Moyen	Moyen	Moyen	Moyen	Nombre de marqueurs nécessaire pour pouvoir révéler une association pour les marqueurs indépendants
Puissance statistique	Moyenne	Faible	Elevée	Elevée	Elevée	Elevée	Puissance de révéler des associations entre marqueurs en ségrégation et les phénotypes en tenant compte du fait que les marqueurs ont des fréquences alléliques différentes
Dégradation du LD	Lent	Rapide	Lent	Intermédiaire	Intermédiaire	Intermédiaire	Dégradation moyenne du déséquilibre de liaison sur l'ensemble du génome.
Résolution cartographique	Faible	Elevée	Intermédiaire	Elevée	Elevée	Elevée	Précision de la cartographie des loci causatifs
Effets du fond génétique	Faible	Elevée	Faible	Faible	Faible	Faible	Impact de la variabilité du fond génétique sur la détection des associations phénotypes vs marqueur
Complexité du développement (coût, temps)	Variable (F2 vs RIL...)	Faible	Faible	Elevée	Elevée	Elevée	Coût et temps d'investissement pour développer le design cartographique
Développement de nouvelle diversité	Faible	Nul	Nul	Elevée	Elevée	très élevée	Estimation du niveau de nouvelle diversité développée lors du développement du dispositif
Transférabilité des résultats dans un fond génétique plus large	Faible	Moyen - élevée	Elevée	Elevée	Elevée	Elevée	Possibilité de transférer et d'utiliser les résultats dans d'autres fonds génétiques
Pertinence pour la sélection	Faible	Faible	Faible	Intermédiaire	Elevée	Elevée	Possibilité d'intérêt direct pour les programmes de sélection

\* Méthode puissante de cartographie, utilisée pour d'identifier les variants génétiques associés à des caractères et / ou des maladies au sein des groupes ethniques dont les génomes résultent d'un mélange récent de deux ou plusieurs populations ancestrales géographiquement distinctes.

relativement équilibrées assurant une bonne puissance de détection (lorsque les effectifs sont suffisamment importants). En revanche dans la majeure partie des cas, ces populations présentent un déséquilibre de liaison à longue distance qui ne permet pas une localisation précise des zones chromosomiques d'intérêt.

D'autres études ont été réalisées via les approches de génétiques d'association en utilisant des panels de diversité. Ces études s'ancrent sur une diversité génétique plus importante que les populations biparentales et modelée par de très nombreux événements de recombinaison. Ces propriétés apportent une résolution de cartographie plus précise que les populations biparentales mais pâtiennent en revanche pour certaines régions du génome de fréquences alléliques très déséquilibrées qui réduisent en conséquence la puissance de détection. Récemment, l'émergence de dispositifs multiparentaux qui combinent les avantages de ces deux dernières stratégies ont permis la détection de QTL lié à la longueur du cycle et à la hauteur des plantes au sein des populations multiparentales (Mace et al., 2013a, Higgins et al., 2014, Bouchet et al., 2017). Le tableau II-2 montre les avantages et inconvénients des différents types de dispositif utilisés pour la détection de QTL / gènes sous-jacents et pour l'utilisation dans un programme de sélection.

Donc, en utilisant des parents porteurs d'allèles contrastants pour la hauteur et la sensibilité à la photopériode/longueur du cycle, dans le développement de populations multiparentales telles que les populations de type NAM (Nested Association Mapping) et BCNAM (Backcross-Nested Association Mapping), cela permettra de capturer plus de diversité et de recombinaisons alléliques contrairement à une population biparentale ; ce qui permet d'améliorer la puissance statistique pour la détection des régions chromosomiques contrôlant ces caractères.

Par ailleurs, afin de répondre à certaines limites liées aux populations conventionnelles biparentales, d'autres types de population multiparentales à base d'intercroisements multiples de génération avancée appelées « MAGIC : Multiparent Advanced Generation Inter-Cross » ont été développées pour la première fois chez la plante modèle d'*Arabidopsis* (Kover et al., 2009, Huang et al., 2011), le riz (Bandillo et al., 2013, Li et al., 2013, Leung et al., 2015, Meng et al., 2016), le maïs (Holland, 2015, Dell'Acqua et al., 2015), le blé (Huang et al., 2012, Mackay et al., 2014), l'orge (Sannemann, 2015), la tomate (Pascual et al., 2015). Ces dispositifs complexes créés à partir de multiples inter-croisements de 4, 8 ou 16 lignées parentales combine les points forts des approches d'analyse de liaison en populations biparentales et de cartographie d'association en termes de diversité allélique, de puissance de détection de QTL ainsi que de résolution de cartographie.

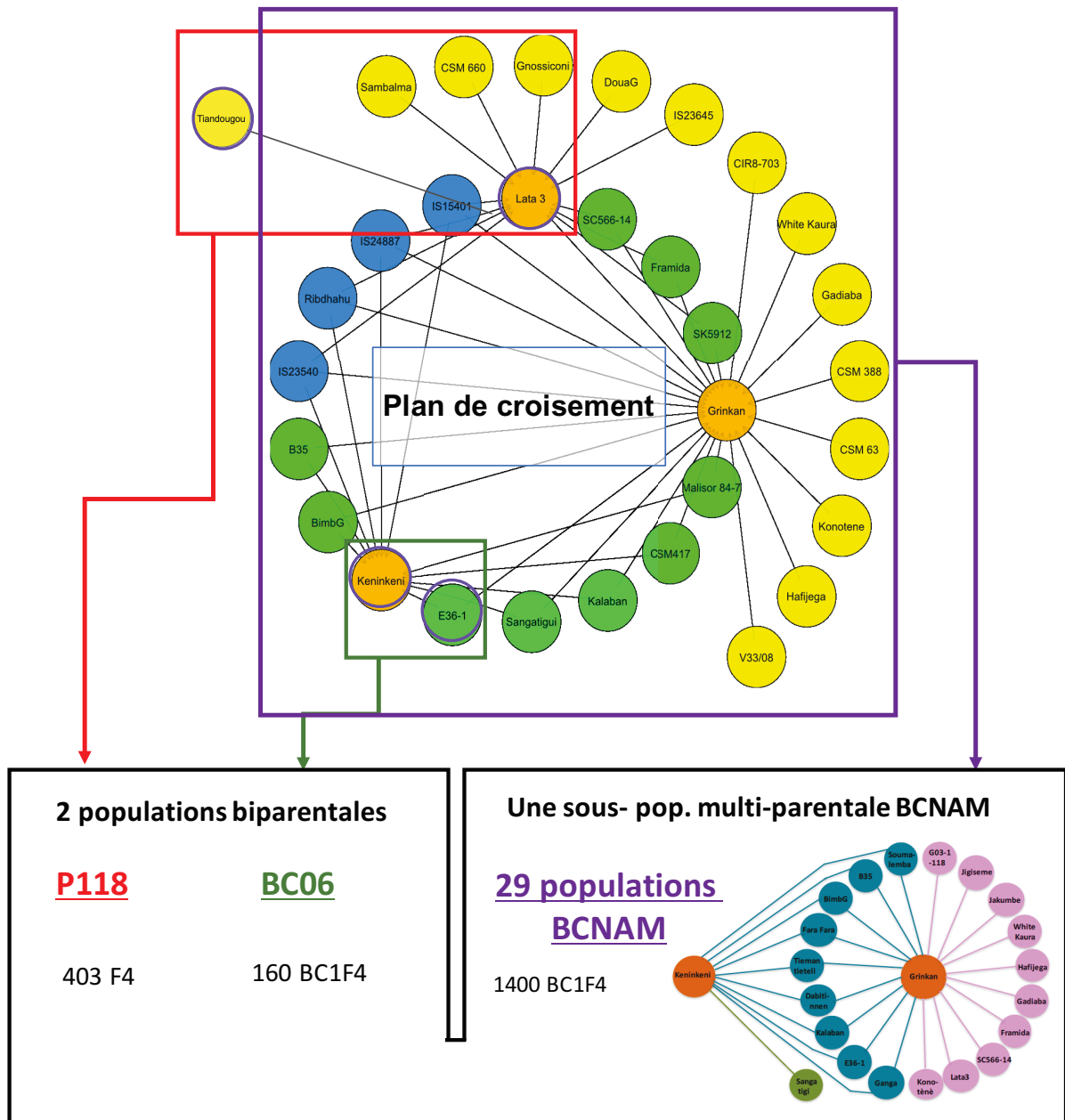


Figure II- A : Dispositif BCNAM mobilisé (2 populations biparentales et une sous-population multi-parentale BCNAM) pour analyser le déterminisme génétique de la floraison / sensibilité à la photopériode et hauteur des plantes des différentes familles BCNAM.

Comme tout dispositif de croisement, les MAGIC présentent en plus de ces atouts, certaines limites liées principalement au choix des lignées parentales qui doivent être sélectionnées de façon rigoureuse afin de conserver une diversité génotypique plus large.

Dans la présente étude, afin d'améliorer notre compréhension du déterminisme génétique de la floraison / sensibilité à la photopériode et hauteur des plantes, trois types de matériels génétiques (Figure II-A) ont été mobilisés :

- ⇒ une première population biparentale (Tiandougou X Lata3) en ségrégation (P118) issue d'un programme de Sélection Récurrente Assistée par Marqueurs moléculaires « SRAM » (ou MARS pour Marker Assisted Recurrent Selection). Cette population diverge de celles qui ont été traditionnellement utilisées pour analyser le déterminisme génétique de la floraison et de la hauteur en raison des caractéristiques des lignées parentales qui ont été utilisées. En effet, ces deux lignées sont sensibles à la photopériode et présentent des tailles relativement proches. Ce contexte devrait permettre d'identifier de nouvelles régions chromosomiques d'intérêt qui ont été jusque-là masquées par les effets de gènes majeurs qui présentaient des allèles contrastants au niveau des différents gènes impliqués.
- ⇒ une seconde population biparentale de type BCNAM développée à partir du parent récurrent « Keninkeni » et du parent donneur « E36-1 ». L'intérêt de cette population réside dans la différence phénotypique qui existe entre les deux parents : à la différence de la population 118, les deux lignées parentales sont variables pour leur sensibilité à la photopériode et leur hauteur, mais elles ne devraient pas présenter d'allèles différents aux niveaux des gènes majeurs déjà identifiés.
- ⇒ Enfin, un dispositif de croisements multiparental de type BC – NAM (Mace and Jordan, 2011) développé à partir de deux parents récurrents et de 29 parents donneurs a été utilisé. Cette population a l'intérêt d'offrir une large diversité allélique à chaque locus et une résolution de cartographie élevée. Les deux parents récurrents de ce dispositif sont faiblement sensibles à la photopériode, ils présentent des cycles et des tailles moyennes adaptés à la zone soudano sahélienne. Les 29 parents donneurs représentent une large diversité des sorghos cultivés avec des caractères d'intérêts agronomiques divers : productivité,

adaptation/sensibilité à la photopériode, qualité du grain, teneur en sucre, jutosité. etc.

Toutes ces populations ont été caractérisées pour la durée du cycle semis floraison, la longueur de la tige principale, le nombre total d'entre-nœuds allongés, la longueur des pédoncules et panicules dans différentes conditions de photopériode (à l'exception des populations BCNAM). Les objectifs étaient :

- i) d'identifier les régions génomiques contrôlant la date de floraison/sensibilité à la photopériode et la hauteur des tiges chez le sorgho dans différentes conditions environnementales, et différents fonds génétiques,
- ii) de comparer ces régions avec celles identifiées dans les études précédentes,
- iii) de tester les interactions entre sensibilité à la photopériode et nanisme en fonctions des environnements,
- iv) de proposer des gènes candidats (dont les séquences sont disponibles) sous-jacents à ces QTL et
- v) de discuter des implications futures de ces résultats dans les programmes de sélection en particulier dans un contexte soudano-sahélienne.

Dans la suite de ce chapitre dédié à la floraison, les études menées sur les deux populations biparentales (P118, BC06) puis sur le dispositif BCNAM seront exposées.



## II. Integrating genetic analysis and crop modeling: a major QTL can finely adjust photoperiod-sensitive sorghum flowering

### II.1. Article soumis à Plant Cell and Environment

**Title : Integrating genetic analysis and crop modeling: a major QTL can finely adjust photoperiod-sensitive sorghum flowering**

Baptiste Guitton<sup>1,3\*</sup>, Korotimi Théra<sup>1,3,4\*</sup>, Mohamed Lamine Tékété<sup>4\*</sup>, David Pot<sup>1,3</sup>, Mamoutou Kouressy<sup>4</sup>, Niaba Témé<sup>4</sup>, Jean-François Rami<sup>1,3</sup>, Michel Vaksman<sup>2,3\*</sup>

\* These authors contributed equally to the work.

**Corresponding author: Michel Vaksman** [michel.vaksman@cirad.fr](mailto:michel.vaksman@cirad.fr) CIRAD, BP1813, Bamako, MALI. Tel. +223 77716795

**Short running title:** Genetic control of photoperiod-sensitive sorghum flowering.

**Highlight :** Crop modeling allows us to decompose sorghum flowering into elements controlled by specific genomic regions. A major QTL for critical photoperiod could help to fine-tune varietal adaptation to West-African environments.

#### Author affiliations

<sup>1</sup> CIRAD, UMR AGAP, F-34398, Montpellier, France

<sup>2</sup> CIRAD, UMR AGAP, F-1813, Bamako, Mali

<sup>3</sup> AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France.

<sup>4</sup> Institut d'Economie Rurale (IER), BP 258, rue Mohamed V, Bamako, Mali

Emails:

Baptiste Guitton : [baptiste.guitton@cirad.fr](mailto:baptiste.guitton@cirad.fr)

Korotimi Théra : [korotimi.thera@yahoo.fr](mailto:korotimi.thera@yahoo.fr)

Mohamed Lamine Tékété : [molate111@yahoo.fr](mailto:molate111@yahoo.fr)

David Pot : [david.pot@cirad.fr](mailto:david.pot@cirad.fr)

Mamoutou Kouressy : [nanym63@gmail.com](mailto:nanym63@gmail.com)

Niaba Témé : [niabateme@gmail.com](mailto:niabateme@gmail.com)

Jean-François Rami : [jean-francois.rami@cirad.fr](mailto:jean-francois.rami@cirad.fr)

Michel Vaksman : [michel.vaksman@cirad.fr](mailto:michel.vaksman@cirad.fr)

Date of submission : 27 september 2017

Number of tables : 1

Number of figures : 5

Word count : 4712

37 **Abstract**

38 In West Africa Sudano-sahelian zone, sorghum sensitivity to photoperiod is a major trait for  
39 flowering adjustment toward end of the rainy season. This trait ensures that conditions for seed  
40 development are optimal.

41 A sorghum segregating population derived from a cross between two photoperiod sensitive  
42 parents was evaluated in three different photoperiod conditions. CERES crop model was  
43 applied to decompose the flowering time of each genotype into intrinsic earliness, critical  
44 photoperiod and photoperiod sensitivity. Phenology and model derived variables were used for  
45 QTL detection.

46 A major QTL affecting critical photoperiod was identified whereas only independent minor  
47 QTLs were found for intrinsic earliness and photoperiod sensitivity. Candidate gene analysis  
48 in the major QTL region allowed us to propose a candidate gene involved in the circadian clock  
49 (*ELF3*) as a key regulator of flowering time in photoperiod-sensitive sorghum. Our results  
50 provide critical informations to allow the development of photoperiod sensitive genotypes  
51 specifically adapted to climate variability encountered in Sudano-Sahelian zone of West Africa.  
52 Improving the understanding of the genetic control of flowering time in sorghum is thus an  
53 important step toward breeding climate resilient varieties and taking up the challenge of climate  
54 smart agriculture.

55 **Key words:** Sorghum, Photoperiod-sensitivity, Flowering, Crop modeling, QTL, Candidate  
56 genes.

57

## 58           **1. Introduction**

59 Sorghum (*Sorghum bicolor* (L.) Moench) is a C4 grass cultivated for grain, forage and biomass  
60 in a wide range of climatic and geographic conditions. In dryland systems, and particularly in  
61 West-Africa, this crop has a predominant role in food security for millions of rural families.  
62 Sorghum is a photoperiodic short-day plant (Garner and Allard, 1923b) for which flowering  
63 occurs when day length becomes shorter than a critical photoperiod.

64 Photoperiod-sensitive sorghum landraces that are cultivated in Africa can be seen as the result  
65 of a long massal selection conducted by African farmers that contributed to fine-tune their  
66 adaptation to specific environmental and management conditions. Indeed, flowering time is a  
67 major ecologic and agronomic trait as it controls sorghum adaptation to environments by  
68 adjusting vegetative and reproductive growth phases to local biotic and abiotic constraints.

69 The Sudano-Sahelian climate is characterized a very high within and across year rainfall  
70 variability. Most sorghum in West AFrica are photoperiod sensitive and flower in the 20 days  
71 preceding the average ending date of the rainy season regardless of the sowing date (Kouressy  
72 et al., 2008d). This trait improves simultaneously (i) biomass production by optimizing the  
73 duration of the vegetative period, (ii) between year yield stability by mitigating drought  
74 (preventing late flowering) and avoiding midge and bird attacks (grouping flowering) and (iii)  
75 grain quality by limiting mold development (preventing early flowering) (Kouressy et al.,  
76 2008b).

77 Sorghum breeding programs eliminated photoperiod-sensitivity from elite grain sorghum  
78 germplasm mainly in order to develop varieties having a broader geographical adaptation  
79 (Stephens et al., 1967b, Chantereau et al., 2001b, Swaminathan, 2006a). However, in West-  
80 Africa, the selection for early maturity has often proven to be an inadequate goal as the problem  
81 is not the short duration of the growing season, but the variability of beginning and the end of  
82 seasons in both time and space. Furthermore, forecasted increase of climate variability requires  
83 the development of breeding strategies that focus on crop adaptation. Nowadays, the  
84 development of high yielding photoperiod-sensitive varieties adapted to the Sudano-Sahelian  
85 climate becomes a priority of dryland cereals breeding programs in West-Africa. In addition,  
86 photoperiod sensitivity drew recently the attention of breeders to develop late-flowering  
87 sorghum(Murphy et al., 2014a) to increase biomass yield for biofuels production (Olson et al.,  
88 2012).

89 The genetic basis of flowering time has been extensively studied in sorghum since 18 QTL and  
90 association mapping studies focusing on sorghum flowering time were inventoried in the  
91 literature (Lin et al., 1995a, Crasta et al., 1999, Hart et al., 2001, Kebede et al., 2001, Chantereau

92 et al., 2001a, Kim, 2004, Parh, 2005, Brown et al., 2006, Feltus et al., 2006, Murray et al.,  
93 2008a, Ritter et al., 2008, Srinivas et al., 2009, El Mannai et al., 2011, Mace et al., 2013a,  
94 Bangbol Sangma, 2013, Higgins et al., 2014, Shiringani et al., 2010, Upadhyaya et al., 2013,  
95 Yang et al., 2014a). From these studies, it may be concluded that flowering is controlled by  
96 several major maturity genes and modulated by a relatively large number of loci with small  
97 effects (Mace et al., 2013b). However, most of these analyses were based on populations  
98 derived from photoperiod insensitive parents or from sensitive by insensitive crossing,  
99 neglecting mapping populations developed from the crossing of different photoperiod sensitive  
100 parents.

101 Sorghum flowering time is mainly controlled by temperature and photoperiod (Mace et al.,  
102 2013b) among the four known flowering genetic pathways (Boss et al., 2004, Mouradov et al.,  
103 2002). For short day plants, photoperiod pathway repress flowering based on the duration of  
104 daylight in the daily cycle when photoperiod becomes too long.

105 To date, six major maturity genes (*Ma<sub>1</sub>-Ma<sub>6</sub>*) have been described in sorghum (Quinby, 1967b,  
106 Rooney and Aydin, 1999b, Morgan and Finlayson, 2000). Recessive alleles in *Ma<sub>1</sub>*, *Ma<sub>2</sub>*, *Ma<sub>3</sub>*  
107 and *Ma<sub>4</sub>* produce early-flowering phenotypes (Quinby, 1967b). More recently, *Ma<sub>5</sub>* and *Ma<sub>6</sub>*  
108 have been genetically characterized (Rooney and Aydin, 1999b, Brady, 2006, Mullet et al.,  
109 2016), they interact epistatically to repress flowering. Among the first four loci, *Ma<sub>1</sub>* causes the  
110 largest delay in flowering time. Both *Ma<sub>2</sub>* and *Ma<sub>4</sub>* have been shown to be temperature sensitive  
111 (Quinby, 1966b, Major et al., 1990b). Three of these genes have been cloned and their  
112 interactions described allowing the proposition of a pathway model (Murphy et al., 2014a, Yang  
113 et al., 2014a). Positional cloning studies enabled to demonstrate that *Ma<sub>1</sub>* encodes *PRR37*, a  
114 flowering repressor (Murphy et al., 2011a), *Ma<sub>3</sub>* gene was shown to be *PHYTOCHROME B*, a  
115 plant photoreceptor (Childs et al., 1997b), and *Ma<sub>6</sub>* has been identified as *SbGhd7*, a floral  
116 repressor regulated by the circadian clock and light signaling (Murphy et al., 2014a). Another  
117 *PHYTOCHROME* gene, *PHYC*, was proposed as a candidate gene for *Ma<sub>5</sub>* based on sequence  
118 alignment and allelic variation (Yang et al., 2014a). It has been then proposed that in long day  
119 conditions, *Ma<sub>3</sub>* regulates *Ma<sub>6</sub>* and *Ma<sub>1</sub>* that, in turn, both repress the expression of the grass  
120 floral integrator *EARLY HEADING DATE 1 (Ehd1)* which usually activates floral inductors  
121 (homologs of the *FLOWERING LOCUS T*, *SbCN8*, *SbCN12* and *SbCN15*) leading to delayed  
122 flowering. In these conditions, floral induction is also dependent of the circadian clock output.  
123 In short days, expressions of the floral repressors *Ma<sub>1</sub>* and *Ma<sub>6</sub>* are reduced resulting in floral  
124 initiation with the condition that the plants have satisfied other requirements for flowering  
125 (Murphy et al., 2014a, Yang et al., 2014a).

126 In photoperiod-sensitive sorghum varieties, maturity and morphology are strongly impacted by  
127 the sowing date and hence difficult to predict. Crop models are commonly applied to assist  
128 plant breeding by integrating physiological and biochemical understanding, along with  
129 agronomic practices, environment and genetic information (Messina et al., 2006). Modeling,  
130 mainly based on the CERES model (Major, 1980), has been used in sorghum to predict the  
131 effect of photoperiod on crop development. This model is based on a simple linear adjustment  
132 between photoperiod and the length of the vegetative phase (Chantereau et al., 2001b, Bangbol,  
133 2013, Sanon et al., 2014). After emergence, the shortest thermal time required to reach panicle  
134 initiation is known as the Basic Vegetative Phase (BVP) or intrinsic earliness (Belliard, 1982,  
135 Major et al., 1990b). During this phase, floral induction does not occur whatever the  
136 photoperiod conditions. The CERES sorghum model considers that below a critical photoperiod  
137 (P2O), the duration of the vegetative stage is constant and is equal to BVP. Above P2O, the  
138 duration of the vegetative stage increases as a linear function of day length.  
139 The objective of the current study is to analyze the genetic architecture of flowering time in  
140 photoperiod sensitive sorghum. In addition to direct measurements of flowering time, genetic  
141 analysis of eco-physiological model parameters shed a new light on the interpretation of  
142 flowering time QTLs and their use in plant breeding programs.

143

## 144 **2. Materials and methods**

### 145 **2.1. Plant materials**

146 A F<sub>3</sub> population was derived from the cross between Tiandougou and Lata3, two photoperiod-  
147 sensitive sorghum elite varieties from respectively IER (Institut d'Economie Rurale) and  
148 ICRISAT (International Crops Research Institute for the Semi-Arid Tropics) breeding  
149 programs. In normal sowing conditions (i.e. June-July), Tiandougou (127 days) is slightly later  
150 than Lata3 (120days). The population was developed by crossing the two parents in 2008. A  
151 single F<sub>1</sub> plant was selected and selfed to produce the F<sub>2</sub> generation. Single F<sub>2</sub> plants were  
152 advanced to F<sub>3</sub> generation in off-season. Rows of 10 F<sub>3</sub> plants were used and a single F<sub>3</sub> plant  
153 was randomly selected in each row to produce the F<sub>4</sub> generation. A total of 400 F<sub>3:5</sub> bulks of  
154 seeds were produced from 10 F<sub>4</sub> plants and used for phenotyping experiments.

### 155 **2.2. Field evaluation**

156 The mapping population was evaluated at Sotuba station (12°39'N, 7°56'E) of IER using an  
157 augmented block design. The Sotuba station is representative of the Sudano-Sahelian climate  
158 characterized by a rainy season in summer (May to November), with an average rainfall of

159 890mm. Monthly average maximal temperature ranges from 30 to 39 and the minimal is  
160 between 16 and 25.°C.

161 Experiments were conducted at three sowing dates corresponding to different photoperiod  
162 conditions. The population was observed in long and intermediate day length (usual cropping  
163 season, sowing 19 June and 15 July 2011) and short day length (off-season, sowing 3 October  
164 2011). Short day observations were rather conducted in October than December to avoid cold  
165 temperatures that could interact with photoperiod sensitivity (Vaksmann et al., 1998).  
166 Experimental trials were composed of 29 blocks of 14 F<sub>3:5</sub> families and were augmented with  
167 the two parents as checks. Each elementary plot consisted of 40 plants arranged in two lines of  
168 4 m.

169 Phenology was measured as the date when the flag-leaf's ligule was visible and fully emerged.  
170 This stage marks the end of leaf expansion and anthesis begins about 10 days later (Kouressy  
171 et al., 2008a). In sorghum, this observation is more accurate and repeatable than the scoring of  
172 the average flowering date commonly practiced in agronomic experiments (Tarumoto et al.,  
173 2003). Dates of flag-leaf ligulation were recorded when 50% of the plants in a plot had reached  
174 the stage. The number of Days from sowing To Flag-leaf Ligulation (DTFL) was calculated for  
175 each family and each sowing date.

176

### 177 **2.3. Environmental parameter calculation**

178 Thermal time after emergence was computed using an algorithm developed by Jones and Kiniry  
179 (1986), considering that growth speed increases as a linear function of temperature between a  
180 base and an optimal temperature, and then decreases linearly between optimal and maximal  
181 temperature. Cardinal temperatures were 11°C for base temperature (T<sub>b</sub>) (Lafarge and  
182 Hammer, 2002), 34.8°C for optimum temperature (T<sub>opt</sub>) and 44°C for maximum temperature  
183 (T<sub>max</sub>) (Ritchie and Alagarswamy, 1989a). The resulting thermal time per day was used to  
184 calculate the progress of developmental processes. The Thermal Time from emergence to flag  
185 leaf ligulation (TTFL) was computed for each family and each sowing date. Thermal time to  
186 panicle initiation (TTPI) was derived from TTFL using the formula proposed by Folliard et al.  
187 (2004). Day length calculation starts and ends when the sun is 6° below the horizon to account  
188 for photoperiod response during dawn and twilight (Aitken, 1974). At Sotuba station,  
189 experimental day length varies from 12h09 on the 21<sup>st</sup> December to 13h39 on the 21<sup>st</sup> June.

190

191

192

193       **2.4. Statistical analyses**

194 Phenotypic data were adjusted taking into account the augmented experimental design using  
195 the DAU.test function (package ‘*agricolae*’) in the R environment (R: The R Project for  
196 Statistical Computing). Normality of the variables was assessed using the Shapiro-Wilk  
197 normality test (shapiro.test function, package ‘*stats*’).

198

199       **2.5. Crop modeling**

200 The varietal response to photoperiod was modeled using the sorghum linear CERES model  
201 (Major et al., 1975, Alagarswamy and Ritchie, 1991) in which the photothermal time  
202 accumulation parameter was replaced by a critical photoperiod threshold parameter below  
203 which sorghum panicle initiation occurs (Folliard *et al.* 2004). Three genetic parameters were  
204 used to parametrize the model: the Basic Vegetative Phase (BVP), the critical photoperiod  
205 (P2O) and the photoperiod sensitivity (P2R).

206 The method used to calculate these parameters was previously presented by Chantereau *et al.*  
207 (2001). For each family, the TTPI was plotted for the three sowing dates against the photoperiod  
208 (day length) at panicle initiation date (Fig. 1). In practice, BVP was calculated from the minimal  
209 duration of the vegetative phase observed on the three sowing dates. In most cases BVP  
210 corresponded to the third sowing in October so TTPI3 and BVP values are virtually identical.  
211 The photoperiod sensitivity P2R was estimated as the slope of the line drawn between the points  
212 related to the sowing dates of June and July. The critical photoperiod P2O, corresponds to the  
213 day length at the intersection of the two lines.

214

## 215        **2.6. Genetic analysis**

216 DNA was extracted from 400 F<sub>4</sub> families, for which a bulk of 10 F<sub>4</sub> plants was used, using the  
217 MATAB method (Risterucci et al., 2000). Two hundred and 228 SNP markers were used. The  
218 information regarding these SNPs is accessible on the SNP genotyping service web page of the  
219 Integrated Breeding Platform<sup>1</sup>. SNP genotyping was outsourced to LGC Genomics.

220 The genetic map was built using Mapmaker (Lander et al., 1987). Linkage groups were  
221 determined using the “group” command with a LOD threshold of 3.0 and maximum distance  
222 of 50 cM. Loci were ordered in each group using the "order" command with the default  
223 parameters of the function. When several local orders were equally probable, the one in  
224 agreement with the expected order deduced from the physical positions of the markers was  
225 kept. The map distances for the final ordered linkage groups were computed using the Haldane  
226 mapping function.

227 QTL mapping was performed using R (R\_Development\_Core\_Team, 2008) and the *R-qt1*  
228 package (Broman and Sen, 2009a). The Multiple Interval Mapping method (MIM) was applied  
229 with a mapping step of 1 cM. To identify the multiple QTL model with maximal LOD score  
230 while controlling false positive rates, the *stepwiseqt1* function was applied that performs  
231 forward/backward model selection using a penalized likelihood approach to compare different  
232 model sizes, with penalties on QTL and pairwise interactions (Manichaikul et al., 2009). LOD  
233 penalties were derived for each trait from the result of 1000 permutations of a two-dimensional  
234 genome scan with a two-QTL model allowing covariates (functions *scantwo* and  
235 *calc.penalties*). The model optimizing the penalized LOD score criterion was fitted with the  
236 *fitqt1* function to get QTLs estimated effects. For each detected QTL, positions with maximum  
237 likelihood were identified (function *refineqt1*) before approximate 95% Bayesian credible  
238 intervals were calculated (function *bayesint*). The QTL map was drawn using the Spidermap  
239 software (Rami, unpublished).

240

---

<sup>1</sup><https://www.integratedbreeding.net/482/communities/genomics-crop-info/crop-information/gcp-kaspar-snp-markers>



## 241        **2.7.    *In silico* mapping of candidate genes**

242    Positions of genetic markers were retrieved from the *Sorghum bicolor* genome sequence  
243    assembly v3.1 (<http://genome.jgi.doe.gov>) using SNP flanking sequences and the Bowtie2  
244    program (Langmead and Salzberg, 2012). The region flanking the Chromosome 3 major QTL  
245    was investigated for candidate genes. Markers flanking the QTL (SB03075 and SB03085)  
246    were considered as the confidence interval on the physical map and their positions on the  
247    chromosome (48,834,006 and 52,478,805 bp, respectively) were used to identify genes  
248    present in this interval. Gene annotation information from Phytozome v11 for the *Sorghum*  
249    *bicolor* genome v3.1  
250    (<http://genome.jgi.doe.gov/pages/dynamicOrganismDownload.jsf?organism=Sbicolor#>) was  
251    used for gene ontology categorization using the best hit *Arabidopsis* name in the TAIR web  
252    tool for functional categorization (<http://www.arabidopsis.org/to/bulk/go/index.jsp>).  
253

## 254        **3.    Results**

255    The two parents exhibited differences for TTPI measurements and CERES model parameters.  
256    The photoperiod sensitivity of Lata3 was higher than the one of Tiandougou (P2R = 2446 and  
257    1275 °C days/h respectively). These values can be considered as high photoperiod sensitivities  
258    and correspond respectively to increases of 150 and 80 days of the duration of the vegetative  
259    phase per hour of photoperiod increase. The critical photoperiod of lata3 was higher than that of  
260    Tiandougou (P2O = 13.25 h and 12.99 h respectively). All the traits exhibited transgressive  
261    segregation in the population (Fig. 2). Transgression was particularly important for P2R and  
262    BVP values. Some families exceeded the value of 5000°Cdays/h for P2R. These extreme values  
263    are partly due to the CERES linear model as some families flower around the same time for  
264    June and July plantings and the values of P2R can then become very high.

265    TTPI1 and TTPI2 were strongly correlated ( $r=0.85$ ,  $p\text{-value}<0.001$ ) (Fig. 3) whereas their  
266    correlation with BVP was not significant ( $r=0.043$ ,  $p\text{-value}=0.492$  for TTPI1 and  $r=0.059$ ,  $p\text{-value}=0.239$  for TTPI2). P2O was significantly correlated to P2R ( $r=0.53$ ,  $p\text{-value}<0.001$ ), and  
267    slightly to BVP ( $r=0.17$ ,  $p\text{-value}=0.001$ ) while the correlation for BVP to P2R was not  
268    significant ( $r=0.058$ ,  $p\text{-value}=0.25$ ). P2O was highly correlated to TTPI2 ( $r=0.91$ ,  $p\text{-value}<0.001$ ) and to TTPI1 ( $r=0.63$ ,  $p\text{-value}<0.001$ ).  
269

271

272

273 **3.1. QTL detection**

274 A total of 16 QTLs spanning eight chromosomes were detected for all the traits under study.  
275 Ten QTLs were detected for Thermal Time to panicle initiation (TTPI) for the two sowing dates  
276 (TTPI1, TTPI2) and six QTLs for model parameters (BVP, P2O, P2R) (Table 1).

277 A major region was identified at 74 cM on chromosome 3 where three QTLs exhibiting high  
278 LOD score values were detected for TTPI1, TTPI2 and P2O (LOD score values of 70.5, 80.4  
279 and 59.1 respectively). The percentages of phenotypic variation explained by these QTLs were  
280 45.2%, 58.8% and 49.4% respectively. At this position, the Tiandougou allele increased TTPI1  
281 and TTPI2 of 66°Cd and 72°Cd respectively and decreased P2O coefficient of 0.11h.

282 In addition to this major locus, several other QTLs with smaller effects were detected for all  
283 traits except P2O for which only the major QTL of chromosome 3 (74 cM) was significant. For  
284 TTPI1, 6 QTLs were identified on chromosomes 2, 3, 4, 7, 8, and 10 explaining from 2.4% to  
285 7.1% of phenotypic variance. The multiple QTL model, including an interaction term between  
286 the two QTLs on chromosomes 2 and 4, explained 63.9% of the total phenotypic variance for  
287 this trait. Two QTLs for TTPI2 on chromosomes 2 and 10 were detected, explaining 2.8% and  
288 2.2% of phenotypic variance respectively, and collocating with two QTLs for TTPI1. Three  
289 QTLs for P2R were detected on chromosomes 3, 7, and 9, explaining 5.8%, 5.6% and 3.7% of  
290 the phenotypic variance respectively. For all of them, the Lata3 parent contributed positive  
291 alleles. Finally, two QTLs were detected for BVP, on chromosomes 6 and 9, explaining 4.8%  
292 and 5.7% of phenotypic variance and having positive contributions from Lata3 and Tiandougou  
293 respectively.

294 **3.2. Candidate gene selection by *in silico* mapping**

295 The region of the major QTLs locus on chromosome 3 involving the P2O, TTPI1 and TTPI2  
296 traits was investigated to identify possible candidate genes involved in the variation of these  
297 traits. This region, flanked by SB03075 and SB03085 markers, represented 3,644,799 bp and  
298 included 159 predicted genes of which 124 (78%) were annotated in the Phytozome v11  
299 database. Gene ontologies (GO) analysis allowed to allocate these genes to 14 main categories  
300 (Fig. 3) with the “cellular” and “metabolic processes” being the largest ones. The  
301 developmental process ontology, accounted for 22 annotations, representing 15 genes. Among  
302 those 15 genes, only two were involved in flowering regulation. The first one,  
303 Sobic.003G196300, homologous to AT3G24870, is involved in the “regulation of  
304 photoperiodism, flowering” ontology (GO:2000028) and is located at 2,054,278 bp from the  
305 LOD peak. The second one, Sobic.003G191700, homologous to AT2G25930 and

306 corresponding to the *EARLY FLOWERING 3* gene (*Elf3*), is involved in the “entrainment of  
307 circadian clock” (GO:0009649), “photoperiodism, flowering” (GO:0048573), “circadian  
308 rhythm” (GO:0007623) and “regulation of flower development” (GO:0009909) ontologies and  
309 is located at 1,191,464 bp from the LOD peak.

310

#### 311 **4. Discussion**

312 This study use a bi-parental population to implement genetic analysis of maturity in photoperiod  
313 sensitive sorghum. Flowering time was decomposed into photoperiod response model  
314 parameters. Most of the previous genetic studies focusing on flowering time in sorghum  
315 involved either highly contrasting parents for photoperiod sensitivity (Lin et al., 1995b, Rooney  
316 and Aydin, 1999b, Chantreau et al., 2001b, Zou et al., 2012) that identified major genes alleles  
317 removing (or strongly reducing) photoperiod-sensitivity or photoperiod insensitive parents that  
318 identified in most cases QTLs associated with the temperature component of flowering (Mace  
319 et al., 2013b). No genetic study has been focused on fine regulation of photoperiod-sensitivity.  
320 In this study, the use of a mapping population derived from two photoperiod-sensitive parents  
321 enabled to deepen the understanding of the genetic determinisms of flowering time and  
322 photoperiod sensitivity. A major QTL involved in the genetic control of critical photoperiod  
323 (P2O) was identified on chromosome 3. Although P2O values seemed nearly similar between  
324 the two parents (15 min difference), this faint difference induced a variation of flowering time  
325 reaching one week, such variation being extremely important in terms of adaptation to local  
326 conditions.

327

##### 328 **4.1. Monitoring critical photoperiod for climate smart varieties in West Africa**

329 Since the late 1960s, sub Saharan Africa has experienced drought events of unprecedented  
330 severity in recorded history (Zeng, 2003). As a result of climate change, rain distribution  
331 changes can potentially affect the occurrence of drought, while changes in temperature can  
332 affect flowering time. Even if climate models are unclear in predicting the future distribution  
333 of African rainfall, an increase in climate variability and a succession of periods of drought and  
334 flooding are expected. Elite grain sorghum varieties developed by breeding programs have been  
335 mainly selected for photoperiod insensitivity and early flowering. However, this approach  
336 failed to produce efficient varieties for the sub-Saharan African context due to the variability  
337 of environmental conditions in this region. Contrastingly, sorghum landraces are examples of  
338 adaptive selection to climate variability. In West Africa, the landraces adapted to the North

339 (dry) zone have a high critical photoperiod, low photoperiod sensitivity and flower earlier than  
340 the landraces adapted to the South (wet) zone characterized by a lower critical photoperiod and  
341 a higher photoperiod sensitivity (Sanon et al., 2014).

342 It is also considered that climate change, and in particular temperature increase, may lead to an  
343 increase of development rates resulting in a shorter life cycle and a decrease of photosynthesis  
344 and crop productivity (Craufurd and Wheeler, 2009, Ainsworth and Ort, 2010). In this context,  
345 photoperiod sensitivity appears as a key trait to face climate change since flowering depends  
346 mainly on photoperiod, an astronomical phenomenon that will remain unchanged regardless of  
347 climatic change.

348 The few studies that have investigated the variation of the critical photoperiod trait in sorghum  
349 varieties found values of P2O varying from 10.2 h to 17.5 h (Craufurd et al., 1999, Chantreau  
350 et al., 2001b, Abdulai et al., 2012, Sanon et al., 2014, Ritchie and Alagarswamy, 1989c).  
351 However, comparison with our P2O values is difficult because genetic coefficients of main  
352 development models are not stable with latitude, invalidating data obtained from multi-local  
353 experimentations (Abdulai et al., 2012). In addition, the photoperiod calculation method varies  
354 according to studies (including or excluding dawn and dusk), which can result in up to one hour  
355 difference on daylength. However, the existence of varieties with lower or higher P2O values  
356 (late and early maturing varieties, respectively) suggests the presence of other alleles / genes  
357 coding for this trait. Their identification will be critical to improve sorghum climate adaptation.  
358 The major QTL for critical photoperiod allows the development of alternative breeding options  
359 based on improved variety selection for various environments. Until now, it was thought that  
360 maturity and photoperiod sensitivity were closely linked and that early varieties were  
361 necessarily less sensitive to photoperiod (Clerget et al., 2007). Our results show that changes  
362 in critical photoperiod make possible to modify the earliness without changing photoperiod  
363 sensitivity. This finding opens up new avenues for the development of early maturing varieties  
364 well suited for northern Mali while preserving photoperiod sensitivity and therefore stability of  
365 crop production.

366

#### 367 **4.2. Candidate gene search identified ELF3 as downstream regulator of photoperiod** 368 **response in sorghum**

369 In the major QTL region controlling the critical photoperiod variability (chromosome 3, 74  
370 cM), the gene ontology categorization approach (Ashburner et al., 2000) allowed the detection  
371 of 15 genes corresponding to the “developmental process” class which is relevant to the

372 flowering time process. Within these 15 genes, two were of particular interest since they have  
373 been shown to be involved in photoperiodism and flowering in other species.

374 Sobic.003G191700 is one of the two sorghum orthologs of *EARLY FLOWERING 3 (ELF3)* for  
375 which functional evidences support its implication in the regulation of circadian rhythms and  
376 photoperiodic flowering in both rice (Fu et al., 2009, Zhao et al., 2012a, Yang et al., 2013) and  
377 *Arabidopsis* (Hicks et al., 1996, McWatters et al., 2000, Hicks et al., 2001, Reed et al., 2000,  
378 Liu et al., 2001, Covington et al., 2001, Kim et al., 2005, Yoshida et al., 2009). These studies  
379 considered that *ELF3* is a time taker that modulates resetting of the circadian clock and  
380 integrates both temperature and photoperiod signals. Furthermore, different independent *ELF3*  
381 alleles impacting flowering dynamics were identified in rice, two as QTLs (*early flowering 7*  
382 (*ef7*) (Saito et al., 2012) and *heading date 17 (hd17)* (Matsubara et al., 2012, Matsubara et al.,  
383 2008), two insertion mutants (Zhao et al., 2012a, Yang et al., 2013) and transgenic RNAi lines  
384 (Zhao et al., 2012a). These results showed that *ELF3* is involved in photoperiod sensitivity in  
385 rice and that this gene is a key component of the circadian clock through its regulatory effect  
386 on several downstream genes.

387 Several sorghum orthologs of these “downstream genes” have been identified as major maturity  
388 genes. Under both short and long-day conditions, *Os-ELF3-1* promotes flowering by negatively  
389 regulating the floral repressor gene *Ghd7* (Saito et al., 2012), which is the ortholog of Ma<sub>6</sub> in  
390 sorghum (Murphy et al., 2014a). Zhao et al. (2012a) proposed that *OsELF3-1* regulates light  
391 input into the clock by binding to *PHYTOCHROME B* (Ma<sub>3</sub> in sorghum) and thus forming a  
392 complex that regulates flowering. In addition, the same study showed that *PRR37*, also known  
393 as Ma<sub>1</sub> in sorghum (Murphy et al., 2011a), is negatively regulated by *Os-ELF3-1* (Zhao et al.,  
394 2012a).

395 It is also important to mention, that the second sorghum ortholog of *ELF3* (Sobic.009G257300)  
396 is localized in the vicinity of the QTL interval for the Basic Vegetative Phase on the  
397 chromosome 9. These results are consistent with the rice results in which genetic analysis of a  
398 late heading-time mutant line has been used to demonstrate that the *OsELF3-2* gene (the second  
399 ortholog of *ELF3* in rice) conferred an extremely long Basic Vegetative Phase and increased  
400 photoperiod sensitivity under long photoperiod (Yuan et al., 2009). Furthermore, considering  
401 the regulation of *PRR37* by *Os-ELF3-1* (Zhao et al., 2012a), it is interesting to underline that  
402 in our study the sorghum Ma<sub>1</sub> gene (Sobic.006G057866) is located within the QTL interval for  
403 the basic vegetative phase on chromosome 6, starting at 40,304,883 bp near the SB06034  
404 marker (40,227,313 bp).

405 Circadian clock genes functions appear to be well conserved among species and they underlie  
406 QTL with positive effect on key agricultural traits, particularly flowering time but also yield  
407 and biomass production (Bendix et al., 2015). In rice, *OsELF3-2* mutant found application in  
408 programs to breed for rice suitable to low latitudes (Yuan et al., 2009). In barley, *ELF3* loci  
409 conferring reduced or non-existent responses to photoperiod have also been used to breed early  
410 maturity cultivars for short growing seasons at different latitudes in the world (Laurie et al.,  
411 1995, Zakhrabekova et al., 2012, Faure et al., 2012). The location of one of the *ELF3* sorghum  
412 ortholog (Sobic.003G191700) within a major QTL affecting the critical photoperiod  
413 (chromosome 3) in addition with the observed proximity of Sobic.009G257300 (the second  
414 ortholog of *ELF3*) with a QTL affecting the basic vegetative phase (chromosome 9) are  
415 consistent with the previous results reported in *Arabidopsis*, rice and barley.

416 As mentioned previously, a second flowering gene was identified in the Chromosome 3 critical  
417 photoperiod region, it corresponds to Sobic.003G196300 which is similar to AT3G24870  
418 (*ATEAF1B*), that has been shown to be necessary for proper timing of transition to flowering  
419 in *Arabidopsis* (Bieluszewski et al., 2015). Although, phenotypical analysis of *Arabidopsis*  
420 mutants showed decreased expression of the *Flowering Locus C (FLC)*, which prevents  
421 flowering, and early flowering, this unique study provides only a few elements to support the  
422 photoperiodic control of flowering by this gene.

423

424

425 **5. Conclusions**

426 This approach combining crop modeling and genetic analysis enabled to reach a better  
427 ecophysiological understanding of the major QTL identified in our analysis. We demonstrated  
428 that it is possible to decompose flowering time into different physiological parameters, which  
429 are controlled by specific genomic regions.

430 The major QTL detected on chromosome 3 controlling critical photoperiod collocate with an  
431 ortholog of the *ELF3* gene for which functional evidences have been provided in *Arabidopsis*  
432 and rice. Further analysis of this locus using Genome Wide Association Studies (GWAS),  
433 genetic analysis in multi-parental design, and isogenic lines would allow to precise its map  
434 position to support the role of this candidate gene as well as to catalogue allelic variation. This  
435 gene, probably acting as an upstream regulator of other known maturity genes in sorghum in  
436 the photoperiodic flowering pathway, extends breeding perspectives by increasing allele  
437 combination possibilities with already known sorghum maturity genes.

438 In this study, plant modeling allowed to reach a better ecophysiological understanding of the  
439 detected QTLs, but in a medium to long-term perspective the expectations linked to the use of  
440 combined modeling and quantitative genetic approaches go far beyond. Integrating modeling  
441 and molecular genetics should allow the design of breeding ideotypes according to climate  
442 change scenarios capitalizing on the prediction ability of ecophysiological modeling and on the  
443 precision and efficiency of marker assisted breeding.

444 The understanding of photoperiod sensitivity genetic bases opens important sorghum breeding  
445 perspectives by supplying innovative tools and genes to fine-tune plant adaptation to climate  
446 variability and climate change.

447

448 **6. Acknowledgements**

449 This work was supported by a grant from the Generation Challenge Programme (Grant Number  
450 G4008.48).

451

## References

- Abdulai AL, Kouressy M, Vaksmann M, Asch F, Giese M, Holger B. 2012. Latitude and Date of Sowing Influences Phenology of Photoperiod-Sensitive Sorghums. *Journal of Agronomy and Crop Science* 198, 340-348.
- Ainsworth EA, Ort DR. 2010. How Do We Improve Crop Production in a Warming World? *Plant Physiology* 154, 526-530.
- Aitken Y. 1974. *Flowering Time, Climate and Genotype: The Adaptation of Agricultural Species to Climate Through Flowering Responses*: Melbourne University Press.
- Alagarswamy G, Ritchie JT. 1991. Phasic Development in CERES-Sorghum Model. *Hodges, T. (Ed.), Predicting Crop Phenology, CRC Press, Boca Raton*, 143–152.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene Ontology: tool for the unification of biology. *Nat Genet* 25, 25-29.
- Bangbol SH. 2013. Genetic characterization of flowering time in sorghum. Doctor of Philosophy, University of Queensland, 158.
- Belliard J. 1982. analyse génétique et physiologique du système de contrôle photopériodique de la floraison chez le mil africain. Doctorat d'état, Université de paris-Sud, Orsay, 120.
- Bendix C, Marshall CM, Harmon FG. 2015. Circadian Clock Genes Universally Control Key Agricultural Traits. *Molecular Plant* 8, 1135-1152.
- Bieluszewski T, Galganski L, Sura W, Bieluszewska A, Abram M, Ludwikow A, Ziolkowski PA, Sadowski J. 2015. AtEAF1 is a potential platform protein for Arabidopsis NuA4 acetyltransferase complex. *BMC Plant Biology* 15, 75.
- Boss PK, Bastow RM, Mylne JS, Dean C. 2004. Multiple Pathways in the Decision to Flower: Enabling, Promoting, and Resetting. *The Plant Cell* 16, S18-S31.
- Brady JA. 2006. Sorghum Ma5 and Ma6 maturity genes. Doctor of philosophy, Texas A&M University, 142.
- Broman KW, Sen S. 2009. *A Guide to QTL Mapping with R/qlt*.
- Brown PJ, Klein PE, Bortiri E, Acharya CB, Rooney WL, Kresovich S. 2006. Inheritance of inflorescence architecture in sorghum. *Theor Appl Genet* 113, 931-942.
- Chanterneau J, Trouche G, Rami JF, Deu M, Barro C, Grivet L. 2001. RFLP mapping of QTLs for photoperiod response in tropical sorghum. *Euphytica* 120, 183-194.
- Childs KL, Miller FR, Cordonnier-Pratt MM, Pratt LH, Morgan PW, Mullet JE. 1997. The Sorghum Photoperiod Sensitivity Gene, Ma3, Encodes a Phytochrome B. *Plant Physiology* 113, 611-619.
- Clerget B, Rattunde HFW, Dagnoko S, Chanterneau J. 2007.) An easy way to assess photoperiod sensitivity in sorghum: Relationships of the vegetative-phase duration and photoperiod sensitivity. *SAT eJournal*.
- Covington MF, Panda S, Liu XL, Strayer CA, Wagner DR, Kay SA. 2001. ELF3 Modulates Resetting of the Circadian Clock in Arabidopsis. *The Plant Cell* 13, 1305-1316.
- Crasta OR, Xu WW, Rosenow DT, Mullet J, Nguyen HT. 1999. Mapping of post-flowering drought resistance traits in grain sorghum: association between QTLs influencing premature senescence and maturity. *Molecular Genetics and Genomics* 262, 579-588.
- Craufurd PQ, Mahalakshmi V, Bidinger FR, Mukuru SZ, Chanterneau J, Omanga PA, Qi A, Roberts EH, Ellis RH, Summerfield RJ, Hammer GL. 1999. Adaptation of sorghum: characterisation of genotypic flowering responses to temperature and photoperiod. *TAG Theoretical and Applied Genetics* 99, 900-911.



- Craufurd PQ, Wheeler TR. 2009. Climate change and the flowering time of annual crops. *J. Exp. Bot.* 60, 2529-2539.
- El Mannai Y, Shehzad T, Okuno K. 2011. Variation in flowering time in sorghum core collection and mapping of QTLs controlling flowering time by association analysis. *Genetic Resources and Crop Evolution* 58, 983-989.
- Faure S, Turner A, Gruszka D, Christodoulou V, J Davis S, Korff M, A Laurie D. 2012. *Mutation at the circadian clock gene EARLY MATURITY 8 adapts domesticated barley.*
- Feltus FA, Hart GE, Schertz KF, Casa AM, Kresovich S, Abraham S, Klein PE, Brown PJ, Paterson AH. 2006. Alignment of genetic maps and QTLs between inter- and intra-specific sorghum populations. *Theoretical and Applied Genetics* 112, 1295-1305.
- Folliard A, Traore PCS, Vaksman M, Kouressy M. 2004. Modeling of sorghum response to photoperiod: a threshold-hyperbolic approach. *Field Crops Research* 89, 59-70.
- Fu C, Yang XO, Chen X, Chen W, Ma Y, Hu J, Li S. 2009. OsEF3, a homologous gene of Arabidopsis ELF3, has pleiotropic effects in rice. *Plant Biology* 11, 751-757.
- Garner WW, Allard HA. 1923. Further studies in photoperiodism, the response of the plant to relative length of day and night. *J Agric Res* 23: 871-920. 23, 871-920.
- Hart GE, Schertz KF, Peng Y, Syed NH. 2001. Genetic mapping of Sorghum bicolor (L.) Moench QTLs that control variation in tillering and other morphological characters. *TAG Theoretical and Applied Genetics* 103, 1232-1242.
- Hicks KA, Albertson TM, Wagner DR. 2001. EARLY FLOWERING3 Encodes a Novel Protein That Regulates Circadian Clock Function and Flowering in Arabidopsis. *The Plant Cell* 13, 1281-1292.
- Hicks KA, Millar AJ, Carré IA, Somers DE, Straume M, Meeks-Wagner DR, Kay SA. 1996. Conditional Circadian Dysfunction of the Arabidopsis early-flowering 3 Mutant. *Science* 274, 790-792.
- Higgins RH, Thurber CS, Assaranurak I, Brown PJ. 2014. Multiparental Mapping of Plant Height and Flowering Time QTL in Partially Isogenic Sorghum Families. *G3: Genes/Genomes/Genetics* 4, 1593-1602.
- Jones CA, Kiniry JR. 1986. *CERES-Maize : A simulation model of maize growth and development*: Rexa A&M University Press, College Station, Texas.
- Kebede H, Subudhi PK, Rosenow DT, Nguyen HT. 2001. Quantitative trait loci influencing drought tolerance in grain sorghum (Sorghum bicolor L. Moench). *TAG Theoretical and Applied Genetics* 103, 266-276.
- Kim J-S. 2003. Genomic analysis of sorghum by fluorescence in situ hybridization., Texas A&M University, 136.
- Kim W-Y, Hicks KA, Somers DE. 2005. Independent Roles for EARLY FLOWERING 3 and ZEITLUPE in the Control of Circadian Timing, Hypocotyl Length, and Flowering Time. *Plant Physiology* 139, 1557-1569.
- Kouressy M, Dingkuhn M, Vaksman M, Clement-Vidal A, Chantereau J. 2008a. Potential contribution of dwarf and leaf longevity traits to yield improvement in photoperiod sensitive sorghum. *European Journal of Agronomy* 28, 195-209.
- Kouressy M, Dingkuhn M, Vaksman M, Heinemann AB. 2008b. Adaptation to diverse semi-arid environments of sorghum genotypes having different plant type and sensitivity to photoperiod. *Agricultural and Forest Meteorology* 148, 357-371.
- Kouressy M, Traoré SB, Vaksman M, Grum M, Maikano I, Soumaré M, Traoré PS, Bazile D, Dingkuhn M, Sidibé A. 2008c. Adaptation des sorghos du Mali à la variabilité climatique. *Cahiers Agricultures* 17, 95-100.

- Lafarge TA, Hammer GL. 2002. Predicting plant leaf area production:: shoot assimilate accumulation and partitioning, and leaf area ratio, are stable for a wide range of sorghum population densities. *Field Crops Research* 77, 137-151.
- Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L. 1987. MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1, 174-181.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Meth* 9, 357-359.
- Laurie DA, Pratchett N, Snape JW, Bezant JH. 1995. RFLP mapping of five major genes and eight quantitative trait loci controlling flowering time in a winter × spring barley (*Hordeum vulgare* L.) cross. *Genome* 38, 575-585.
- Lin YR, Schertz KF, Paterson AH. 1995. Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population. *Genetics* 141, 391-411.
- Liu XL, Covington MF, Fankhauser C, Chory J, Wagner DR. 2001. ELF3 Encodes a Circadian Clock-Regulated Nuclear Protein That Functions in an Arabidopsis PHYB Signal Transduction Pathway. *The Plant Cell* 13, 1293-1304.
- Mace ES, Hunt CH, Jordan DR. 2013. Supermodels: sorghum and maize provide mutual insight into the genetics of flowering time. *Theoretical and Applied Genetics*, 1-19.
- Major DJ. 1980. Photoperiod response characteristics controlling flowering of nine crop species. *Canadian Journal of Plant Science* 60, 777-784.
- Major DJ, Johnson DR, Tanner JW, Anderson IC. 1975. Effects of Daylength and Temperature on Soybean Development. *Crop Sci.* 15, 174-179.
- Major DJ, Rood SB, Miller FR. 1990. Temperature and Photoperiod Effects Mediated by the Sorghum Maturity Genes. *Crop Sci* 30, 305-310.
- Manichaikul A, Moon JY, Sen Ś, Yandell BS, Broman KW. 2009. A Model Selection Approach for the Identification of Quantitative Trait Loci in Experimental Crosses, Allowing Epistasis. *Genetics* 181, 1077.
- Matsubara K, Kono I, Hori K, Nonoue Y, Ono N, Shomura A, Mizubayashi T, Yamamoto S, Yamanouchi U, Shirasawa K, Nishio T, Yano M. 2008. Novel QTLs for photoperiodic flowering revealed by using reciprocal backcross inbred lines from crosses between japonica rice cultivars. *Theoretical and Applied Genetics* 117, 935-945.
- Matsubara K, Ogiso-Tanaka E, Hori K, Ebana K, Ando T, Yano M. 2012. Natural Variation in Hd17, a Homolog of Arabidopsis ELF3 That is Involved in Rice Photoperiodic Flowering. *Plant and Cell Physiology* 53, 709-716.
- McWatters HG, Bastow RM, Hall A, Millar AJ. 2000. The ELF3zeitnehmer regulates light signalling to the circadian clock. *Nature* 408, 716-720.
- Messina CD, Boote KJ, Löffler C, Jones JW, Vallejos CE. 2006. Model-assisted genetic improvement of crops. In: Wallach D, Makowski D, Jones JW, eds. *working with Dynamic crop Models*. Amsterdam: Elsevier, 309-335.
- Morgan PW, Finlayson SA. 2000. Physiology and genetics of maturity and height. In: Smith CW, Frederiksen RA, eds. *Sorghum: Origin, History, Technology, and Production*. New York: Wiley Series in Crop Science, 240-242.
- Mouradov A, Cremer F, Coupland G. 2002. Control of Flowering Time: Interacting Pathways as a Basis for Diversity. *Plant Cell* 14, S111-130.
- Mullet JE, Rooney WL, Klein PE, Morishige D, Murphy R, Brady JA. 2016. Discovery and utilization of sorghum genes (Ma5, Ma6). Vol. US 216/0340684 A9.

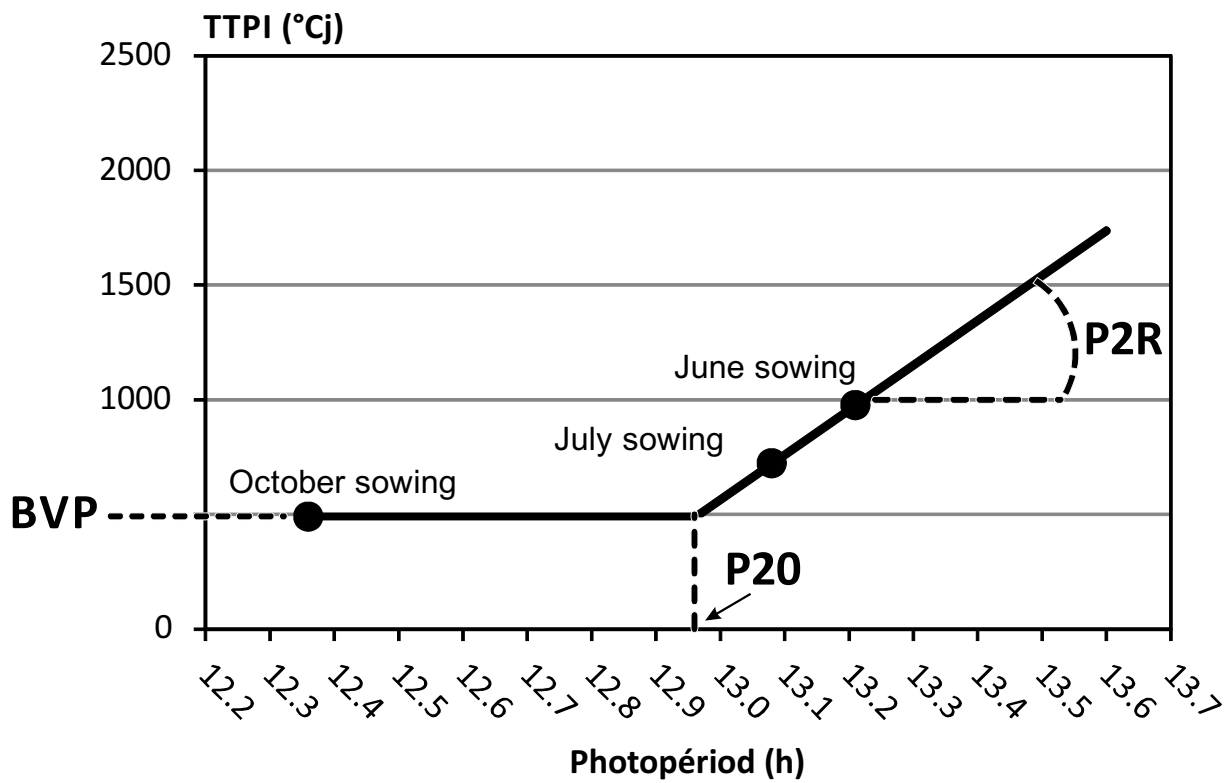
- Murphy RL, Klein RR, Morishige DT, Brady JA, Rooney WL, Miller FR, Dugas DV, Klein PE, Mullet JE. 2011. Coincident light and clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum. *Proceedings of the National Academy of Sciences*.
- Murphy RL, Morishige DT, Brady JA, Rooney WL, Yang S, Klein PE, Mullet JE. 2014. Ghd7 (Ma6) Represses Sorghum Flowering in Long Days: Ghd7 Alleles Enhance Biomass Accumulation and Grain Production. *Plant Gen.* 7, 1-10.
- Murray SC, Sharma A, Rooney WL, Klein PE, Mullet JE, Mitchell SE, Kresovich S. 2008. Genetic Improvement of Sorghum as a Biofuel Feedstock: I. QTL for Stem Sugar and Grain Nonstructural Carbohydrates. *Crop Sci* 48, 2165-2179.
- Olson SN, Ritter K, Rooney W, Kemanian A, McCarl BA, Zhang Y, Hall S, Packer D, Mullet J. 2012. High biomass yield energy sorghum: developing a genetic model for C4 grass bioenergy crops. *Biofuels, Bioproducts and Biorefining* 6, 640-655.
- PARH, D. K. 2005. DNA-based markers for ergot resistance in sorghum. PhD thesis. University of Queensland, School of Land and Food Sciences.
- Quinby JR. 1966. Fourth Maturity Gene Locus in Sorghum. *Crop Sci.* 6, 516-518.
- Quinby JR. 1967. The maturity genes of sorghum. In: Norman AG, ed. *Advances in Agronomy XIX*. New York: Academic Press Inc, 267-305.
- R\_Development\_Core\_Team. 2008. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
- Reed JW, Nagpal P, Bastow RM, Solomon KS, Dowson-Day MJ, Elumalai RP, Millar AJ. 2000. Independent Action of ELF3 and phyB to Control Hypocotyl Elongation and Flowering Time. *Plant Physiology* 122, 1149.
- Risterucci AM, Grivet L, N'Goran JAK, Pieretti I, Flament MH, Lanaud C. 2000. A high-density linkage map of Theobroma cacao L. *Theoretical and Applied Genetics* 101, 948-955.
- Ritchie JT, Alagarswamy G. 1989a. Genetic Coefficients for CERES Models. In: Virmani SMaT, H L S and Alagarswamy, ed. *Reserch bulletin n°12: International Crops Research Institute for the Semi-Arid Tropics*, 27-29.
- Ritchie JT, Alagarswamy G. 1989b. Simulation of Sorghum and Pearl Millet Phenology *Modeling the Growth and Development of Sorghum and Pearl Millet: Res.Bull. n°12, ICRISAT*, 24-29.
- Ritter K, Jordan D, Chapman S, Godwin I, Mace E, Lynne McIntyre C. 2008. Identification of QTL for sugar-related traits in a sweet × grain sorghum (&i>Sorghum bicolor L. Moench) recombinant inbred population. *Molecular Breeding* 22, 367-384.
- Rooney WL, Aydin S. 1999. Genetic control of a photoperiod-sensitive response in Sorghum bicolor (L.) moench. *Crop Sci* 39, 397-400.
- Saito H, Ogiso-Tanaka E, Okumoto Y, Yoshitake Y, Izumi H, Yokoo T, Matsubara K, Hori K, Yano M, Inoue H, Tanisaka T. 2012. Ef7 Encodes an ELF3-like Protein and Promotes Rice Flowering by Negatively Regulating the Floral Repressor Gene Ghd7 under Both Short- and Long-Day Conditions. *Plant and Cell Physiology* 53, 717-728.
- Sanon M, Hoogenboom G, Traoré SB, Sarr B, Garcia AGy, Somé L, Roncoli C. 2014. Photoperiod sensitivity of local millet and sorghum varieties in West Africa. *NJAS - Wageningen Journal of Life Sciences* 68, 29-39.
- Shiringani AL, Frisch M, Friedt W. 2010. Genetic mapping of QTLs for sugar-related traits in a RIL population of Sorghum bicolor L. Moench. *Theor Appl Genet* 121, 323–336.
- Srinivas G, Satish K, Madhusudhana R, Nagaraja Reddy R, Murali Mohan S, Seetharama N. 2009. Identification of quantitative trait loci for agronomically important traits and their association

- with genic-microsatellite markers in sorghum. *TAG Theoretical and Applied Genetics* 118, 1439-1454.
- Stephens JC, Miller FR, Rosenow DT. 1967. Conversion of Alien Sorghums to Early Combine Genotypes. *Crop Sci.* 7, 396-396.
- Swaminathan MS. 2006. An Evergreen Revolution. *Crop Sci* 46, 2293-2303.
- Tarumoto I, Yanase M, Iwahara Y, Kuzumi Y, Morikawa T, Kasuga S. 2003. Inheritance of a Thermo-sensitivity Gene Controlling Flower Initiation in Sorghum. *Breeding Science* 53, 353-357.
- Upadhyaya H, Wang Y-H, Gowda CLL, Sharma S. 2013. Association mapping of maturity and plant height using SNP markers with the sorghum mini core collection. *Theoretical and Applied Genetics* 126, 2003-2015.
- Vaksmann M, Chantereau J, Bahmani I, Ag Hamada M, Chartier M, Bonhomme R. 1998. Influence of night temperature on photoperiod response of a west african guinea sorghum landrace. In: Ratnadass A. (éd.) CJé, Gigou J. (éd.). ed. *Amélioration du sorgho et de sa culture en Afrique de l'Ouest et du Centre. Atelier de Restitution du Programme Conjoint sur le Sorgho*. Bamako, Mali: CIRAD, 23-28.
- Yang S, Murphy RL, Morishige DT, Klein PE, Rooney WL, E. MJ. 2014. Sorghum Phytochrome B Inhibits Flowering in Long Days by Activating Expression of SbPRR37 and SbGHD7, Repressors of SbEHD1, SbCN8 and SbCN12. *Plos One* 9, 1-12.
- Yang Y, Peng Q, Chen G, Li X, Wu C. 2013. OsELF3 Is Involved in Circadian Clock Regulation for Promoting Flowering under Long-Day Conditions in Rice. *Molecular Plant* 6, 202-215.
- Yoshida R, Fekih R, Fujiwara S, Oda A, Miyata K, Tomozoe Y, Nakagawa M, Niinuma K, Hayashi K, Ezura H, Coupland G, Mizoguchi T. 2009. Possible role of EARLY FLOWERING 3 (ELF3) in clock-dependent floral regulation by SHORT VEGETATIVE PHASE (SVP) in *Arabidopsis thaliana*. *New Phytologist* 182, 838-850.
- Yuan Q, Saito H, Okumoto Y, Inoue H, Nishida H, Tsukiyama T, Teraishi M, Tanisaka T. 2009. Identification of a novel gene ef7 conferring an extremely long basic vegetative growth phase in rice. *Theoretical and Applied Genetics* 119, 675-684.
- Zakhrabekova S, Gough SP, Braumann I, Müller AH, Lundqvist J, Ahmann K, Dockter C, Matyszczak I, Kurowska M, Druka A, Waugh R, Graner A, Stein N, Steuernagel B, Lundqvist U, Hansson M. 2012. Induced mutations in circadian clock regulator Mat-a facilitated short-season adaptation and range extension in cultivated barley. *Proceedings of the National Academy of Sciences of the United States of America* 109, 4326-4331.
- Zeng N. 2003. Drought in the Sahel. *Science* 302, 999-1000.
- Zhao J, Huang X, Ouyang X, Chen W, Du A, Zhu L, Wang S, Deng XW, Li S. 2012. OsELF3-1, an Ortholog of *Arabidopsis* EARLY FLOWERING 3, Regulates Rice Circadian Rhythm and Photoperiodic Flowering. *Plos One* 7, e43705.
- Zou G, Zhai G, Feng Q, Yan S, Wang A, Zhao Q, Shao J, Zhang Z, Zou J, Han B, Tao Y. 2012. Identification of QTLs for eight agronomically important traits using an ultra-high-density map based on SNPs generated from high-throughput sequencing in sorghum under contrasting photoperiods. *Journal of Experimental Botany* 63, 5451-5462.

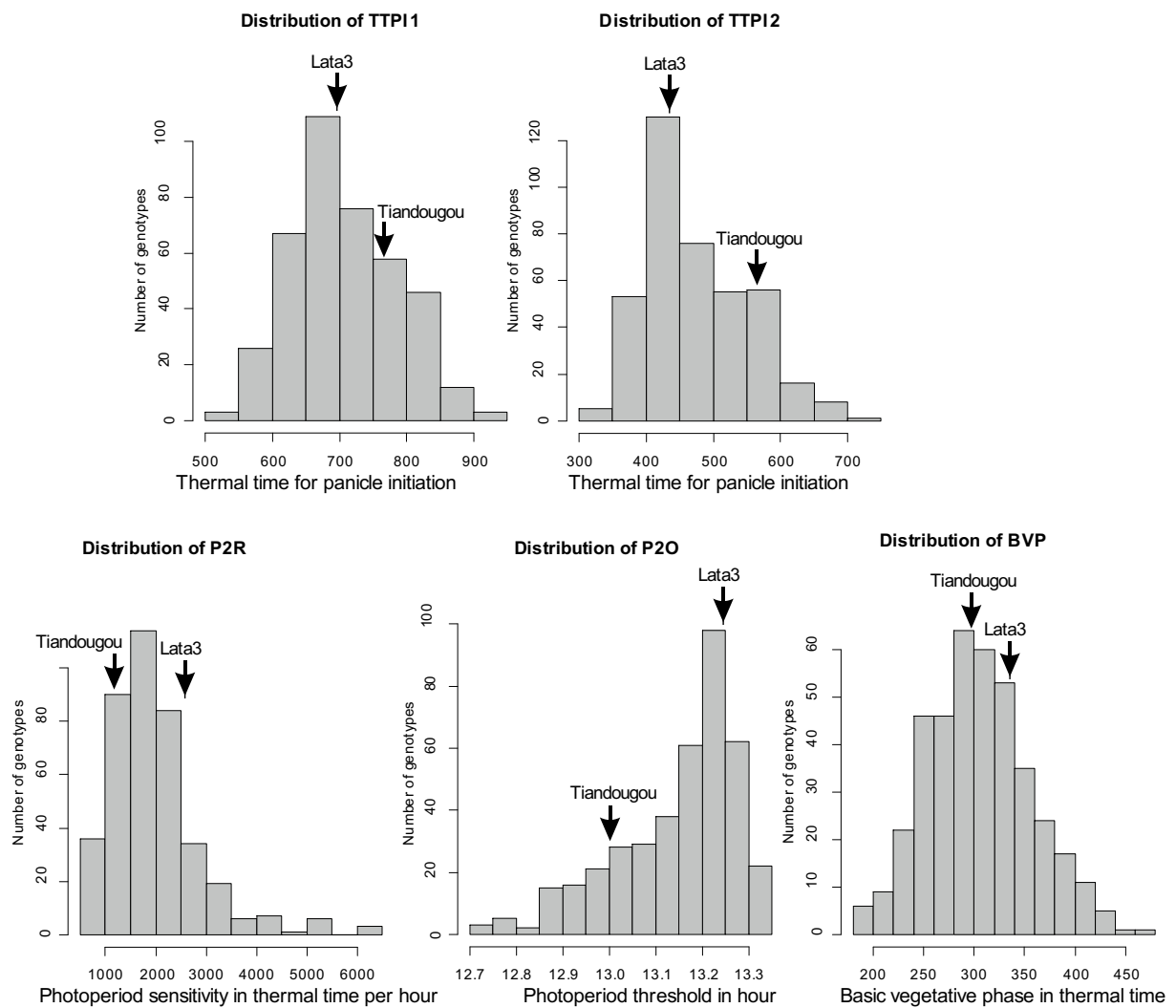
**Table 1.** Parameters associated to the quantitative trait loci (QTL) detected in the Sorghum ‘Tiandougou’ x ‘Lata3’ population by multiple QTL mapping (MQM) for phenological and model derived variables: Thermal Time to panicle initiation (TTPI) for the 3 sowing dates (TTPI1, TTPI2 and TTPI3), Basic Vegetative Phase (BVP), critical photoperiod (P2O) and photoperiod sensitivity (P2R).

Trait	CHR	LOD score	Position (cM)	95% Bayesian credible interval (cM)	Total R2 (%)	Additive effect	Position on the Sorghum genome						
							Closest marker	Left marker	Right marker	Closest marker			
<b>TTPI1</b>	2	15.5	135	131-139	7.1%	-22.5	SB02139	SB02118	SB02145	69504054	65052311	70854896	
	3	70.5	74	73-75	45.2%	66.0	SB03075	SB03075	SB03085	48834006	48834006	52478805	
	3	7.0	130	125-137	3.0%	16.4	SB03131	SB03115	SB03149	64831343	60950542	68814938	
	4	8.7	76	62-84	3.8%	-9.5	SB04090	SB04082	SB04108	52736206	50515899	56368786	
	7	6.6	90	84-102	2.9%	-17.0	SB07090	SB07077	SB07108	59613944	57506979	62866313	
	8	5.6	24	18-27	2.4%	7.3	SB08023	SB08009	SB08028	3669490	13022298	4374962	
	10	6.3	67.5	62-75	2.7%	-13.8	SB10062	SB10055	SB10083	19451149	11760684	49923290	
	2x4*	4.3	-	-	1.8%	3.5	-	-	-	-	-	-	
	<b>TTPI2</b>	2	6.0	134	127-166	2.8%	-16.0	SB02132	SB02118	SB02171	67990256	65052311	76894769
		3	80.4	74	73-74	58.8%	72.2	SB03075	SB03075	SB03085	48834006	48834006	52478805
10		4.9	66	56-71	2.2%	-12.4	SB10062	SB10043	SB10079	19451149	8537969	48798729	
3		59.1	74	73-75	49.4%	-0.11	SB03075	SB03075	SB03085	48834006	48834006	52478805	
<b>P2O</b>	3	5.7	68	54-75	5.8%	-260.4	SB03069	SB03048	SB03085	42474125	11243730	52478805	
	7	5.6	88	82-96	5.6%	-244.8	SB07085	SB07077	SB07096	58732836	57506979	60740850	
<b>P2R</b>	9	3.8	21	9-36	3.7%	-217.7	SB09014	SB09009	SB09030	2423487	1508004	6175775	
	6	4.5	23	15-42	4.8%	-12.4	SB06027	SB06014	SB06057	32521645	3465631	46043053	
<b>BVP</b>	9	5.3	105	103-117	5.7%	14.4	SB09127	SB09122	SB09145	55954750	55056523	58738356	

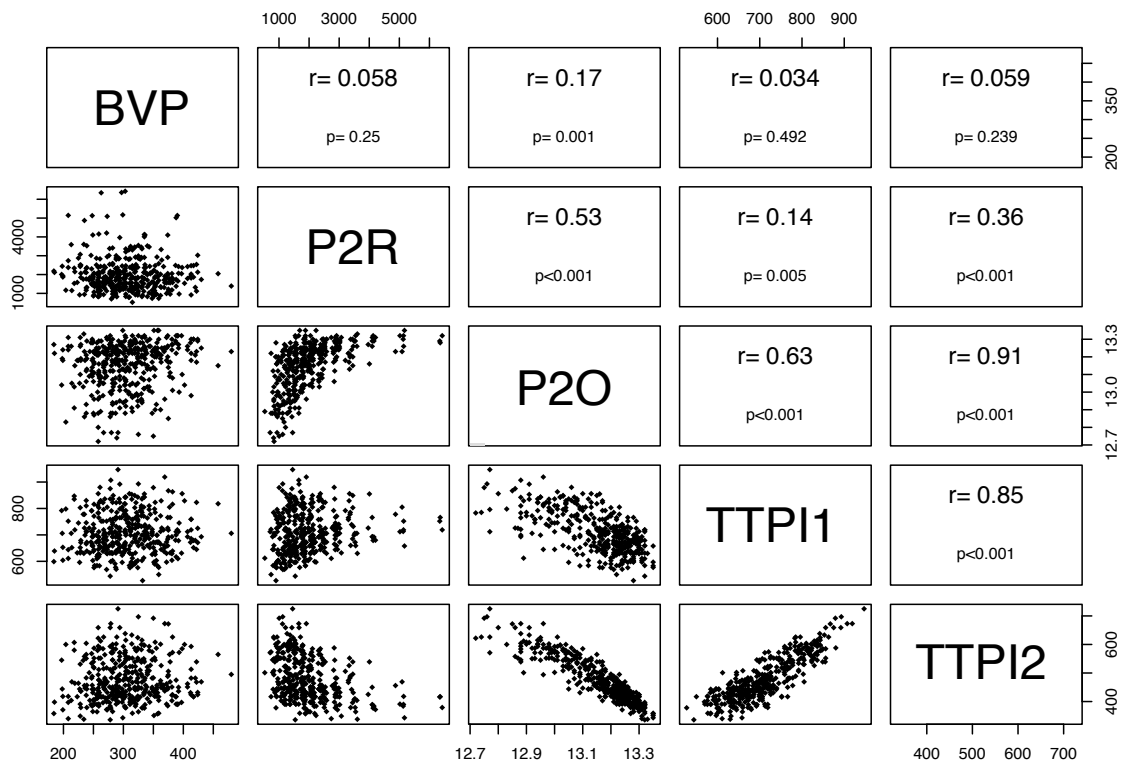
\* Interaction between the QTL on chr 2 and the QTL on chr 4 (p<0.01).



**Fig. 1.** Response to photoperiod for the CERES model. Thermal time for panicle initiation (TTPI) is plotted against photoperiod at panicle initiation. Data from the parent Lata3 were used to establish the figure.

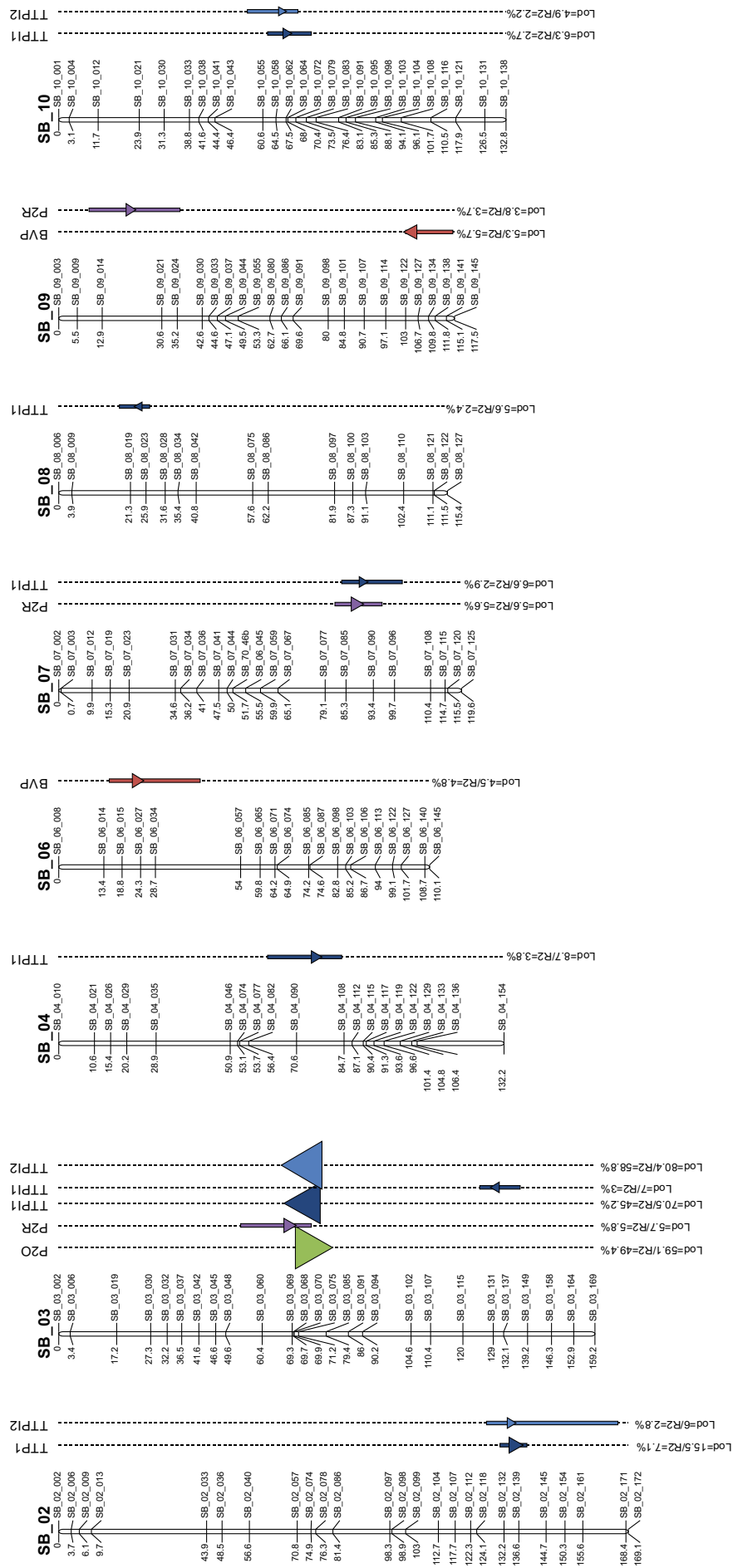


**Fig. 2.** Phenotypic distribution of Thermal Time from emergence to Panicle Initiation for the two sowing dates (TTPI1, TTPI2), and of model parameters (Basic Vegetative Phase (BVP), critical photoperiod (P2O) and photoperiod sensitivity (P2R)). TTPI3 distribution is not displayed since BVP and TTPI3 values can be considered identical. Arrows indicate average phenotypic values of population's parents, 'Tiandougou' and 'Lata3'.

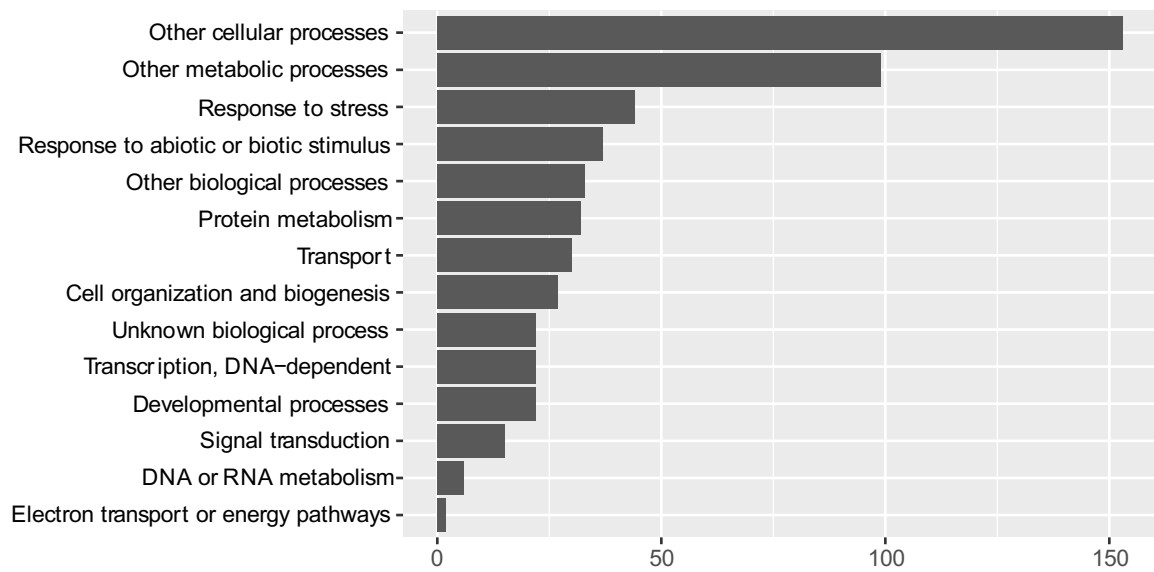


**Fig. 3.** Correlation matrix plot between variables. The figure displays in the upward half matrix correlation coefficients ( $r$ ) between variables with associated  $p$ -values ( $p$ ) and in the downward half matrix, scatter plot between variables.





**Fig. 4.** QTLs detected on the 'Tiandougou' x 'Lata3' genetic map of for Thermal Time to Panicle Initiation at 2 sowing dates (TTPI1 and TTPI2), Basic Vegetative Phase (BVP), critical photoperiod (P2O) and photoperiod sensitivity (P2R). Each QTL is represented by a rectangle spanning the QTL confidence interval and a triangle located at the LOD peak with an area proportional to the R<sup>2</sup>. Upward triangles represent a positive effect from 'Tiandougou' alleles and downward triangles a positive effect from 'Lata3' alleles.



**Fig. 5.** Gene ontology (GO) distribution by second-level GO terms for biological processes for predicted genes located in the chromosome 3 quantitative trait locus (QTL) confidence interval.

## **II-2. Principaux résultats et discussion de l'article « Integrating genetic analysis and crop modeling: a major QTL can finely adjust photoperiod-sensitive sorghum flowering »**

### ***II.2.1. Analyse de la population P118 : Intégration de l'analyse génétique et de la modélisation du développement***

***Un QTL majeur de photopériode critique P20 permet d'ajuster finement la floraison des sorghos photopériodiques.***

Le climat soudano-sahélien est caractérisé par une très grande variabilité interannuelle des précipitations. La plupart des variétés locales sont photopériodiques et fleurissent dans les 20 jours qui précèdent la date moyenne de fin des pluies de leur zone d'origine, indépendamment de la date de semis (Kouressy et al., 2008d). Ce caractère améliore simultanément (Kouressy et al., 2008b) :

- (i) la production de biomasse en optimisant la durée de la période végétative,
- (ii) la stabilité du rendement en atténuant les risques de sécheresse (en prévenant les floraisons trop tardives) et en évitant les attaques d'oiseaux et de cécidomyie (en groupant la floraison)
- (iii) la qualité du grain en limitant le développement des moisissures (en évitant des floraisons trop précoces).

Les programmes d'amélioration du sorgho ont éliminé le photopériodisme du matériel élite de façon à développer des variétés ayant une plus large adaptation géographique (Stephens et al., 1967b, Chantereau et al., 2001b, Swaminathan, 2006a). Cependant, en Afrique de l'Ouest, la sélection pour la précocité a montré ses limites car le problème est moins la courte durée de la saison des pluies que la variabilité, dans le temps et dans l'espace, des débuts et fins de saison. Améliorer la compréhension du déterminisme génétique de la floraison des céréales sèches est le premier pas pour développer des variétés résilientes à la variabilité et au changement climatique.

Une population a été développée à partir de deux variétés élites provenant des programmes d'améliorations au Mali : Tiandougou et Lata3. Ces deux variétés sont photopériodiques, dans des conditions normales de semis (juin), le cycle de Tiandougou est légèrement plus long (127 days) que celui de Lata3 (120 days). Au total les semences de 400 familles F<sub>3:5</sub> ont été produites et utilisées pour les expérimentations de phénotypages.

Les expérimentations ont été menées à 3 dates de semis correspondants à différentes conditions de photopériodes. La population a été étudiée en jours longs et intermédiaires (saison normale de culture, semis du 19 juin et du 15 juillet) et en conditions de jours courts (en contre-saison, semis du 3 octobre 2011).

### ***II-2.2. La modélisation permet de décomposer le caractère complexe de floraison.***

La modélisation, principalement basé sur le modèle CERES, a été utilisée pour prédire les effets de la photopériode sur la phase végétative. Ce modèle est basé sur un simple ajustement linéaire entre la photopériode et la longueur de la phase végétative (Bangbol, 2013, Chantereau et al., 2001b, Sanon et al., 2014). La plus courte durée de la période végétative correspond au BVP (Basic Vegetative Phase) ou précocité intrinsèque (Belliard, 1982, Major et al., 1990b). Pendant cette phase, l'induction florale ne peut pas se produire quelles que soient les conditions de photopériode. Le modèle CERES considère que lorsque la durée du jour est inférieure à une photopériode critique (P2O), la durée de la phase végétative est constante et égale au BVP. Lorsque la durée du jour devient supérieure à P2O, la durée de la phase végétative augmente comme une fonction linéaire de la photopériode. Le modèle CERES original (Ritchie and Alagarswamy, 1991) a été modifié, la notion d'accumulation de temps photothermique a été remplacée par un seuil photopériodique en dessous duquel l'initiation paniculaire du sorgho se produit (Folliard et al., 2004).

### ***II-2.3. Un QTL majeur à effet faible permet d'envisager un réglage fin de la sensibilité à la photopériode.***

Un bulk de 10 plantes F4 a été utilisée pour extraire l'ADN de chaque famille. Au total, 228 marqueurs SNP ont été utilisés. La méthode "Multiple Interval Mapping (MIM)" a été utilisée pour l'analyse génétique avec un pas de cartographie de 1 cM. Seize QTL répartis sur 8 chromosomes ont été détectés. Dix QTL ont été détectés pour la durée de la période végétative (en temps thermique) et les 2 dates de semis (TTPI1, TTPI2) et 6 QTL pour les paramètres du modèle (BVP, P2O, P2R).

Une région majeure a été identifiée à 74 cM sur le chromosome 3 où les QTL de TTPI1, TTPI2 et de P2O montrent des LOD élevés (respectivement des valeurs de LOD de 70.5, 80.4 and 59.1). Le pourcentage de variation phénotypique expliqué par ces QTL était respectivement de

45.2%, 58.8% et 49.4%. A cette position l'allèle de Tiandougou augmente TTPI1 et TTPI2 respectivement de 66°Cd et 72°Cd et diminue P2O de 0.11h.

L'analyse des gènes candidats présents dans la région génomique du QTL majeur nous a amené à proposer un gène impliqué dans l'horloge circadienne (ELF3) comme un des principaux régulateurs de la floraison des sorghos photopériodiques.

Dans cette étude, l'utilisation d'une population de cartographie issue de deux parents photopériodiques permet d'approfondir la compréhension des déterminismes génétiques de la floraison et de la sensibilité à la photopériode. Jusqu'à présent, "aucune étude génétique" n'avait porté sur la régulation fine de la sensibilité à la photopériode du sorgho. On s'était surtout intéressé aux gènes majeurs à effet très fort ayant pour fonction de supprimer (ou de réduire fortement) la sensibilité à la photopériode.

Bien que les valeurs de P2O entre les deux parents semblent presque similaires (différence de 15 min), cette légère différence induit une variation de la date de floraison d'environ une semaine. Une telle variation est extrêmement importante en termes d'adaptation aux conditions locales.

L'identification d'un QTL majeur de photopériode critique permet de développer de nouvelles options d'amélioration du sorgho basées sur une sélection de variétés améliorée photopériodiques adaptées aux différents environnements climatiques rencontrés en Afrique Soudano-Sahélienne. Jusqu'à présent, on pensait que la maturité et la sensibilité à la photopériode étaient étroitement liées et que les variétés précoces étaient nécessairement moins sensibles à la photopériode (Clerget et al., 2007). Nos résultats montrent que les changements dans la photopériode critique permettent de modifier la précocité sans changer la sensibilité à la photopériode. Cette constatation ouvre de nouvelles voies pour le développement de variétés précoces, bien adaptées au nord du Mali, tout en préservant la sensibilité à la photopériode et donc la stabilité de la production végétale.

## **Conclusion**

Cette approche combinant la modélisation des cultures et l'analyse génétique a permis de parvenir à une meilleure compréhension écophysiological du QTL majeur identifié. Nous avons démontré qu'il est possible de décomposer le temps de floraison en différents paramètres physiologiques, qui sont contrôlés par des régions génomiques spécifiques.

Le QTL majeur détecté sur le chromosome 3 contrôlant la photopériode critique repose sur un orthologue du gène *ELF3* pour lequel des preuves fonctionnelles ont été fournies chez *Arabidopsis* et le riz. Des analyses plus approfondies de ce locus à l'aide des études d'association (GWAS), d'études de populations multiparentales et des lignées isogéniques permettraient de préciser sa position cartographique pour étayer le rôle de ce gène candidat et en cataloguer les variations alléliques.

A moyen terme, les attentes liées à l'utilisation de la modélisation combinée et des approches de génétiques quantitatives devrait permettre de concevoir des idéotypes tenant compte des scénarios de changement climatique en valorisant la capacité de prédiction de la modélisation écophysiological et la précision et l'efficacité de la sélection assistée par les marqueurs moléculaires.

La compréhension des bases génétiques de la sensibilité à la photopériode ouvre ainsi des perspectives importantes de sélection du sorgho en fournissant des outils et des gènes innovants pour affiner l'adaptation des plantes à la variabilité climatique et au changement climatique.

### **III. Analyse du déterminisme génétique de la floraison et de la croissance en hauteur au sein d'une population biparentale de type rétrocroisement avancé (BC1F4)**

Cette analyse a pour objectif d'approfondir notre compréhension du déterminisme génétique de la date de floraison et de la hauteur des plantes au sein d'une population de rétrocroisement avancé (advanced backcross) caractérisée en condition de jours longs (semis de Juin et Juillet) et en condition de jours courts (semis de Septembre).

L'avantage qu'apporte ce type de population « AB-QTL » par rapport à la population P118 (population F4) étudiée dans la section précédente, résulte du fait qu'elle permet de combiner l'analyse QTL avec le développement des variétés en identifiant et en introgressant simultanément les allèles favorables au QTL (pour des caractères agronomiques important) d'un fond génétique non adapté (sauvage ou locale) vers un génotype cultivé (Tanksley and Nelson, 1996). Avec la stratégie AB-QTL, les travaux de phénotypage s'effectuent sur des populations de génération avancée en backcross (BC2 ou BC3) lesquelles sont proches du parent récurrent cultivé (car ayant reçue un petit fragment du génome provenant du parent donneur sauvage). Un autre avantage de la stratégie AB-QTL c'est qu'elle ne nécessite que

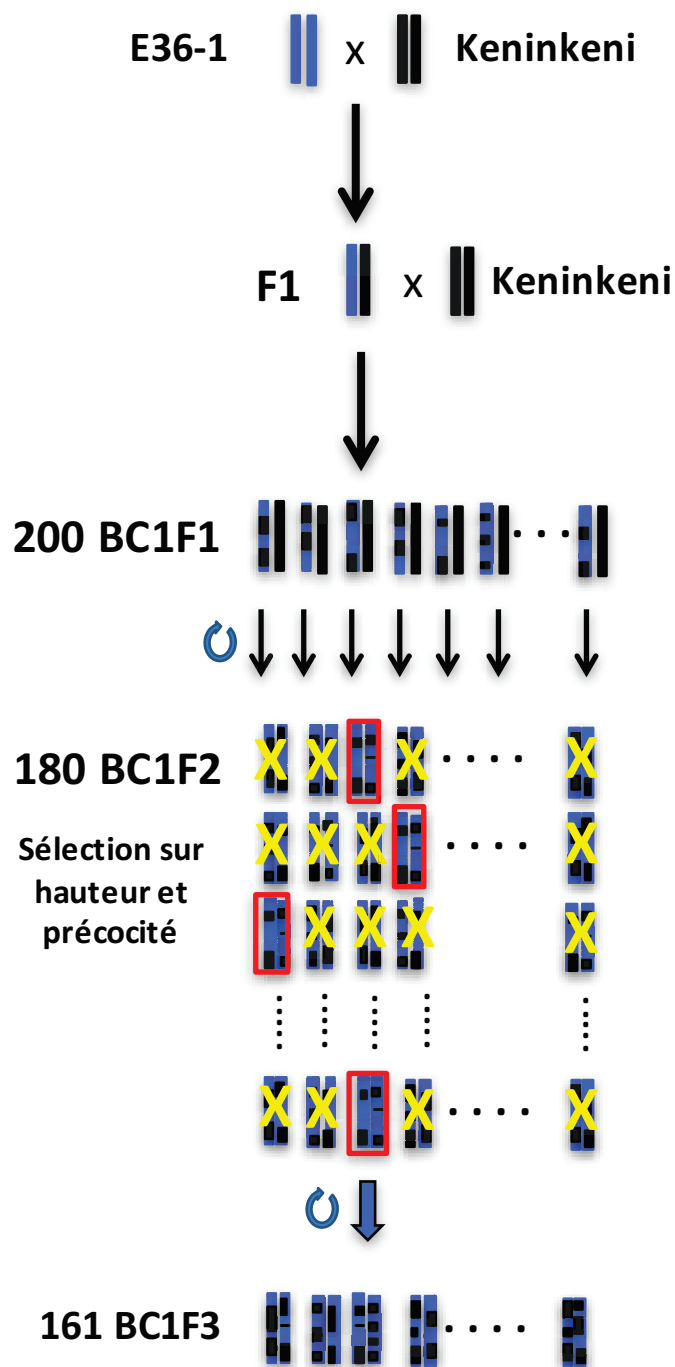


Figure II-5a : Schéma de développement de la population BC06 issue du parent récurrent Keninkeni et du parent donneur E36-1 pour l'obtention des familles BC1F3.

quelques générations de rétrocroisement pour générer des lignées quasi-isogéniques (fond génétique homogène) qui peuvent être évaluées et où utilisées dans un schéma de sélection variétale.

Par ailleurs, cette stratégie développée pour la première fois chez la tomate (Tanksley et al., 1996) fut par la suite appliquée avec succès chez beaucoup d'autres espèces tel que le riz (Moncada et al., 2001, Septiningsih et al., 2003, Thomson et al., 2003, Xiao et al., 1998), le maïs (Ho et al., 2002), le blé (Huang et al., 2003), l'orge (Pillen et al., 2004, Von Korff et al., 2005, Von Korff et al., 2006), le haricot (Blair et al., 2006) afin d'identifier les QTL impliqués dans la variation des caractères d'intérêts agronomiques.

### **III.1. BC06 : Matériels et méthodes**

#### ***III.1.1- Développement de la population BC06 : E36-1 (parent donneur) X Keninkeni (parent récurrent)***

La population « BC06 » a été développée à partir du croisement entre un parent donneur (E36-1) et le parent récurrent (Kéninkéni). La variété E36-1, encore appelée « Zera-Zera » est de type botanique Caudatum avec de gros grains et est originaire d'Éthiopie. Cette variété présente des tiges juteuses et sucrées avec une concentration en sucres solubles de la matière sèche atteignant jusqu'à 19 degrés Brix ou « °B » (données non publiées). Le degré Brix correspond à la fraction de saccharose dans un liquide ou le pourcentage de matière sèche soluble (1°B = 1 g de matière sèche soluble pour 100g de solution).

La variété Keninkeni a été développée par le programme Sorgho de l'IER (Mali) dans le cadre d'un projet de valorisation de la diversité génétique des sorghos du Mali (FFEM : Fond Français pour l'Environnement Mondial). Keninkeni est une variété élite adaptée à la zone Soudano-sahélienne dont la pluviométrie varie entre 800 – 1000 mm par an. Cette variété améliorée a des panicules d'aspect Guinea avec une bonne qualité des grains. Elle est issue d'un programme de sélection impliquant 19 parents de types botaniques différents (guinéa, caudatum, durra et kafir).

Ces deux génotypes parentaux sont de tailles moyennes (hauteur pouvant atteindre 1.45m) et moyennement sensibles à la photopériode (leurs coefficients de sensibilité à la photopériode « Kp » varie entre 0.4 - 0.5). Ils ont un cycle « semis- maturité physiologique du grain » de 110 jours en moyenne dans le contexte de semis précoces réalisés dans les conditions Maliennes.

A l'issue des premiers croisements (Figure II-5a) réalisés durant la contre saison de culture 2010 (janvier 2010), plus de 200 plantes F1 ont été obtenues. Pendant la saison normale de



culture (Juin 2010), chaque plante F1 a été rétro-croisée (Backcross) avec le parent récurrent Keninkeni afin d'obtenir des BC1F1. Environ 200 plantes BC1F1 ont été retenues afin de conserver une certaine diversité génétique du matériel. Chaque plante BC1F1 a ensuite été autofécondée (Juin 2011) pour obtenir des populations BC1F2. Ces populations BC1F2 ont été soumises à une légère pression de sélection sur la taille et le cycle (les plantes trop grandes/courtes, et très précoces ont été supprimées). Une seule plante BC1F2 a été conservée par famille (descendance issue du croisement) et autofécondée pour obtenir des BC1F3 pendant la contre saison 2011. Les semences BC1F3 ont été alors mélangées (bulk) pour servir à la réalisation des premières analyses de phénotypage (juin 2012) sur les caractères agromorphologiques (phénologie, cycle, rendement...etc.).

### ***III.1.2. Site expérimental et conditions environnementales***

Le phénotypage de toutes les populations étudiées dans le cadre de cette thèse a été réalisé à la station de recherche agronomique de Sotuba (Bamako, 12°39'N et 7°56'W altitude : 381m). Le climat est de type soudano sahélien avec deux saisons : une saison sèche qui s'étend de novembre à mai et une saison pluvieuse de type monomodale (une seule saison pluvieuse par an, ou hivernage) qui se déroule de mai à octobre. La texture des sols rencontrés sur le site de Sotuba est de types limono argileux et sablo argileux.

### ***III.1.3. Dispositif expérimental***

La caractérisation de la population BC06 a été effectuée sur 4 périodes différentes : les hivernages 2014 et 2015 et les contre saisons de culture 2013 et 2015.

#### **III.1.3.1- Hivernage 2014**

Durant la saison hivernale 2014, les familles BC1F4 (de chaque semis) de la population BC06 ont été randomisées dans un dispositif en bloc augmenté appelé « Augmented design ». Ce type de dispositif a été introduit pour la première fois par Federer (1956) afin de répondre à un besoin de mise en place de nouveaux dispositifs expérimentaux pertinent pour l'évaluation de nouvelles variétés de Canne à sucre. Il est principalement utilisé par les programmes de sélection pour l'évaluation de nombres importants de nouveaux génotypes lorsque les quantités de semences disponibles sont limitées pour la mise en place d'essais avec des répétitions des génotypes mais aussi pour des évaluations phénotypiques multi-environnementales (Crossa, 2012, Federer, 1961). Dans un dispositif en bloc augmenté, les variétés analysées ne sont pas répétées alors que des variétés « témoin » sont répétées de nombreuses fois sur la totalité de l'essai pour contrôler la variabilité spatiale à l'intérieur de l'essai.

Dans notre expérimentation, les 161 familles et témoins (E36-1, Keninkeni et Grinkan) ont été affectés de façon aléatoire dans 12 blocs. Les 11 premiers blocs étaient constitués de 17 traitements comprenant 14 familles BC1F4 et 3 témoins ; tandis que le dernier bloc ne comportait que 10 traitements correspondant à 7 familles et 3 témoins. Au total, pour l'ensemble du dispositif, 197 traitements ont été analysés pour chaque date de semis.

Chaque traitement était constitué d'une ligne de 4m. L'écartement entre les lignes était de 75 cm et l'écartement entre les poquets au sein de chaque ligne était de 40 cm. Les allées entre les bandes (sous-parcelle) faisaient une largeur de 1.50 m. Ces dimensions de parcelles/ traitement sont valables pour les 3 autres essais (hivernage 2015, contre saison 2013 et 2015).

### III.1.3.2- Hivernage 2015

Durant la saison hivernale de culture 2015, une expérience factorielle comportant deux facteurs à étudier a été mise en place « dispositif en split-plot ». Le facteur principal est la date de semis avec deux niveaux de variation (semis de Juin et de Juillet) et le facteur secondaire correspond aux 161 familles et aux 3 variétés témoins (164 niveaux de variations). Le dispositif comportait deux répétitions. Le nombre total de traitements correspondant à l'ensemble du dispositif est de 680 entrées (161 familles x 2 blocs x 2 semis et 3 témoins x 3 répétitions x 2 blocs x 2 semis).

### III.1.3.3- Contre saisons 2013 et 2015

Durant les contre-saisons 2013 (semis du 20/9/2013) et 2015 (semis du 21/9/2015), les mêmes 161 familles et témoins ont été randomisés dans des dispositifs en bloc augmenté sans répétition. Une irrigation supplémentaire a été apportée sur ces essais. Chaque essai comportait 11 blocs avec au total 18 traitements (15 familles + 3 témoins) dans chacun des blocs à l'exception des derniers blocs qui en présentaient 17 (14 familles et 3 témoins) et 14 traitements (11 familles + 3 témoins) respectivement pour les essais 2013 et 2015. Le nombre total de traitements dans les deux essais était de 197 et 193 respectivement pour les essais 2013 et 2015.

### ***III .1.4. Caractérisation phénotypique : Phénologie et croissance***

Les 161 familles BC1F4 de la population BC06 ont été caractérisées sur l'ensemble des essais pour les variables suivantes :

- Date d'émergence de la feuille drapeau (FD), observée sur les 5 premiers plants de chaque parcelle élémentaire,
- Nombre total de feuilles vertes à maturité sur la tige principale (NBFV),

- Longueur de la tige principale (LTIG) mesurée de la base du collet au sommet du dernier entre-nœud situé sous le pédoncule (en cm),
- Nombre total d'entre-nœuds allongés de la tige principale (NEN),
- Longueur du pédoncule de la tige principale (LPED) (en cm),
- Longueur de la panicule (LPAN) de la tige principale (en cm).

Ces mesures ont été réalisées sur 5 plantes de chaque parcelle élémentaire au stade de maturation du grain (à l'exception de la variable de date d'apparition de la feuille drapeau).

- Le nombre total de feuilles développées (NBTF) a été mesuré uniquement sur les essais de 2015.

A partir de ces variables mesurées, les variables suivantes ont été calculées :

- La durée entre le semis et l'émergence de la feuille drapeau (SFD) exprimée en nombre de jours a été calculée comme suit :  $SFD = FD - \text{Date de semis}$ .
- Le coefficient de photopériodisme ( $K_p$ ) de chaque famille a été calculé selon la formule présentée par (TRAORE et al., 2001). Ce coefficient de sensibilité à la photopériode est défini comme le rapport du raccourcissement de la durée « semis feuille drapeau » entre les deux dates de semis ( $SFD1 - SFD2$ ) et la différence de nombre de jours entre le premier et le second semis ( $D1$  et  $D2$ ).

$$K_p = \frac{SFD1 - SFD2}{|D1 - D2|}$$

avec  $SFD1$  et  $SFD2$  les durées entre le semis et l'émergence de la feuille drapeau pour les deux dates de semis  $D1$  et  $D2$ . Dans le cadre de cette étude, la durée du cycle semis-feuille drapeau est exprimée en nombre de jours.

Ce coefficient  $K_p$  varie entre 0 et 1. Pour les géotypes très sensibles à la photopériode,  $K_p$  est proche de 1 car le raccourcissement de la période végétative est égale au décalage de semis (Kouressy et al., 2008c). Pour les géotypes non sensibles à la photopériode,  $K_p$  est proche de 0 ce qui indique que, la durée du cycle n'est pas influencée par la date de semis.

- La longueur moyenne des entre-nœuds (LMEN), calculée à partir de la longueur de la tige principale et du nombre total d'entre-nœuds allongés :  $LMEN = LTIG / NEN$ ,

### ***III.1.5 Analyse des données phénotypiques***

#### **III.1.5.1- Vérification des données**

Pour l'ensemble des jeux de données, une vérification des données a été réalisée sur chacune des variables à l'aide du logiciel R version 3.3.2 (R development Core Team, 2016) . La distribution de chaque caractère a été analysée ainsi que les corrélations inter-caractères, cela a permis de corriger certaines données aberrantes. En outre, les statistiques descriptives (moyenne, distribution, écart-type et coefficient de variation phénotypique) ont également été calculées.

#### **III.1.5.2- Ajustements des données en prenant en compte la structure des dispositifs expérimentaux**

##### *Calcul des données ajustées intra-essai*

Les données ajustées aux variations environnementales de chaque essai et de chaque date de semis ont été utilisées directement pour les analyses phénotypiques (corrélations...) et les analyses de détection de QTL.

- Pour les essais où un dispositif en blocs augmentés a été utilisé (contre saisons 2013 et 2015 et saison normale de culture 2014), les moyennes ajustées ont été calculées grâce à la fonction « DAU.test » du package R « *agricolae* ».
- Pour les données phénotypiques obtenues en saison normale de culture 2015 sur le dispositif en split-plot, les moyennes phénotypiques de chaque famille pour chaque date de semis ont été obtenues par « extraction » des valeurs génétiques (BLUP : Best Linear Unbiased Prediction) à partir du modèle linéaire mixte suivant à l'aide du package R « lme4 » :

$$Y_{ij} = \mu + G_i + B_j + e_{ij} \quad (1)$$

avec  $Y_{ij}$  la valeur phénotypique mesurée sur la famille  $i$  dans le bloc  $j$  ;  $\mu$  la moyenne générale de la population ;  $G_i$  l'effet aléatoire de la famille  $i$  ;  $B_j$  l'effet du bloc  $j$  déclaré en effet fixe et  $e_{ij}$  la résiduelle du modèle.

##### *Calcul des données ajustées inter-essais*

En complément des données ajustées de chaque essai et de chaque date de semis, la combinaison des données phénotypiques issues des deux années (saisons normales de culture

2014 et 2015) de chaque date de semis a permis d'obtenir des estimations d'effets génétiques (i.e. BLUP) de chacune des familles en utilisant le modèle suivant :

$$Y_{ij} = \mu + A_i + G_j + e_{ij} \quad (2)$$

avec  $Y_{ij}$  la valeur phénotypique mesurée sur la famille  $i$  pour l'année  $j$  ;  $A_j$  l'effet de l'année  $j$  déclaré en fixe,  $G_i$  : l'effet aléatoire de la famille et  $e_{ij}$  la résiduelle du modèle.

A partir de ce modèle, nous avons extrait les valeurs génotypiques de chaque famille (i.e. les BLUP) pour chacune des variables mesurées. Ces données ont ensuite été utilisées pour l'analyse des données phénotypiques et les analyses de détection de QTL.

### III.1.5.3- Corrélations phénotypiques

Les corrélations entre les différents caractères mesurés sur les 161 familles BC1F4 communes aux deux années d'expérimentation, ont été calculées sur la base des valeurs génétiques (i.e. BLUP) obtenues pour les semis de Juin des deux années de culture normale (2014 et 2015). Toujours sur la base des valeurs Blup, les mêmes corrélations ont été calculées entre les 2 essais de contre saisons (2013 et 2015) mais également entre les essais de saison normale et de contre saison.

### III.1.5.4- Estimation des variabilités génétiques

Les coefficients de variation phénotypique (CVP %) et génotypique (CVG %), ainsi que les héritabilités des différents caractères mesurés pour les deux dates de semis (Juin et Juillet) de la saison normale de culture 2015 ont été calculés à partir du modèle d'analyse de variance présenté ci-dessous :

$$Y_{ij} = \mu + G_i + B_j + e_{ij} \quad (3)$$

Où  $Y_{ij}$  est la valeur phénotypique mesurée sur la famille  $i$  dans le bloc  $j$ ;  $G_i$  l'effet du génotype (famille)  $i$ , et  $e_{ij}$  la résiduelle du modèle.

A partir des résultats de ce modèle d'analyse, les CVP, CVG et les héritabilités ont été calculés de la manière suivante :

$$CVP = \sqrt{\sigma_p^2 / \bar{x}} * 100 \quad (4)$$

$$CVG = \sqrt{\sigma_G^2 / \bar{x}} * 100 \quad (5)$$

Avec  $\sigma_p^2$  la variance phénotypique :  $\sigma_p^2 = \sigma_G^2 + CMe$  ;

Où  $\sigma_G^2$  représente la variance génotypique :  $\sigma_G^2 = (CMg - CMe) / k$

Avec CMg et CMe ; le carré moyen génotypique et résiduel ; et k le nombre de répétition de l'essai.

Connaissant la variance phénotypique et génotypique, l'héritabilité au sens large ( $h_{sl}^2$ ) a été calculée selon la formule suivante :

$$H_{sl}^2 = \sigma_P^2 / \sigma_G^2 \quad (6)$$

Où  $\sigma_P^2$  et  $\sigma_G^2$  représentent des variances phénotypiques et génotypiques respectivement.

### III.1.5.5- Analyse des effets de date de semis et des interactions « Génotype x Date de semis »

Enfin, afin d'évaluer les effets des dates de semis et des interactions entre dates de semis et famille BC1F4 le modèle suivant a été utilisé :

$$Y_{ijk} = \mu + G_i + S_j + (G * S)_{ij} + e_{ijk} \quad (7)$$

Où  $Y_{ijk}$  est la valeur phénotypique mesurée sur la famille i dans le semis j;  $G_i$  l'effet du génotype (famille) i,  $S_j$  est l'effet de la date de semis j et  $e_{ijk}$  la résiduelle du modèle. Seules les données de 2015 (2 dates de semis) ont été analysées avec ce modèle d'analyse.

### **III.1.6 – Génotypage de la population BC06**

En amont de la réalisation du premier backcross et des différents cycles d'autofécondation impliqués dans le développement de la population BC06, des tests de conformité des hybrides et des populations ont été effectués avec quelques marqueurs.

Pour le génotypage de l'ensemble des 161 familles de la population BC06, environ 30 graines de chaque famille BC1F3 ont été semées en serre pour effectuer les prélèvements d'ADN.

Pour chaque famille, un mélange de feuilles provenant de 20 plantes a été utilisé.

Le génotypage a été effectué selon la technique de génotypage par séquençage (GBS : Genotyping By Sequencing) avec l'enzyme de restriction ApeK1 en utilisant le protocole développé par Elshire (Elshire et al., 2011).

Le génotypage par séquençage a été effectué en quatre étapes principales : 1) la digestion des ADN génomiques (individuellement) par l'enzyme de restriction ApeKI afin de réduire la complexité du génome, 2) la ligation de l'ADN avec deux adaptateurs (un adaptateur avec code-barres unique à chaque échantillon et un second adaptateur sans code-barres) ; 3) le multiplexage des individus: les échantillons d'ADN sont mélangés dans un seul tube puis

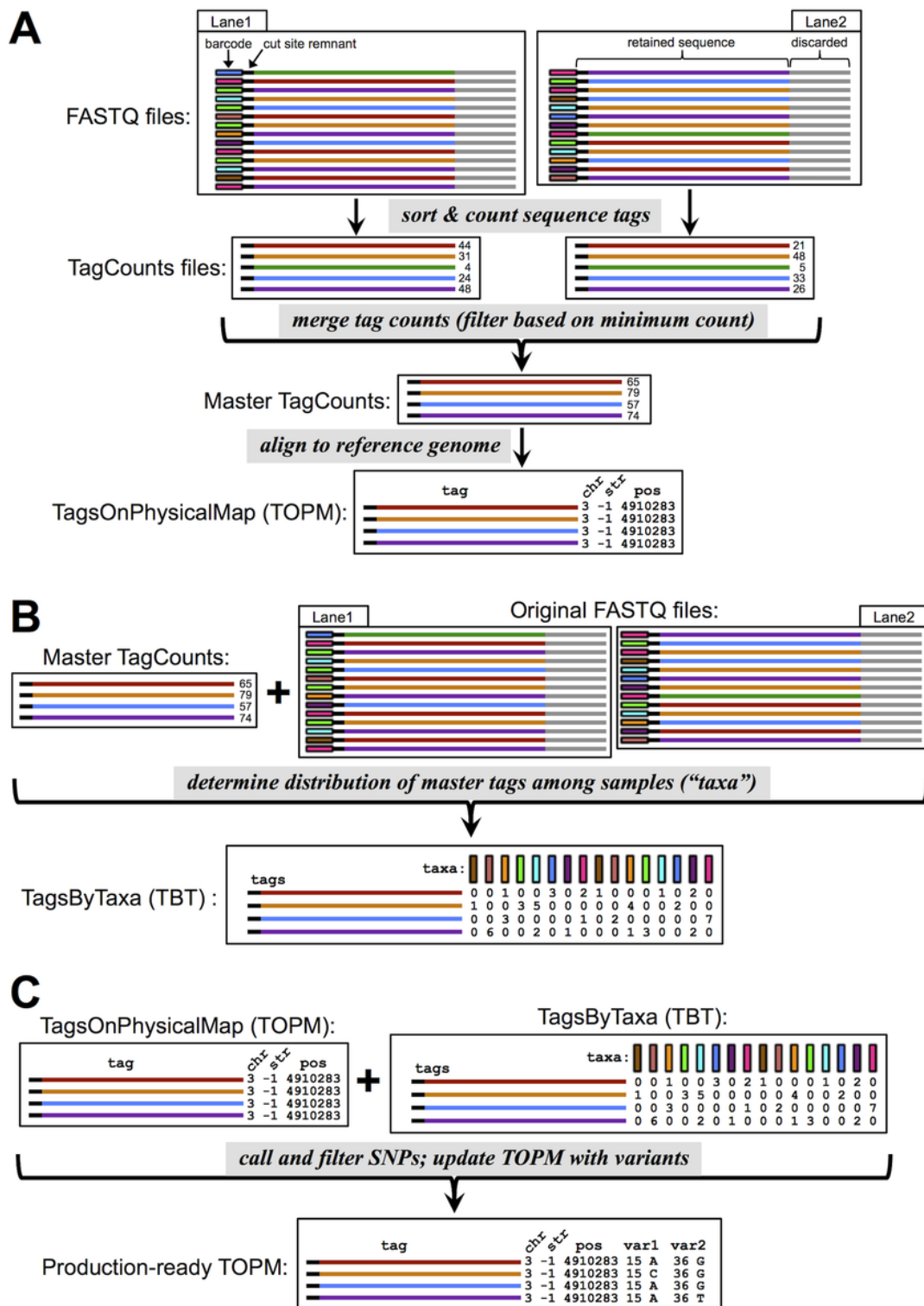


Figure II-5b : Représentation schématique du pipeline d'analyse TASSEL- GBS d'après Glaubitz et al. (2014). A : regroupement des séquences en tag, B : distribution des Masters Tags parmi les échantillons et C : identification de polymorphisme d'un nucléotide (SNP) et filtration des sites polymorphes.

subissent une amplification par PCR, la préparation des banques et le séquençage des multiplex est réalisée via la plateforme Illumina (dans le cas de notre étude des multiplex de 384 échantillons d'ADN différents ont été effectués) ; enfin 4) l'utilisation des codes-barres pour assigner les séquences aux échantillons et générer des fichiers contenant les séquences obtenues pour chaque échantillon.

Au total, pour chacune des 161 familles séquencées, 480 000 séquences d'ADN (reads) ont été produites.

Bien que la technologie GBS soit rapide et économique, l'un des inconvénients de l'utilisation de cette technique réside dans le nombre important de données manquantes obtenues ; qui peuvent être due (dans notre cas d'étude) soit à un problème de qualité d'ADN, ou des problèmes de dosages des codes-barres.

En outre, les deux lignées parentales de la population BC06 ont été génotypées en plusieurs répétitions : 8 répétitions pour le parent récurrent « Keninkeni » et 9 répétitions pour le parent donneur E36-1. Ce qui a permis d'obtenir des données plus solides pour ces dernières ; c'est à dire une bonne couverture de séquençage : 900 million de séquences obtenues pour Keninkeni et 60 millions pour E36-1.

Une fois le contrôle de qualité des données de séquences terminé, un traitement bio-informatique des séquences d'ADN a été effectué (Figure II-5b) en utilisant le pipeline d'analyse TASSEL-GBS (version 4.0) développé par Glaubitz et al. (2014). Le traitement des données de séquences a été réalisé de la manière suivante :

- 1) le regroupement des séquences en Tag (séquence unique) : les séquences sont d'abord triées en fonction des codes-barres ; avec élimination des séquences de mauvaise qualité ou de taille très courte. Une liste de séquence appelée « Master TagCount » est générée et par la suite alignée sur le génome de référence ;
- 2) la distribution des Masters Tags parmi les échantillons. La liste des « Masters TagCount » est ajoutée au fichier originale FASTQ pour indiquer le nombre de fois où chaque tag de la liste Master Tag est observé ; ces résultats sont stockés dans un fichier TBT (TagsByTaxa) ;
- 3) l'identification de polymorphismes d'un nucléotide (SNP) et filtration des sites polymorphes en fonction de la fréquence des allèles mineurs encore appelé « MAF : Minimum Allele Frequency » afin de réduire les erreurs de séquençage.



A l'issue de ces différentes étapes de traitement bio-informatique des données, une matrice de données composée de 52 409 marqueurs SNP a été générée.

Un premier filtre basé sur le polymorphisme des deux parents a été effectué sur cette matrice grâce à des procédures implémentées au niveau du logiciel R, suite à laquelle 31 892 SNP ont été sélectionnés. Puis un second filtre a été réalisé en deux étapes : 1) le recodage de la matrice en allèles parentaux (AA= homozygotes hérités du parent récurrent Keninkeni, BB= homozygotes du parent donneur E36-1 et H = hétérozygotes) ;

2) la sélection des SNP sur la base de la matrice recodée en fonction de leur taux de données manquantes, la fréquence de leur allèle mineur, leur taux d'hétérozygotie.

Enfin, un dernier filtre a été mis en place sur les individus en se basant sur leur similitude avec le parent récurrent (les individus trop proches du parent récurrent ont été supprimés) et/ ou avec le parent donneur.

Suite à ces différents filtres, une matrice de génotypes contenant 14 076 marqueurs SNP et 148 individus a été produite.

Une imputation des données manquantes a ensuite été réalisée à partir de cette dernière matrice de données SNP à l'aide de l'algorithme d'imputation « FSFHap » (Full-Sib Family Haplotype) implémenté dans Tassel version 5.0.

La matrice obtenue suite à cette imputation contenait des données manquantes, en effet FSFHap assigne aux points de cassure de recombinaisons entre haplotypes des données manquantes. Grâce à la haute densité de génotypage, les génotypes à ces points de cassures ont été inférés via R/qtl.

Au terme de la phase d'imputation, la matrice de données finales obtenue était composée de 144 individus avec 12 233 marqueurs SNP.

La carte génétique de cette population BC06 a été obtenue en utilisant les informations de la carte de la population P118 (cf. section II.) qui dispose de 228 marqueurs Kaspar sur 400 individus. Cette carte P118 est utilisée comme carte de référence afin de convertir les distances physiques en distance génétique grâce à une fonction R nommée « ZiplinR » en faisant des projections.

### III.1.7. Stratégie de détection de QTL au sein de la population BC06

La recherche de QTL a été effectuée sur les valeurs phénotypiques ajustées aux variations micro-environnementales obtenues sur chacune des 144 familles BC1F4 pour chaque essai (2013, 2014 et 2015). En outre une détection QTL sur les données blup basées sur les essais des saisons normales de cultures 2014 et 2015 pour chaque date de semis a aussi été effectuée. La détection a aussi été effectuée sur les données blup obtenues sur la base des deux essais de contre saison (2013 et 2015) (Tableau II-3).

**Tableau II- 3-** : Types de données phénotypiques et dispositifs expérimentaux utilisés pour l'analyse des différents essais.

<b>Jeu de données phénotypique</b>	<b>Essais</b>	<b>Ajustement</b>
2013_contre saison	2013	Bloc augmenté
2014_semis1	2014_Semis1	Bloc augmenté
2014_semis2	2014 Semis 2	Bloc augmenté
2015_semis1	2015_Semis1	Split splot
2015_semis2	2015 Semis 2	Split Splot
2014-2015_semis1	2014_Semis1 + 2015_Semis1	Blup

La détection de QTL par analyse de liaison est basée sur l'identification de locus dont la variation allélique est associée à la variation phénotypique du caractère analysé. Parmi les méthodes de détection de QTL qui existent en analyse de liaison (Single Marker Analysis « SMA » (Tanksley et al., 1982), Single Interval Mapping « SIM »(Lander and Botstein, 1989), Composite Interval Mapping « CIM » (Zeng, 1993) (Zeng, 1993), multiple QTL mapping « MQM » (Jansen, 1994) et Multiple Interval Mapping « MIM » (Kao et al., 1999)) ; nous avons choisi d'utiliser la méthode de cartographie de QTL multiples « MQM » de Jansen et al., (1994) pour la recherche de QTL sur la population BC06.

L'approche MQM combine les atouts du modèle linéaire généralisé avec ceux de la cartographie de QTL en SIM. C'est donc une méthode possédant une puissance statistique plus élevée que beaucoup d'autres méthodes de cartographie d'intervalle (Arends et al., 2010). Elle permet de s'affranchir des effets des principaux QTL détectés (en utilisant les informations des marqueurs les plus proches comme cofacteurs) ainsi que les effets d'interactions. Néanmoins, dans notre analyse, ces effets d'interactions entre QTL n'ont pas été pris en compte dans le modèle à cause de l'instabilité des résultats issus du modèle intégrant les interactions.

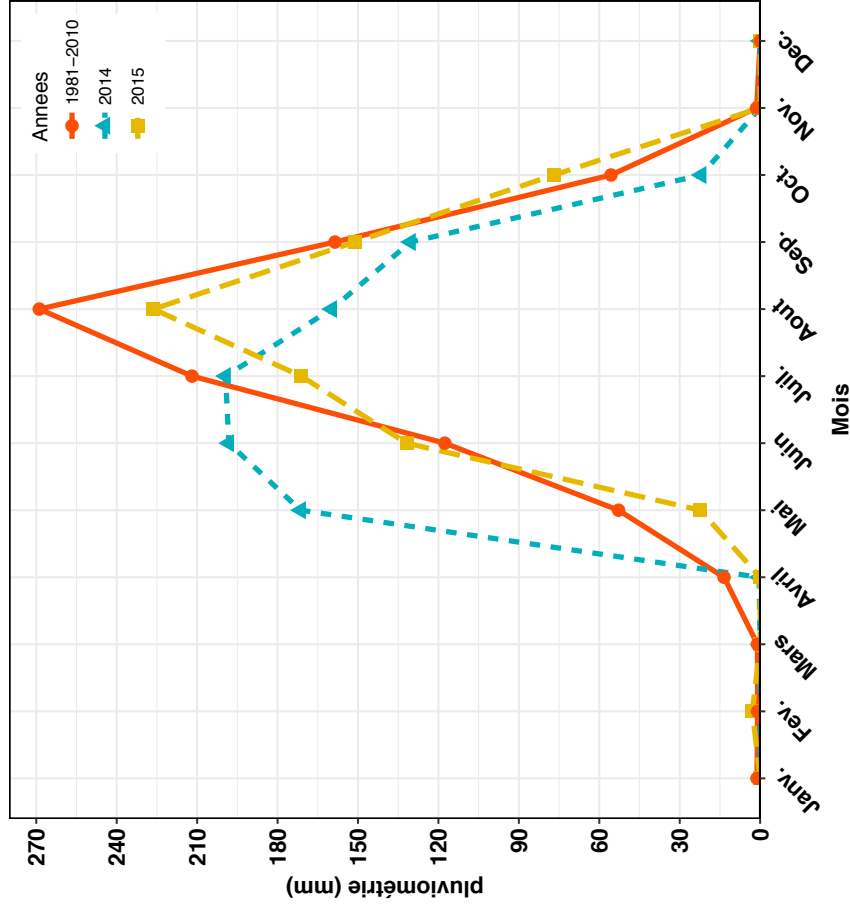
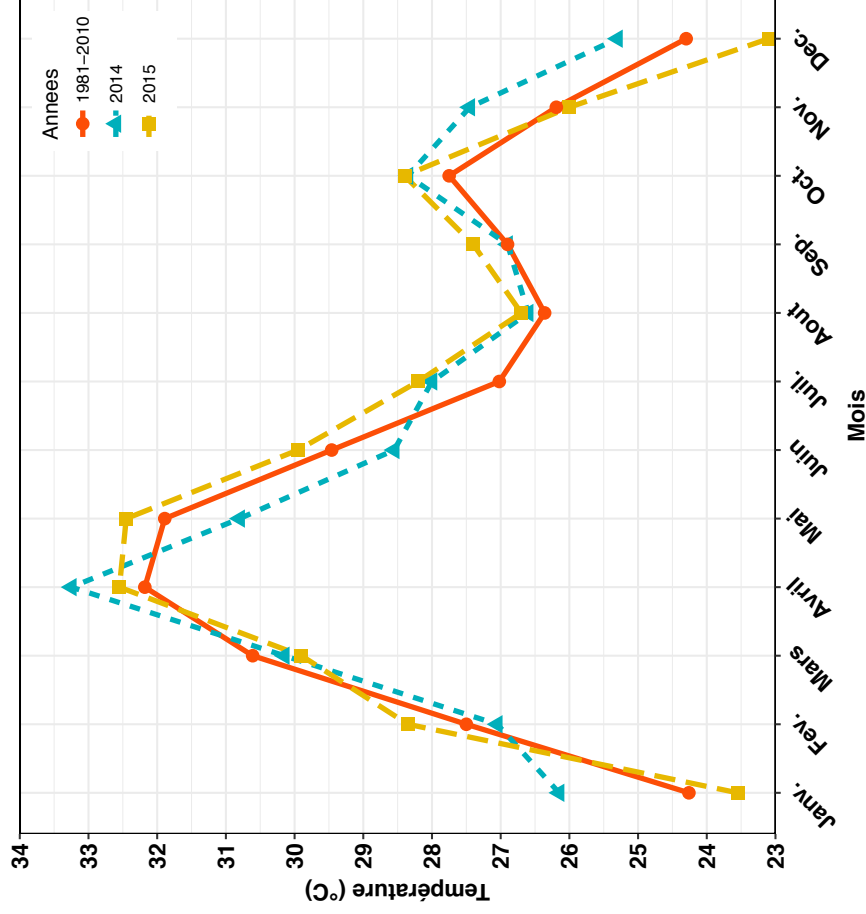
Un autre avantage de la méthode MQM, par rapport à la méthode CIM et à beaucoup d'autres méthodes, est la réduction des risques d'erreurs de type I (alpha) et II (Beta). Ces risques d'erreurs de type I et II, correspondent aux risques de rejeter l'hypothèse nulle alors qu'elle est vraie (erreurs de type I) ou inversement (erreurs de type II).

L'analyse a été effectuée à l'aide du package R/qtl développé par Broman et al. (2003). Ce package a été conçu pour les études de cartographie QTL dans les croisements expérimentaux tel que les populations de backcross (BC), les populations de type F2, et de lignées recombinantes (RIL). R/qtl est donc bien adaptée à notre population d'étude de type backcross car nous avons la possibilité d'indiquer le type de population dont on dispose grâce aux options « BC. gen » et « F. gen » (utilisé uniquement pour les croisement de types « bcsft ») de la fonction read.cross.

Grâce aux récents travaux de Arends et al. (2010), sur la cartographie de QTL multiples, nous disposons maintenant de la toute première procédure automatique et libre d'accès permettant des recherches de QTL au travers de la stratégie MQM intégrée dans R/qtl nommé « R/qtl-MQM » avec beaucoup d'autres fonctionnalités telles que la parallélisation des analyses, les tests de permutations pour la définition des seuils de détection, et la visualisation des effets d'interaction.

Pour notre étude, nous avons adopté une approche de sélection de modèle par étape (Forward/backward) à l'aide de la fonction « stepwiseqtl » implémentée dans le package R/qtl (Broman et al., 2003, Arends et al., 2010). La fonction Stepwiseqtl vise à d'identifier le meilleur modèle à QTL multiples en optimisant le LOD score du modèle global (Manichaikul et al., 2009).

Avant d'effectuer cette analyse, un scan complet du génome à deux dimensions (fonction « scantwo » ou test de modèles à deux QTL) a d'abord été réalisé en utilisant la méthode de régression de Haley et Knott (1992). Cette analyse permet de tester la présence de deux QTL situés sur le même groupe de liaison. Un LOD seuil pour déclarer que les QTL détectés sont significatifs a ensuite été défini pour chaque caractère selon la méthode décrite par Churchill et al., (1994) en appliquant 1000 permutations avec un risque global d'erreur de première espèce (alpha) de 5% à l'échelle du génome complet. Ce risque correspond à la probabilité de détecter des QTL alors qu'il n'y en a pas, en d'autres termes, le risque qu'un QTL détecté soit un faux positif.



**Figure II-6 :** Evolution des précipitations (à gauche) et des températures (à droite) mensuelles enregistrées sur le site de Sotuba en 2014 et en 2015 et leurs moyennes sur la période 1981-2010.

Enfin le modèle à QTL multiples le plus pertinent (i.e. celui présentant la valeur de LOD pénalisé la plus élevée) a été sélectionné.

A l'issue de l'analyse MQM, les différents paramètres du modèle obtenus sont :

- les positions des QTLs détectés
- leurs effets d'additivité et de dominance
- les pourcentages de la variation phénotypique ( $r^2$ ) qu'ils expliquent

Parmi les nombreuses méthodes proposées pour déterminer l'intervalle de confiance des QTL détectés, nous avons choisi celle décrite par Broman and Sen (2009b) et par Manichaikul et al. (2009) (i.e. LOD-1.5).

Une représentation graphique de tous les QTL obtenus, a été construite en utilisant l'outil Spidermap version 1.7 (Rami, non publié).

## **III.2. Résultats**

### ***III.2.1. Variabilité des caractères et héritabilités***

Au total, 9 caractères ont été analysés dans 6 environnements différents correspondant à trois années d'essais : un premier essai en contre saison 2013 (semis en Octobre), un second essai en saison normale 2014 (semis de Juin et Juillet) puis un troisième essai en 2015 (semis de Juin, Juillet et Septembre). Les quantités de pluies recueillies au cours de la réalisation de ces différents essais (des jours précédant le semis jusqu'à la récolte) ainsi que les températures moyennes mensuelles enregistrées durant les saisons normales de culture 2014 et 2015 sont présentées au niveau de la Figure II-6.

Les analyses conjointes des 6 essais mis en place, ont mis en évidence l'importance de l'effet date de semis sur les phénotypes observés. En effet, comme illustré au sein de la Figure II-5 les données se groupent en fonction des dates de semis.

Les résultats de l'ensemble des essais ont montré que le parent donneur E36-1 est plus précoce de 5 et 10 jours que le parent récurrent Keninkeni respectivement pour les semis les plus précoces de 2014 (24 juin) et 2015 (21 juin). Par contre pour les deux autres semis tardifs (16 juillet 2014 et 2015), c'est le parent Keninkeni qui s'est montrée plus précoce. Avec un coefficient de sensibilité à la photopériode ( $K_p$ ) de 0.56 (mesuré uniquement sur l'essai de saison normale de culture 2015), Keninkeni est plus sensible à la variation de la durée du jour

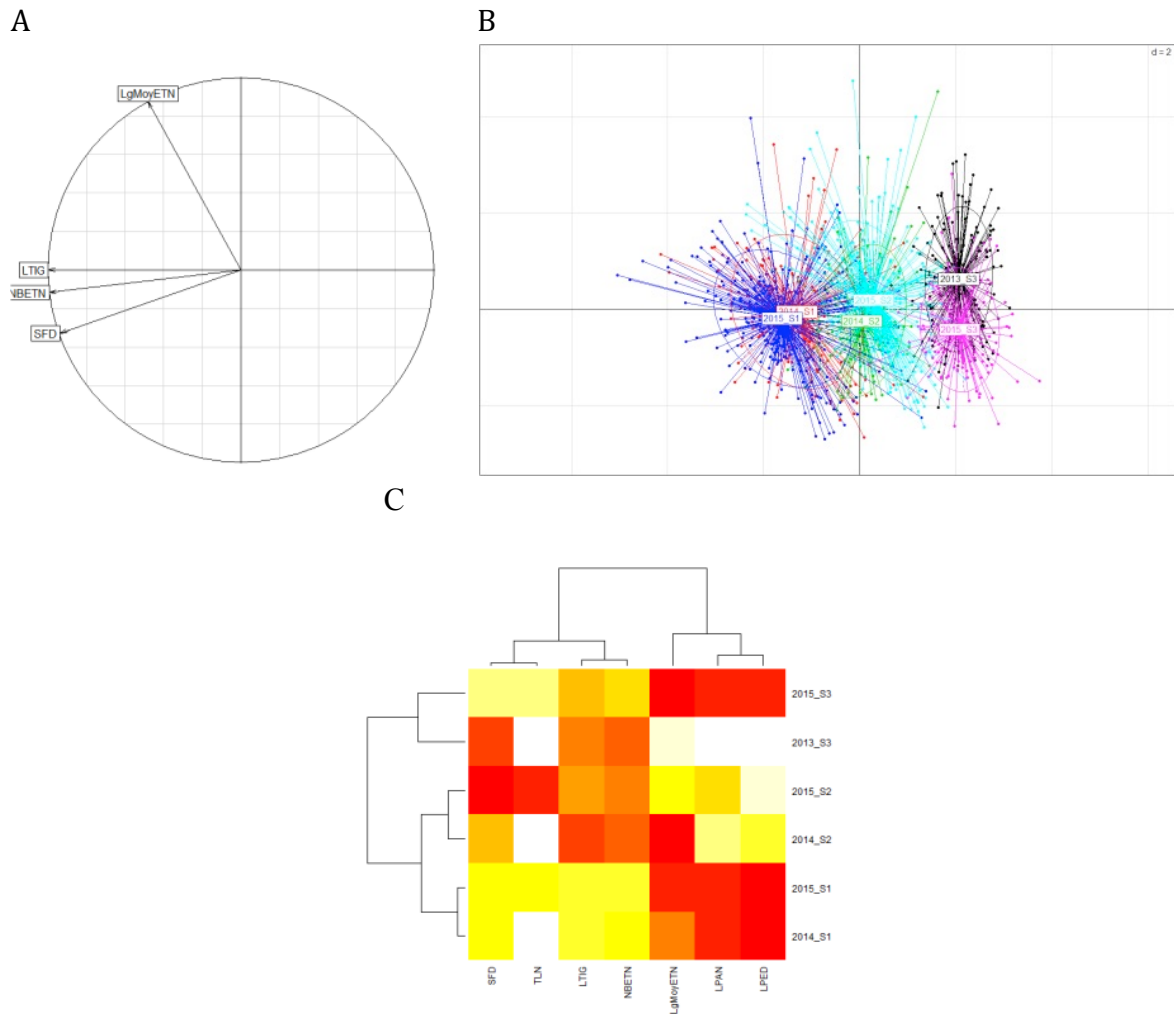


Figure II-7 : Analyses conjointes des 6 essais mis en place avec la population BC06. A) Analyse en composantes principales basée sur les quatre caractères communs aux 6 essais (SFD, LTIG, NBETN, lgmoyETN), les deux premières composantes principales expliquent 99 % de la variabilité phénotypique. L'axe 1 est principalement défini par la longueur totale de la tige, le nombre d'entre-nœuds et la durée du cycle semis-feuille drapeau, tandis que l'axe 2 dépend principalement de la longueur moyenne des entre-nœuds. B) Sur le plan composé des deux premières composantes principales, les valeurs phénotypiques des différentes familles BC1F4 ont été positionnées en fonction des différents essais. Il apparaît que les données se groupent principalement en fonction des dates de semis en mettant une évidence une transition claire entre la première date de semis sur la gauche du graphique vers la contre-saison sur la droite en passant par la seconde date de semis. C) la prise en compte des données relatives au nombre total de feuilles développées et aux longueurs du pédoncule et de la panicule en complément des quatre caractères analysés en ACP ont permis de mettre en évidence la même tendance. En effet, l'analyse de clusterisation hiérarchique a mis en évidence 3 groupes principaux d'essais (en lignes) correspondant aux différentes dates de semis (du semis 1 à la base du clustering vers le semis 3 (contre saison) en haut). Les données en rouge correspondent à des valeurs élevées, tandis que celles en jaune correspondent à des valeurs basses. Les données en blanc correspondent à des données manquantes.

Tableau II-4 : Statistiques descriptives (moyenne et écart-type), coefficients de variation phénotypique et génotypique (CVP% et CVG%), héritabilité et test des effets « date de semis » et « Genotype X Date de semis » pour l'année 2015 (modèle 7) des principaux caractères mesurés sur les 161 familles de la population BC06, cultivées durant les saisons normales de culture 2014 et 2015 (semis de juin et juillet) et pendant la contre saison 2013 et 2015 (semis d'octobre et de septembre) à Bamako (Mali).

Variables	Semis	Semis 2014 et 2013				Semis 2015				ANOVA(modèle7)					
		Moyenne ± Ecart-type		Moyenne ± Ecart-type		Moyenne ± Ecart-type		CVP% CVG% H2		ANOVA(modèle7)		H2			
		E36-1	Kkni	Familles	E36-1	Kkni	Familles	S	F	S*	F*	S*	F*		
<i>SFD</i>	Juin	68.7±1.8	78.5±1.9	75.77±8.37	Juin	75.4±2.6	80±2.4	77.7±8.6	11.0	10.0	0.9	***	***	***	0.9
	Juil.	77.5±4	71.8±1.9	70.1±6.6	Juil.	67.2±2.6	63.6±3.6	63.7±6.3	9.7	8.3	0.9				
	Oct.	-	-	53.8±5.6	Sept.	60.5±2.7	56.2±1.91	58.9±5.5	-	-	-				
<i>Kp</i>	Juin	-0.4±0.2	0.28±0.1	0.31±0.2	Juin	0.28±0.1	0.6±0.1	0.49±0.2	38.0	19.2	0.4				-
	Juil.	-	-	-	Juil.	-	-	-	-	-	-				
	Oct.	-	-	-	Sept.	-	-	-	-	-	-				
<i>NBTFD</i>	Juin	-	-	-	Juin	25.5±2.4	28.1±1.3	28.2±2.5	8.7	7.4	0.8	***	***	***	0.9
	Juil.	-	-	-	Juil.	21.4±2.7	21.6±1.5	21.5±2.2	10.0	7.8	0.8				
	Oct.	-	-	-	Sept.	17.6±1.2	16.1±1.5	18.5±7.7	-	-	-				
<i>LTIG</i>	Juin	158±10.9	163.6±8.1	161±36	Juin	154.6±11	167.7±15	168.9±38.5	22.5	20.6	0.9	***	***	***	NS
	Juil.	105.6±14.4	101±12.5	111.6±28	Juil.	123.2±14	112.3±8.1	119.1±33.5	27.6	24.0	0.9				
	Oct.	-	-	66.8±17.6	Sept.	64.2±2	44.4±3.6	53.5±14.5	-	-	-				
<i>NBETN</i>	Juin	13.9±0.7	16.1±0.7	15.4±2.1	Juin	14.9±0.7	16.8±1.1	16.4±1.9	11.8	10.0	0.8	***	***	***	0.9
	Juil.	11.7±0.9	11.4±1.1	11.5±2.1	Juil.	11.5±0.9	11.8±0.8	11.5±1.8	15.4	13.1	0.8				
	Oct.	-	-	6.2±1.2	Sept.	7.8±1.1	5.8±0.6	6.4±1.4	-	-	-				
<i>LgMoyETN</i>	Juin	11.4±0.7	10.2±0.6	10.5±2.2	Juin	10.4±0.9	10±0.8	10.3±2.1	19.8	17.6	0.9	NS	***	NS	NS
	Juil.	9.1±0.9	8.8±0.8	9.9±2.6	Juil.	10.6±0.6	9.5±0.6	10.4±2.6	24.7	21.1	0.8				
	Oct.	-	-	10.9±2.4	Sept.	8.1±1.5	7.7±0.8	8.6±2.02	-	-	-				
<i>LPED</i>	Juin	34.6±4.2	60.8±4.75	47.7±9.7	Juin	32.4±2.2	58.7±5.1	48.3±8.7	17.8	15.3	0.8	*	***	NS	NS
	Juil.	28.1±3.7	60.4±6.9	48.7±10.9	Juil.	30.9±3.1	62.1±2.5	49.3±9.1	18.2	14.7	0.8				
	Oct.	-	-	-	Sept.	31.05±5	51.2±4.	43.8±9.9	-	-	-				
<i>LPAN</i>	Juin	21.3±1.2	41.3±1.9	33.9±3.8	Juin	23.7±1.1	40.3±3.4	34.5±4.5	12.2	9.7	0.8	NS	***	NS	NS
	Juil.	22.8±3.6	42.1±2.9	35.6±4.5	Juil.	20.4±1.8	40.5±2.3	34.4±4.8	13.5	9.8	0.7				
	Oct.	-	-	-	Sept.	22.2±2	34.3±1.6	29.6±7.3	-	-	-				

\*: Significatif au seuil de probabilité de 5%, \*\*: Significatif à 1% et \*\*\*: Significatif au seuil de probabilité de 0.1%. NS: non significatif

que E36-1 ( $K_p=0.28$ ), ce qui explique la forte diminution de la durée de son cycle entre les semis de juin et juillet (17 jours) pour l'essai de 2015.

La valeur moyenne du cycle semis-floraison (estimée au travers de la durée entre le semis et la feuille drapeau : SFD) pour les 161 familles de la population est inférieure à celle du parent récurrent Keninkeni de 3 jours et supérieure (de 2 à 7 jours) à la moyenne du parent donneur E36-1 dans les deux environnements. Dans l'ensemble, les descendance sont plus sensibles à la variation de la durée du jour que le parent donneur E36-1. En ce qui concerne les paramètres relatifs à la production de biomasse (hauteur des tiges, nombre total de feuilles développées, nombre et longueur moyenne des entre-nœuds), la moyenne de l'ensemble de la population pour chacun de ces caractères est similaire à celle du parent récurrent Keninkeni. Par contre pour les caractères liés à la longueur du pédoncule et de la panicule, la moyenne des descendance est similaire dans les deux environnements et supérieure à celle du parent donneur E36-1.

De manière générale, pour l'ensemble des caractères analysés dans les trois semis de 2015, les valeurs des coefficients de variation phénotypique (CVP %) sont similaires à celles des coefficients de variation génotypique (CVG%) (cf. modèles (4) et (5)) à l'exception de la sensibilité à la photopériode pour laquelle les CVP (38 %) sont deux fois plus grands que les CVG (19 %). Cette différence entre variance phénotypique et génotypique pour le  $K_p$ , indique un fort effet de l'environnement sur la sensibilité à la photopériode contrairement aux autres variables phénologiques et agromorphologiques. Les plus grandes valeurs de CVP ont été observées pour les caractères de hauteur des tiges, de longueur moyenne d'entre-nœuds (semis de juillet) et de sensibilité à la photopériode ( $> 20\%$ ), alors qu'elles sont moins élevées pour les autres variables agronomiques.

L'analyse de variance (modèle 7) a permis de mettre en évidence un effet hautement significatif ( $P < 0.001$ ) de la date de semis sur la plupart des variables mesurées. Par contre, l'effet de la date de semis a été non significatif sur la longueur moyenne des entre-nœuds et sur la longueur de la panicule (Tableau II- 4). Des différences très significatives ( $P < 0.001$ ) entre les géotypes (familles) ont été également observées pour tous les caractères. Les interactions « Famille\*Semis » sont significatives seulement pour les variables de durée du cycle semis-floraison ( $P < 0.001$ ), de nombre total de feuilles développées et de nombre total d'entre-nœuds allongés ( $P < 0.05$ ).

De fortes héritabilité au sens large (cf. modèle 6,  $H^2_{sl} > 0.80$ ) ont été obtenues pour les différents caractères agronomiques mesurés.



Tableau II-5 : Corrélations phénotypiques entre les caractères phénologiques et agromorphologiques sur la base des résultats Blup obtenus sur les 161 familles phénotypées en juin durant la saison normale de culture 2015 à Bamako (Mali).

<i>Variables</i>	<i>SFD</i>	<i>Kp</i>	<i>TLN</i>	<i>LTIG</i>	<i>NBETN</i>	<i>LgMoyETN</i>	<i>LPED</i>
<i>Kp</i>	0.76***						
<i>TLN</i>	0.83***	0.64***					
<i>LTIG</i>	0.42***	0.22**	0.44***				
<i>NBETN</i>	0.80***	0.59***	0.89***	0.53***			
<i>LgMoyETN</i>	0 NS	-0.1 NS	-0.04 NS	0.85***	0.02 NS		
<i>LPED</i>	-0.11 NS	-0.16 NS	-0.13 NS	0.26**	-0.1 NS	0.38***	
<i>LPAN</i>	0.02 NS	-0.07 NS	0.02 NS	0.13 NS	0.09 NS	0.11 NS	0.54***

\*: Significatif au seuil de probabilité de 5%, \*\*: Significatif à 1% et \*\*\*: Significatif au seuil de probabilité de 0.1%. NS: non significatif.

Tableau II-6 : Corrélations de Pearson entre les années (2013, 2014 et 2015) pour les différentes variables agronomiques sur la base des données moyennes ajustées (saison normale de culture 2014 et contre saison 2013 et 2015) et/ ou à partir des estimations des valeurs génétiques « Blup » (saison normale de culture 2015).

<i>Variables</i>	Semis de juin r(2014-2015)	Semis de juillet r(2014-2015)	Semis de septembre r(2013-2015)
<i>SFD</i>	0.89***	0.81***	0.32***
<i>Kp</i>	0.26**	-	-
<i>LTIG</i>	0.89***	0.06 NS	0.03 NS
<i>LPED</i>	0.765***	0.06 NS	-
<i>LPAN</i>	0.56***	0.55***	-
<i>NBETN</i>	0.75***	0.72***	0.44***
<i>LgMoyETN</i>	0.86***	0.78***	0.67***

\*: Significatif au seuil de probabilité de 5%, \*\*: Significatif à 1% et \*\*\*: Significatif au seuil de probabilité de 0.1%. NS: non significatif

Par ailleurs, l'estimation des corrélations génotypiques entre les caractères phénologiques et agromorphologiques sur la base des résultats Blup obtenus sur le semis de juin de l'essai 2015, ont montré de fortes corrélations positives ( $P < 0.001$ ) entre la durée du cycle semis floraison et la sensibilité à la photopériode (Tableau II-5). Ces deux variables sont également corrélées ( $P < 0.01$ ) aux variables de production de biomasse (NBTFD, LTIG et NEN).

Par contre aucune corrélation significative n'a été mise en évidence entre la durée du cycle, le nanisme (LMEN) et les caractères de longueur du pédoncule et de la panicule (LPED et LPAN). De fortes corrélations positives ( $P < 0.01$ ) ont été observées entre la hauteur des tiges, la longueur moyenne des entre-nœuds et la longueur du pédoncule.

En complément des corrélations intra année entre caractères, les corrélations phénotypiques inter-années pour un caractère donné ont été calculées. Ces corrélations sont dans la plupart des cas significatives et positives (Tableau II-6).

### ***III.2.2. Identification des QTL au sein de la population BC06***

Afin de définir des QTL uniques, nous avons considéré que lorsque la distance entre les pics de LOD de 2 QTL situés sur un même chromosome était inférieure ou égale à 10cM, les 2 QTL correspondaient à une région identique. Au total, pour l'ensemble des 9 variables analysées dans les différents essais, 53 régions génomiques uniques ont été identifiées. Un résumé de ces QTL est présenté dans l'annexe (Annexe 2 et suite).

#### ***III.2.2.1- Des régions chromosomiques à effet fort associées à la variation de la durée semis floraison et à la sensibilité à la photopériode***

Au total, 9 QTL ont été identifiés pour les caractères relatifs à la date de floraison. Ces QTL expliquent individuellement entre 3.64 et 54 % de variation de la date de floraison (Annexe1 et Figures II.7). Parmi ces QTL, le QTL à effet majeur identifié à 1.33cM sur le chromosome 6, explique à lui seul jusqu'à 54% de variation de la durée semis-feuille drapeau (SFD) en saison normale de culture (semis de juin et juillet). En contre saison, ce QTL explique en revanche, seulement 5.45 % de variance phénotypique de SFD.

Sur le chromosome 3 (12 cM) se trouve deux QTL à effets majeurs expliquant respectivement 31 et 26 % de la variation phénotypique de la durée du cycle semis-feuille drapeau. D'autres QTL à effets forts ( $R^2$  variant entre 15 et 17 %) sont identifiés dans la même zone génomique. Les autres QTL à effets mineur, sont localisés sur les chromosomes 1, 2, 4, 8, 9 et 10.

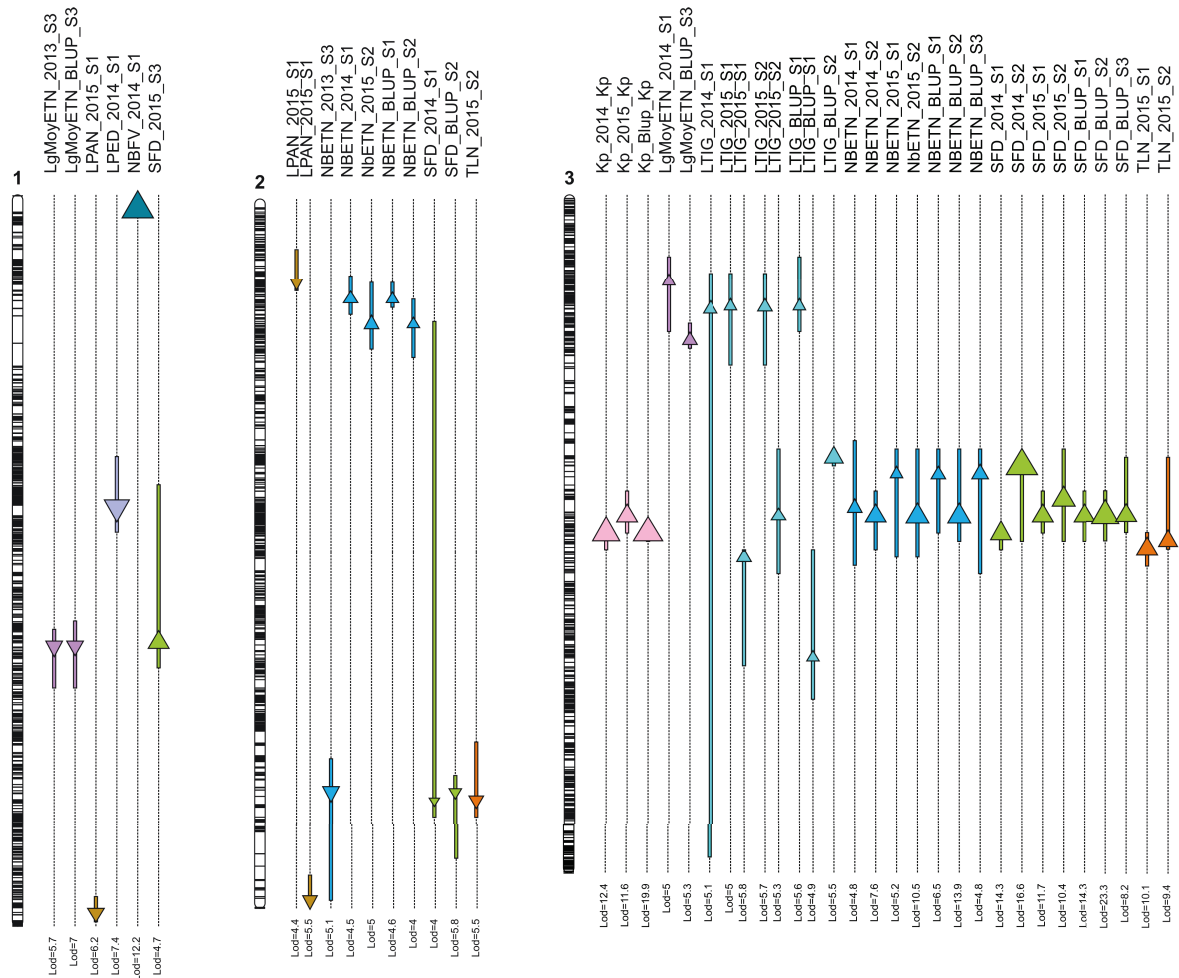


Figure II-7 : Représentation graphique des QTL détectés dans les 6 environnements pour les 9 variables étudiées. Chaque QTL est représenté par un triangle localisé au niveau du pic du QTL et indiquant le sens de l'effet additif (vers le haut : effet positif provenant du parent donneur E36-1, vers le bas : effet positif apporté par le parent récurrent keninkeni) et d'une barre représentant l'intervalle de confiance à 95% calculé par le LOD-1.5. La taille du triangle est proportionnelle à la part de variation phénotypique expliquée par le QTL ( $R^2$ ). Les valeurs de Lod sont représentées en bas de chaque QTL. Chaque couleur représente les QTL détectés pour un caractère

Concernant la sensibilité à la photopériode (Kp), 5 QTL ont été détectés sur les chromosomes 3, 4, 6, 7 et 8 parmi lesquels 2 QTL à effets majeurs (LOD= 19 et 29). Ces QTL majeurs localisés sur les chr 3 (et chr6, colocalisent avec les QTL identifiés pour la date de floraison ; et expliquent 27 et 49 % de variation de la sensibilité à la photopériode respectivement pour les chr 3 (80 cM) et 6 (1.33 cM). Par ailleurs, ce QTL majeur de floraison/sensibilité à la photopériode cartographié sur le chr 3 a été également détecté au sein de la population « P118 » (cf. chapitre II. II-article). Les autres QTL mineurs situés sur les chromosomes 4, 7 et 8 expliquent seulement une faible part de variation phénotypique de ce caractère ( $R^2$  variant entre 4.79 et 5.96 %).

Il est intéressant de noter que dans la zone du chromosome 3 comprise entre 64.85 - 80.85 cM), tous les allèles de tardiveté de la floraison et de sensibilité à la photopériode (Kp) des familles BC1F4, sont apportées par le donneur E36-1 ; alors qu'au niveau du chr6 (1.33 cM), la tardiveté est apportée par le parent récurrent Keninkeni (Figures II-7, annexe1).

#### III.2.2.2- Régions chromosomiques associées à la hauteur des tiges et à ses composantes

Sept QTL localisés sur les chromosomes 3, 6 et 7 ont été identifiés pour la hauteur des tiges (LTIG) pour les semis de juin (2014) et juillet (2015). Entre 5.20 et 54 % de variance phénotypique de la hauteur des tiges sont expliquées par ces derniers (Annexe1). Sur le chr7 se trouvent deux QTL majeurs de hauteur en répulsion avec des valeurs d'additivité très élevées (jusqu'à -30 et +42 cm). Ces deux QTL se situent à 22 cM de distance l'un de l'autre et expliquent chacun entre 23 et 54 % (pour le blup S1) de variation phénotypique de la longueur de la tige (Figures II-7). D'autres QTL à effets mineurs ont été détectés sur les chr3, 6 et 7 et expliquent une faible part de la variation de la longueur de la tige dans les différents environnements de culture.

Au total, 7 QTL pour le nombre total d'entre-nœuds allongés (NEN) sur la tige principale ont été détectés expliquant entre 4.49 à 45 % de la variation phénotypique. Parmi ces QTL, on retrouve le QTL majeur du chromosome 6 (LOD= 28) expliquant à lui seul plus de 45 % de variance phénotypique du nombre d'entre-nœuds allongés. Sur le même chromosome, à 74 cM du QTL majeur, d'autres régions génomiques sont associées à la variation du nombre d'entre-nœuds allongés ( $R^2= 5.36 - 12\%$ ) ; en plus des autres zones génomiques à effets faibles localisées sur les chr2, 3, 7 et 8.

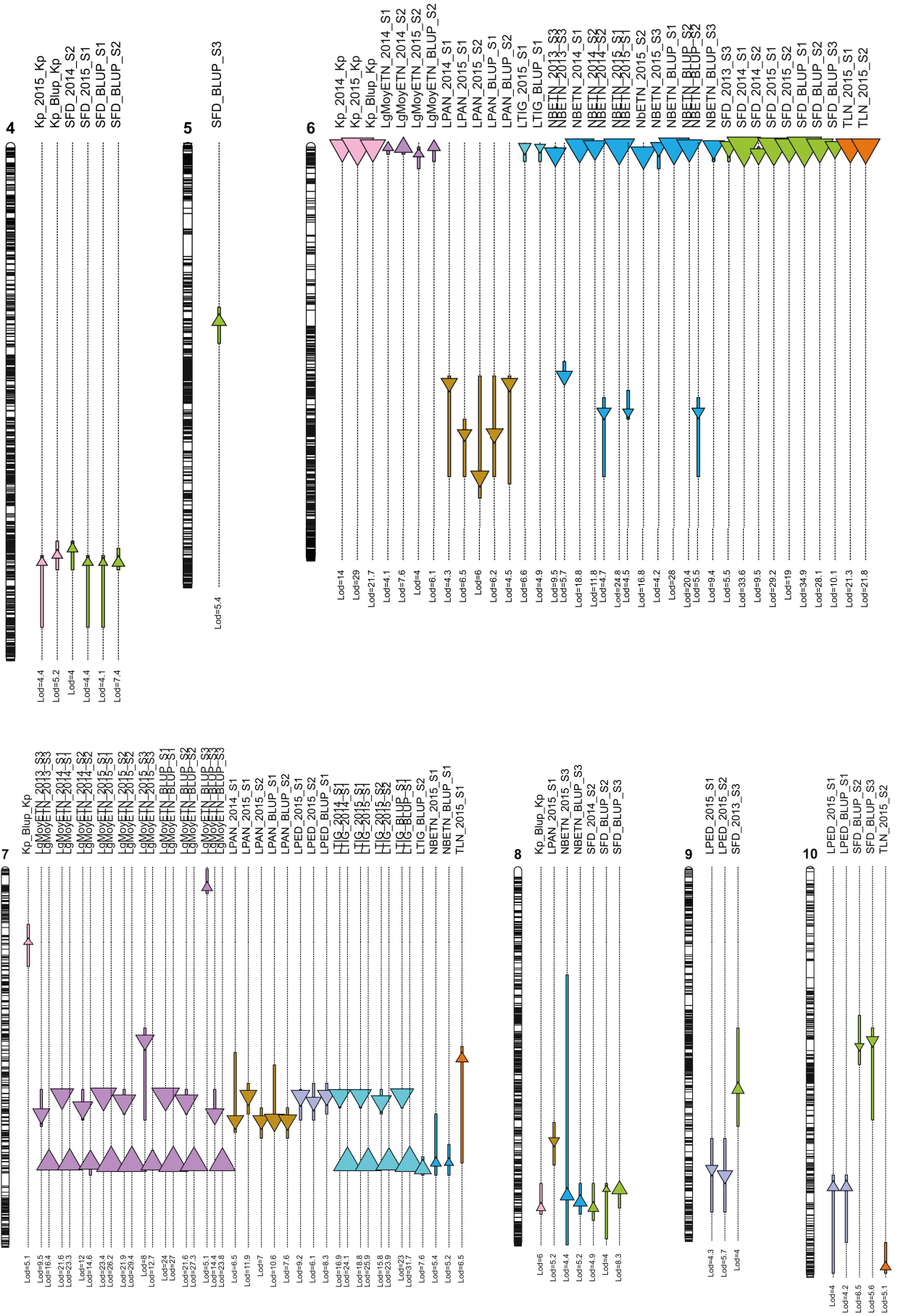


Figure II-7: Suite

En ce qui concerne la longueur moyenne des entre-nœuds (indicateur du nanisme des plantes), 6 QTL ont été détectés, dont 2 QTL majeurs en répulsion (les mêmes que ceux détectés pour la longueur de la tige). Ces QTL majeurs localisés dans deux zones génomiques du chr7, expliquent jusqu'à 54% de variation de la longueur moyenne des entre-nœuds. Sur les chromosomes 1, 3, 6 et 7, se trouvent des QTL à effets mineur (LOD= 3.97-9.5), mais expliquant une part assez importante de variance phénotypique de LMEN ( $R^2= 4 - 18\%$ ). Comme la hauteur des tiges, les allèles conférant un phénotype « nain » sont apportés par le parent donneur E36-1.

Cinq QTL associés à la variation du nombre total de feuilles développées entre le semis et la maturation du grain, ont été identifiés sur les chromosomes 2, 3, 6, 7 et 10, parmi lesquels un QTL majeur sur le chr6 (1.33 cM). La plus grande part de variance phénotypique de NBTFD est expliquée par ce QTL ( $R^2=37\%$ ) ; tandis que les autres régions génomiques expliquent seulement de faibles parts de sa variation ( $R^2= 6 - 8\%$ ) à l'exception du chr3 ( $R^2= 14\%$ ). Par contre, un seul QTL a été identifié pour la variable de nombre total de feuilles vertes à maturité ( $R^2= 32\%$ ).

Quant aux caractères relatifs à la longueur du pédoncule et de la panicule, 13 QTL à effets plus ou moins faibles (Lod= 3.96 – 11.95) ont été identifiés dans tous les environnements dont 5 QTL de LPED et 8 QTL de LPAN. Dans l'ensemble, ces QTL expliquent une part de variation assez importante de LPED et de LPAN ( $R^2$  variant entre 5.86 et 25 %)

### III.2.2.3- Stabilité des QTL en fonction de la date de semis

De manière générale la plupart (50%) des QTL identifiés dans cette étude sont stables entre les différentes dates de semis de la saison normale de culture (S1 ; juin et S2 : juillet), contrairement au semis de contre saison (S3 ; septembre) pour lequel la majorité de ces QTL ne s'exprime pas (10%). Seuls les QTL impliqués dans la variation phénotypique de la durée du cycle « semis - floraison » (SFD, Figures II.8, II.9), du nombre d'entre-nœuds allongés (NEN, Figure II.8, II.10), du nanisme (longueur moyenne d'entre-nœuds : LMEN, Figures II.8, II.11) et du nombre total de feuilles développées (NBTF) ont été détectés dans cette condition de culture particulière.

Parmi les 52 QTL identifiés dans les différents environnements pour l'ensemble des variables analysées (à l'exception de NBFV mesurée uniquement pour le semis de juin 2014), 9 sont impliqués dans la variation de la durée semis feuille drapeau. Parmi ceux-ci, 3 sont stables entre les dates de semis de juin et juillet tandis qu'un seul QTL est stable entre les 3 semis. Trois

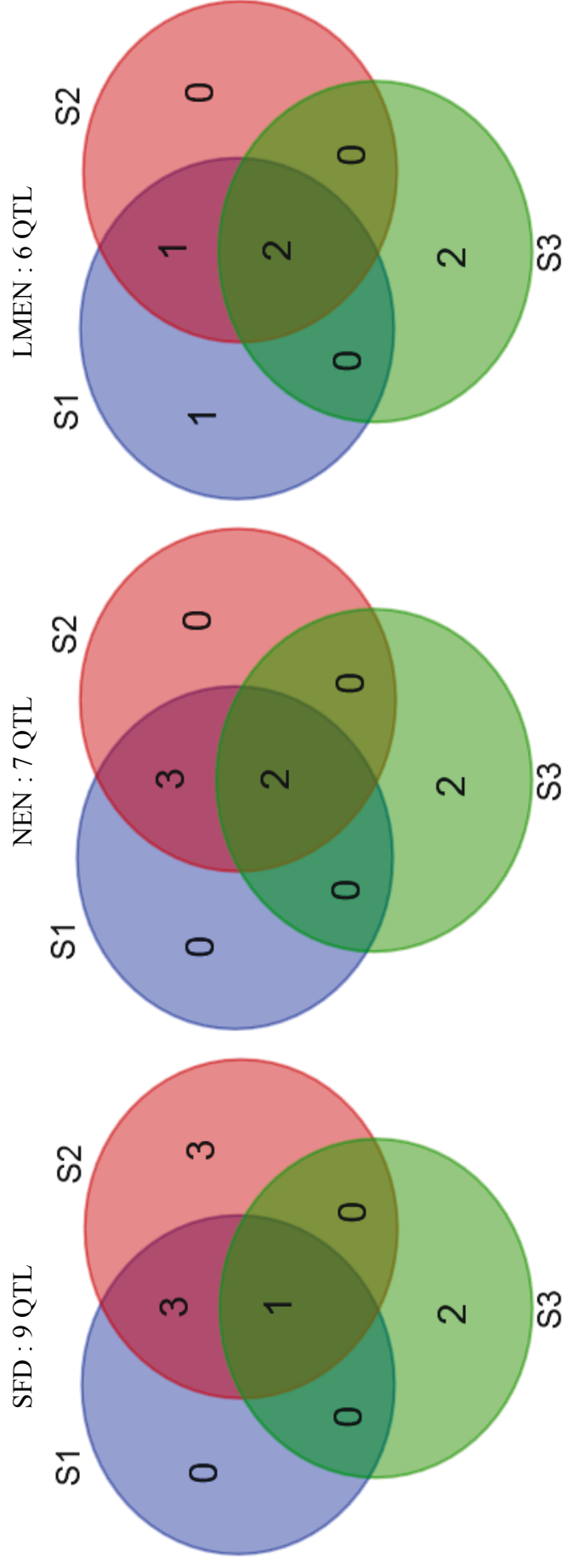


Figure II-8 : Stabilité des QTL entre les 3 dates de semis pour les caractères de longueur moyenne des entrenœuds (LMEN), de nombre d'entrenœuds allongés (NEN) et de durée du cycle « Semis-Floraison » (SFD) représentée au travers de diagrammes de Venn. Le semis S3 présente le plus grand nombre de QTL spécifique (27 % (6/22)), tandis que presque la moitié des QTL sont détectés simultanément pour les semis S1 et S2 (54 % (12/22)). L'ensemble des QTL détectés sur les deux années d'essais a été considéré pour effectuer cette analyse, les QTL correspondant à chaque date de semis ont donc pu être détectés sur une des deux années uniquement, sur les deux années simultanément ou sur les valeurs de blup calculées à partir des 2 années.

QTL sont spécifiques au semis de juin et seulement 2 QTL sont spécifiques au semis de contre saison et qui sont localisés sur les chromosomes 1 et 9 (Figure II-8, II.9).

Concernant les QTL de nombre total d'entre-nœuds allongés, un total de 7 QTL a été identifié. Deux QTL ont été détectés dans tous les environnements, 3 QTL sont stables entre les semis de Juin et de Juillet (Chromosome 2) et les 2 derniers QTL sont spécifiques au semis de Septembre (Figure II.8, II.10).

Quant aux QTL de longueur moyenne des entre-nœuds, ceux-ci sont moins stables à travers les environnements : seulement 2 QTL sur les 6 identifiés sont communs aux trois dates de semis et correspondent à la zone génomique du chromosome 7 où se trouve le gène majeur de hauteur/nanisme. En plus, une autre région chromosomique (chr6) commune aux semis de Juin et Juillet a été détectée, alors que les 3 autres QTL restant sont spécifiques au semis de Juin (1 QTL) et de Septembre (2 QTL) (Figures II.8, II.11).

Parmi les 7 QTL détectés pour la hauteur des tiges, 4 QTL sont communs aux semis de juin et juillet (impliquant 4 régions génomiques des chromosomes 3 et 7) et 3 QTL spécifiques aux semis de juin (2 QTL) et juillet (1QTL) (Annexe 3).

En ce qui concerne la stabilité des QTL de longueur du pédoncule et de la panicule (QTL détectés seulement dans l'environnement de saison normale de culture (semis de Juin et Juillet), sur les 4 QTL de LPED, 2 QTL sont communs aux deux semis, alors que, seulement 2 QTL sur les 8 QTL de LPAN sont stables à travers les environnements. Sur les 6 QTL restant 5 QTL sont spécifiques aux semis de Juin et un seul QTL propre au semis de Juillet (Annexe 3).

Concernant les QTL de nombre total de feuilles développées (variable mesurée uniquement pour les semis de Juin et Juillet 2015), seulement 2 QTL stables entre les deux dates de semis ont été mis en évidence (Annexe 3).



### SFD

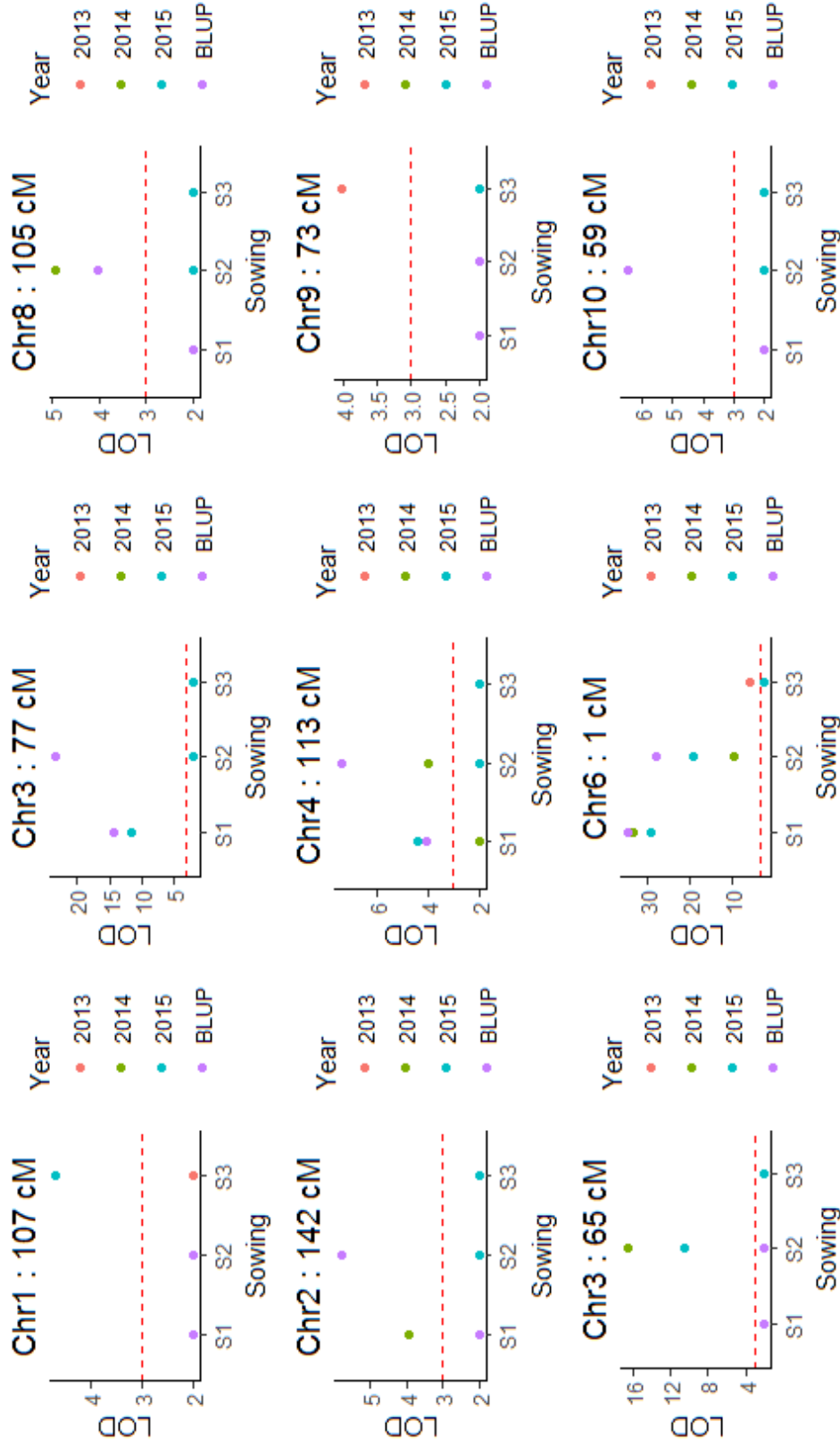


Figure II-9 : Stabilité de la significativité des QTL détectés pour les différentes dates de semis pour le caractère de durée du cycle « Semis-Floraison » (SFD). Les QTL détectés au sein des différents essais sont indiqués avec des couleurs différentes (2013 en rouge, 2014 en vert et 2015 en bleu). Les QTLs détectés pour les valeurs de Blup « inter-années- intra semis » sont indiqués en violet. Les QTL présentant des LOD en dessous du seuil de significativité ont arbitrairement été représentés avec une valeur de LOD de 2. La ligne rouge pointillée représente le seuil de LOD de 3 qui caractérise la valeur de LOD significative minimale qui a été considérée entre les différents caractères.

### III.3. Discussion

#### *III.3.1- Durée du cycle « Semis – Floraison » : De nombreux QTL validés entre études et un nouveau QTL majeur mis en évidence sur le chromosome 3*

Dans la présente étude, 9 régions chromosomiques impactant la durée du cycle « semis-feuille drapeau » ont été détectées. Trois régions à effet fort ont été mises en évidence. Deux d'entre elles (chr3 entre 60 et 68 cM et chr6 entre 5.6 et 10 cM) ont déjà été mises en évidence dans des études précédentes (Parh, 2005, Mace et al., 2013a, Bangbol Sangma, 2013) tandis que la région située sur le chromosome 3 à 76.8 cM (Figure II.8, Annexe1) a été détectée pour la première fois dans le cadre de cette étude au sein des différentes populations (P118, BC06 et BCNAM) . Cette région a été mise en évidence pour les semis S1 et S2, mais n'affecte pas la date de semis en condition de contre saison. Elle présente aussi des effets sur la sensibilité à la photopériode, le nombre d'entre-nœuds et la longueur des tiges (Figure II.9, II.10).

Parmi les 6 autres régions, toutes à l'exception de celle située sur le chromosome 4 (113 cM) ont déjà été détectées.

Le QTL majeur du chr6 qui impacte d'une part la durée « semis –floraison » et d'autre part la sensibilité à la photopériode se situe dans la région du gène majeur de maturité Ma6 déjà identifié chez le sorgho (Rooney and Aydin, 1999a, Murphy et al., 2014b, Yang et al., 2014b). Ce gène est impliqué dans le contrôle génétique de la sensibilité à la photopériode en condition de jour long. Murphy et al. (2014b) ont mis en évidence les intérêts de ce gène sur l'augmentation de la production de biomasse par son action de tardiveté de la floraison.

Les autres QTL de floraison à effets plus faibles ont déjà été identifiés au sein d'études précédentes :

- chr1 « 107 cM » (Srinivas et al., 2009, El Mannai et al., 2011, Mace et al., 2013a)
- chr2 « 142 cM » et chr4 « 113.8 cM » : Bangbol Sangma, 2013
- chr7 « 62.7 cM » : (Srinivas et al., 2009, Shiringani et al., 2010, Mace et al., 2013a), qui a été détecté uniquement pour le nombre total de feuilles sur l'essai de 2015 dans le semis1
- chr8 « entre 105.4 et 115.4 cM », chr9 « 72.7 cM » et chr10 « 58.6 cM (Crasta et al., 1999, Parh, 2005, Mace et al., 2013a).

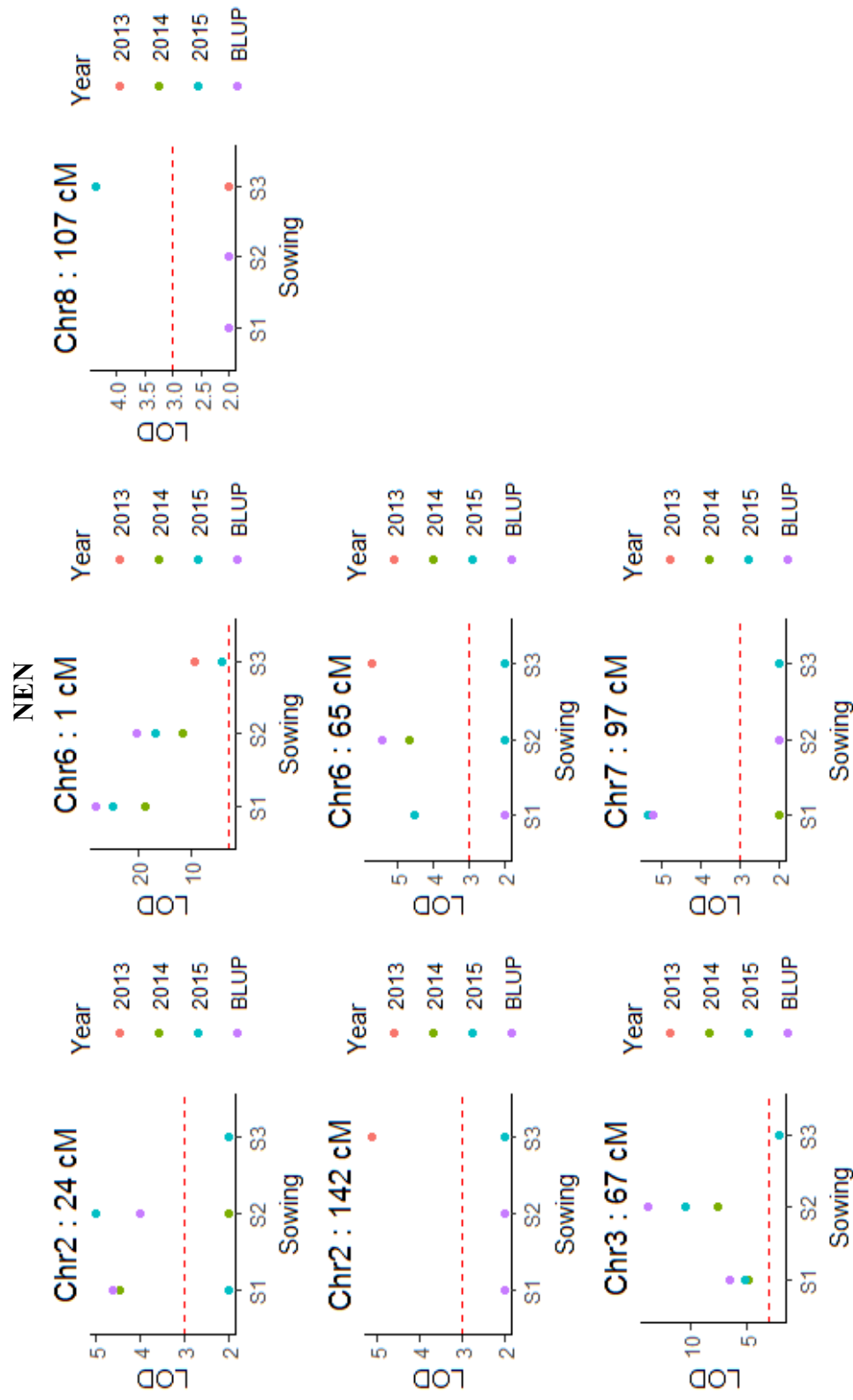


Figure II-10 : Stabilité de la significativité des QTL détectés pour les différentes dates de semis pour le caractère de nombre moyen d'entre-nœuds allongés (NEN). Les QTL détectés au sein des différents essais sont indiqués avec des couleurs différentes (2013 en rouge, 2014 en vert et 2015 en bleu). Les QTLs détectés pour les valeurs de Blup « inter-années- intra semis » sont indiqués en violet. Les QTL présentant des LOD en dessous du seuil de significativité ont arbitrairement été représentés avec une valeur de LOD de 2. La ligne rouge pointillée représente le seuil de LOD de 3 qui caractérise la valeur de LOD significative minimale qui a été considérée entre les différents caractères.

### ***III.3.2- La hauteur et ses composantes : des gènes majeurs confirmés et plusieurs nouvelles régions à effet plus faibles mises en évidence.***

L'analyse du déterminisme génétique de la hauteur de la tige et de ses composantes dans la population BC06 a permis la mise en évidence de 7 zones chromosomiques impactant les composantes du rendement en biomasse végétative.

Deux zones à effets majeurs ont été détectées sur le chr7 à 22 cM de distance (72.7 et 94.7 CM). Le premier QTL localisé à 74.7 cM sur le chr7 est proche du QTL de hauteur (qHT7.1) cartographié par Li et al. (2015); tandis que le second QTL identifié à 96.7 cM sur le chr7 colocalise avec le gène majeur de nanisme dw3 (Casady, 1967, Campbell et al., 1975, Multani et al., 2003, Li et al., 2015). Dw3 est l'un des principaux gènes de nanisme utilisé par les programmes de conversion Américaine pour réduire la taille des sorghos. Dans la population BC06, pour les QTL de hauteur du chr 7, tous les allèles conférant un phénotype « grande taille » proviennent du parent récurrent Keninkeni.

Les autres régions génomiques à effets mineurs localisées sur le chr3 (S3\_6268239, S3\_54721296 et S3\_58744591) et sur le chr 6 (S6\_367564) ont également été cartographiées par les études antérieures (Ritter et al., 2008, Shiringani et al., 2010, Madhusudhana and Patil, 2013). La zone mise en évidence sur le chromosome 6 impacte non seulement la longueur totale de la tige mais aussi le nombre total d'entre-nœuds et de nombre de feuilles développées.

De nombreuses autres zones chromosomiques à effets mineurs associées à la production de biomasse (NEN, LMEN, NBTFD), aux longueurs du pédoncule et de la panicule (LPED et LAPN) et au nombre total de feuilles vertes à maturité (NBFV) ont été identifiées dans la population BC06. Une large majorité (60%) correspond à des zones n'ayant pas encore été mises en évidence.

### ***III.3.3- Co-localisations détectées entre les caractères***

L'analyse des 161 familles BC1F4 dans les 6 environnements de cultures a mis en évidence des zones génomiques uniques associées à la variation de plusieurs caractères. C'est le cas du QTL majeur de durée du cycle semis feuilles drapeau (SFD) identifié sur le chr6 qui colocalise avec les QTL de sensibilité à la photopériode (Kp), de hauteur des tiges et ses composantes (NEN et LMEN) et du nombre total de feuilles développées. Ce QTL explique entre 30 et 49.28 % de variation du Kp, entre 37 et 45% de variance phénotypique de NEN, entre 36 à 37% de NBTFD et entre 4 et 14 % de variation de LMEN uniquement dans les semis réalisés en saison normale de culture (juin et juillet) des années 2014 et 2015.

## LMEN

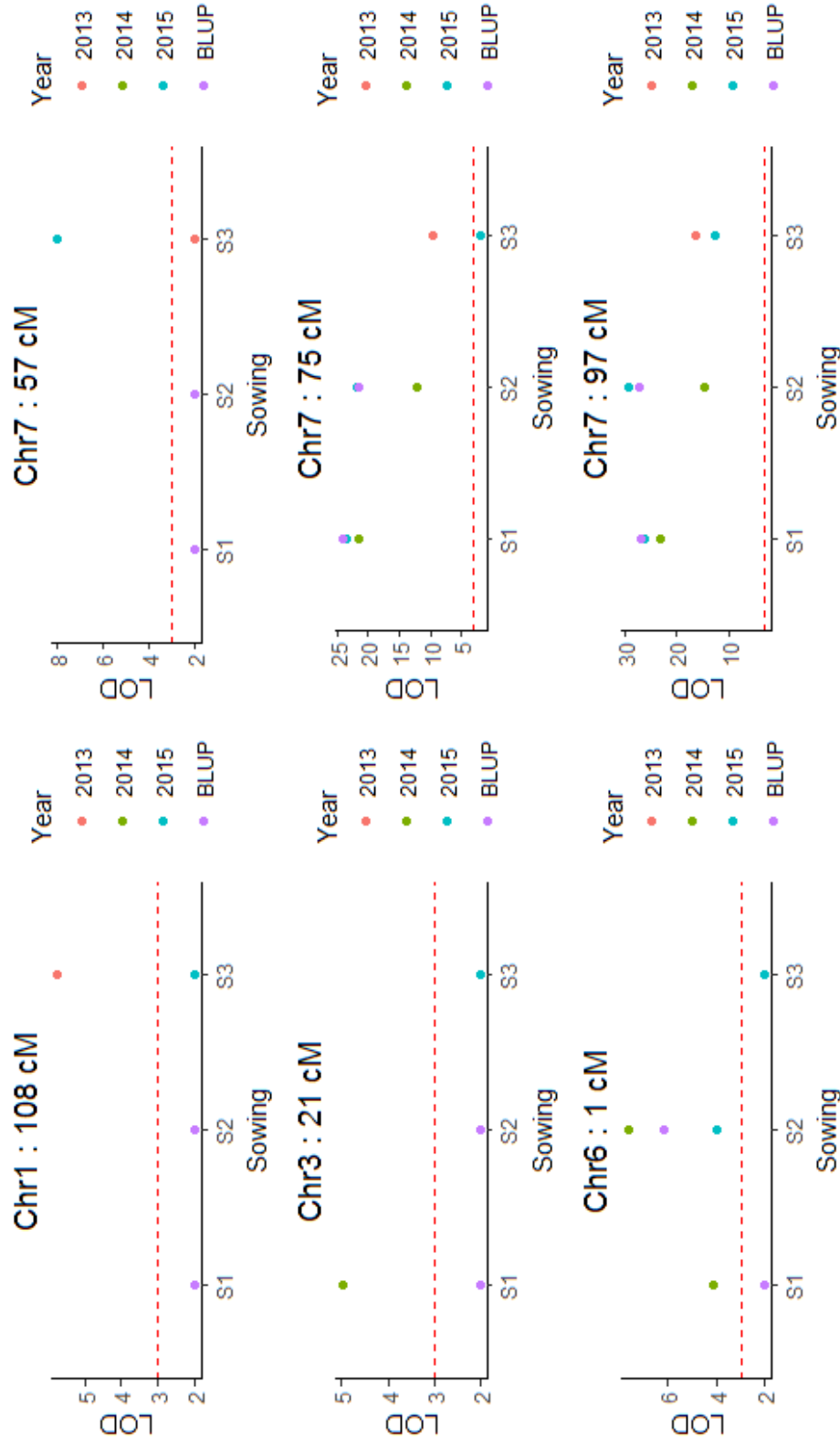


Figure II-11 : Stabilité de la significativité des QTL détectés pour les différentes dates de semis pour le caractère de longueur moyenne des entrenœuds allongés (LMEN). Les QTL détectés au sein des différents essais sont indiqués avec des couleurs différentes (2013 en rouge, 2014 en vert et 2015 en bleu). Les QTLs détectés pour les valeurs de Blup « inter-années- intra semis » sont indiqués en violet. Les QTL présentant des LOD en dessous du seuil de significativité ont arbitrairement été représentés avec une valeur de LOD de 2. La ligne rouge pointillée représente le seuil de LOD de 3 qui caractérise la valeur de LOD significative minimale qui a été considérée entre les différents caractères.

La colocalisation du QTL majeur de floraison du Chr6 avec la hauteur des tiges pourrait résulter de l'effet indirect du gène de maturité *ma6* qui impacte le nombre d'entre-nœud et donc la longueur des tiges. Dans la littérature, une hypothèse alternative est proposée, cette colocalisation pourrait être due à une liaison très forte entre le gène de floraison *Ma1* du chr6 et le gène nanisme *Dw2* du chr7 (Klein et al., 2008, Lin et al., 1995a). Ce phénomène est connu sous le nom de pléiotropie, c'est dire la capacité d'un seul gène à induire plusieurs effets phénotypiques.

Les QTL majeurs de hauteur et de nanisme (LMEN) cartographiés sur le chr7 à 55156 et 58358 Mb, colocalisent avec les QTL de LPED et de LPAN (Rami et al., 1998, Hart et al., 2001), alors que sur le même chromosome 7, à seulement 3Mb de ces QTL majeurs (58358 Mb) colocalisent les QTL de LTIG/LMEN avec les QTL de NEN. Cette zone génomique est également associée à la variation de la longueur de la panicule (56386 Mb). Selon la littérature, ce QTL majeur de hauteur du chr7 (58358 Mb) est très proche du QTL cartographié par Srinivas et al. (2009) à 58616 Mb.

La zone génomique du ch4 est associée à la variation de la durée du cycle « semis- floraison » et à la sensibilité à la photopériode en saison normale de culture.

#### *III.3.4-. Stabilité des QTL détectés entre les dates de semis*

En se basant sur les QTL détectés pour les caractères de longueur moyenne des entre-nœuds (LMEN), de nombre d'entre-nœuds allongés (NEN) et de durée du cycle « Semis-Floraison » (SFD) (Figure II-8), on observe que presque la moitié des QTL est détectée simultanément pour les semis S1 et S2 (48 % (12/25)) tandis que le semis S3 présente le plus grand nombre de QTL spécifiques (24 % (6/25)). Ces résultats indiquent une régulation fine du déterminisme génétique de la croissance des plantes entre les semis S1 et S2 et un remaniement majeur de son architecture génétique en condition de contre saison.

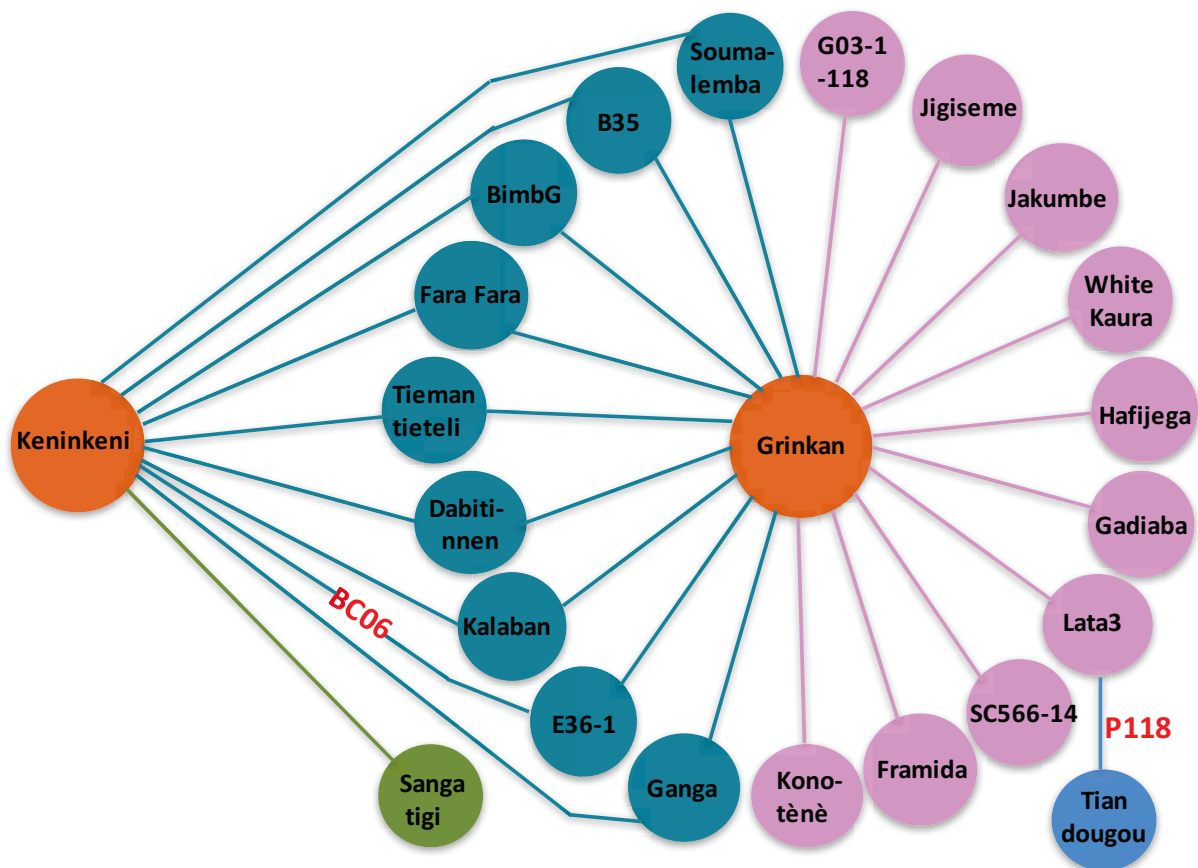


Figure II-12 : Dispositif multi-parental utilisé dans le cadre de cette thèse. Vingt-neuf populations ont été développées à partir des deux parents récurrents élités : Grinkan et Keninkeni (indiqués en orange) et 20 lignées parentales donneuses. Neuf lignées parentales ont été croisées avec les deux parents récurrents (indiquées en bleu), 10 lignées ont été croisées exclusivement avec le parent Grinkan (indiquées en rose) et une exclusivement avec le parent Keninkeni (indiquée en vert).

## **IV. Analyse du déterminisme génétique de la durée du cycle et de la croissance en hauteur sur la population « GCP-BCNAM »**

### **IV.1. Le dispositif « GCP-BCNAM » : des ambitions pour « l'analyse génétique » et la « sélection »**

Un dispositif multiparental de type Backcross Nested Association Mapping « BC-NAM » (Jordan et al, 2011), été développé dans le cadre d'une collaboration entre l'IER et le CIRAD au travers d'un projet financé par le Generation Challenge Programme. Les objectifs poursuivis, au travers de ce plan de croisement, étaient double, d'une part développer des populations pertinentes pour les actions de sélection, et d'autre part permettre une analyse fine du déterminisme génétique des caractères cibles de la sélection.

### **IV.2. Matériels et Méthodes**

#### ***IV.2.1- Matériel végétal***

##### **IV.2.1.1- GCP-BCNAM : des ambitions et des contraintes**

Au lancement des travaux de croisements, 33 parents donneurs et 2 parents récurrents ont été considérés. Néanmoins l'ensemble des combinaisons n'ont pas été menées à leurs termes en raison principalement :

- d'échecs survenus lors de certains croisements initiaux,
- de différences de parents donneurs entre les croisements F1 et les croisements BC1 révélées lors du contrôle de qualité des croisements effectué sur les différentes populations (parents donneurs trop hétérogènes),
- des proximités entre parents, qui ne justifiaient donc pas la création de certaines populations (par exemple Tiandougou et Grinkan se sont révélés être génétiquement très proches) et donc redondants.

En outre, au départ, il était prévu ne considérer qu'un seul parent récurrent (i.e. Grinkan), qui représente l'idéotype des sélectionneurs dans la région Malienne, mais les sélectionneurs du programme Sorgho de l'IER ont décidé d'ajouter la variété Kéninkeni comme autre parent récurrent en raison de ses caractères agronomiques (morphologie et qualité du grain), complémentaires de ceux de Grinkan. Ce choix explique le faible nombre de croisements (Figure II-12) à base de ce parent Keninkeni.



Tableau II-7 : Identification (Nom et code des accessions), pays d'origine, classification raciale, type botanique et caractéristiques agronomiques des 35 variétés utilisées comme lignées parentales du dispositif GCP BCNAM. Les variétés impliquées dans les populations utilisées dans le cadre de cette thèse sont indiquées dans la colonne « Thera 2017 »

Nom	Code	Pays d'origine <sup>1</sup>	Race	Type botanique	Kp <sup>3</sup>	Hauteur <sup>4</sup>	Durée du cycle <sup>5</sup>	Thera 2017
Seguetana CZ	IS 26057	MLI	G	Locale	Elevé	GT	Moyen.	Non
Jakumbe	CSM 63E	MLI	G	Locale	Faible	TM	Court	Non
Jigi Seme	CSM388	MLI	G	Locale	Elevé	GT	Long	Oui
Gadiaba Dié	IS 25916	MLI	D	Locale	Elevé	GT	Long	Oui
Konoténé	IS 25705	MLI	D	Locale	Elevé	GT	Long	Oui
Tiandougou	98 SB-F2-78	MLI	C/G	Améliorée	Moyen	TM	Moyen.	Non
G03-1-118	V33/08	MLI	C/G	Améliorée	Moyen	TM	Long	Oui
Sanga Tigi	98-BE-F5P-84	MLI	G/C	Améliorée	Moyen	TM	Moyen.	Oui
Dabittinnen	Malisor 84-7	MLI	C/G	Améliorée	Faible	TC	Court	Oui
Kalaban	00KOF5DT19	MLI	C	Locale	Faible	TC	Moyen.	Oui
Tiemantieteli	CSM 417	MLI	G	Locale	Elevé	GT	Long	Oui
Grinkan	02-SB-F4DT-275	MLI	C/G	Améliorée	Moyen	TM	Moyen.	Oui
Keninkeni	V248/08	MLI	G/C	Améliorée	Moyen	TM	Moyen.	Oui
Lata3		MLI	G/C	Améliorée	Elevé	GT	Moyen.	Oui
E36-1, Zera Zera	IS 30469	ETH	C	Locale	Faible	TC	Moyen.	Oui
Framida	IS 8744	ZAF		Locale	Faible	TC	Moyen.	Oui
Soumalemba	IS15401	CMR	G	Locale	Elevé	GT	Long	Oui
Ganga	IS 23540	ETH	C	Locale	Faible	TC	Moyen.	Oui
SC566-14	PI 533871	NGA	C/G	Améliorée	Faible	TC	Court	Oui
Fara Fara	IS 24887	NGA	G	Locale	Elevé	GT	Long	Oui
Hafijega	IS 23645	GMB	G	Locale	Elevé	GT	Long	Oui
Short Kaura	IS10699	NGA	D	Locale	Elevé	GT	Long	Oui
Gnossiconi		BFA	G	Locale	Moyen	TM	Moyen.	Non
Ribdahu		NGA	C	Locale	Elevé	TM	Long	Non
White Kaura	SSV 20043	NGA	C	NA	Elevé	TM	Long	Oui
Sambalma(4)		NGA	G	Locale	Elevé	GT	Long	Non
Lina-3		ICRISAT	G	Locale	Elevé	GT	Moyen.	Non
Soumba	Cirad406	FRA	G/C	Améliorée	Faible	TM	Moyen.	Non
Ngolofing	CSM 660	MLI	G	Locale	Elevé	GT	Moyen.	Non
Tiéblé	CSM 335	MLI	G	Locale	Elevé	GT	Long	Non
Doua-G		MLI	G	Locale	Elevé	GT	Long	Non
L2P9 (1)		ERI	C	Locale	Faible	GT	Court	Non
B35	IS 12555	ETH	D/C	Améliorée	Faible	TC	Court	Oui
IRAT 204	IS 36563	SEN	C	Locale	Faible	TC	Court	Non
BimbG	BB_G_5_34	GIN	G	Locale	Elevé	GT	Long	Oui

<sup>1</sup>Code du pays d'origine: MLI: Mali, ETH: Ethiopia, ZAF: South Africa, CMR: Cameroun, NGA: Nigeria, GMB: Gambia, BFA: Burkina Fasso, ERI: Eritrea, SEN: Senegal and GIN: Guinea

<sup>2</sup>Classification raciale : G : Guinea, D : Durra, C/G : Caudatum / Guinea, C/D : Caudatum / Durra, C : Caudatum, Durra/ Caudatum and G/C : Guinea/ Caudatum.

<sup>3</sup>Faible : coefficient de sensibilité à la photopériode (Kp) <0.4; Moyen : Kp varie entre 0.4 – 0.6 et Elevé : Kp varie entre 0.7 – 1.

<sup>4</sup>TC : taille courte < 1m ; TM : la taille moyenne varie entre 1 et 2m et GT : grande taille > 2m.

<sup>5</sup>Court : < 70 jours ; Moyen : le cycle varie entre 70 et 110 jours et Long : cycle > 110 jours. NB : ces mesures ont été réalisées en conditions de semis de précoce (Juin) dans les conditions Malienne.

#### IV.2.1.2- Variétés parentales du dispositif BCNAM

Trente-cinq variétés provenant de différents programmes de recherches agronomiques Ouest Africains, en particulier de l'Institut d'Economie Rurale du Mali (IER) et de l'Institut international de recherche sur les cultures des zones tropicales semi-arides (ICRISAT) au Mali ont été sélectionnées pour leurs caractères d'intérêts tels que : la productivité (rendement en grains et en biomasse), l'adaptation au climat (sensibilité à la photopériode), la qualité du grain, la tolérance à la sécheresse et la teneur en sucres solubles de la tige. Les origines géographiques et les caractéristiques agronomiques de ces variétés sont présentées dans le Tableau II-V7. Elles correspondent à des variétés locales (22 variétés) et améliorées (13 variétés) d'origine diverses représentant une large part de la diversité des sorghos cultivés d'Afrique. L'arbre phylogénétique (Figure II-13) de ces variétés, construit sur la base de 22 991 marqueurs SNP, illustre les proximités génétiques entre les différents parents impliqués et met en évidence une agrégation de ces derniers en fonction de leurs races botaniques.

Le dispositif final obtenu comprend 47 populations comprenant au total 4 717 familles BC1F4. Au sein de ce dispositif, 29 populations correspondant aux croisements dérivés de 20 parents donneurs et 2 parents récurrents ont été sélectionnées dans le cadre de cette thèse. Ces 29 populations ont été développées à partir de 9 lignées donneuses (colorée en cyan) communes aux deux parents récurrents (GR et KK), de 10 lignées spécifiques à GR et d'un seul donneur spécifique à KK (Figure II-12).

#### IV.2.1.3- Développement des populations BCNAM

Le développement des 29 populations BCNAM utilisées dans cette étude, a été identique à celui présenté pour la population BC06 (Figure II-5a).

### ***IV.2.2. Dispositifs expérimentaux***

L'évaluation des 29 populations BCNAM utilisées dans le cadre de cette thèse a été réalisée dans deux environnements correspondant aux saisons normales de culture 2014 et 2015.

#### IV.2.2.1- Essai 2014

Cet essai avait initialement été mis en place pour servir aux multiplications de semences des différentes familles. Sur les 29 populations BCNAM utilisées dans le cadre de cette thèse, 26 populations totalisant 3 904 familles étaient présentes au sein de dispositifs en blocs augmentés. Chaque dispositif en bloc augmenté était composé d'une population comprenant au moins 50 familles BC1F4 et trois variétés témoins (Grinkan, Keninkeni et Lata) répétées dans tous les

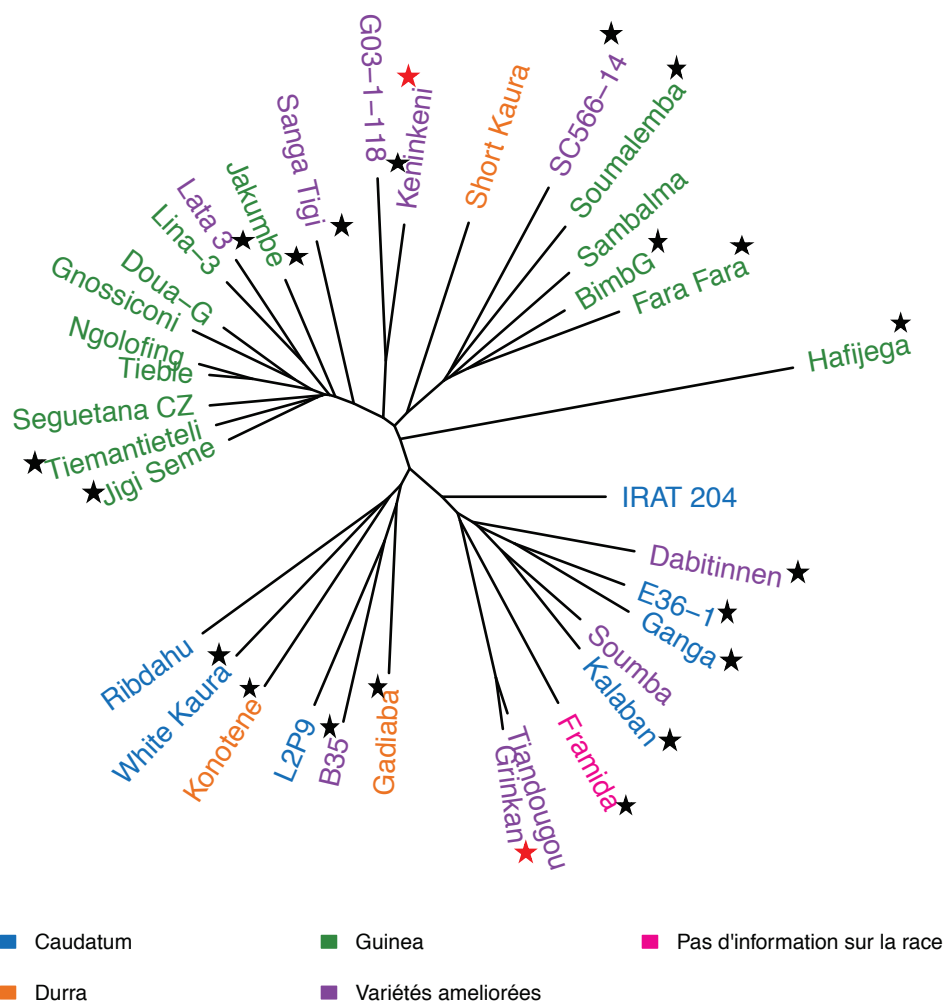


Figure II-13 : Arbre phylogénétique des 35 variétés utilisées pour le développement des populations BCNAM basé sur 22 991 marqueurs SNP analysés avec la méthode de Neighbor joining (NJ) (Saitou & Nei, 1987). Les variétés sont colorées en fonction de leurs races à l'exception des variétés améliorées. Les astérisques indiquent les 22 parents donneurs (en noir) et les 2 parents récurrents (en rouge) utilisés pour la création des 29 populations utilisées dans le cadre de la thèse.

blocs. Chaque essai comportait au total 9 blocs de 17 entrées (14 familles BC1F4 et 3 témoins). Parmi ces 9 blocs, un échantillonnage de 3 blocs successifs totalisant 51 entrées (42 familles + 3 témoins répétés 3 fois) pour chacune des 26 populations a été effectué. Le nombre total de familles échantillonnées dans les 26 populations était donc au final de 1092.

En ce qui concerne les conditions de culture (densité de plantation, écartement entre les lignes, largeur des allées entre chaque parcelle), elles sont identiques à ce qui a été présenté pour la population BC06 (cf. III.1.3).

Les 3 autres populations BCNAM (BC03, BC06 et BC10) ont été analysées dans des dispositifs différents. Les populations BC03 (152 familles BC1F4) et BC10 (148 familles BC1F4) ont été analysées dans des dispositifs « split-plot » à deux dates de semis et deux répétitions chacune. Pour les analyses menées sur l'ensemble du dispositif BCNAM considéré dans le cadre de cette thèse, 47 et 49 familles ont été échantillonnées dans les populations BC03 et BC10 respectivement, uniquement dans la deuxième date de semis (16 juillet) dans les deux blocs. Pour la dernière population, BC06, le dispositif utilisé a déjà été décrit dans le paragraphe III.1.3. Dans le cadre de l'analyse globale du dispositif BCNAM, afin de maintenir un équilibre dans la taille des différentes populations, 50 familles BC1F4 ont été sélectionnées au sein du semis de juillet.

#### ***IV.2.2.2- Essai 2015***

Durant la saison normale de culture 2015, un dispositif en blocs augmentés a été utilisé pour l'évaluation phénotypique de 25 populations BCNAM. Par rapport à l'essai de 2014, outre les 3 variétés élites (Grinkan, Keninkeni et Lata, déjà utilisées en 2014), une variété élite (P114-160) provenant d'une population développée dans le cadre du projet GCP MARS et une variété améliorée A12-79 issue d'un projet de valorisation de la diversité génétique des sorghos du Mali (projet FFEM) ont aussi été utilisées comme témoin. Dans l'essai comprenant les 25 populations BCNAM, le nombre total de familles utilisées dans chacune des populations était de 51.

Chaque population correspondait à un essai en blocs augmentés. Chaque essai comportait 66 entrées dont 51 familles randomisées dans 3 blocs au sein duquel les variétés témoin de l'essai étaient répétées (5 témoins\*3 répétitions= 15 entrées). Le nombre total de traitement de l'ensemble du dispositif était de 1650 entrées dont 1275 familles et 75 répétitions de 5 témoins pour les 25 populations.

La population BC38 a quant à elle été étudiée à deux dates de semis (Juin et Juillet) et 50 familles BC1F4 semées en juillet ont été sélectionnées pour les analyses globales du dispositif BCNAM.

En ce qui concerne les 3 populations additionnelles (BC03, BC06 et BC10), les dispositifs expérimentaux et le nombre total de traitements utilisés restaient les mêmes que ceux utilisés en 2014. Comme pour les essais de 2014, 50 familles BC1F4 ont été sélectionnées.

Pour les analyses statistiques, nous avons considéré un dispositif en bloc augmenté couvrant l'ensemble des populations y compris les populations BC03, BC06, BC10, et BC38 (l'ensemble des dispositifs en blocs augmenté de chaque population a été agrégé en un « macro » essai en blocs augmentés).

#### ***IV.2.3- Caractérisation phénotypique des 29 populations BCNAM : Phénologie et croissance***

Au sein de l'essai 2014, les 1092 familles BC1F4 provenant des 26 populations ont été caractérisées pour leur longueur de tiges (LTIG) et leur nombre total d'entre-nœuds allongés (NEN). Ces mesures ont porté sur 5 plants de chaque parcelle élémentaire au stade de maturité des grains. La longueur de la tige a été mesurée de la base de la tige principale au sommet du dernier entre-nœud situé sous le pédoncule. A partir de ces deux variables, la longueur moyenne des entre-nœuds a été calculée en divisant la longueur de la tige principale par le nombre total d'entre-nœuds allongés (LMEN).

La caractérisation phénotypique des 3 populations additionnelles a porté sur les mêmes variables phénologiques et agromorphologiques à l'exception de la population BC06, pour laquelle des caractères additionnels ont aussi été évalués (cf. III.1).

Pour l'essai 2015, 183 familles additionnelles par rapport à l'essai 2014 ont été caractérisées. En complément des mesures déjà effectuées en 2014, qui ont été répétées en 2015, la date d'émergence de la feuille drapeau (FD), les longueurs du pédoncule (LPED) et de la panicule (LPAN) ont été acquises sur les 5 premières plantes de chaque parcelle. La date d'apparition de la FD a été notée lorsque la ligule de la feuille drapeau des 5 premières plantes devenait visible. Cette mesure a permis de calculer la durée entre le semis et l'émission de la feuille drapeau (SFD) en nombre de jours. Cette méthode d'observation de la phénologie à partir de la date de FD est plus précise qu'une simple observation de la date de floraison qui dépend plus de l'observateur (incertitude pouvant aller jusqu'à 10 jours, (Vaksmann communication personnelle)).

#### IV.2.4. Analyse des données phénotypiques

Toutes les analyses statistiques y compris les analyses QTL ont été réalisées en utilisant le logiciel R (R development Core Team, 2016). Les analyses statistiques de base (moyenne et écart-type) ont été effectuées pour chaque caractère mesuré dans les deux environnements (2014 et 2015). Des calculs de moyennes ajustées ont été effectués sur les données brutes collectées sur les 1092 et 1275 familles BC1F4 évaluées dans les essais de 2014 et 2015 respectivement, en utilisant les dispositifs en blocs augmentés. Concernant les 3 autres populations (BC6, BC3, BC10, BC38), les calculs de moyennes ajustées ont été réalisées essai par essai car les dispositifs étaient différents (cf. III.7).

Un examen des données (distribution de chaque caractère et corrélations entre les caractères) a été réalisé afin de corriger les données. Dans le but de vérifier la stabilité des résultats en fonction des années, nous avons effectué des tests de corrélations entre les deux années pour les différentes variables mesurées.

Une première analyse de variance (modèle 1) a été réalisée sur chaque année d'essai afin d'estimer les effets génétiques et résiduels sur chaque caractère et dans chaque environnement sur la base des génotypes témoins répétés dans l'essai :

$$Y_{ij} = \mu + P_i + B_j + e_{ij} \quad (1) : \text{pour chaque année}$$

$$Y_{ij} = \mu + A_j + G_i' + e_{ij} \quad (2) : \text{pour les deux années}$$

Avec  $Y_{ij}$  la valeur phénotypique mesurée sur le génotype  $i$  dans le bloc  $j$  ;

$\mu$  : la valeur moyenne de la population,

$P_i$  et  $G_i'$  indiquent les effets de la variété témoin  $i$  et du génotype (famille)  $i'$  respectivement,

$B_j$  : effet du bloc  $j$ ,

$A_j$  : effet de l'année  $j$ , et

$e_{ij}$  : la résiduelle du modèle.

A partir des résultats d'Anova (1), la répétabilité des différents caractères a été estimée dans chaque environnement sur la base des valeurs obtenue sur les témoins selon la formule suivante :

$$\text{rep} = \frac{\sigma_G^2}{(\sigma_G^2 + \sigma_R^2)} \quad (3)$$

Avec  $\sigma_G^2 = 1/k$  ( $CM_g - CM_r$ ) où  $k$  représente le nombre de répétitions de chaque génotype (témoin) dans l'essai ( $k=76$ ) pour les essais de 2015 et 2014 ;  $CM_g$  et  $CM_r$  correspondent aux carrés moyens génotypiques et résiduels obtenues à partir de l'analyse d'anova et  $\sigma_R^2 = CM_r$ .

Les données phénotypiques (variables de hauteur des tiges, de nombre total et longueur moyenne d'entre-nœuds allongés) collectées sur les 1006 familles BC1F4 communes aux deux essais (2014 et 2015), ont été analysées en utilisant un modèle linéaire mixte à l'aide de la méthode « lme4 » du package R, en combinant les deux années :

$$Y_{ij} = \mu + A_j + G_i + e_{ij} \quad (4)$$

Où  $Y_{ij}$  est la valeur phénotypique mesurée sur la famille  $i$  dans l'année  $j$ ;  $\mu$  : est la moyenne générale de la population,  $A_j$  correspond à l'effet de l'année  $j$  déclaré en effet fixe,  $G_i$  : l'effet aléatoire de la famille  $i$  et  $e_{ij}$  la résiduelle du modèle.

Sur la base des résultats du modèle (4), la variabilité environnementale a été estimée et les effets génétiques (Best linear unbiased predictors : BLUP) de chaque famille ont été extraits pour chacune des variables mesurées dans les deux années d'expérimentation.

L'héritabilité au sens large ( $h_{SL}^2$ ) des familles a été également estimée en utilisant les résultats du modèle 2 selon la formule ci-dessous :

$$h_{SL}^2 = \frac{\sigma_G^2}{(\sigma_G^2 + \sigma_R^2)} \quad (5)$$

Où  $\sigma_G^2$  correspond à la variance génotypique et  $\sigma_R^2$  la variance résiduelle.

Les corrélations intra-caractères (inter-années) et inter-caractères (intra-année) ont été calculées sur la base des valeurs moyennes ajustées obtenues sur chaque famille dans chaque essai suite aux analyses en bloc augmenté.

#### ***IV.2.5. Génotypage des populations BCNAM***

Les populations BCNAM ont été génotypées en suivant la stratégie décrite pour la population BC06 (cf. III.1.6). En moyenne, pour chaque population, 11 200 marqueurs SNP ont été générés. Au total, pour l'ensemble des 29 populations BCNAM, 51 545 SNP ont été identifiés.

#### ***IV.2.6. Identification des zones chromosomiques impliquées dans le contrôle de la durée du cycle et de la croissance aérienne***

Pour analyser le déterminisme génétique des caractères relatifs à la durée entre le semis et la floraison et la production de biomasse chez le sorgho, nous avons réalisé des études d'associations (Genome Wide Association Study : GWAS) entre les 51 545 marqueurs SNP développées sur l'ensemble des populations BCNAM et la variation phénotypique de ces

caractères. Ces recherches d'association ont été effectuées sur les valeurs moyennes phénotypiques ajustées obtenue sur chacune des familles BC1F4 et pour chaque essai. Par ailleurs, les valeurs génétiques (Blup) obtenues par combinaison des deux essais (2014 et 2015), uniquement sur les familles communes aux deux essais ont été également utilisées.

L'identification des SNP associés à la variation des différents caractères mesurés a été entreprise via l'utilisation d'un modèle linéaire mixte incorporé dans le package NAM développé par (Xavier et al., 2015) et implémenté dans le logiciel d'analyse statistique R. Ce package, conçu initialement pour les études d'association dans les panels de type NAM (Nested Association Mapping), peut aussi être utilisé dans le cas de populations multi-parentales de type backcross NAM tel que la nôtre.

Un des problèmes majeurs rencontrés dans les études d'associations au sein des panels de diversité est la forte structuration des populations. Cependant, grâce à l'utilisation des modèles linéaires mixtes (MLM), qui prennent en compte la structure des populations (Yu et al., 2006) ce problème de structuration des populations a été résolu.

Dans le modèle MLM, la structuration en groupe de la population est définie en tant que covariable et l'apparement comme une matrice de variance-covariance des effets génétiques (méthode Q + K). La formule ci-dessous illustre le modèle de base proposé par Yu et al. (2006) :

$$y = \mu + Qv + X\beta + Zu + e$$

Avec  $y$  le phénotype,

$\mu$  la moyenne générale,

$Qv$  le vecteur des effets structure,

$X\beta$  le vecteur des SNP tous les deux considérés comme des effets fixes,

$Zu$  l'effet polygénique et  $e$  la résiduelle (effets aléatoires).

La structure de la population a été prise en compte comme effet fixe en utilisant l'option « Fam » de regroupement de sous-population correspondant à chaque population/famille BC1F4. L'argument « fixed » de la fonction gwas/NAM offre la possibilité de traiter les marqueurs en effet fixes et par conséquent, évalués à travers le test statistique de Wald sous l'hypothèse nulle selon laquelle l'effet du marqueur est nul. Pour déclarer la significativité du QTL, nous avons appliqué un taux de faux positif de 1% correspondant à une valeur de  $-\log_{10}$  p-value de 2.

Afin de définir les QTL uniques, les SNP proches physiquement (distance < à 50 Kb) présentant des associations significatives avec les caractères ont été considérés comme correspondant à un QTL unique.



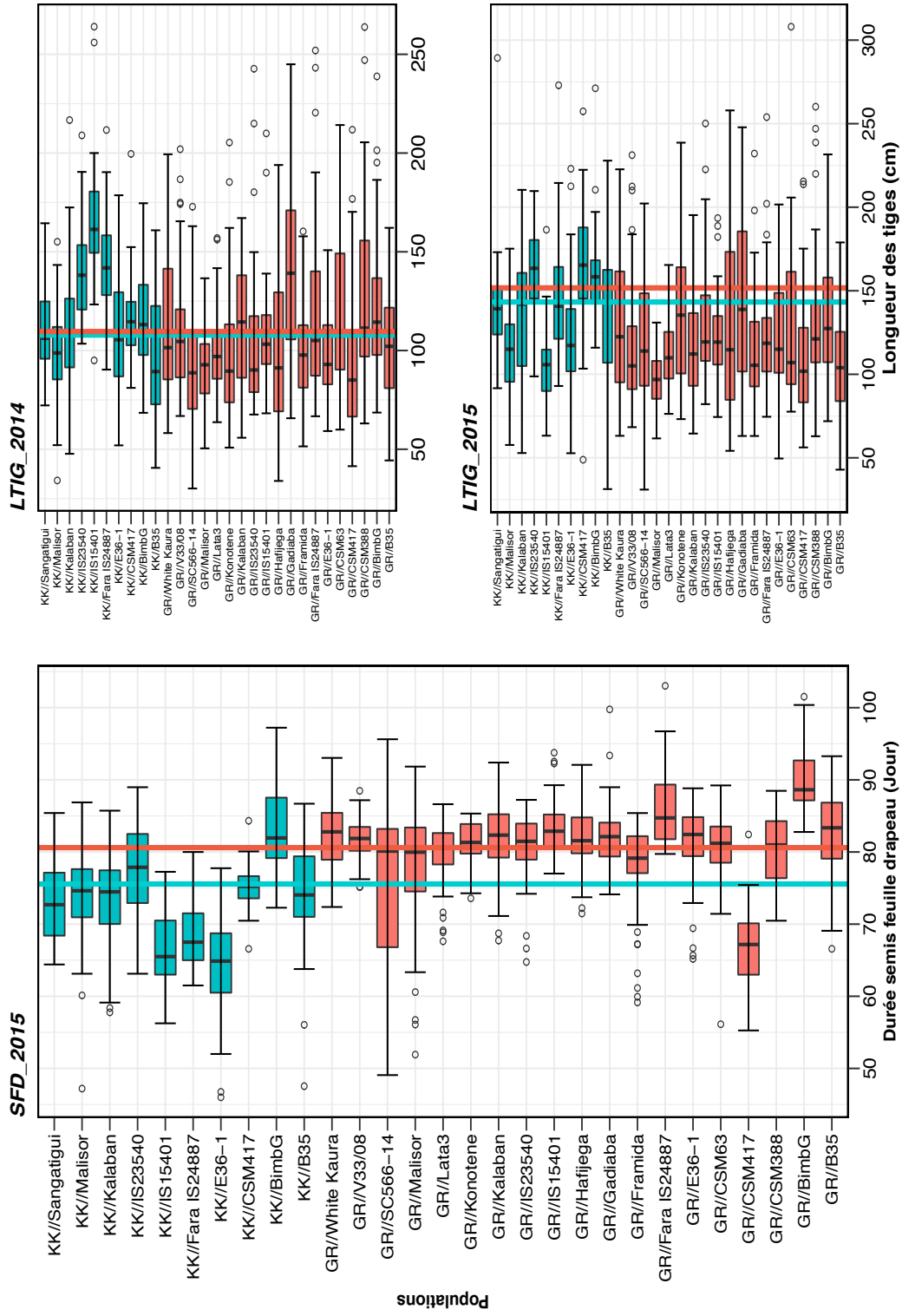


Figure II-14 : Variabilité inter et intra population des 29 populations BCNAM pour la durée « semis feuille drapeau » (SFD) dans l'essai de saison normale de culture 2015 (à gauche) et la longueur des tiges (LTIG) dans les environnements 2014 et 2015 (à droite). Les barres verticales représentent les valeurs moyennes des parents récurrents Kenenkeni (couleur cyan) et Grinkan (couleur orange).

### **IV. 3. Résultats**

#### ***IV. 3.1. Analyse phénotypique des 29 populations BCNAM***

##### **IV. 3.1.1- Variabilité inter et intra population des caractères mesurés et calcul d'héritabilité**

De manière générale, pour chacune des variables mesurées dans les deux environnements de culture (2014 et 2015), il existe une grande variabilité phénotypique entre familles issues d'une même population mais également entre populations développées à partir d'un même parent récurrent (Grinkan ou Keninkeni).

La Figure II-14 illustre ce résultat pour la variable de durée du cycle entre le semis et la feuille drapeau (SFD) mesurée uniquement durant la saison normale de culture 2015 et la longueur des tiges (LTIG) mesurée dans les deux essais (2014 et 2015).

Les populations à base du parent Grinkan ont un cycle plus long que celles développées à partir du parent récurrent Keninkeni à l'exception de deux populations : KK//BimbG (BC30), dont les descendances sont très similaires au parent donneur (BimbG), qui est très tardif (cycle long), et la population GR//CSM417 (BC38) au sein de laquelle les familles ont un cycle aussi court que celles à base de Keninkeni.

On observe une structuration des populations par rapport aux parents récurrents GR et KK ; et les moyennes générales des familles issues des populations à base de GR dépassent celles du parent récurrent GR (hormis les populations GR//Framida et GR//CSM417) contrairement aux populations développées à partir du parent KK pour lesquelles les moyennes générales de durée du cycle des familles sont plus courtes que celle du parent KK (à l'exception des populations KK//IS23540 et KK//BimbG) (Figure II-14).

Globalement, les durées du cycle SFD observées entre les 1466 familles relatives aux 29 populations, varient de 50 à 105 jours en saison normale de culture.

En outre, pour les caractères de longueur des tiges (LTIG) et ses composantes (NEN et LMEN), l'analyse de variabilité inter et intra population a révélé une stabilité environnementale de ces caractères au niveau des différentes familles en fonctions des années.

La corrélation observée ( $r=0.73$ ,  $p\text{-value}=0.0001$ ) entre les hauteurs de tiges pour les deux années d'expérimentation confirme cette tendance (Figure II-15).

Dans l'environnement 2014, pour toutes les variables mesurées (LTIG, NEN et LMEN) les valeurs moyennes générales (Tableau II-8) de l'ensemble des familles BC1F4 relatives aux 29

Tableau II-8 : Moyenne, coefficient de variation phénotypique (CVP%) et génotypique (CVG%), l'analyse de variance ANOVA et héritabilités des différents caractères agromorphologiques

Variables	2014				2015				2014-2015					
	GR		Familles		GR		Familles		Familles		Analyse de variance			
	Moy.	KK	Moy.	rep.	Moy.	rep.	Moy.	rep.	Moy.	CVP%	CVG%	ANNEE	Famille	h <sup>2</sup>
SFD	-	-	-	-	81.11 ±2.6	75.55 ±2.4	77.7 ±8.6	0.77	-	-	-	-	-	-
LTIG	109 ±18.7	108 ±15.1	112 ±36.9	0.81	151.74 ±28.9	143.25 ±14.3	127 ±39.7	0.79	120 ± 38.8	31.38	26.65	***	***	0.84
NBETN	12.52 ±1.6	12 ±1.1	12.8 ±2.4	0.79	15.37 ±1.4	14.96 ±1.1	44.4 ±11.5	0.38	14.2 ±2.7	28.84	26.37	***	***	0.61
LgMoyETN	8.83 ±1.5	9 ±1.1	8.82 ±2.5	0.54	9.93 ±1.1	9.58 ±1	32.3 ±6.2	0.5	8.55 ±2.5	16.3	10.7	***	***	0.91
LPED	-	-	-	-	39.48 ±4.3	62.21 ±7,5	15.2 ±2.5	0.79	-	-	-	-	-	-
LPAN	-	-	-	-	30.29 ±3.3	41.29 ±3.7	8.44 ±2.5	0.57	-	-	-	-	-	-

populations BCNAM se rapprochent de celles des deux parents récurrents (GR et KK) à l'exception de la hauteur des tiges des familles (plus élevée de 4cm par rapport à GR et KK). Par contre en 2015, pour les mêmes caractères analysés, les valeurs phénotypiques moyennes de LTIG des deux parents dépassent celles des familles de 16 à 24 cm respectivement pour KK et GR tandis que pour les caractères de NEN et de LMEN les familles sont similaires aux deux parents récurrents. En plus de ces 3 variables, la durée du cycle semis feuille drapeau (SFD) ainsi que les longueurs du pédoncule et de la panicule ont également été observées en 2015 et ont montrées des différences moyennes phénotypiques entre les familles et les parents GR et KK. De manière générale dans cet environnement, les familles sont plus précoces que le parent récurrent GR (4 jours) et légèrement plus tardive que KK (de 2 jours). Les valeurs moyennes phénotypiques des longueurs du pédoncule et de la panicule des 1641 familles BC1F4 sont nettement inférieures à celles à deux parents récurrents entre (28-50 cm pour GR) et entre (22-33 cm pour KK).

Les valeurs de répétabilité sont élevées (rep. variant entre 0.51 et 0.81) sur tous les caractères analysés pour les deux environnements de culture (Tableau II-8) à l'exception de la longueur du pédoncule (rep. =0.38). Pour l'ensemble des 3 variables évaluées pour les deux années d'expérimentations, de fortes valeurs de répétabilité ont été obtenue ( $h^2_{SL} > 0.6$ ). Ces résultats indiquent un déterminisme génétique fort de ces caractères avec des effets environnementaux faibles.

Les coefficients de variation phénotypique ( $CV_P\%$ ) et génotypique ( $CV_G\%$ ) sont élevés pour les caractères LTIG ( $CV_P=31\%$  et  $CV_G=26\%$ ) et NEN ( $CV_P=28$  et  $CV_G=26\%$ ) contrairement à la variable de LMEN pour laquelle les  $CV_P\%$  et  $CV_G\%$  sont plus faibles (16 et 10%).

Par ailleurs, l'analyse combinée de variance sur les données phénotypiques obtenues en 2014 et 2015 a montré des effets génétiques (effet famille) et environnementaux (effet ANNEE) hautement significatifs ( $P < 0.001$ ) sur l'ensemble des caractères mesurés durant les saisons normales de culture 2014 et 2015.

#### IV.3.1.2- Corrélations phénotypiques entre les caractères de phénologie et de croissance des familles BCNAM en fonction des années

L'analyse des corrélations entre les caractères observés dans chaque environnement ont révélé de forte corrélations positives ( $r$  entre 0.36 et 0.83 ;  $P < 0.0001$ ) entre la hauteur des tiges et ses composantes (NEN et LMEN) (Figure II-15). La même tendance a été observée dans les deux essais (2014 et 2015).

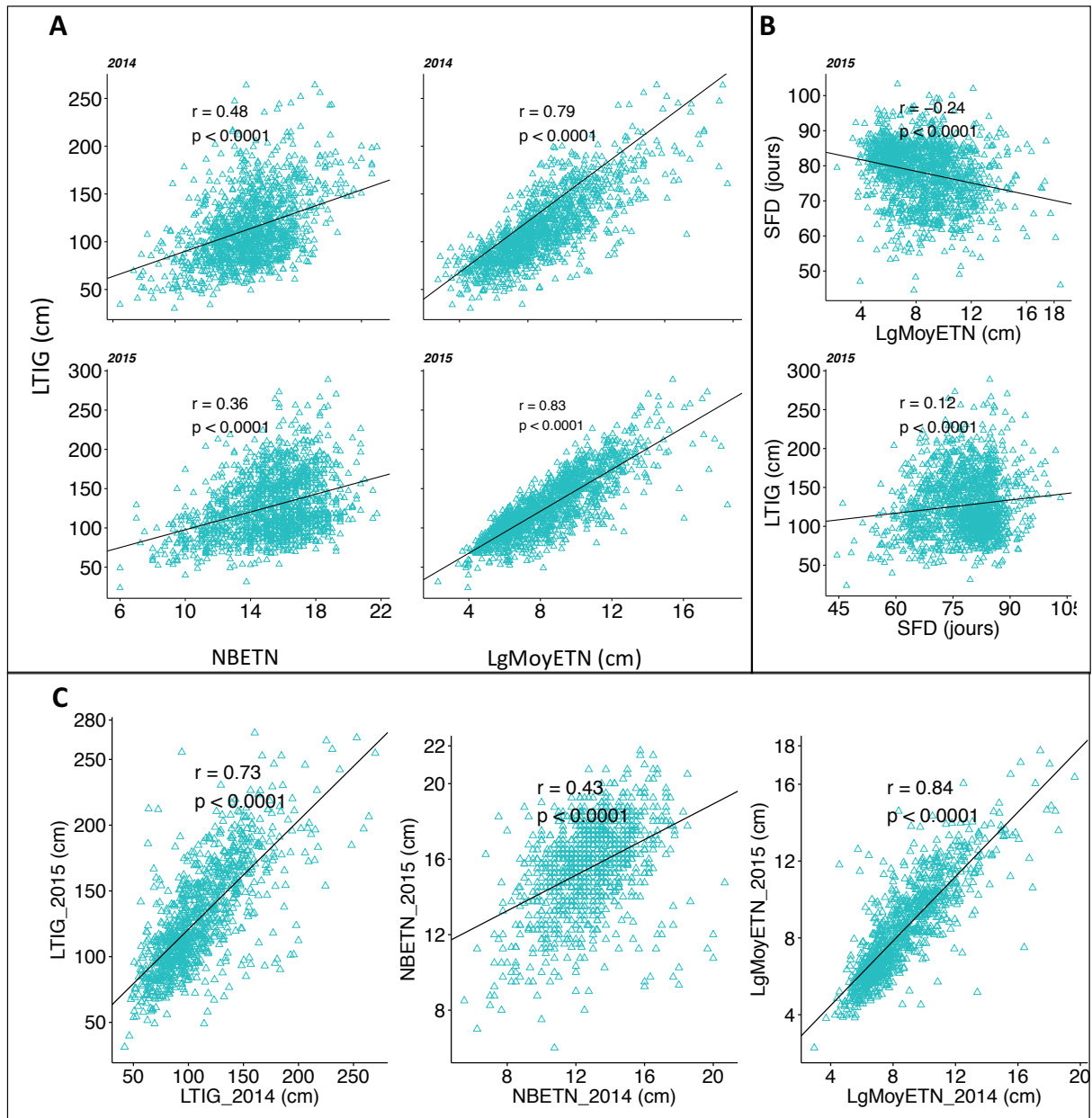


Figure II-15 : Corrélations phénotypiques entre les caractères agromorphologiques mesurés sur les différentes familles BC1F4 des 29 populations BCNAM dans chacun des deux environnements. A : corrélation entre la hauteur des tiges (LTIG) et ses composantes (nombre total d'entre-nœuds allongés (NEN) et la longueur moyenne des d'entre-nœuds (LMEN)) pour les années 2014 et 2015. B : corrélation entre la durée du cycle semis feuille drapeau (SFD) avec LTIG et LgMoyETN dans l'environnement 2015. C : les corrélations entre essai (année) pour un même caractère donné (LTIG et LgMoyETN).

Par ailleurs, la hauteur des tiges est faiblement corrélée avec la durée du cycle semis feuille drapeau (SFD) ( $r=0.12$  ;  $P<0.0001$ ) mesurée uniquement dans l'essai de 2015. Par contre, SFD est négativement corrélé à la longueur moyenne des entre-nœuds ( $r= -0.24$  ;  $P<0.0001$ ), indiquant qu'une augmentation de la durée du cycle entraîne une accumulation du nombre d'entre-nœuds allongés mais de taille courte. Les corrélations phénotypiques entre deux années sur les valeurs moyennes ajustées pour chacune des 3 variables communes aux deux années d'expérimentations, sont élevées et significatives ( $r>0.70$  ;  $P<0.0001$ ).

#### ***IV.3.2-Analyse d'association « phénotype-génotype »***

Les résultats des analyses de cartographie d'association réalisées sur l'ensemble des caractères dans les deux environnements (2014 et 2015) en utilisant la méthode de régression linéaire multiple (MLM) sont indiqués dans la Figure II-16 (et en Annexe 4). A partir de ces résultats, une synthèse du nombre de régions génomiques uniques ainsi que des QTL majeurs détectés pour les différents caractères a été réalisée (Figure II-17).

Cent vingt-quatre régions génomiques uniques ont été détectées sur la base des deux environnements analysés. Parmi celles-ci, 12 régions uniques présentent des valeurs de LOD très élevés (entre 10.54 et 39.91), expliquant respectivement entre 3.15 et 6.45 de variation de SFD, entre 112 et 191 % pour LTIG, entre 0.10 et 0.43 % de variation de NEN, entre 0.31 et 1.38 % de variation de LMEN et 4.54% de variation phénotypique expliquée par LPED ont été mises en évidence.

##### **IV.3.2.1- Des régions génomiques à effet majeur sont associées à la variation de la durée du cycle et de la hauteur des tiges et de ses composantes**

Pour la durée du cycle « semis -feuille drapeau » (SFD), variable mesurée uniquement dans l'environnement de culture 2015, 10 régions uniques ont été détectées dont 3 à effet majeur localisées sur les chromosomes 3 (51 Mb) et 6 (0.6 et 38 Mb). Ces QTL majeurs, avec des valeurs de LOD variant entre 13.3 et 39.91, expliquent entre 3.15 et 6.45 % de variation phénotypique de ce caractère. D'autres associations à effets faibles ont été détectées sur les chromosomes 1, 2, 3, 4, 7 et 10, lesquelles contribuent entre 1.92 et 3.33 % de la variance phénotypique de ce caractère.

Vingt-quatre zones chromosomiques contribuent à la variation de la longueur des tiges (LTIG) dans les différents environnements. Parmi ces 24 zones, deux régions localisées sur le chr 7 à 3 Mb de distance (55 Mb avec  $LOD=36.6$  et 58 Mb avec  $LOD=22.78$ ), expliquent plus de 100% de variation phénotypique de la longueur des tiges dans les environnements de culture 2014

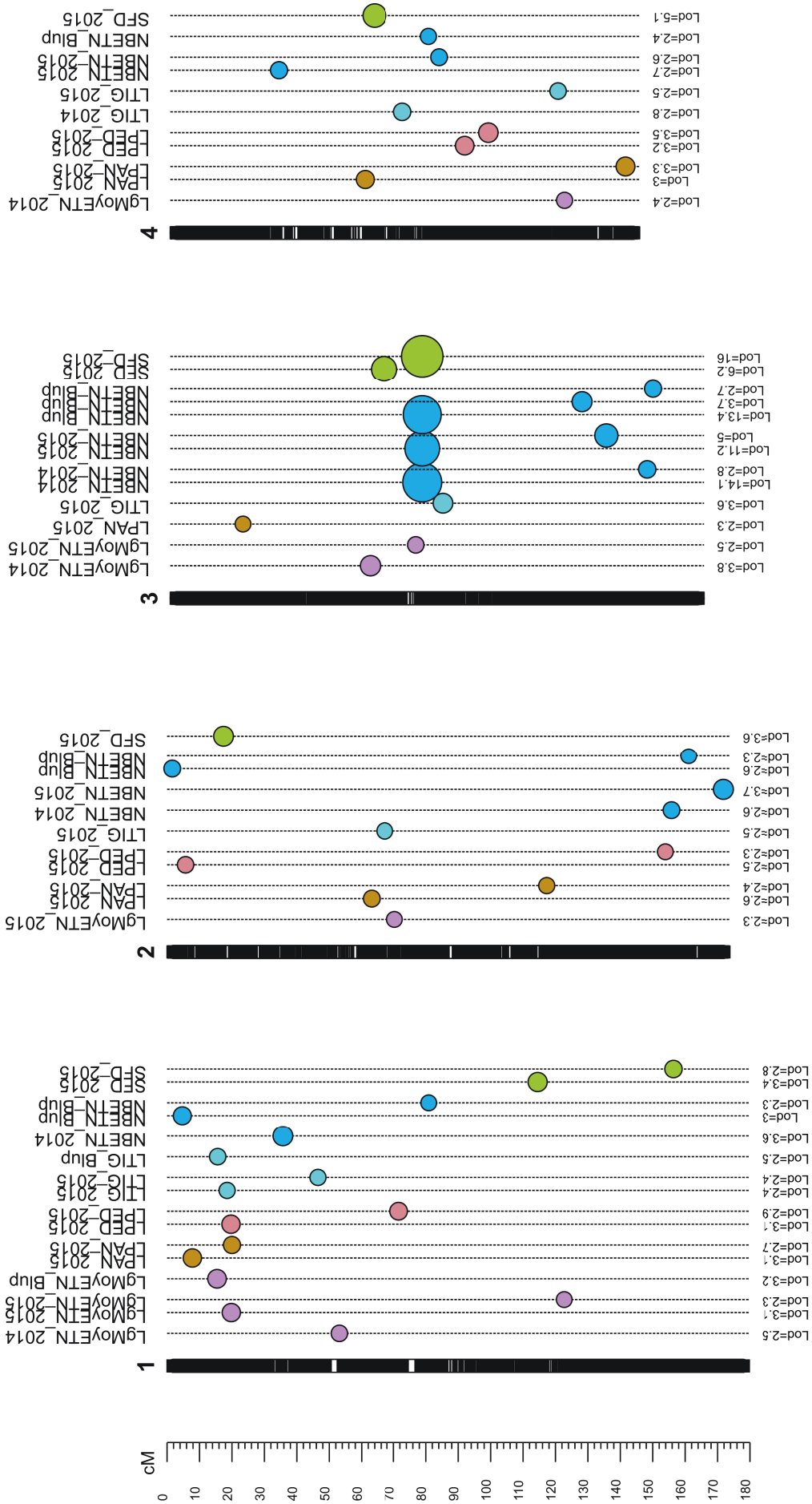


Figure II-16 : Représentation graphique des QTL détectés dans les deux années de saison normale de culture (2014 et 2015) et de l'analyse blup (2014-2015) pour l'ensemble des variables analysées. Chaque couleur représente un caractère. La taille des cercles indique la valeur de «  $-\log(p\text{-value})$  ».

( $R^2=191.56$ ) et 2015 ( $R^2=112.95$ ). Ces deux régions présentent des effets alléliques opposés. De plus, le QTL majeur localisé sur le Chr3 à 55Mb a été également détecté sur la base des blup interannuels ( $R^2=125.47$ ).

Sur le même chr7, d'autres associations significatives ont été détectées pour la longueur de la tige (LTIG) dans les deux environnements. Ces SNPs contribuent à une forte part de variance phénotypique de ce caractère ( $R^2=27.62$  et  $57.21\%$ ).

La variation de la longueur des tiges est expliquée par d'autres régions génomiques significatives avec des LOD faibles (2.35 et 5.73 %) par contre expliquant une grande part de variation phénotypique de ce caractère ( $R^2$  variant entre 5.95 et 88.03).

Pour le nombre total d'entre-nœuds allongé (NEN), au total, 34 zones chromosomiques uniques ont été identifiées. Les plus fortes associations significatives avec le NEN, ont été trouvées sur les chr3 (51 Mb) et chr6 (40 et 0.6 Mb). La zone chromosomique du chr3 explique entre 0.10 et 0.28% de variation phénotypique de NEN (LOD=11.23 et 14.2), alors que celles du chr 6 expliquent respectivement 0.28 et 0.48 % de variance phénotypique de NEN dans les deux environnements de culture (LOD=13.35 et 21.72). Les 31 zones génomiques restantes ont des valeurs de LOD très faibles (entre 2.57 et 8.83 % de  $R^2$ ) et sont faiblement associées à la variation phénotypique de NEN ( $R^2=0.01$  et 0.34).

La variation phénotypique de la longueur moyenne des entre-nœuds (LMEN) est expliquée par 25 régions génomiques uniques dont 3 à effets majeurs localisées sur le chr7. Les deux premiers QTL majeurs (55 Mb) de cette région chromosomique se trouve à seulement 256 Kb de distance avec des associations très fortes (LOD= 38.14 et 23.82 %) et expliquent respectivement 1.38 et 0.31 % de variance phénotypique associée à ce caractère ; tandis que le 3<sup>ème</sup> QTL majeur (LOD=21.22 et 22.85) localisé à 58 Mb contribue seulement à 0.59 et 0.39 % de  $R^2$  dans les environnements de saison normales de culture 2014 et 2015 respectivement.

A quelques mégabases de ces QTL majeurs, se trouve d'autres associations significatives avec la LMEN, par contre avec des LOD faibles (3.79 et 5.03 % de variation phénotypique) pour les environnements 2015 et 2014-2015.

Les vingt et deux zones génomiques restantes expliquent seulement une part assez faible de variation de LMEN ( $R^2$  variant entre 0.08 et 0.37 %).



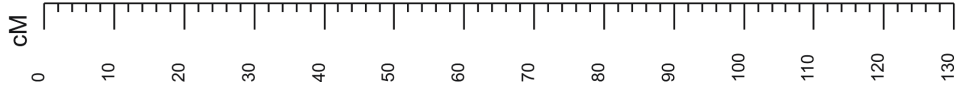
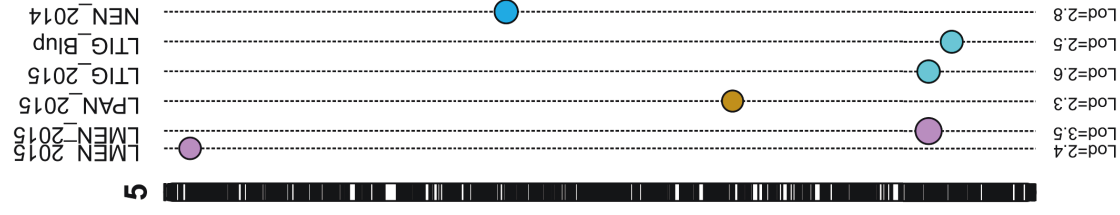
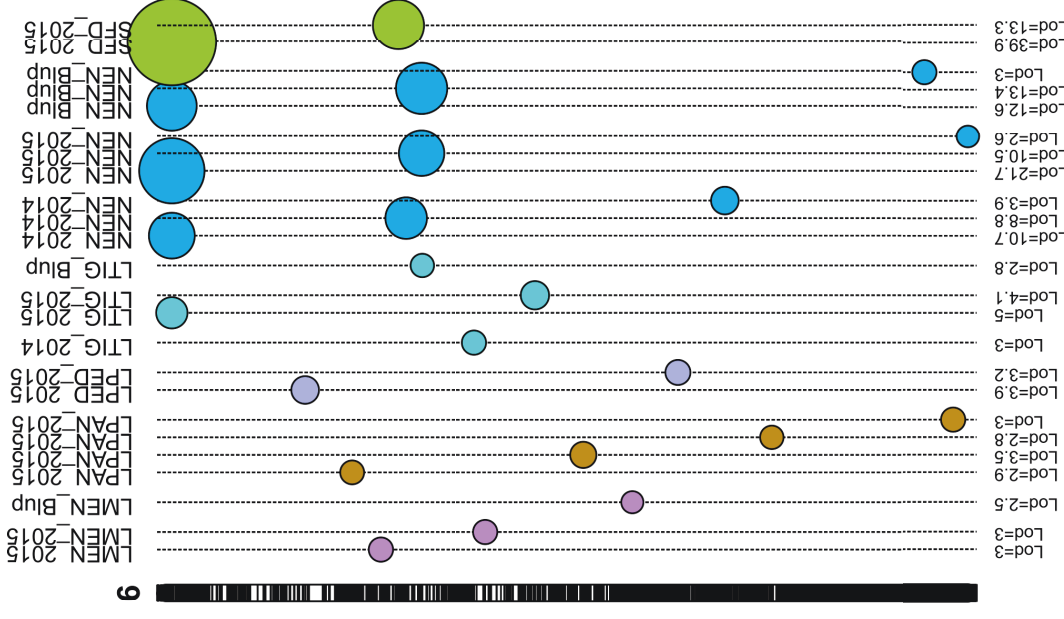
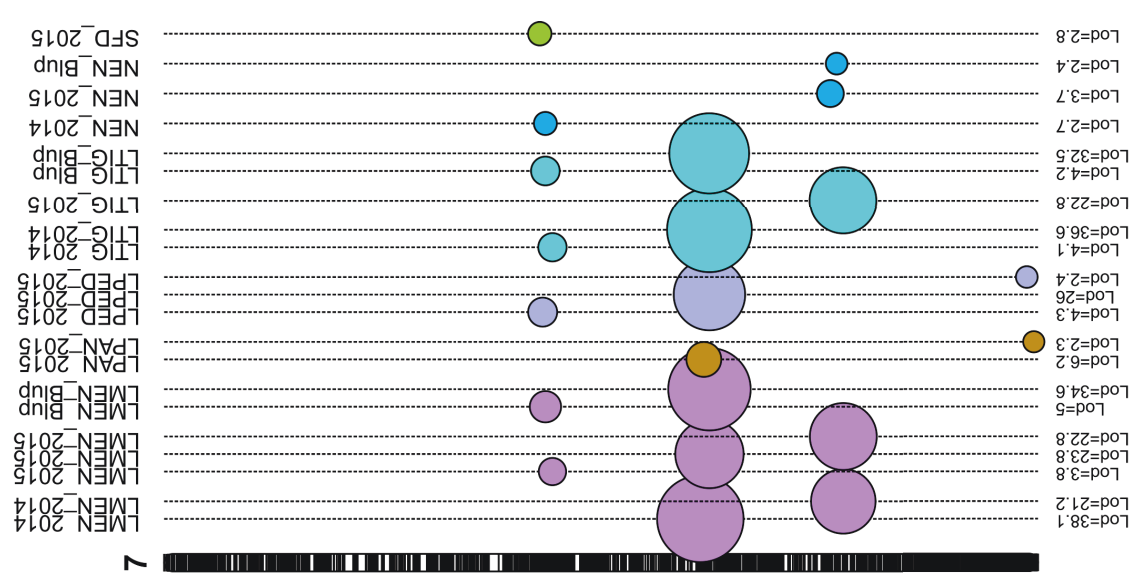


Figure II-16 : suite

#### IV.3.2.2- Les longueurs du pédoncule et de la panicule sont majoritairement contrôlées par des régions génomiques à effet faibles

Au total, 14 régions génomiques sont associées à la variation de la longueur du pédoncule (LPED) évaluée uniquement dans l'environnement de saison normale de culture 2015.

Pour ce caractère, une seule région génomique du chr7 (55 Mb) présente la plus forte association significative (LOD= 26) et contribue à la plus grande part de variation phénotypique de LPED ( $R^2= 4.54\%$ ) dans l'environnement 2015. En plus, deux autres zones génomiques situées à 63 et 92 Mb de distance du QTL majeur sont également associées à la variation de LPED.

La longueur de la panicule (LPAN) est contrôlée par 18 zones génomiques avec des valeurs de LOD relativement faibles. Entre 0.17 et 1.24 % de la variation phénotypique de LPAN sont expliqués par ces QTL. Par contre, aucun QTL à effet majeur n'a été détecté pour ce caractère.

#### IV.3.2.3- Stabilité des régions chromosomiques détectées en fonction des années

Seules trois variables (LTIG, NEN et LMEN) ont été analysées sur les 2 années de culture (2014 et 2015). Pour ces 3 variables, 85 zones chromosomiques ont été détectées. Dix-neuf zones sont spécifiques à l'environnement 2014, 29 à l'environnement 2015 et 22 ont été détectées sur la base de l'analyse des BLUP (2014-2015). Les quinze zones génomiques restantes sont communes aux deux années (3 zones), communes à 2014 et blup (4 zones), à 2015 et blup (6 zones) où aux trois environnements (2 zones) (Figure II-18).

Sur les 85 QTL détectés, 24 sont impliqués dans le contrôle génétique de la longueur des tiges dont 4 zones spécifiques à l'environnement 2014, 12 zones sont spécifiques à l'environnement 2015, 7 zones sont spécifiques au blup, et seulement 2 zones commune à 2014 et blup. Cette zone génomique du chr7 (55Mb), commune à 2014 et blup est spécifique à la hauteur des plantes et implique des loci avec des associations très fortes (LOD=32).

Par ailleurs, aucune zone génomique commune n'a été détectée entre 2014 et 2015, entre 2015 et blup puis entre les 3 environnements.

En ce qui concerne les QTL relatifs au nombre total d'entre-nœuds, sur les 34 régions génomiques détectées, 10 sont spécifiques à 2014, 7 à 2015, 12 au blup, 3 zones communes entre 2015 et blup, contre seulement 2 zones génomiques communes aux deux années qui sont localisées sur les chromosomes 3 (51 Mb) et 6 (0.6 Mb). Ces deux régions chromosomiques stables, impliquent chacune des loci à effet majeurs associées aux variations de NEN. Cette zone est proche du gène majeur de maturité Ma1 localisé sur le chr6.

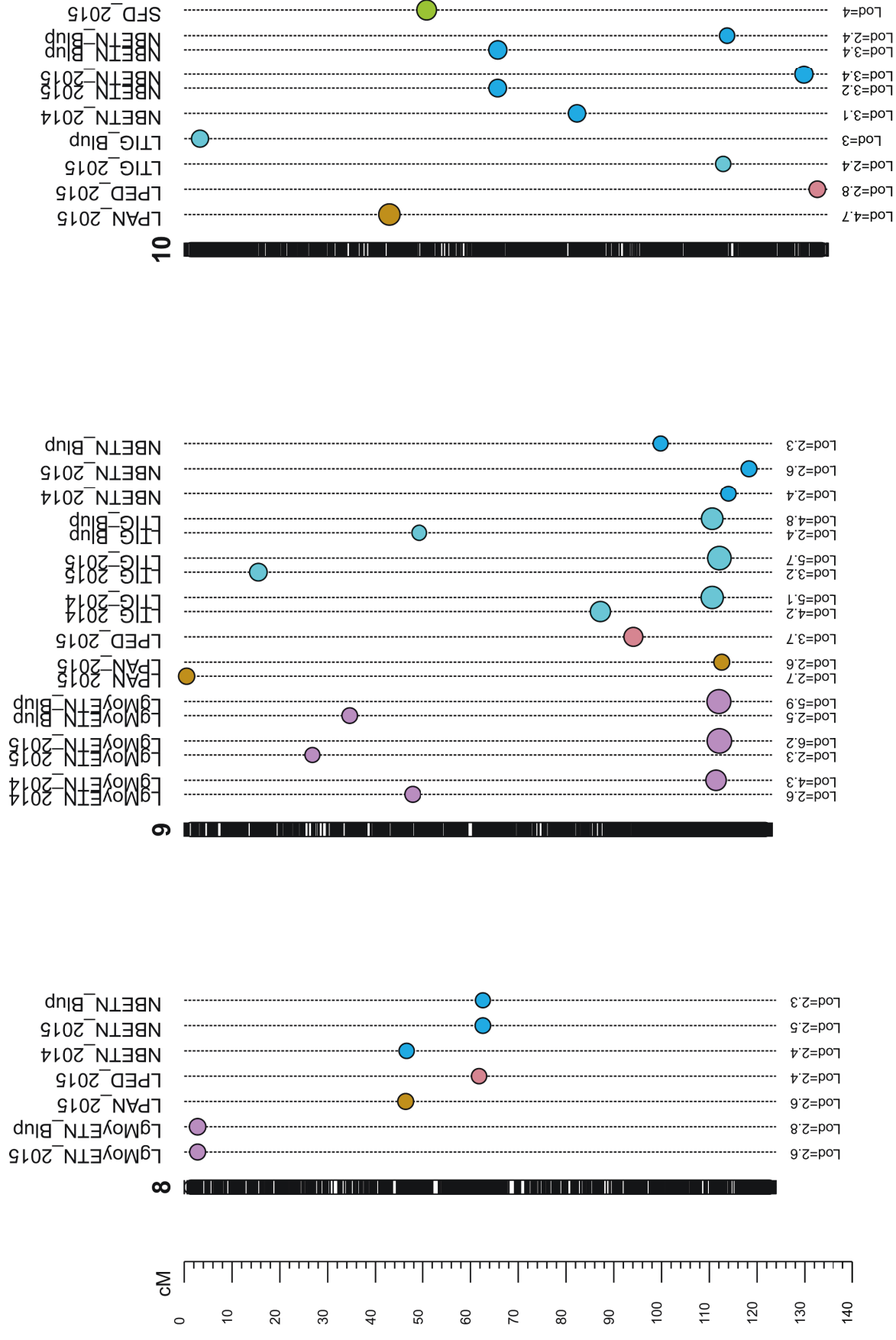


Figure II-16 : suite

Aucune zone génomique commune n'a été détectée entre 2014 et 2015 puis entre 2015 et le blup.

Parmi les 24 zones génomiques détectées pour la longueur moyenne des entre-nœuds, 6 sont spécifiques à l'environnement 2014, 11 spécifiques à 2015, 5 au blup, 2 entre 2015 et blup et une seule région génomique du chr7 est commune aux deux années. Cette zone chromosomique à effet majeur localisée à 58 Mb est proche de l'emplacement du gène de nanisme *dw3* localisé sur le chr7.

#### IV.3.2.4- Une même région génomique associée à la variation de la hauteur des tiges et de ses composantes

Dans cette étude, des colocalisations ont été mises en évidence (Figure II-16). Le QTL majeur de durée du cycle semis feuille drapeau du chr6 (618 Kb) colocalise avec la longueur des tiges (environnement 2015) et le nombre total d'entre-nœuds dans les deux environnements. D'autres colocalisations ont été trouvées entre LTIG et NEN dans une autre zone génomique du chromosome 6 localisée à 40Mb.

La zone génomique du chr3 (51Mb) est associée à la fois à la variation de la durée du cycle semis-feuille drapeau mais aussi à celle de NEN. Une autre région génomique à 58Mb sur le chr5, est associée à la fois à la variation de la LMEN et du NEN. Des colocalisations multiples entre les QTL majeurs de nanisme (LMEN), et hauteur (LTIG) avec la longueur du pédoncule (LPED) ont été détectées à 55 et 58 Mb sur le chromosome 7 (Figure II-16, Annexe 4).

### IV.4. Discussion :

#### *IV.4.1- Comparaison des régions chromosomiques identifiées et la bibliographie*

Dans cette étude, plusieurs zones génomiques associées à la variation de la durée du cycle « semis floraison » ont été mises en évidence sur la base de l'essai 2015. A l'exception du QTL situé à 156.38 cM sur le chr1, toutes les associations trouvées pour la SFD ont été détectées dans les études antérieures (Bangbol Sangma, 2013, Crasta et al., 1999, El Mannai et al., 2011, Mace et al., 2013a, Mace and Jordan, 2010, Parh, 2005, Srinivas et al., 2009). Sur les 10 associations trouvées pour la durée du cycle semis floraison, 3 régions génomiques à effet majeur étaient significativement associées à la variation de ce caractère. Le premier QTL majeur localisé à 51Mb (77.94 cM) sur le chr3, est le même que le QTL de durée du cycle semis floraison détectée au sein de la population biparentale « BC06 » analysée dans la section

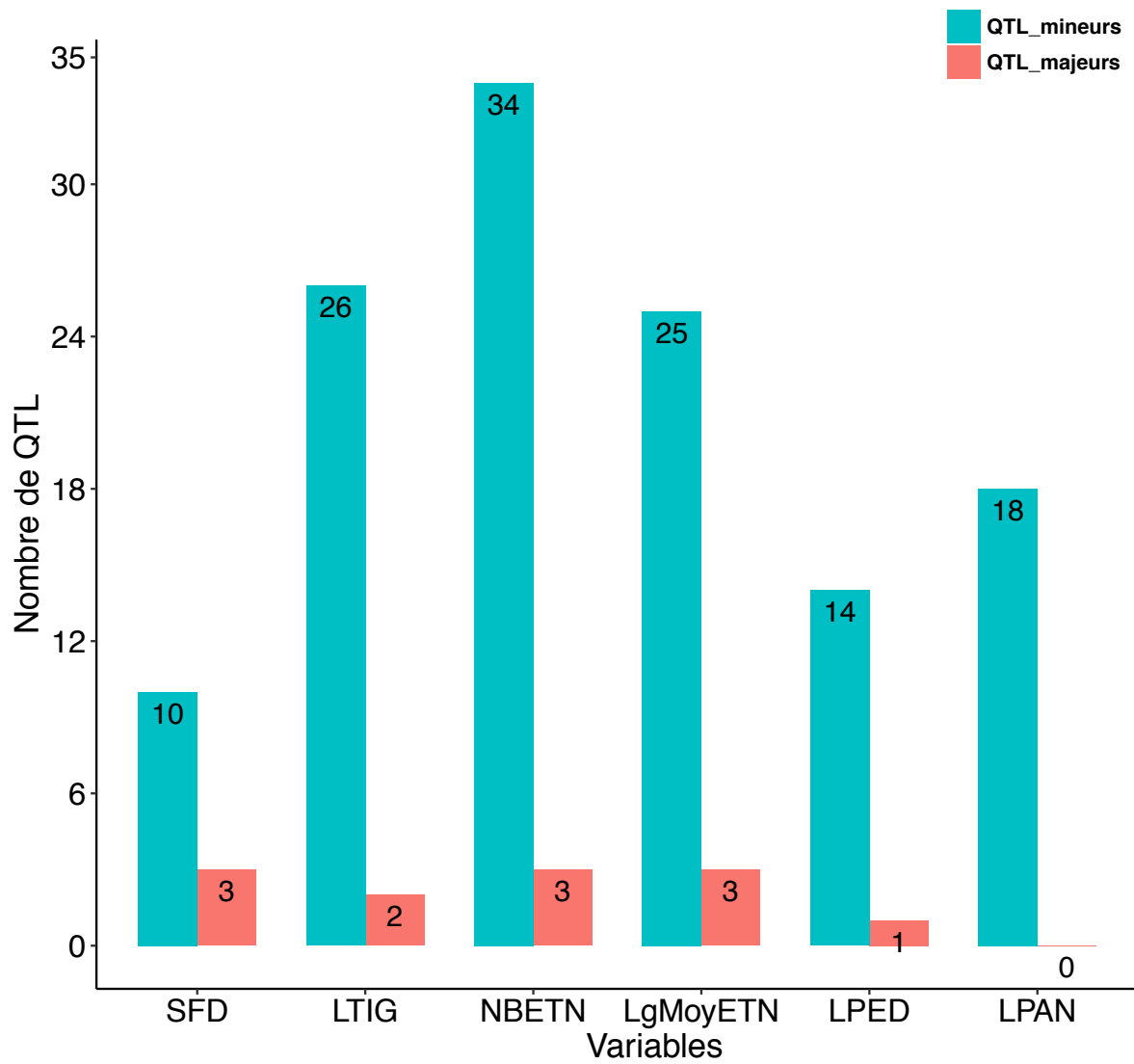


Figure II-17 : Nombres de QTL à effet mineur (couleur cyan) et majeur ( $r^2_p > 10\%$  : couleur orange) détectés pour les différents caractères évalués sur les familles BC1F4 dans les deux environnements de saison de culture (2014 et 2015).

précédente. Les deux autres zones chromosomiques à effets majeurs, détectés à 38 et 0.618 Mb sur le chr6 sont communes celles révélées dans les études précédentes (Bangbol Sangma, 2013, Mace et al., 2013a) ; mais également à celle identifiée dans la population BC06.

Vingt-six zones chromosomiques dont deux à effets majeurs sont associées à la variation de la longueur des tiges dans les deux environnements de culture (2014 et 2015) et dans l'analyse blup. La majorité de ces associations (70 %) étaient spécifiques à cette étude. Les deux QTL à effets majeurs localisés sur le chr7 à 3 Mb de distance (55 et 58 Mb) ont déjà été détectés dans des études antérieures. En outre, il est intéressant de noter que ces deux QTL majeurs de hauteur des tiges détectés sur le chr7 dans les environnements de culture 2014 et 2015 sont en oppositions.

Le premier QTL majeur (Chr7, 55 Mb) correspond à qHT7.1 un QTL majeur de hauteur identifié par Li et al. (2015), et le second au gène de nanisme *dw3* déjà connu chez le sorgho, identifié par Karper (1932) et, séquencé et cloné par Multani et al. (2003). Le même gène *dw3* a été identifié comme étant le « Sb07g023730 » localisé à 58Mb correspondant à 105 cM au niveau de la carte consensus développée par Mace et al. (2009).

Par ailleurs, ces résultats sont en parfait accord avec ceux obtenus dans l'étude réalisée par Li et al. (2015).

Les 30% des régions génomiques restantes situées sur les chr1 (56 et 72 Mb), chr3 (53 Mb), chr6 (0.618, 40, et 42 Mb), et le chr07 (14, et 98 Mb) sont très proches des associations détectées dans de nombreuses études (Ritter et al., 2008, Kebede et al., 2001, Madhusudhana and Patil, 2013, Brown et al., 2008, Shiringani et al., 2010).

Trois zones chromosomiques avec des effets majeurs sur la longueur moyenne des entre-nœuds (LMEN) ont été détectées. Ces zones correspondent aux mêmes régions génomiques détectées pour la longueur des tiges (55 et 58 Mb sur le chr7) dans les deux environnements. Sur le chromosome 9, trois régions génomiques situées sont proches du gène de nanisme *dw1* précédemment cartographié (Brown et al., 2008, Thurber et al., 2013).

Concernant le nombre total d'entre-nœuds (NEN), sur les 34 QTL détectés, 3 présentent des effets majeurs associés à la variation de ce caractère. La zone génomique localisée sur le chromosome 3 à 40Mb, est très proche du gène de nanisme *dw1* déjà identifié dans les études antérieures (Lin et al., 1995a, Klein et al., 2008). Les deux autres zones chromosomiques avec des effets majeurs sur NEN, correspondent aux zones précédemment identifiées pour la durée du cycle semis floraison.

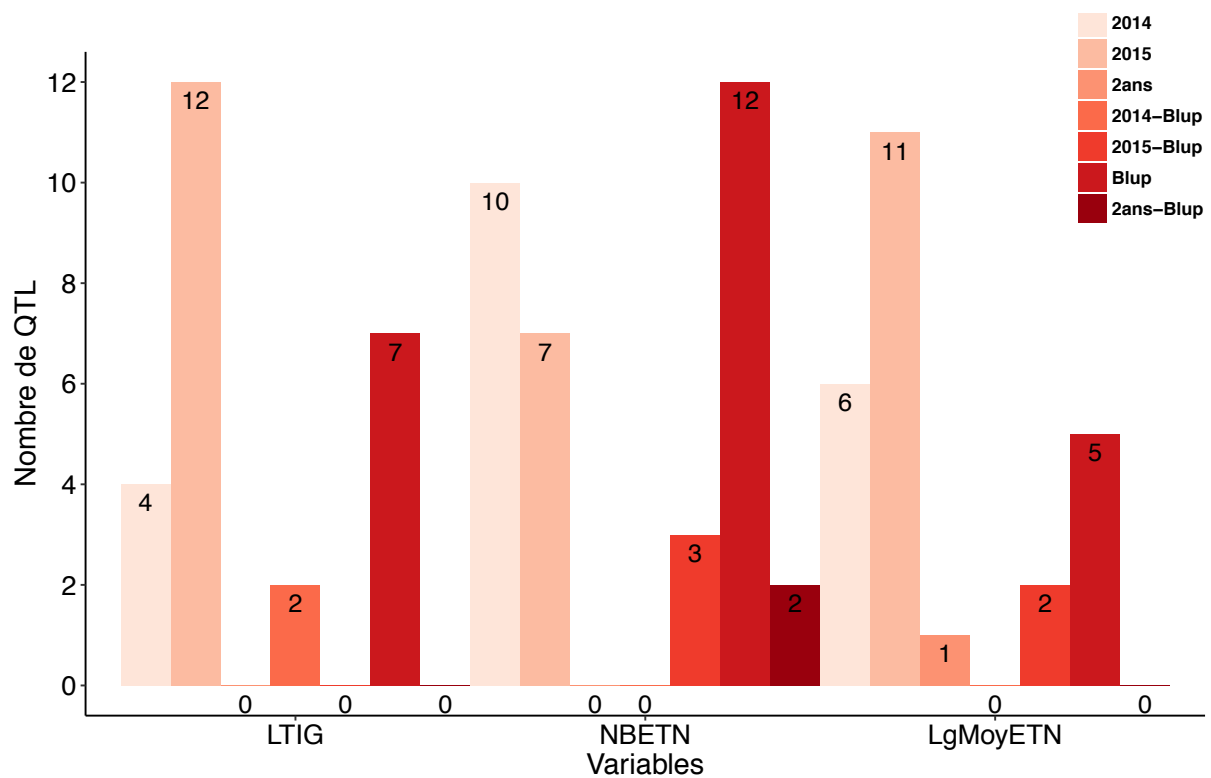


Figure II-18 : Stabilité des régions génomiques détectées dans les deux années pour les variables de longueur des tiges (LTIG), du nombre total d'entre-nœuds (NEN) et de la longueur moyenne des entre-nœuds (LMEN) mesurées sur les différentes familles BC1F4 relatives aux 29 populations BCNAM.

Une seule région chromosomique à effet majeur (à 55 Mb sur le chr7) est associée à la variation de la longueur du pédoncule (LPED).

De nombreuses autres zones chromosomiques avec des effets mineurs sur la longueur du pédoncule et de la panicule (LPED et LAPN), ont été identifiées dans la présente étude. La majorité de ces associations sont spécifiques à cette étude.

#### ***IV.4.2- Colocalisations entre QTL de caractères différents***

Le QTL majeur de durée du cycle semis floraison (SFD) du chromosome 3 (51 Mb) colocalise avec le nombre total d'entre-nœuds allongés (NEN) dans les deux environnements (y compris le blup). La même colocalisation entre SFD et NEN a été détectée sur le chr6 à 0.61 Mb. Ce QTL majeur explique entre 10.54 et 39.91 % de variation phénotypique associée à ces caractères.

En plus de ces associations/colocalisations, une autre région chromosomique du chr6, impliquée dans la variation de NEN colocalise avec la hauteur. Cette colocalisation est probablement due à l'effet indirecte du gène de maturité *ma6* qui joue sur la taille et donc sur le nombre d'entre-nœuds / longueur des tiges.

Sur le chromosome7, deux régions génomiques (55 et 58 Mb) à effets majeurs sont associées à la fois à la variation de la longueur des tiges mais aussi à celle des longueurs moyennes d'entre-nœuds (LMEN) et du pédoncule (LPED). Par ailleurs, les études réalisées par (Rami et al., 1998, Hart et al., 2001) ont rapportées la même colocalisation entre hauteur et longueur du pédoncule.

D'autres zones chromosomiques à effets mineurs localisées sur les chr5 (58 Mb) et chr7 (98 Mb), sont associées à la variation de la hauteur et du nanisme (dans l'environnement 2015) et /ou au nombre total d'entre-nœuds (dans l'environnement 2014 et blup).

#### ***IV.4.3. Stabilité des QTL détectés en fonction des années***

Seulement deux zones génomiques à effets majeurs, associées à la variation du nombre total d'entre-nœuds sont stables entre les deux années (détectées en 2014, 2015 et blup). Une seule région chromosomique est associée à la variation de la longueur moyenne des entre-nœuds dans les deux environnements de culture (2014 et 2015). Cinq zones génomiques dont deux à effets majeurs situées sur les chr3 et chr7 sont associées à la variation phénotypique de NEN et de LMEN et sont stables entre 2015 et blup.

En ce qui concerne les QTL relatifs à la hauteur des tiges, deux régions génomiques dont une à effet majeur sont communes à 2014 et au blup.

De façon générale, dans l'ensemble, 82% des zones génomiques sont spécifiques à une année (2014 ou 2015) où spécifique au blup (2014-2015).



## **Chapitre III : Relations entre production et qualité de la tige : variabilité génétique et impact de la date de semis**

---

### **I- Variabilité génétique de la composition de la biomasse**

Comme présenté au sein la revue bibliographique, les différentes voies de valorisation de la biomasse (alimentation animale, bioénergie, biomatériaux) requièrent en premier lieu un fort rendement en biomasse par hectare. Dans ce contexte, le second chapitre de ce document a eu pour objectif d'analyser le déterminisme génétique des différentes composantes de la production de biomasse (essentiellement la longueur du cycle (semis – Floraison) et la hauteur de la tige principale). Or il est admis que la composition de la biomasse et ses propriétés (sa digestibilité notamment) permettent d'orienter cette dernière vers différentes valorisations. Dans ce contexte, une évaluation précise de la variabilité génétique de la composition de la biomasse au sein du sorgho est nécessaire pour orienter au mieux les programmes de sélection. La variabilité génétique de différentes composantes de la biomasse (sucres solubles, carbohydrates structuraux, lignines, digestibilité) a été analysée au sein de populations biparentales (Murray, Rooney, *et al.*, 2008; Murray, Sharma, *et al.*, 2008; Ritter *et al.*, 2008; Shiringani, Frisch and Friedt, 2010; Shiringani and Friedt, 2011) et de panels de diversité (Murray *et al.*, 2009; Trouche *et al.*, 2014; Burks *et al.*, 2015a; Brenton *et al.*, 2016). Ces analyses ont mis en évidence des héritabilités au sens large relativement élevées (>0.5) pour les différentes composantes avec des interactions GXE très souvent significatives et fortes (Murray, Rooney, *et al.*, 2008; Trouche *et al.*, 2014). Néanmoins ces analyses ont souvent été menées sur des géotypes pas ou peu sensibles à la photopériode et permettant donc leur culture en conditions tempérées. En outre, les effets de la taille des plantes et de la variabilité de la durée de cycle sur la composition de la biomasse ont été assez peu explorés à l'exception de Trouche *et al.* (2014) qui ont analysés les corrélations entre qualité et production de la biomasse pour deux classes de précocité différentes. Cette analyse a notamment mis en évidence des différences de corrélations (production vs qualité) entre les deux classes de précocité. Dans ce contexte, il semble important i) d'affiner l'évaluation de la diversité génétique de la qualité de la biomasse en intégrant des géotypes sensibles à la photopériode aux analyses et ii) d'explorer plus finement les interactions entre hauteur, précocité et composition de la biomasse.

## **II- Impacts des conditions de culture sur la composition de la biomasse**

Outre la variabilité génétique, les conditions de culture impactent aussi significativement le potentiel de production de biomasse et sa qualité. Les effets de la fertilisation (Russo and Fish, 2012), de l'irrigation et de la densité de semis (Rocateli *et al.*, 2012) ainsi que de la date de récolte (Ayub *et al.*, 2003; Atis *et al.*, 2012; Zhao *et al.*, 2012) sur la qualité de la biomasse ont été analysés. Néanmoins ces analyses ont principalement portées sur la production de sucres solubles dédiés à la production d'éthanol de première génération et seules quelques études ont aussi abordées les teneurs en carbohydrates structuraux et en lignines (Zhao *et al.*, 2012). En outre les effets des dates de semis ont elles aussi été peu explorées en ce qui concerne les carbohydrates structuraux et les lignines et ce malgré leurs importances d'une part en conditions semis arides où elles peuvent conditionner le succès ou non d'une culture et en conditions tempérées où des dates de semis plus tardives peuvent être planifiées dans le cadre de la mise en place de production en cultures dérobés (semis tardifs). Il est donc dans ce contexte important d'évaluer les effets de décalages de la date de semis sur l'ensemble des composantes de la biomasse et notamment sur les corrélations entre production et qualité de la biomasse.

## **III- Vers une compréhension plus fine de la variabilité de la composition de la biomasse**

Afin d'affiner notre compréhension des relations entre la variabilité génétique de la production et de la qualité de la biomasse, 6 essais impliquant 35 géotypes sélectionnés pour couvrir les différentes tailles de plantes, de précocité et de sensibilité au photopériodisme ont été mis en place. Différentes dates de semis au sein d'une même saison de végétation ont été effectuées afin d'analyser la réponse des différents géotypes à ces variations d'itinéraires cultureux. Outre la longueur de la tige principale et ses composantes (la longueur moyenne des entrenœuds et leur nombre) la composition biochimique de la biomasse au stade de récolte des grains a été analysée. Ces analyses de composition ont porté sur la teneur en sucres solubles des tiges, les teneurs en carbohydrates structuraux (cellulose et hémicelluloses), les lignines et les propriétés de digestibilité de la biomasse.

L'analyse des résultats obtenus fait l'objet d'un article en cours de préparation pour soumission à Field Crops Research qui est présenté dans la suite de ce document. Les résultats principaux ainsi qu'une discussion sont présentés à la suite de l'article.

## **IV- Article en cours de finalisation à soumettre à Field Crop Research**

1                   **Genotype and sowing date effects on the production and**  
2                   **composition of sorghum stems, disentangling the contributions of**  
3                   **dwarfism, earliness and photoperiod sensitivity**

4  
5 Théra Korotimi<sup>1,2,3</sup>, Tékété Mohamed<sup>3</sup>, Doumbia Mohamed<sup>3</sup>, Kouressi Mamoutou<sup>3</sup>, Maïga  
6 Aoua<sup>3</sup>, Bonnal Laurent<sup>4,5</sup>, Bastianelli Denis<sup>4,5</sup>, Rouan Lauriane<sup>1,2</sup>, Eric Gozé<sup>1,2</sup>, Verdeil Jean-  
7 Luc<sup>1,2</sup>, Baptiste Christelle<sup>1,2</sup>, Trouche Gilles<sup>1,2</sup>, Luquet Delphine<sup>1,2</sup>, Rami Jean-François<sup>1,2</sup>,  
8 Vaksman Michel<sup>1,2,3</sup>, Tiémé Niaba<sup>3</sup>, Pot David<sup>1,2</sup>

9  
10 <sup>1</sup> CIRAD, UMR AGAP, F-34398 Montpellier, France

11 <sup>2</sup> AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France.

12 <sup>3</sup> Institut d'Economie Rurale (IER), Laboratoire Sol Eau Plante de Sotuba, P.O. Box 262,  
13 Bamako, Mali

14 <sup>4</sup> CIRAD, UMR SELMET, F-34398 Montpellier, France

15 <sup>5</sup> SELMET, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France.

16  
17 \*Corresponding author: Tel.: +33 (0) 4 67 61 71 50

18 Email address: [david.pot@cirad.fr](mailto:david.pot@cirad.fr) (David Pot)

19  
20  
21  
22  
23 Key words: Sorghum, stem composition, heritability, sowing date, dwarfism, earliness,  
24 photoperiod sensitivity  
25  
26

## 27 I- Introduction

28 Sorghum (*Sorghum bicolor* [L.] Moench) is the fifth most important cereal crop in the world  
29 both in terms of planted area and grain production (FAOSTAT, 2017). This species that belong  
30 to the Poaceae family is the staple food of more than 500 million people in over 30 countries  
31 worldwide. It constitutes the main source of energy and protein for the poorest people in these  
32 countries. Sorghum grains and vegetative biomass are also used for feed. In developed countries  
33 like United States and Europe, vegetative biomass (stem and leaves) is produced in large  
34 quantities (4.475 000 tons in 2015 in the United States for instance  
35 (<https://www.nass.usda.gov>)) and mainly used for forage, hay, silage and pasture  
36 (<http://www.sorghumcheckoff.com/all-about-sorghum>).

37 In Africa, in most of the sorghum-based production systems, vegetative biomass residues  
38 obtained after the harvest of mature grains are used for animal feed during the shortage period  
39 of fodder.

40 Beside its interests for food and feed, sorghum has been recently identified as a relevant  
41 feedstock for bioenergy and biomaterial production. Indeed, its high biomass yield potential,  
42 its short cycle, its adaptation to low input crop management and its flexibility in planting  
43 periods (Berenji and Dahlberg, 2004, p. 4; Vermerris *et al.*, 2007; Carpita and McCann, 2008;  
44 Paterson *et al.*, 2009; Olson *et al.*, 2012; Mace *et al.*, 2013) constitute relevant attributes. For  
45 the different uses of sorghum vegetative biomass (food, feed, energy, biomaterials...), biomass  
46 yield per unit area is the primary trait of interest (Bayu, Rethman and Hammes, 2005; Murray,  
47 Rooney, *et al.*, 2008; Trouche *et al.*, 2014; Maucieri *et al.*, 2016; May *et al.*, 2016).

48 However, the biomass quality that can be summarized as the biochemical and physical  
49 properties of the vegetative biomass also significantly impacts the different end-use products  
50 properties. Although plant height which is a predictor of biomass yield has been extensively  
51 studied in sorghum, our knowledge of the genetic determinism of biomass quality and its  
52 relations with biomass yield remains in its infancy.

53 It is classically accepted that plant height in sorghum is directly influenced by the internodes  
54 length that depends on the dwarfism genes and by the duration of the cycle through the  
55 photoperiod sensitivity genes. Internode length is under the dependence of four major dwarf  
56 genes called: *dw1*, *dw2*, *dw3* and *dw4* (Windscheffel *et al.*, 1973, Quinby and Martin, 1954).  
57 Regarding the genetic determinism of cycle duration, six maturity loci mainly corresponding to  
58 photoperiod sensitivity genes have been identified in sorghum (*Ma1* to *Ma6* : (Quinby, 1967;  
59 Quinby, Hesketh and Voigt, 1973; Rooney and Aydin, 1999).

60 In addition to these major genes, several QTL and GWAS studies have been performed that  
61 allowed the detection of several QTL for both internode length and cycle duration (Brown *et al.*,  
62 2008; Mace and Jordan, 2011; Mace, Hunt and Jordan, 2013; Higgins *et al.*, 2014; Zhao *et al.*,  
63 2016).

64 The description of the genetic architecture of biomass quality is much less advanced although  
65 a great genetic variability has been described in sorghum (Vandenbrink *et al.*, 2010, 2013;  
66 Trouche *et al.*, 2014). In terms of genetic architecture of the biomass composition, some QTL  
67 and GWAS analyses have been recently performed either on the soluble sugar content of the  
68 biomass (Murray *et al.*, 2008a; Ritter *et al.*, 2008; Shiringani, Frisch and Friedt, 2010;  
69 Shiringani and Friedt, 2011; Burks *et al.*, 2015) or on the structural carbohydrates components  
70 (lignin, cellulose, hemicellulose : (Murray, Rooney, *et al.*, 2008; Brenton *et al.*, 2016b).

71 However, only a limited understanding of the variability of the biomass composition in relation  
72 with the dwarfism and earliness effects is currently available. Trouche *et al.* (2014) reported  
73 genotypic correlations between plant height, cycle duration and the biomass composition on a  
74 panel encompassing around 100 non photoperiodic accessions. To refine our understanding of

75 the correlations between plant growth and biomass composition related traits, it would be  
76 relevant to explore the contributions of the different dwarfism classes to the biomass  
77 composition variability instead of considering only the plant height which is defined, as  
78 mentioned earlier by the internode length and also their number.

79 In addition, although genetic correlations between cycle duration and biomass composition  
80 traits have been already analyzed revealing different behaviors of the biomass composition  
81 traits (positive correlations with the soluble sugar content, and the in vitro organic matter  
82 digestibility and a negative correlation with the cell wall content of the biomass assessed  
83 through the NDF (Trouche *et al.*, 2014), a better understanding of the effect of the cycle  
84 duration on the biomass composition would be relevant to monitor more efficiently breeding  
85 programmes.

86 It is also important to mention here that most of the analyses that addressed the correlations  
87 between biomass production (i.e cycle duration and growth components) have been performed  
88 in non-photoperiodic accessions possibly limiting the genetic variability analyzed.

89 However, if the genetic control explain a large part of the phenotypic variability of the biomass  
90 composition (heritabilities and repetabilities generally over 0.8 (Trouche *et al.*, 2014; Burks *et*  
91 *al.*, 2015; Brenton *et al.*, 2016a)) significant effects of external factors as fertilizers (Russo and  
92 Fish, 2012), irrigation, tillage practices (Rocateli *et al.*, 2012), plant density (Carmi *et al.*, 2006)  
93 and harvest date (Atis *et al.*, 2012, Ayub *et al.*, 2002, ÇAKMAKÇI *et al.*, 1999, Glamoclija *et*  
94 *al.*, 2011, Zhao *et al.*, 2012b) have been reported. Although, the effects of the sowing date have  
95 been largely documented on the biomass and grain production (Poornima, Geethalakshmi and  
96 Leelamathi, 2008; Burks *et al.*, 2013; Rao *et al.*, 2013; Almodares and Hoseini, 2016), its  
97 effects on the quality of the biomass have not been yet clearly analyzed. Sugar yield variability  
98 in relation with the sowing date have been explored in temperate (Burks *et al.*, 2013) and semi-  
99 arid conditions (Poornima, Geethalakshmi and Leelamathi, 2008; Rao *et al.*, 2013; Almodares  
100 and Hoseini, 2016; Tovignan *et al.*, 2016) but no global analysis of the different components  
101 of the biomass have been performed yet. In addition most of the analyses aiming to address the  
102 sowing effects on the biomass composition (soluble sugar yield mainly until now) have been  
103 performed with relatively small diversity panels including principally non photoperiod sensitive  
104 genotypes (with the exception of Tovignan *et al.*, 2016).

105 Taking into account, these knowledge gaps, the main objectives of this study were to: 1)  
106 evaluate the contributions of the dwarfism and earliness effects on the genetic variability of the  
107 stem biomass composition on a panel of 35 genotypes selected to encompass most of the  
108 cultivated sorghum diversity (i.e. sensitive and non sensitive to the photoperiod) and 2) analyze  
109 the effects of the sowing date on the biomass quality and the specific impacts of the photoperiod  
110 sensitivity variability in this context. To reach these objectives, the panel of genotypes was  
111 cultivated over 6 different growing conditions (including different sowing dates) and  
112 characterized for its growth, cycle duration and biomass quality (i.e soluble sugar, structural  
113 carbohydrates contents and the biomass and cell wall digestibilities.

114  
115

## 116 **II- Materials and methods**

### 117 **II.1- Plant material**

118 The experimental material was composed of a set of 35 sorghum varieties (*Sorghum bicolor*  
119 (L.) Moench) from different agronomic research programs of West Africa, in particular IER  
120 (Institute of Rural Economic) and ICRISAT Mali (International Crops Research Institute for  
121 the Semi-Arid Tropics). These varieties cover a large portion of the diversity of cultivated  
122 sorghums. The panel of 35 sorghum varieties include both landraces and improved varieties  
123 selected for their variable photoperiod sensitivities, durations of cycle, stem length, grain yield  
124 and quality. The geographical origins and agronomic characteristics of the 35 genotypes are  
125 indicated in Table 1.

126

### 127 **II.2- Experimental design**

128 The field experiments were conducted in Bamako, Mali at the Sotuba agronomical research  
129 station of the “Institut d’Economie Rurale” (12°39’N, 7°56’E), which is located at 381m of  
130 altitude and that corresponds to a Sudano-sahelian savannah climate type.

131 The 35 varieties were evaluated during the 2010, 2014 and 2015 rainy seasons. With the  
132 exception of 2010, in which only one sowing date was performed, the 35 sorghum varieties  
133 were tested in split-plot designs at two sowing dates (2010: June 24th, 2014 : June 16th and  
134 July 16th, 2015 : June 21st and July 20th ) in two replicates. The main treatments for all trials  
135 were the sowing date and the genotypes. Each experimental plot for these trials included 40  
136 plants, sown by hand in four rows of 3.6m-long, with 0.75m spacing between rows and 0.40m  
137 between the hills on each row. Fertilization included 150 kg/ha of DiAmmonium Phosphate  
138 (DAP) before sowing and 150 kg/ha of Urea supplied during vegetative growth.

139 During the off-season 2015, an additional sowing date was performed in September 21<sup>st</sup>. In this  
140 experiment, the enotypes were allocated into a randomized complete block design with two  
141 replicates. This trial included 20 plants sown by hand on two rows of 3.6m-long, with 0.80m  
142 spacing between the rows, 0.40m between the hills on each row and 1m between each plot.  
143 Irrigation was provided for this trial according to the plants needs.

### 144 **II.3- Phenotypic characterization**

#### 145 **II.3.1- Phenology and growth**

146 For all trials, 4 plants in each experimental plot were used to evaluate the number of leaves  
147 produced from the seedling to the flag leaf stage (TLN: Total Leaf Number), the total number  
148 of green leaves at the harvest stage (NGL: number of green leaves), and the stem length (STEL)  
149 that was measured from the base of the main stem to the base of peduncle.

150 The total number of elongated internodes (INTN) was counted after removing all the leaves  
151 and sheath from the main stem. The Mean internode length (MITNL) was then calculated by  
152 dividing the stem length by the total number of elongated internodes.

153 The duration from the sowing to the flag leaf appearance (DTFL: Duration To Flag Leaf)  
154 expressed in days was also recorded. This information was used to calculate the coefficient of  
155 photoperiodism (Kp) according to Traore et al 2001 (TRAORE et al., 2001). This calculation  
156 was performed on the 2014 and 2015 trials considering only the first two sowing dates. This  
157 coefficient is defined as the ratio of duration from sowing to flag leaf between the first and  
158 second sowing (DFTL1 and DFTL2) by the difference of the number of day between first and  
159 second sowing (D1 and D2).

$$160 \quad Kp = \frac{DTFL1 - DTFL2}{|D1 - D2|}$$

161 This coefficient  $K_p$  varies between 0 and 1. For highly photosensitive genotypes, this  
162 coefficient is closer to 1 because the shortening of the vegetative period is equal to the seedling  
163 offset (Kouressy et al., 2008c).

### 164 ***II.3.2- Stem Biomass composition and structure***

#### 165 ***II.3.2.1- Measurement of sugar content: Brix value***

166 For all rainy season trials, four plants from each experimental plot were harvested at grain  
167 physiological maturity to measure the sugar concentration in the stalks (through the Brix  
168 measurement). These measurements of Brix were realized on all the genotypes from the 2015  
169 trial, and only for 33 and 30 genotypes in 2010 and 2014 trials respectively. The handheld  
170 refractometer (Atago PR 32a, ATAGO CO, LTD, Tokyo, Japan) was used to measure the Brix  
171 (a soluble solids) contained in the juice extracted on the four stems after crushing in a three-roll  
172 sorghum press (ICRISAT; Jagadish No. 6, Gujarat, India).

173 With the exception of the 2010 rainy season trial, the juice weight (SJW) and volume (SJW)  
174 were measured respectively with an electronic balance and a graduated cylinder (250ml). Fresh  
175 weight of the stems (without leaf sheath) before juice extraction (SFW), were recorded in g.  
176 The fresh weight after extraction was also measured and used to calculate the stem extractable  
177 moisture content (SMEX).

178 The stem sugar concentration (g/g of stem) was calculated according to the BRIX and Stem  
179 Juice weight (SJW).

#### 180 ***II.3.2.2- Stem biochemical analysis***

181 For each planting date, 4 plants from each experimental plot were selected at the grain maturity  
182 stage for stem biochemical composition analysis. These sampling were carried out on all the 35  
183 genotypes in 2010 and 2014 trials and only on 26 genotypes in 2015 trial.

184 The stem biochemical analyses were performed according to the protocol presented by Trouche  
185 et al 2014. The availability of NIRS calibrations based on reference laboratory measurements  
186 (chemical composition, food value, etc.) made possible the NIRS prediction of our samples.  
187 The following section describes firstly the set of reference analyses, the calibration process and  
188 finally the NIRS prediction process.

##### 189 ***II.3.2.2.1- Kjeldhal, Van Soest and Aufrere analysis***

190 The crude protein (CP) content was determined using the standard method of nitrogen content  
191 (N) determination of Kjeldhal following the formula:  $CP = N \times 6.25$  (Kjeldahl, 1883).

192 Regarding the description of the fiber fraction, the Van Soest method (Van Soest, Robertson et  
193 al. 1991) was used to estimate the relative contents of the different fiber fractions ; Neutral  
194 Detergent Fiber (NDF), Acid Detergent Fiber (ADF) and Acid Detergent Lignin (ADL). The  
195 NDF fraction includes all cell wall components (cellulose, hemicellulose and lignin); ADF  
196 contains the cellulose and lignin; and the ADL fraction corresponds to lignin.

197 In addition, we calculated the relative of the NDF in lignin, cellulose and hemicellulose using  
198 the following formulae's:

- 199     ▪ Lignin content,  $LIG = [ADL / NDF] * 100$
- 200     ▪ Cellulose content,  $CEL = [(ADF-ADL)/NDF]*100$
- 201     ▪ Hemicelluloses content,  $HEM = [(NDF-ADF)/NDF] * 100$

202 For the determination of the solubility of the biomass, Aufrère et al (2007) method was used.  
203 This enzymatic method, based on the digestion by pepsin-cellulase allows to predict the in-vitro  
204 organic matter digestibility (IVOMD) and in-vitro dry matter digestibility (IVDMD). Based on  
205 these predicted values, the in-vitro NDF digestibility (IVNDFD) that represents the potential  
206 degradation of fiber in ruminants was calculated assuming that the non-NDF part of plants  
207 material was completely digestible using the following formulae:

$$208 \quad IVNDFD = 100 \times (IVOMD - ((100 - MM) - NDF)) / NDF$$

209

210

#### II.3.2.2.2- Stem NIRS measurements

211 To predict the different biochemical parameters described in the precedent section, NIRS  
212 analyses were performed on dried samples using a Tango instrument (Bruker Optics, Hong  
213 Kong Ltd). A total of 1086 dried stem samples from the rainy and off-season trials were finely  
214 ground at 1mm. Each sample was scanned two times at 2 nm intervals over the wavelength  
215 ranging from 400 to 2500 nm. The two spectra obtained were averaged. The system had already  
216 been calibrated for analytic parameters listed in the table3. The available calibrations are based  
217 on reference measurements obtained on at least 450 samples (Table S1).  
218 Approximately 100 samples originating from the different trials of this study were selected  
219 for reference analyses in order to update the preexisting NIRS calibrations.

220 The prediction equations for the analytical parameters were calculated from linear regression  
221 and the validation of the results of predictions were determined by the following parameters:  
222 standard error of calibration (SEC), the multiple calibration coefficient of determination (RSQ),  
223 the standard error of cross validation (SECV), the multiple coefficient of determination of cross  
224 validation (R2) and the ratio performance/validation (RPD) (Table S1).



225

#### II.4- Statistical analyses

226 All statistical analyses were performed using R version 3.2.4 (R development Core Team,  
227 2016). Means, range, and coefficient of phenotypic variation were first calculated on the two  
228 replicates from each sowing date for each year to check the consistency of the results.

229

#### Variance components analyses

230 As a first step, the year, genotype and genotype\*year effects were analyzed for the first sowing  
231 date of the three years of experiments according to the following model

$$232 Y_{ijk} = \mu + Y_j + G_i + B_k/Y_j + (GY)_{ij} + \varepsilon_{ijk} \text{ (model 1)}$$

233 Where  $Y_{ijk}$  the phenotypic value of genotype  $i$  measured in block  $k$  for year  $j$ ;  $Y_j$ : year  $j$  ( $j=3$ ),  
234  $G_i$ : the random effect of genotype  $i$  ( $i=1$  to 35),  $B_k/Y_j$ : the fixed effect of block  $k$  nested in year  
235  $i$ ,  $GY$ : the random interaction between the genotype and the year and  $\varepsilon_{ijk}$  the residual error ( $\sigma^2$ ),  
236  $\sigma^2$ ).

237 This model was used to calculate the broad sense heritabilities and the coefficient of genetic  
238 variations (CGv) of the different traits. The Best linear unbiased predictor of the genotype  
239 effects were also derived from this model.

240 As a second step, the contributions of the dwarfism and earliness effects to the variability of  
241 the analyzed traits among the first sowing date of the three years of experiment were explored.  
242 Each genotype was first allocated to different classes of mean internode length (i.e. dwarfism,  
243 Short : <10 cm ( $n = 10$ ), Medium : 10-15 cm ( $n = 13$ ), Long : >15 cm ( $n = 12$ )) and earliness  
244 (Early : <75 ( $n = 13$ ), Medium : between 75 and 90 ( $n = 9$ ), Late : >90 days ( $n = 13$ )). Internode  
245 length and earliness classifications were based on the results from the first sowing dates (June)  
246 of the three years of experiments.

247 Following these genotypes characterizations, dwarfism and earliness effects were analyzed on  
248 the first sowing dates of the three years of experiment using the following model:

249

$$250 Y_{ijklm} = \mu + Y_j + D_j + E_k + G_l + B_l/Y_j + (G/D*Y)_{ijk} + F_m + \varepsilon_{ijklm} \text{ (model 2)}$$

251

252 Where  $Y_{ijklm}$  is the phenotypic value of genotype  $k$  measured in block  $l$  for year  $j$ ;  $Y_j$  the effect  
253 of the year  $i$  ( $i=1$  to 3),  $Dwj$ , the dwarfism effect ( $j=1$  to 3),  $E_k$  : the effect of the earliness class  
254 ( $k=1$  to 3),  $G_l$ : the genotype  $k$  ( $i=1$  to 35),  $B_m/Y_j$ : the fixed effect of block  $m$  nested in year  $I$ , ,  
255  $G*Y$ : the interaction effect of the genotype  $l$  by year  $i$  and  $\varepsilon_{ijklm}$  the residual error ( $\sigma^2$ ).

256 As a third step, taking advantage of the different dates of sowing performed on the different  
257 trials, the variability of the traits were analyzed over the three years of experiments using the  
258 following model :

259

$$260 Y_{ijklm} = \mu + Y_i + S_j + (B/Y/S)_{kij} + PPS_l + (PPS*S)_{lj} + G_m + (G*S)_{mj} + \varepsilon_{ijklmn} \text{ (model 3)}$$

261

262 Where  $Y_{ijklmn}$  is the phenotypic value of genotype  $m$  measured in block  $k$  for year  $I$  in sowing  
263 date  $j$ ;  $Y_i$  the effect of the year  $i$  ( $i=1$  to 2),  $S_j$  the sowing date effect ( $j=1$  to 97),  $PPS_l$  : the  
264 Photoperiod sensitivity class ( $l = 1$  to 3),  $(PPS*S)_{lj}$  the interaction effect of the photoperiod  
265 sensitivity class with the sowing date effect,  $G_m$ : the genotype effect ( $m=1$  to 35),  $(G*S)_{mj}$  the  
266 interaction effect of the genotype with the sowing date and  $\varepsilon_{ijklmn}$  the residual error ( $\sigma^2$ ). The  
267 effect of the photoperiod sensitivity classes of the genotypes was also included in the model as  
268 photoperiod sensitivity is expected to impact the response of the genotypes to the different  
269 sowing dates. The Photoperiod sensitivity classes (Low : <0.4 ( $n = 10$ ), Moderate : between 0.4  
270 and 0.6 ( $n = 8$ ), High : >0.6 ( $n = 17$ ), Table 1) were based on the  $K_p$  estimations from the 2014  
271 and 2015 experiments.

272 Multiple comparisons between means were performed using Tukey HSD (Honestly Significant  
273 Difference analysis). The critical probability threshold p-value for accepting the hypotheses  
274 ( $H_0/H_1$ ) used to define statistically significant results was fixed at  $P = 0.05$ .

#### 275 **Correlations analyses**

276 Correlations between the values of the same traits over the three years of experiments (first  
277 sowing date) were computed based on the adjusted genotypic values of the different trials. To  
278 allow a global visualization of the correlation structure between the traits and of the similarities  
279 between the genotypes a hierarchical clustering based on the BLUP from the model 1 from the  
280 35 genotypes was performed using the hclust function in R.

281 Correlations between the phenotypic traits at the first sowing date were calculated on the basis  
282 of the BLUP obtained for the 35 genotypes using the mixed linear model 1. In addition  
283 correlations between the different sowing dates were also calculated based on the adjusted  
284 genotypic values of the different sowing dates.

### III- Results

#### III.1- Phenotypic and genetic variability of phenology and stem biomass composition related traits

Twenty six traits were measured over 3 years for a set of 35 genotypes over one (in 2010) to three sowing dates (in 2015). Different precipitation patterns were observed for the three years although the temperature patterns were stable over years (Figure S1). The year 2015 was the driest one with only 702.8 mm of cumulated water received compared to 859.4 mm in 2014 and 976.7 mm in 2010. In addition, a peculiar pattern of precipitation was observed in 2014, with a large amount of precipitation between May and July and a relatively dry August month. In order to assess the genotypic variability available between the different genotypes, the dataset obtained from the June sowing dates of the 3 years of trial were considered. Almost the same levels of phenotypic variability (PCV%) were observed over the three different years of field experiment (Table 2). Only a few traits deviated from this general trend (NGL, BRIX, SJW, SSC and CP) The increase of the range of variation observed in 2015 for these traits is likely due to the lower amount of water received for this cropping season (Figure S1) in combination with the variability of the staygreen trait in the panel of genotypes analyzed. Indeed, a severe senescence at maturity was observed in 2015 (mean number of green leaves of 1.91 in 2015 whereas it reached 6.99 and 4.68 respectively in 2010 and 2014) probably increasing the variability of some of the traits observed between the Staygreen and non-staygreen genotypes.

Whereas large variabilities (PCV >30 %) were observed for the phenological, agromorphological, proteins, sugars, histological (only 2014 experiment) and stem digestibility related traits, smaller variabilities (PCV < 10 % in average) were observed for the fiber components. The large reduction of variabilities observed between 2014 and 2015 for the histological traits can partly be attributed to the reduction of the number of genotypes analyzed between 2014 (22) and 2015 (6).

Biomass production relation traits exhibited highly significant genotype effects with broad sense heritabilities over 0.96 with the exception of the number of green leaves (NGL :  $H^2_{bs} = 0.46$ ) and the stem fresh weight (SFW :  $H^2_{bs} = 0.75$ ) (Table 3). The soluble sugars related traits presented heritabilities ranging from 0.58 for the soluble sugar content of the stem (SSC) to 0.74 for the stem moisture extractable percentage (SMEX). The same range of heritabilities were observed for the Protein matter content, the Van Soest and Aufrere related traits (from 0.52 for NDF to 0.73 for Hem). The genotypic coefficient of variations were relatively high (i.e. greater than 15 %) for most of the traits with the exception of the fiber proportion (NDF) and its components (lignin, cellulose, hemicellulose contents) for which genotypic coefficient of variation usually lower than 10 % were observed.

In addition with the analysis of the global genotypic effects (model 1), the genotypic variance was further decomposed to assess the effects of the dwarfism and earliness classes (Model 2). The dwarfism effect was highly significant for all the traits with the exception of the number of green leaves at maturity (NGL) (Table 3). This effect explained between 6 and 27 % of the phenotypic variability of the biomass composition related traits (Figure 1). The earliness effect followed a divergent pattern, highly significant effects were detected for the biomass production traits whereas it was almost never significant or presented really weak effects on the biomass quality traits. Deviations from this general trend were observed for the BRIX and the Stem moisture extractable content (SMEX) which were highly impacted by the earliness classes. The genotype by year (GxY) effect was significant for most of the traits with the exception of the stem length (which is highly heritable:  $h^2=0.99$ ) and of the sugar related traits which are dependent on the moisture content at the harvest (Stem fresh weight and Stem extractable

moisture content). The structural biomass composition related traits exhibited highly significant GxY effects with phenotypic explained variation ranging from 14 to 28 %. In agreement with the highly significant genetic effects detected taking into account the 3 years of experiments the correlations obtained between the adjusted means of each year were most of the time highly significant (Table S2). A few exception to this general trend were observed. The between years correlations for the number of green leaves obtained for the 3 years of experiments were never significant. In addition, a few correlations involving the fiber components and the fiber digestibility were non significant for the 2014 and 2015 comparisons.

### **III.2- Dwarfism and cycle duration effects on the agromorphological and biomass composition related traits**

#### ***III.2.1- Dwarfism and cycle duration impact differentially biomass yield and composition***

The dwarfism and earliness effects have been characterized on all the analyzed traits (Table 3). In order to illustrate the observed trends, six representative traits of the different phenotypic categories have been used (Figure 2). STEL and DTFL represent the production and phenology categories whereas NDF, SSC, IVOMD and IVNDFD are representative of the biomass composition traits. As shown in Figure 3 (that will be detailed later) these traits also represent the three clusters of traits (P1, P2 and P3) that have been detected through a hierarchical clustering analysis. The estimator of the cell wall content of the biomass (NDF) increases gradually with the mean length of the internode, whereas the digestibilities of the organic matter of the whole biomass (IVOMD) and of the cell walls (IVNDF) follow opposite patterns, presenting lower values for long internode genotypes. Although for a few biomass quality related traits, the earliness effects are significant, the proportion of phenotypic variances are always extremely weak (lower than 5% with the exception of SMEX (22.52 %), Table 3). The three groups of genotypes identified through the hierarchical clustering analysis illustrate these results (Figure 3). The “genotype” group G1 includes mainly short and medium genotypes with early to medium flowering that present high biomass and cell wall digestibilities and low cell wall contents (NDF). The genotype group “G2” encompass short to medium genotypes with medium to late flowering that also present high biomass and cell wall digestibilities. The last genotype group “G3” contains mainly medium and tall genotypes that present low biomass and cell wall digestibilities with a high cell wall content. This last genotype group include all the earliness classes.

#### ***III.2.2- Impacts of Dwarfism and Earliness classes on the relationships between agro-morphological and stem composition related traits***

The hierarchical clustering analysis (Figure 3) provides a global picture of the correlations between the different phenotypic traits (see also Table S3 for the main pairwise correlations). The trait cluster “P1” indicates that digestibility related traits (IVDMD, IVOMD, IVNDFD) are strongly correlated with the hemicellulose content of the cell walls (Hemi), the crude protein content and the extractable stem moisture content (SMEX) (correlations greater than 0.5). Trait cluster P2 underlines the strong relationships between the stem length, the mean internode length and the richness of the biomass in cell wall ( $r_{\text{STEL-NDF}} = 0.35^*$ ), lignin ( $r_{\text{STEL-Lig}} = 0.63^{***}$ ) and cellulose contents ( $r_{\text{STEL-Cell}} = 0.35^*$ ). The last cluster of traits (P3) encompass phenology (DTFL, Kp, INTN, TLN) and sugar related traits (Brix, SSC). As expected, an opposition, confirmed by significant negative correlations, has been detected between the trait clusters P1 and P2 opposing biomass digestibility related traits, plant height and cell wall content richness in lignin and cellulose. Stem length is negatively correlated with the biomass digestibilities ( $r_{\text{STEL-IVOMD}} = -0.46^{**}$ ,  $r_{\text{STEL-IVNDFD}} = -0.57^{***}$ ).

Interestingly a positive correlation was detected between the BRIX and the Stem length ( $r = 0.56^{***}$ ), but non significant correlation was observed between the stem length and the stem soluble sugar content ( $r = 0.08^{ns}$ ) likely resulting from the strong negative correlation between stem length and the stem moisture extractable proportion ( $r = -0.76^{***}$ ).

To reach a better understanding of the correlations between biomass production and biomass quality related traits, correlation analyses were also performed independently for each dwarfism and earliness classes (Figure 4). Drastic evolutions of the correlations between the stem length and the biomass quality related traits were observed between the different dwarfism classes. Most of the significant correlations for the whole panel disappeared when each dwarfism class was considered. At the opposite stable correlations were observed over the different earliness classes.

### ***III.3- Sowing date effects on phenological, agro-morphological and stem biomass composition trait***

Variance components analysis allowed identifying the main sources of variation for the different traits when the genotypes were sown at different sowing dates (Figure 5 and TableS4). Overall, the sowing date, the photoperiod sensitivity classes (PPS), and the genotype effects are the major contributors to the variability of the traits. The specific impacts of these components vary according to the considered trait. Phenology (DTFL) and biomass production related traits (INTN, TLN, STEL and SFW) are mainly affected by the sowing dates, the photoperiod sensitivity classes and the genotype effects. Sugar related traits are mainly controlled by the genotype effect whereas structural carbohydrates components and their digestibilities are mainly driven by the sowing dates and the genotype effects.

As expected the sowing date, decreased the biomass production components and the duration of the cycle (Figure 6). Delayed sowing reduced the cell wall content of the biomass (NDF) and increased its soluble sugar content (SSC) leading to an improvement of the organic matter (IVOMD) and cell wall digestibilities (IVNDFD).

The Photoperiod sensitivity classes have highly significant effects on the biomass production related traits with phenotypic explained variation reaching 30 % (Figure 5). Although, they were also significant for most of the biomass composition related traits, their impacts were much lower ( $r^2$  between 1 and 5 %).

Although the proportions of variance explained by the interaction between the PPS classes and the sowing dates are much lower than the main effects of the sowing dates, PPS classes and genotype effects, the effects of this interaction are highly significant for most of the traits (Table S4). As expected the stem length and cycle duration responses of the different PPS classes to the different sowing dates were extremely different. Highly photoperiod sensitive genotypes reduced strongly their stem length (Figure 7) and flowering time (Figure S2) whereas non photoperiod sensitive genotypes were weakly affected. The responses of the biomass quality related traits to the interaction between the PPS classes and the sowing dates diverged significantly from this pattern. The high photoperiod sensitivity genotypes presented a more stable composition than the low and moderate photoperiod sensitive genotypes as observed according to the regression coefficient obtained on the figures 7 and S2 for the Neutral detergent fiber content (NDF), the soluble sugar content of the stem and the organic matter digestibility which are lower than than the ones obtained for the low and moderate photoperiod sensitivity classes.

As a last step the impact of the sowing dates on the correlations between the Stem length and the biomass composition related traits were analyzed. The last sowing date significantly modified the correlation matrix structure whereas the earliest sowing only slightly modified the intensities of the correlations (Figure 8).

## Discussion

### **The high genetic variability of biomass composition related traits in sorghum: a breeding reservoir for future biomass value chains**

In agreement with the bibliography, our results underlined the large genetic variability available for the key components of the stem quality in sorghum. These results are highly relevant in the context of the development of sorghum varieties adapted to the different biomass value chains that have already been developed or that are currently under development. We observed higher values of lignin contents (ADL up to 11.3 and lignin content of the NDF reaching 14.95%) than the ones previously reported (ADL of 8.7 (Trouche *et al.*, 2014) and 10.5 (Brenton *et al.*, 2016b) and lignin content of the NDF of 10.8 (Trouche *et al.*, 2014). Although the lignin content is negatively correlated to the digestibility of the biomass (also reported in this analysis) and thus not relevant for animal feed or methane production it is likely to be interesting for most of the biomaterial applications. For the other traits, the same or slightly lower ranges of variability were reported in the literature (Trouche *et al.*, 2014; Burks *et al.*, 2015; Brenton *et al.*, 2016b) Broad sense heritabilities over 0.52 were obtained based on the three years of experiments for the different biomass composition related traits. These values are comparable with the ones reported in the literature (Murray, Sharma, *et al.*, 2008; Trouche *et al.*, 2014; Brenton *et al.*, 2016b). Although the coefficient of genetic variability reported for the different structural carbohydrates are relatively limited (generally lower than 10 % for NDF, ADF, ADL, Lig, Cell and Hem) it is interesting to note that the ones obtained for the integrative biomass composition quality traits (i.e. the Dry matter, organic matter and NDF digestibilities) are higher (over 15%) Taken together, the heritabilities and genetic variabilities obtained underlined the ability to develop varieties of sorghum presenting specific compositions that could fit to the expectations of the different end-uses.

### **Some antagonistic relationships between the Stem length and the structural carbohydrate components**

Although, the dry matter yield of the vegetative biomass was not estimated in our study, the stem length (which is the main component of the plant height) is usually significantly correlated to the dry weight biomass yield (Trouche *et al.*, 2014; Brenton *et al.*, 2016b). The correlations observed between the stem length and the biomass composition related traits in our study were quite variable.

Regarding soluble sugar production, as already reported (Murray, Sharma, *et al.*, 2008; Trouche *et al.*, 2014; Burks *et al.*, 2015), a positive correlation was observed between the stem length and the BRIX ( $r = 0.56^{***}$ ). Nevertheless the correlation between the stem length and the stem soluble sugar concentration was not significant ( $r = 0.008^{ns}$ ) and only a really weak effect of the dwarfism class was detected on the stem soluble sugar concentration. These results contrast with the positive correlation ( $r=0.69$ ) observed by Burks *et al.* (Burks *et al.*, 2015) between the plant height and the sugar yield by weight. One possible explanation of this discrepancy could probably be the genetic composition of the two panels analyzed: Burks *et al.* enriched their panel in sweet sorghum ideotype whereas no such type of selection was applied to our panel. Breeding for high sugar yielding genotypes has thus to be carefully tuned according to the diversity available in the breeding programme considered.

A negative relationship was observed between the stem length and the animal feed quality with negative correlations between the stem length and the digestibilities of the whole biomass ( $r=-0.56^{***}$ ,  $-0.46^{**}$  respectively for IVDMD and IVOMD) and of the neutral detergent fiber ( $r=-0.57^{***}$ ) also illustrated by the significant effects of the dwarfism classes on these traits. If

these results are consistent with the ones reported for the In vitro digestibility of the NDF ( $r=-0.49$  to  $-0.56$  depending on the year), they diverged for the ones concerning the In vitro organic matter digestibility (non significant or weak positive correlations (both reported by Trouche *et al.*, 2014)). In addition, strong positive correlations were observed between the stem length and the lignin content.

Analyses of the correlation structure between the biomass production and biomass quality related traits for the different dwarfism classes highlighted major modifications of the traits relationships. These results suggest that if sorghum is a rich reservoir of genetic diversity for biomass quality improvement, breeding efforts should carefully take into consideration that this diversity is not structured in the same way for the different ideotypes (i.e dwarfism classes in this case).

### **Sowing effect significantly impacts the biomass composition related traits**

In combination with breeding efforts, cultural practices can also be mobilized to monitor biomass quality to fit the expectations of the different end-uses. In addition, in semi-arid conditions, the sowing date is of crucial importance to ensure significant grain and fodder production. A decrease of the stem soluble sugar production in relation with the sowing date have been repeatedly reported in temperate and semi-arid conditions (Burks *et al.*, 2013; Rao *et al.*, 2013; Almodares and Hoseini, 2016; Tovignan *et al.*, 2016) in the context of ethanol production. In terms of structural carbohydrates, no information is currently available regarding their when the sowing date is delayed. Based on three different sowing dates, we observed strong effects of the sowing date on the biomass quality. Delayed sowing led to highest feed quality (i.e. increase of dry and organic matter digestibilities) through a decrease of the lignin content (in relation to the dry matter or to the NDF portion) and an important increase of the hemicellulose content. However this biomass quality improvement was also accompanied, as expected, by an important reduction of the stem length (a key component of the biomass production). Although, the percentages of variation explained by the interactions between the photoperiod sensitivity classes and the sowing dates were small, a more stable biomass composition of the photoperiod sensitive genotypes were observed over the different sowing dates.

## **Conclusion**

The analysis of a panel of 35 sorghum genotypes encompassing short, tall, photoperiod-sensitive and insensitive cultivars over 6 experiments with different sowing dates in the Malian semi-arid conditions confirmed the large genetic variability available to develop varieties adapted to the different biomass value chains (animal feed, energy, biomaterials). A large effect of the dwarfism status of the genotypes on the structural carbohydrates (lignin, cellulose, hemicellulose) and on the biomass digestibility was underlined. At the opposite a limited effect of the earliness classes was observed for the same traits. Correlation analyses revealed negative relationships between biomass production (estimated through the stem length) and biomass quality with a positive correlation between stem length and lignin content and negatives ones with the biomass digestibility traits. Furthermore modifications of the correlations structures for the different dwarfism classes were detected suggesting the need to carefully design the breeding schemes depending on the size of the genotypes considered.

Significant impacts of the sowing dates on the structural carbohydrates were observed, delayed sowing leading to more digestible biomass composition.

**Table 1:** Identification, country of origin, racial classification, and agronomic characteristic (photoperiod sensitivity, duration of length cycle and plants height) of the 35 genotypes.

Accession name	Synonym	Country of origin <sup>a</sup>	Race <sup>b</sup>	Type <sup>c</sup>	Plant height <sup>d</sup>	Dwarfism <sup>e</sup>	Earliness <sup>f</sup>	Kp <sup>g</sup>
B35	IS 12555	ETH	D	BM	Short	Short	Early	Low
BimbG	BB_G_5_34	GIN	G	LR	Tall	Medium	Late	High
Dabitinnen	Malisor 84-7	MLI	C	BM	Short	Short	Early	Low
Doua-G		MLI	G	LR	Tall	Long	Late	High
E36-1, Zera	IS 30469	ETH	C	LR	Short	Medium	Early	Low
Fara Fara	IS 24887	NGA	G	LR	Tall	Medium	Late	High
Framida	IS 8744	ZAF	C	LR	Med	Medium	Early	Moderate
G03-1-118	V33/08	MLI	C/G	BM	Med	Short	Late	High
Gadiaba Dié	IS 25916	MLI	D	LR	Tall	Medium	Late	High
Ganga	IS 23540	ETH	C	LR	Short	Long	Early	Low
Gnossiconi		BFA	G	LR	Med	Long	Early	Moderate
Grinkan	02-SB-F4DT- 77	MLI	C/G	BM	Med	Short	Medium	Moderate
Hafijega	IS 23645	GMB	G	LR	Tall	Medium	Late	High
IRAT 204	IS 36563	SEN	C	LR	Short	Short	Early	Low
Jakumbe	CSM 63E	MLI	G	LR	Med	Long	Early	Low
Jigi Seme	CSM388	MLI	G	LR	Tall	Long	Medium	High
Kalaban	00KOF5DT19	MLI	C	LR	Short	Medium	Early	Low
Keninkeni	V248/08	MLI	G/C	BM	Med	Short	Medium	Moderate
Konoténé	IS 25705	MLI	D	LR	Tall	Medium	Late	High
L2P9 (1)		ERI	C	LR	Tall	Long	Early	Low
Lata3		MLI	G	BM	Tall	Medium	Medium	High
Lina-3		ICRISAT ??	G	LR	Tall	Long	Medium	High
Ngolofing	CSM 660	MLI	G	LR	Tall	Long	Medium	Moderate
Ribdahu		NGA	C	LR	Med	Medium	Late	High
Sambalma(4)		NGA	G	LR	Tall	Medium	Late	Moderate
Sanga Tigi	98-BE-F5P- 4	MLI	G	BM	Med	Long	Early	Moderate
SC566-14	PI 533871	NGA	C	BM	Short	Short	Early	Low
Seguetana CZ	IS 26057	MLI	G	LR	Tall	Long	Medium	High
Short Kaura	IS10699	NGA	D	LR	Tall	Short	Late	High
Soumalemba	IS15401	CMR	G	LR	Tall	Medium	Late	High
Soumba	Cirad406	FRA	G/C	BM	Med	Medium	Early	Low
Tiandougou	98 SB-F2-78	MLI	C/G	BM	Med	Short	Medium	Moderate
Tiéblé	CSM 335	MLI	G	LR	Tall	Long	Medium	High
Tiemantieteli	CSM 417	MLI	G	LR	Tall	Long	Late	High
White Kaura	SSV 20043	NGA	C	NA	Med	Short	Late	High

<sup>a</sup> MLI: Mali, ETH: Ethiopia, ZAF: South Africa, CMR: Cameroun, NGA: Nigeria, GMB: Gambia, BFA: Burkina Fasso, ERI: Eritrea, SEN: Senegal and GIN: Guinea.

<sup>b</sup> G: Guinea, D: Durra, C/G: Caudatum / Guinea, C/D: Caudatum / Durra, C: Caudatum, Durra/ Caudatum and G/C: Guinea/ Caudatum.

<sup>c</sup> LR: Landrace and BM: Breeding material

<sup>d</sup> Short: < 1m; Med: medium from 1 to 2m and Tall: > 2m

<sup>e</sup> Short : internode length < 10 cm, 10<Medium<15, Long : >15

<sup>f</sup> Early: < 70 days; Med: from 70 to 110 days and Late: > 110 days, measured in June sowing in the Malian condition.

<sup>g</sup> low: coefficient of photoperiod (Kp) < 0.4; Mod: Moderate Kp from 0.4 to 0.6 and High: Kp from 0.7 to 1



Table 2: Phenotypic variability observed over the three years of experiments for the first sowing dates

Trait Category	Traits (acronym, description, unit)	Mean			Range			PCV%		
		2010	2014	2015	2010	2014	2015	2010	2014	2015
Phenology	DIFL, Duration from sowing to Flag Leaf, days	80.79	87.3	87.09	46-114	47-147	45-137	19.39	25.19	23.68
	TLN , Total Leaf Number	26.12	29.91	30.56	18-34	19-43	20-42	16.93	20.04	17.69
Plant Architecture / Biomass production	NGL, Number of green leaves at maturity	6.99	4.68	1.91	2-12.67	0-9.25	0-7.5	30.19	46.46	108.75
	STEL, Stem length, cm	223.54	232.54	221.97	22.3-397.7	37.8-376.1	26.8-394.3	42.8	39.03	42.51
	INTN, Number of elongated internodes	16.27	19.9	17.52	7 - 27	9 - 30	8 - 27	27.9	25.28	24.86
	MINTL, Mean internode length, cm	13.67	11.68	12.52	2.48-21.62	2.91-18.5	2.44-21.08	36.49	35.33	36.97
	SFW, Stem fresh weight, g	nd	1702.54	1443.05	nd	260.5-3816.3	236.8-3475.3	nd	42.42	47.12
	Brix, Sugar content, °Brix	13.99	17.5	14.69	6-19.2	12.4-21.4	6 - 21	23.22	11.37	26.96
	SJW, Stem juice weight, g	nd	279.72	230.56	nd	43.7-1038.5	5.1-996.4	nd	58.21	83.67
	SSC, Stem sugar concentration, g/stem	nd	10.22	8.99	nd	1.57-38.21	0.09-41.85	nd	61.35	95.76
	SMEX, Stem moisture extractable, %	nd	17.95	15.84	nd	0.03-0.35	0.01-0.39	nd	48.26	60.53
	CP, Crude protein, %DM	4.24	2.93	2.76	1.46-9.94	0-8.07	0.04-10.72	33.44	54.49	79.79
Stem composition	NDF, Stem neutral detergent fiber, %DM	70.87	70.83	69.54	58.69-80.5	55.22-79.68	56.87-79.87	6.46	6.92	6.91
	ADF, Stem acid detergent fiber, %DM	45.38	43.84	43.66	34.22-53.81	34.55-51.71	33.32-52.03	9.12	8.46	9.04
	ADL, Stem acid detergent lignin, %DM	8.14	8.4	7.87	4.88-10.59	5.87-11.03	4-10.55	14.28	13.4	16.98
	Lig, Stem lignin content, %NDF	11.45	11.84	11.29	7.71-13.22	9.01-14.95	6.38-13.86	10.18	9.92	14.04
	Cell, Stem cellulose content, %NDF	52.48	50.01	51.43	47.18-56.2	45.98-54.28	47.59-53.75	2.98	2.67	2.39
	Hem, Stem hemicellulose content, %NDF	36.08	38.15	37.28	32.74-42.44	33.78-42.81	32.96-46	6.35	4.77	6.61
	IVDMD, In Vitro Dry matter digestibility, %DM	31.72	29.41	34.91	18.62-49.34	16.72-43.47	21.14-50.15	19.06	17.7	16.64
	IVOMD, In vitro organic matter digestibility, %OM	29.03	28.32	33.02	15.46-46.31	16.21-43.84	20.01-48.88	20.81	18.85	16.62
	IVNDFD, In vitro NDF digestibility, %NDF	13.79	7.51	7.46	4.67-28.39	0-25.31	0-29.27	35.87	61.61	73.43

Table 3: Coefficients of genetic variation, heritabilities and variance components (expressed as the percentages of phenotypic variation explained by the different effects) of the phenology, agromorphology and biomass quality traits measured over 3 years of experiment for the first sowing dates (June).

Trait Category	Traits	CvG <sup>a</sup>	H <sup>2</sup> bs <sup>a</sup>	Year	Dw	Earliness	Gen	Bloc(Year)	GenxYear
Phenology	DTFL	31.93	0.99	2.32 ***	14.73 ***	57.11 ***	23.46 ***	0.07 ns	1.21 *
	TLN	24.51	0.96	11.91 ***	13.08 ***	50.84 ***	15.97 ***	0.60 ***	4.57 ***
	NGL	33.88	0.46	48.64 ***	0.21 ns	1.85 ***	15.21 ***	0.48 ns	24.06 ***
	STEL	55.74	0.98	0.25 *	61.35 ***	24.54 ***	7.58 ***	1.20 ***	1.75 ns
	INTN	34.46	0.96	9.73 ***	12.19 ***	52.57 ***	18.25 ***	0.33 *	3.65 *
Biomass production	MINTL	49.27	0.98	3.13 ***	82.25 ***	1.08 ***	8.22 ***	0.81 ***	2.23 *
	SFW	46.37	0.75	3.30 **	24.07 ***	15.91 ***	27.03 ***	2.39 *	5.90 ns
	Brix	19.77	0.67	17.46 ***	11.42 ***	4.82 ***	26.31 ***	2.81 ***	24.65 ***
	SJW	63.1	0.59	1.85 *	7.17 ***	3.60 *	46.69 ***	1.36 ns	11.44 ns
	SSC	72.33	0.58	0.65 ns	6.32 ***	1.01 ns	50.06 ***	1.64 ns	16.18 ns
Stem composition	SMEX	58.31	0.74	1.31 *	21.15 ***	22.52 ***	27.28 ***	0.23 ns	8.82 ns
	CP	47.99	0.68	12.95 ***	12.38 ***	1.48 *	33.95 ***	0.76 ns	21.85 **
	NDF	5.12	0.52	1.46 ns	6.14 ***	0.05 ns	34.19 ***	1.75 ns	32.64 **
	ADF	7.77	0.63	3.68 ***	15.19 ***	0.10 ns	32.66 ***	1.05 ns	28.53 ***
	ADL	14.08	0.70	3.01 ***	26.06 ***	1.12 ns	28.71 ***	0.09 ns	25.32 ***
composition	Lig	10.59	0.70	3.08 ***	27.19 ***	1.54 *	25.92 ***	1.02 ns	23.65 **
	Cell	2.55	0.70	36.09 ***	7.52 ***	0.95 *	27.00 ***	1.32 *	14.43 *
	Hem	5.73	0.73	14.16 ***	24.54 ***	0.17 ns	26.69 ***	0.77 ns	17.80 *
	IVDMD	15.84	0.65	12.96 ***	17.40 ***	0.03 ns	27.63 ***	0.17 ns	23.41 **
	IVOMD	16.09	0.61	10.79 ***	13.20 ***	0.13 ns	30.24 ***	0.31 ns	25.72 **
IVNDFD	52.47	0.71	27.08 ***	22.75 ***	0.03 ns	19.43 ***	0.28 ns	18.10 ***	

<sup>a</sup> Broad sense heritabilities and coefficients of genetic variation were estimated according to the Model 1 whereas the others result presented in this table were obtained according to Model 2.

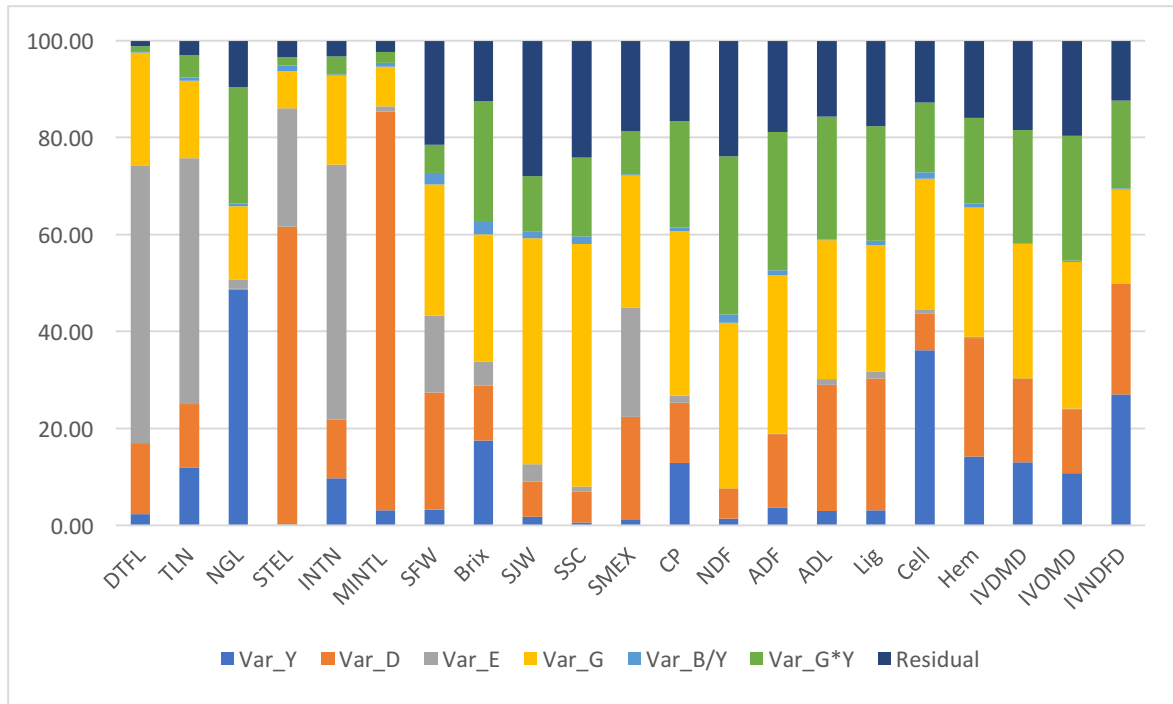


Figure 1: Percentages of phenotypic variation explained by the different effects tested in model 2 according to the data produced for the first sowing date over the three years of experiment. Y stands for Year effect, D for dwarfism, E for Earliness, G for genotypes, B/Y for the block effect nested in the year, and Residual for the residual effect.

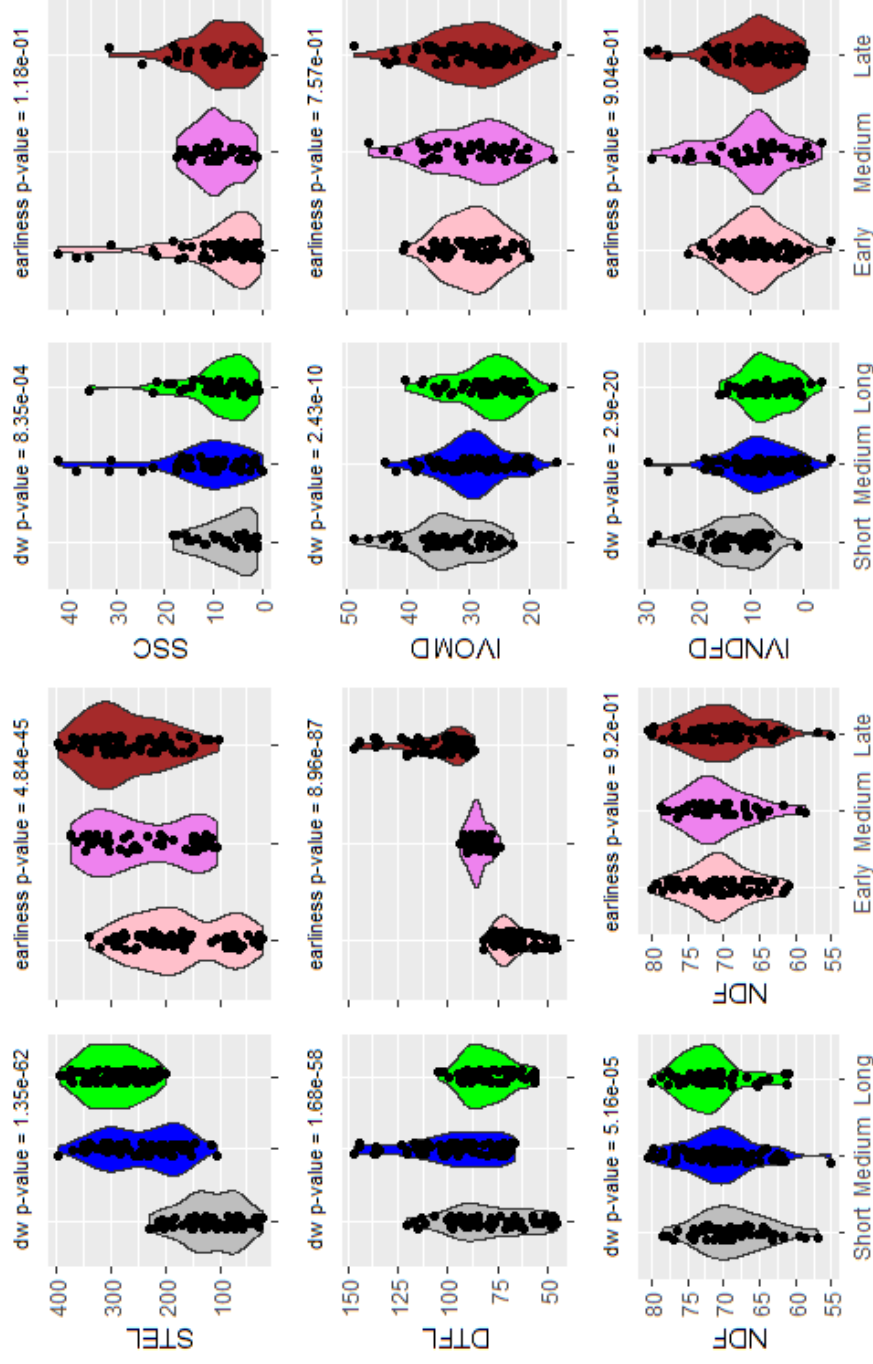


Figure 2: Dwarfism and earliness effects on 6 traits representative of the variability of the different phenotypic categories. Significance of the dwarfism and earliness effects have been evaluated using Model 2. The phenotypic traits presented here have been selected to represent the different phenotypic categories according to the hierarchical clustering presented in Figure 3 in which they are indicated with a red circle.

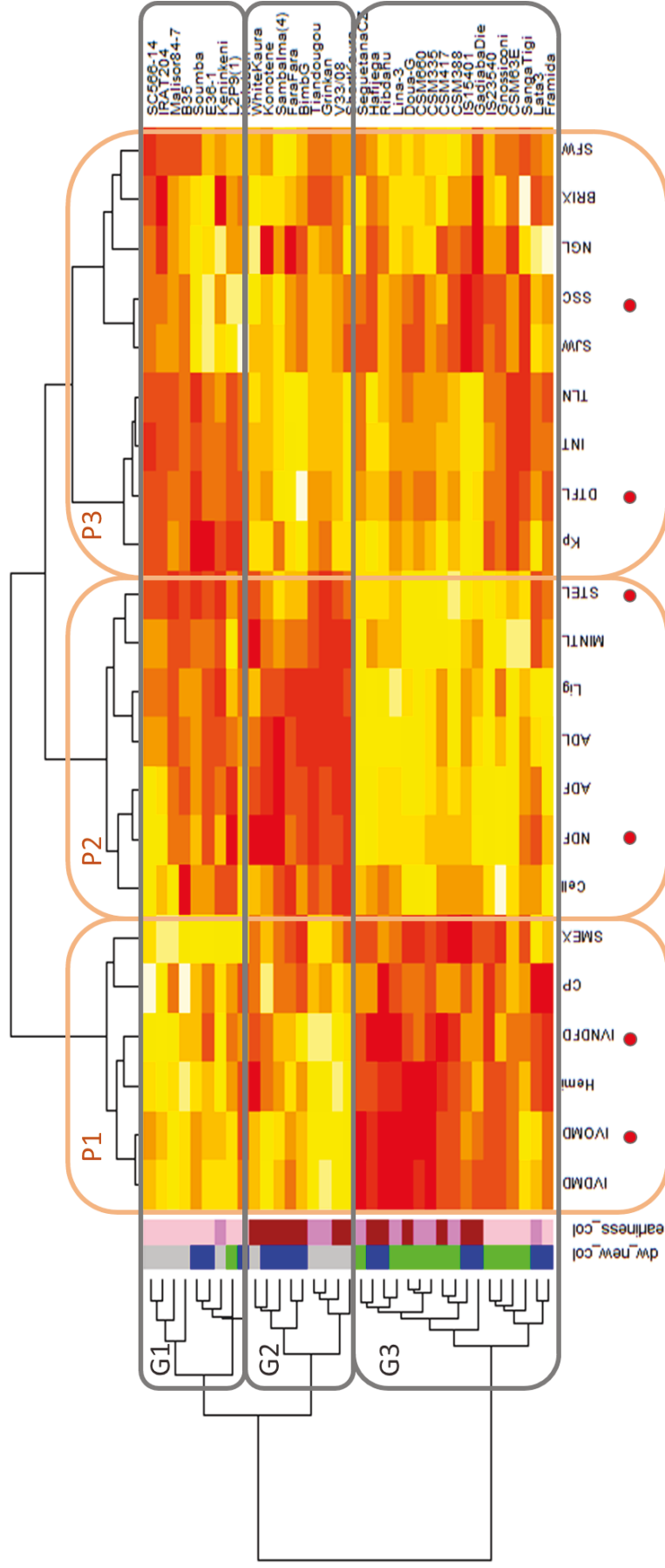


Figure 3: Multi-traits variability of the 35 genotypes analyzed in this study. Dwarfism and Earliness classes are provided on the left hand side. Genotypes presenting short internodes are indicated in grey, medium ones in blue and tall ones are in green. Early sowing dates are indicated in pink, medium ones in violet and late ones in purple. BLUP from the 3 years of experiment for the first sowing dates have been used to perform this clustering analysis. The yellow color corresponds to high values, whereas the red color corresponds to low values. The traits selected to describe the dwarfism and earliness effects in Figure 2 are indicated by a red circle.

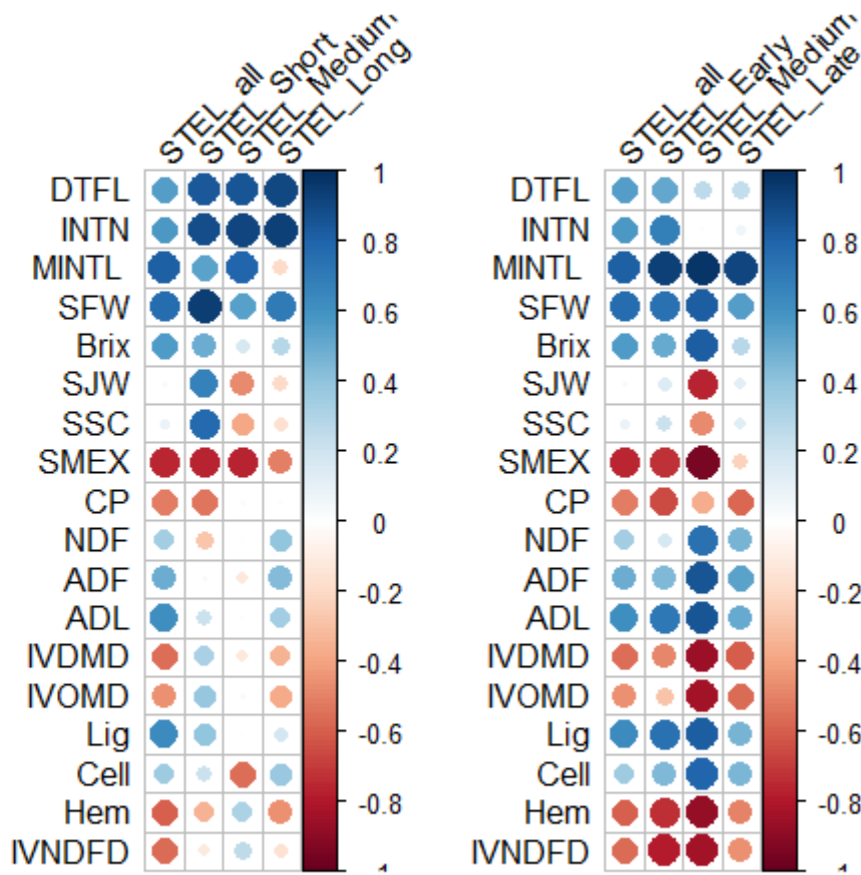


Figure 4: Evolutions of the correlations between the stem length (STEL), cycle duration (DTFL), height components (INTN, MINTL) and biomass quality related traits (the others traits) over the different dwarfism (left side) and earliness classes (right side). The correlations presented with STEL\_all correspond to the ones observed on the whole panel of genotypes (n=35), whereas the ones presented for STEL\_Short (n=10), STEL\_Medium (n=13) and STEL\_Long (n=12) correspond to the different dwarfism classes and the ones presented for STEL\_Early (n=13), STEL\_Medium (n=9) and STEL\_Late (n=13) correspond to the different earliness classes. The 5% significance thresholds of the correlation's values for the different dwarfism classes are 0.64, 0.55 and 0.57 for the Short, Medium and Long classes respectively. For the different earliness classes, the 5% significance thresholds of the correlation values are 0.55, 0.66 and 0.55 for the Early, Medium and Late classes respectively. For the whole set of genotypes, the threshold value to declare a correlation significant at the 5% level is 0.41.

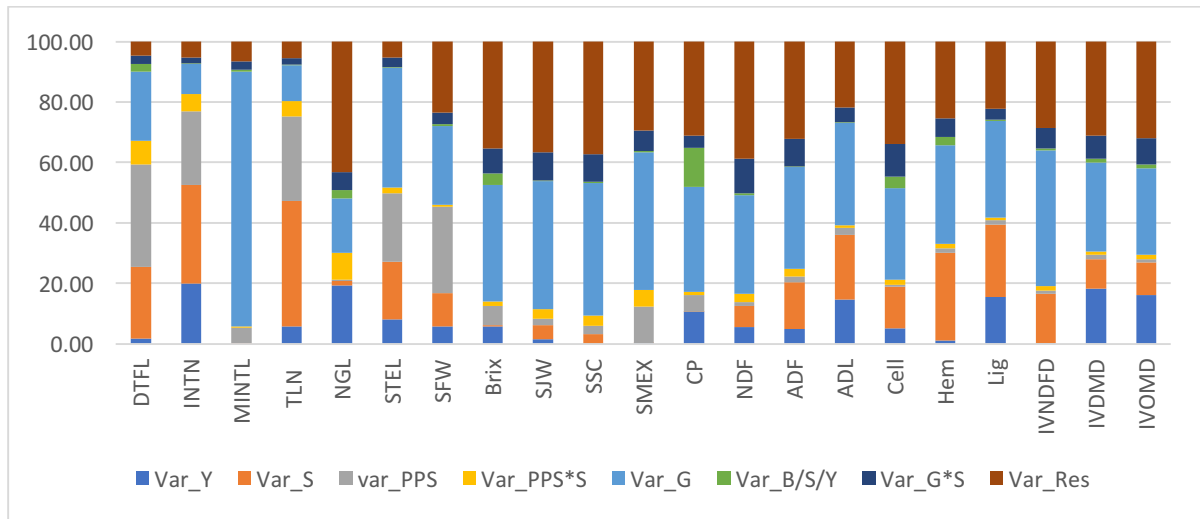


Figure 5: Percentages of phenotypic variation explained by the different effects tested in model 3. Y stands for the Year, S for the sowing date, PPS for the photoperiod sensitivity class, PPS\*S for the interaction between the photoperiod sensitivity class and the sowing date, G for the genotype effect, B/S/Y for the block nested in the sowing date and in the year, G\*S for the interaction between the genotype and the sowing and Residual for the residual effects.

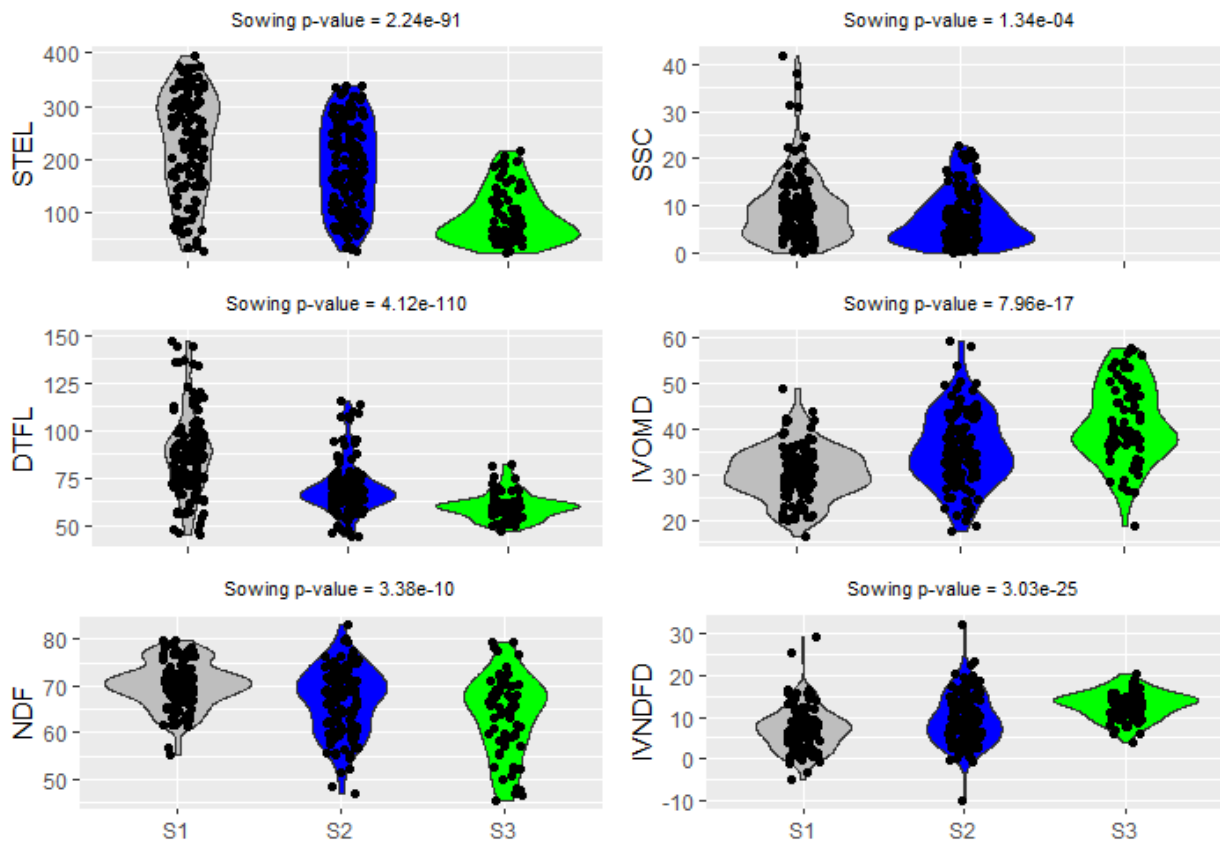


Figure 6: Sowing date effects on 6 traits representative of the variability of the different phenotypic categories. Significance of the sowing effects have been evaluated using Model 3. To facilitate the representation, the sowing date has been considered here as a factor although in the linear model 3 it is considered as a quantitative variable (i.e. it is the number of days from the earliest sowing that is considered). The phenotypic traits presented here have been selected to represent the different phenotypic categories according to the hierarchical clustering presented in Figure 3 in which they are indicated with a red circle.



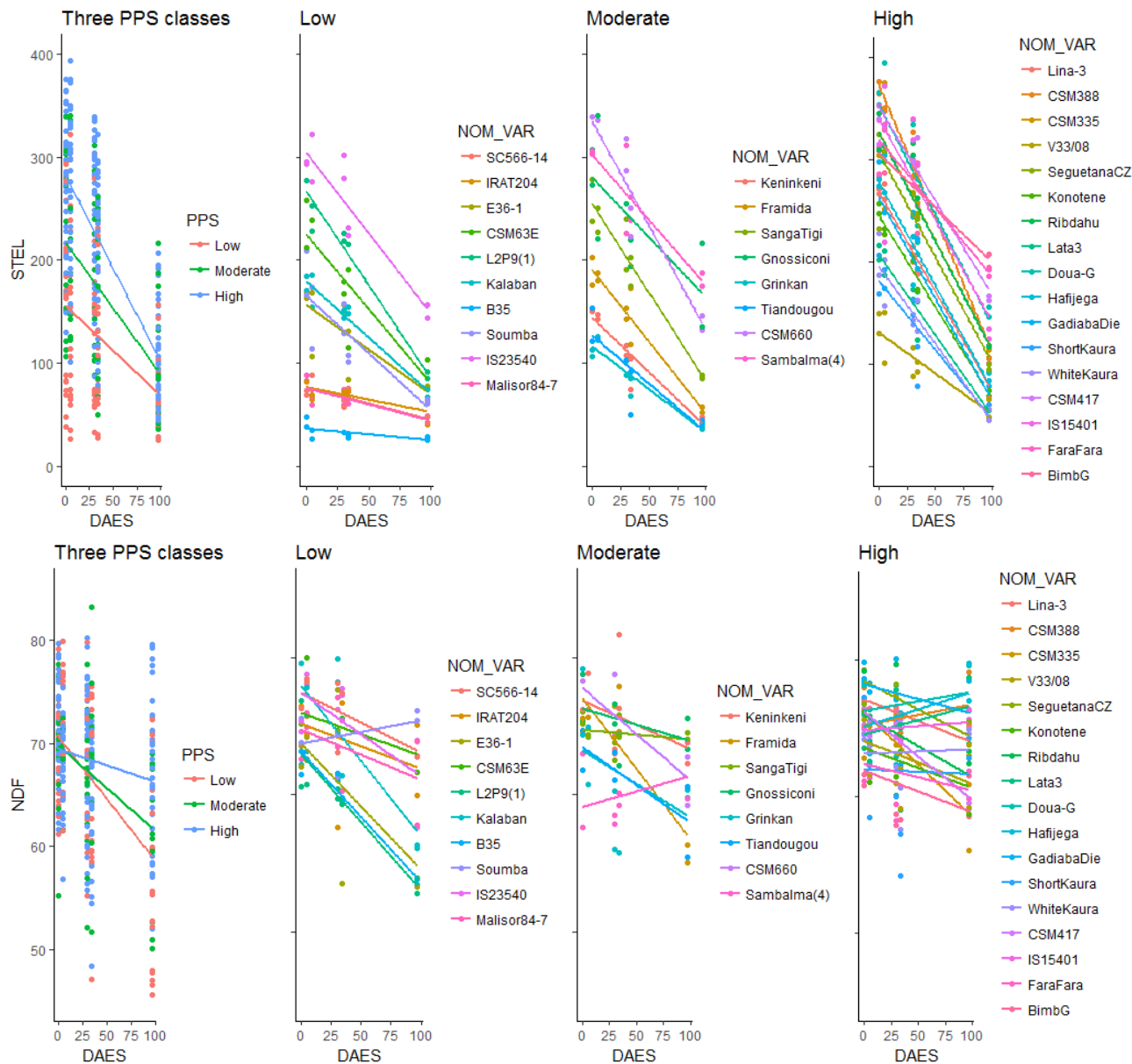


Figure 7 : Effect of the sowing date on the stem length (STEL) and the cell wall content of the biomass (NDF) for the whole set of genotypes (Left side) and for the different classes of Photoperiod sensitivity (Low, Moderate, High).

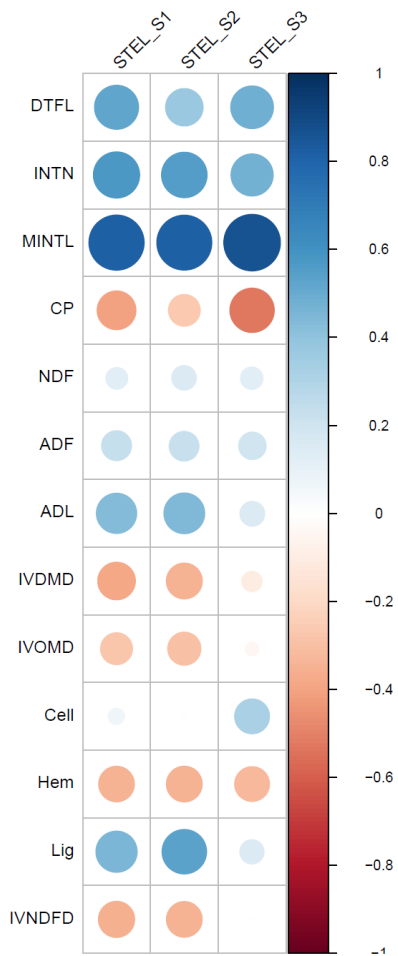


Figure 8: Evolutions of the correlations between stem length (STEL), cycle duration (DTFL), height components (INTN, MINTL) and biomass quality related traits over the three sowing dates. The threshold value to declare a correlation significant at the 5% level is 0.41.

Supplementary material

**Table S1:** Prediction models of the Van Soest and Aufrere stem biomass quality related traits.

Traits	Analyzed collection				Calibration		Validation	
	N	Mean	SD	CV %	SEC	RSQ	SECV	RPD
CP	518	3.73	2.21	59.3	0.43	0.96	0.5	4.4
NDF	455	68.77	10.9	15.85	1.36	0.98	1.57	6.9
ADF	461	41.66	8.58	20.61	1.24	0.98	1.41	6.1
ADL	463	6.46	2.56	39.58	0.56	0.95	0.66	3.9
IVDMD	467	38.48	12.33	32.04	1.69	0.98	1.91	6.5
IVDMD	463	36.47	12.16	33.34	1.74	0.98	1.9	6.4

Table S2 : Inter-annual correlations for the first sowing dates

Trait Category	Traits	r (2010-2014)	r (2010-2015)	r (2014-2015)
Phenology	DTFL	0.98***	0.97***	0.99***
	TLN	0.9***	0.88***	0.96***
Plant	NGL	-0.05	0.3	0.1
Architecture /	STEL	0.98***	0.97***	0.97***
Biomass production	INTN	0.93***	0.92***	0.94***
	MITNL	0.97***	0.97***	0.98***
	SFW	NA	NA	0.83***
	Brix	0.46*	0.54**	0.39*
Stem composition	SJW	NA	NA	0.65***
	SSC	NA	NA	0.55**
	SMEX	NA	NA	0.76***
	CP	0.53**	0.5*	0.61**
	NDF	0.42*	0.43*	0.16
	ADF	0.51**	0.47*	0.38
	ADL	0.61***	0.56**	0.49*
	Lig	0.59***	0.6**	0.53**
	Cell	0.61***	0.56**	0.56**
	Hem	0.68***	0.63***	0.67***
	IVDMD	0.53**	0.55**	0.19
	IVOMD	0.54**	0.51*	0.1
IVNDFD	0.73***	0.47*	0.56**	

\*: Significant at 5% level probability, \*\*: significant at 1% level probability and \*\*\*: significant at 0.1% level probability. Correlations without asterisks are non-significant.

Table S3 : Pairwise correlations beetwen the phenology, biomass production and biomass quality related traits.

Traits	Phenology and biomass production							
	DTFL	TLN	NGL	STEL	INTN	MITNL	SFW	
BRIX	0.39*	0.33	0.52**	0.56***	0.39*	0.43*	0.63***	
SJW	0.06	0.07	0.38*	-0.03	0.09	-0.05	0.5**	
SSC	0.12	0.13	0.37*	0.08	0.15	0.05	0.58***	
SMEX	-	-	0.09	-0.76***	-	-	-0.36*	
	0.5**	0.52**			0.54***	0.56***		
CP	-0.27	-0.24	-0.24	-0.52**	-0.32	-0.46**	-0.48**	
Stem composition	NDF	-0.18	-0.13	-0.22	0.35*	-0.12	0.45**	-0.07
	ADF	-0.17	-0.12	-0.07	0.49**	-0.09	0.62***	0.09
	ADL	-0.07	0	0.12	0.62***	0.03	0.71***	0.25
	Lig	-0.02	0.04	0.27	0.63***	0.08	0.71***	0.36*
	Cell	-0.18	-0.17	0.03	0.35*	-0.13	0.52**	0.18
	Hem	0.12	0.08	-0.18	-0.59***	0.03	-	-0.32
							0.74***	
	IVDMD	0.05	0.02	0.08	-0.56***	0.01	-	-0.16
							0.63***	
	IVOMD	0.16	0.12	0.1	-0.46**	0.11	-	-0.06
						0.57***		
IVNDFD	0.14	0.08	-0.19	-0.57***	0.04	-0.7***	-0.26	

\*: Significant at 5% level probability, \*\*: significant at 1% level probability and \*\*\*: significant at 0.1% level probability. NA: traits not measured for one of the two years considered

Table S4: Variance components (expressed as the percentages of phenotypic variation explained by the different effects) of the phenology, agromorphology and biomass quality traits measured over 3 years of experiment for the different sowing dates.

Trait Category	Traits	Year	Sowing	PPS	PPS*Sowing	Gen	B (Year/Sowing)	Gen*Sowing
Phenology	DTFL	1.68 ***	23.72 ***	33.97 ***	7.83 ***	22.81 ***	2.51 ***	2.93 ***
	TLN	5.79 ***	41.39 ***	27.97 ***	5.20 ***	11.84 ***	0.11 ns	2.25 ***
Plant Architecture / Biomass production	NGL	19.36 ***	1.48 **	0.28 ns	9.04 ***	17.92 ***	2.81 **	5.94 ns
	STEL	7.99 ***	19.10 ***	22.70 ***	1.88 ***	39.62 ***	0.15 ns	3.25 ***
	INTN	19.88 ***	32.58 ***	24.46 ***	5.64 ***	10.12 ***	0.14 ns	1.94 ***
	MINTL	0.06 ns	0.07 ns	5.13 ***	0.53 ***	84.27 ***	0.60 ***	2.84 ***
	SFW	5.73 ***	11.10 ***	28.47 ***	0.59 ns	26.13 ***	0.70 ns	3.81 ns
	Brix	5.68 ***	0.53 ns	6.38 ***	1.46 *	38.47 ***	3.91 ***	8.19 ns
Stem composition	SJW	1.55 **	4.55 ***	2.14 **	3.25 ***	42.33 ***	0.23 ns	9.24 ns
	SSC	0.06 ns	3.10 ***	2.82 **	3.37 ***	43.91 ***	0.41 ns	9.02 ns
	SMEX	0.12 ns	0.00 ns	12.12 ***	5.51 ***	45.62 ***	0.44 ns	6.83 ns
	CP	10.55 ***	0.00 ns	5.61 ***	1.07 *	34.61 ***	13.10 ***	3.92 ns
	NDF	5.44 ***	6.97 ***	1.30 *	2.83 ***	32.68 ***	0.69 ns	11.36 ***
	ADF	4.88 ***	15.46 ***	1.84 **	2.59 ***	33.81 ***	0.21 ns	8.97 **
Stem composition	ADL	14.69 ***	21.37 ***	2.26 ***	0.92 **	33.86 ***	0.13 ns	4.90 *
	Lig	15.41 ***	23.97 ***	1.56 ***	0.79 *	31.98 ***	0.50 ns	3.54 ns
	Cell	5.12 ***	13.73 ***	0.70 ns	1.64 **	30.21 ***	3.88 ***	10.93 ***
	Hem	1.06 **	29.11 ***	1.47 **	1.40 **	32.60 ***	2.76 ***	6.25 **
	IVDMD	18.27 ***	9.75 ***	1.47 **	1.10 *	29.31 ***	1.39 *	7.64 **
	IVOMD	16.02 ***	10.82 ***	1.08 *	1.50 **	28.67 ***	1.28 *	8.58 **
IVNDFD	0.15 ns	16.38 ***	1.12 *	1.37 **	44.90 ***	0.68 ns	6.71 *	

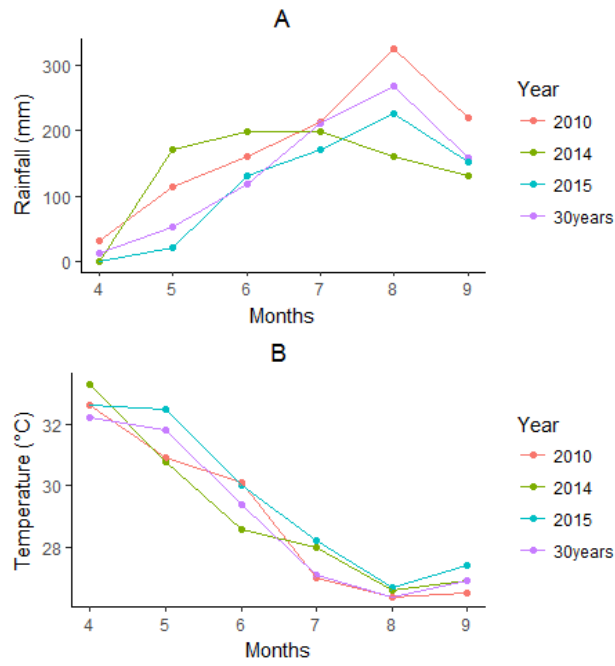


Figure S1: Precipitation (A) and temperature (B) patterns for the three years of experiment and for the last 30 years.

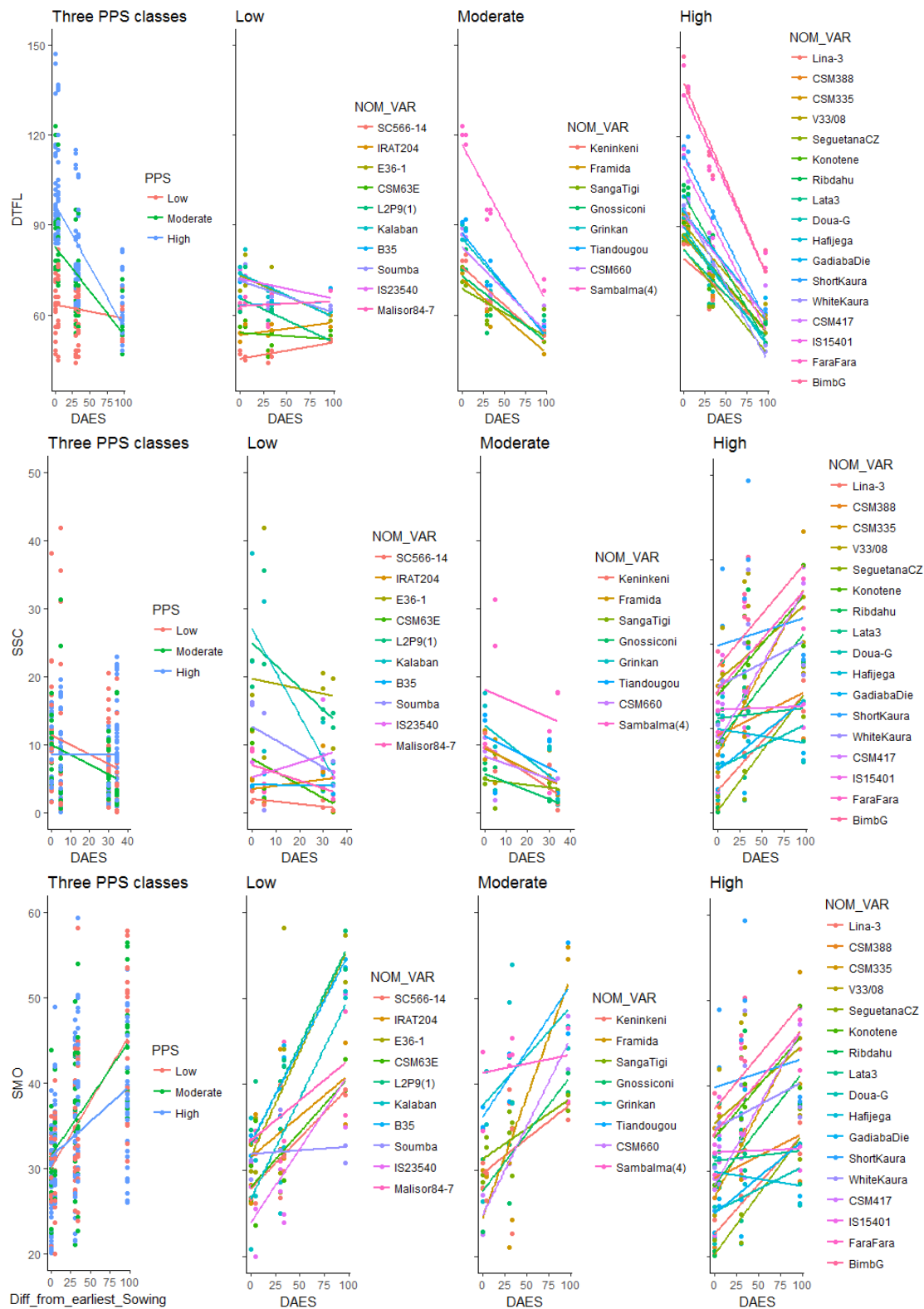


Figure S2: Effects of the sowing dates on the cycle duration on (DTFL), the soluble sugar concentration, of the setm (SSC) and the digestibility of the organic matter for the whole set of genotypes (Left side) and for the different classes of Photoperiod sensitivity (Low, Moderate, High). The stem sugar concentration (SSC) was not available for the latest sowing date.



## References

- AFNOR (1977). "Aliments des animaux. Dosage des cendres brutes." Animal Feed Science and Technology **18-101**: 2pp.
- AFNOR (2009). "Aliments des animaux. Dosage de l'azote et calcul de la teneur en protéines brutes – Partie 2 : Méthode de digestion ne bloc et distillation à la vapeur." Animal Feed Science and Technology(5983-2): 16pp.
- Almodares, A. and Hoseini, S. H. (2016) 'Effect of sowing dates and nitrogen levels for ethanol production from sweet sorghum stalks and grains', *African Journal of Agricultural Research*, 11(4), pp. 266–275. doi: 10.5897/AJAR2015.9765.
- Atis, I., O. Konuskan, M. Duru, H. Gozubenli and S. Yilmaz (2012). "Effect of harvesting time on yield, composition and forage quality of some forage sorghum cultivars." Int. J. Agric. Biol **14**: 879-886.
- Ayub, M., M. A. Nadeem, A. Tanveer and A. Husnain (2002). "Effect of different levels of nitrogen and harvesting times on the growth, yield and quality of sorghum fodder." Asian J. Plant Sci **4**: 304-307.
- Badu-Debrah, A. (1990). The effect of plant population density on growth and yield of maize.
- Bayu, W., N. Rethman and P. Hammes (2005). "Growth and yield compensation in sorghum (*Sorghum bicolor* L. Moench) as a function of planting density and nitrogen fertilizer in semi-arid areas of northeastern Ethiopia." South African Journal of Plant and Soil **22**(2): 76-83.
- Brenton, Z. W., E. A. Cooper, M. T. Myers, R. E. Boyles, N. Shakoob, K. J. Zielinski, B. L. Rauh, W. C. Bridges, G. P. Morris and S. Kresovich (2016). "A genomic resource for the development, improvement, and exploitation of sorghum for bioenergy." Genetics **204**(1): 21-33.
- Brown, P., P. Klein, E. Bortiri, C. Acharya, W. Rooney and S. Kresovich (2006). "Inheritance of inflorescence architecture in sorghum." Theoretical and Applied Genetics **113**(5): 931-942.
- Brown, P. J., W. L. Rooney, C. Franks and S. Kresovich (2008). "Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes." Genetics **180**(1): 629-637.
- Burks, P. S., C. M. Kaiser, E. M. Hawkins and P. J. Brown (2015). "Genomewide Association for Sugar Yield in Sweet Sorghum." Crop Science **55**(5): 2138-2148.
- Burks, P. S. *et al.* (2013) 'The Influence of Hybrid Maturity and Planting Date on Sweet Sorghum Productivity during a Harvest Season', *Agronomy Journal*, 105(1), pp. 263–267. doi: 10.2134/agronj2012.0317.
- ÇAKMAKÇI, S., İ. GÜNDÜZ, M. A. TÜSÜZ, S. ÇEÇEN and B. AYDINOĞLU (1999). "Effects of different harvesting times on yield and quality of sorghum silage (*Sorghum bicolor* L.)." Turkish Journal of Agriculture and Forestry **23**(EK3): 603-612.
- Carmi, A., Y. Aharoni, M. Edelstein, N. Umiel, A. Hagiladi, E. Yosef, M. Nikbachat, A. Zenou and J. Miron (2006). "Effects of irrigation and plant density on yield, composition and in vitro digestibility of a new forage sorghum variety, Tal, at two maturity stages." Animal feed science and technology **131**(1): 121-133.
- Carpita, N. C. (2012). "Progress in the biological synthesis of the plant cell wall: new ideas for improving biomass for bioenergy." Current opinion in biotechnology **23**(3): 330-337.
- Casady, A. (1967). "Effect of a single height gene (Dw3) of *Sorghum vulgare* Pers. on certain culm and leaf blade characteristics." Crop Science **7**(6): 595-598.

- Childs, K. L., F. R. Miller, M.-M. Cordonnier-Pratt, L. H. Pratt, P. W. Morgan and J. E. Mullet (1997). "The sorghum photoperiod sensitivity gene, Ma3, encodes a phytochrome B." *Plant Physiology* **113**(2): 611-619.
- Clerget, B., H. Rattunde, S. Dagnoko and J. Chantreau (2007). "An easy way to assess photoperiod sensitivity in sorghum: Relationships of the vegetative-phase duration and photoperiod sensitivity." *Journal of SAT agricultural research* **3**(1): 1-3.
- FAOSTAT (2015). "United Nations Food and Agriculture Organization Statistical Database. Available Online at: <http://faostat.fao.org/>."
- Fedenko, J. R. (2011). Biomass yield and composition of potential bioenergy feedstocks, University of Florida.
- Fernandez, M. G. S., P. W. Becraft, Y. Yin and T. Lübberstedt (2009). "From dwarves to giants? Plant height manipulation for biomass yield." *Trends in plant science* **14**(8): 454-461.
- Glamoclija, D., S. Jankovic, S. Rakic, R. Maletic, J. Ikanovic and Z. Lakić (2011). "Effects of nitrogen and harvesting time on chemical composition of biomass of Sudan grass, fodder sorghum, and their hybrid." *Turkish Journal of Agriculture and Forestry* **35**(2): 127-138.
- Hanumappa, M., L. H. Pratt, M. M. Cordonnier-Pratt and G. F. Deitzer (1999). "A photoperiod-insensitive barley line contains a light-labile phytochrome B." *Plant Physiol* **119**.
- Higgins, R., C. Thurber, I. Assaranurak and P. Brown (2014). "Multiparental mapping of plant height and flowering time QTL in partially isogenic sorghum families." *G3: Genes| Genomes| Genetics* **4**(9): 1593-1602.
- Izawa, T., T. Oikawa, N. Sugiyama, T. Tanisaka, M. Yano and K. Shimamoto (2002). "Phytochrome mediates the external light signal to repress FT orthologs in photoperiodic flowering of rice." *Genes Dev* **16**.
- Kjeldahl, J. (1883). "A new method for the determination of nitrogen in organic matter." *Z. Anal. Chem* **22**(1): 366-382.
- Kouressy, M., S. Traoré, M. Vaksman, M. Grum, I. Maikano, M. Soumaré, P. S. Traoré, D. Bazile, M. Dingkuhn and A. Sidibé (2008). "Adaptation des sorghos du Mali à la variabilité climatique." *Cahiers Agricultures* **17**(2): 95-100.
- Mace, E., C. Hunt and D. Jordan (2013). "Supermodels: sorghum and maize provide mutual insight into the genetics of flowering time." *Theoretical and Applied Genetics* **126**(5): 1377-1395.
- Mace, E. and D. Jordan (2011). "Integrating sorghum whole genome sequence information with a compendium of sorghum QTL studies reveals uneven distribution of QTL and of gene-rich regions with significant implications for crop improvement." *Theoretical and applied genetics* **123**(1): 169-191.
- Mace, E. S., C. H. Hunt and D. R. Jordan (2013). "Supermodels: sorghum and maize provide mutual insight into the genetics of flowering time." *Theor Appl Genet* **126**.
- Maucieri, C., V. Cavallaro, C. Caruso, M. Borin, M. Milani and A. C. Barbera (2016). "Sorghum Biomass Production for Energy Purpose Using Treated Urban Wastewater and Different Fertilization in a Mediterranean Environment." *Agriculture* **6**(4): 67.
- May, A., V. F. d. Souza, G. d. A. Gravina and P. G. Fernandes (2016). "Plant population and row spacing on biomass sorghum yield performance." *Ciência Rural* **46**(3): 434-439.
- Multani, D. S., S. P. Briggs, M. A. Chamberlin, J. J. Blakeslee, A. S. Murphy and G. S. Johal (2003). "Loss of an MDR Transporter in Compact Stalks of Maize br2 and Sorghum dw3 Mutants." *Science* **302**(5642): 81-84.

- brown midrib 12 (Bmr12), the caffeic O-methyltransferase (COMT) of sorghum." *BioEnergy Research* **5**(4): 855-865.
- Sattler, S. E., A. Saballos, Z. Xin, D. L. Funnell-Harris, W. Vermerris and J. F. Pedersen (2014). "Characterization of novel Sorghum brown midrib mutants from an EMS-mutagenized population." *G3 (Bethesda)* **4**(11): 2115-2124.
- Shiringani, A. L. and W. Friedt (2011). "QTL for fibre-related traits in grain  $\times$  sweet sorghum as a tool for the enhancement of sorghum as a biomass crop." *Theoretical and Applied Genetics* **123**(6): 999-1011.
- Shiringani, A. L., Frisch, M. and Friedt, W. (2010) 'Genetic mapping of QTLs for sugar-related traits in a RIL population of Sorghum bicolor L. Moench', *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, **121**(2), pp. 323–336. doi: 10.1007/s00122-010-1312-y.
- Stephens, J., F. Miller and D. Rosenow (1967). "Conversion of alien sorghums to early combine genotypes." *Crop Science* **7**(4): 396-396.
- TRAORE, S. B., F.-N. REYNIERS, M. VAKSMANN, K. Bather, A. SIDIBE, A. YOROTE, K. YATTARA and M. KOURESSY (2001). "Adaptation à la sécheresse des écotypes locaux de sorghos du Mali." *Science et changements planétaires/Sécheresse* **11**(4): 227-237.
- Tovignan, T. K. *et al.* (2016) 'The sowing date and post-flowering water status affect the sugar and grain production of photoperiodic, sweet sorghum through the regulation of sink size and leaf area dynamics', *Field Crops Research*, **192**(Supplement C), pp. 67–77. doi: 10.1016/j.fcr.2016.04.015.
- Trouche, G., D. Bastianelli, T. C. Hamadou, J. Chantereau, J.-F. Rami and D. Pot (2014). "Exploring the variability of a photoperiod-insensitive sorghum genetic panel for stem composition and related traits in temperate environments." *Field Crops Research* **166**: 72-81.
- Van Soest, P. J., J. B. Robertson and B. A. Lewis (1991). "Methods for Dietary Fiber, Neutral Detergent Fiber, and Nonstarch Polysaccharides in Relation to Animal Nutrition." *Journal of Dairy Science* **74**(10): 3583-3597.
- Vandenbrink, J. P., M. P. Delgado, J. R. Frederick and F. A. Feltus (2010). "A sorghum diversity panel biofuel feedstock screen for genotypes with high hydrolysis yield potential." *Industrial Crops and Products* **31**(3): 444-448.
- Wilson, R. T. (1986). Livestock production in central Mali: Long-term studies on cattle and small ruminants in the agropastoral system, ILRI (aka ILCA and ILRAD).
- Windscheffel, J. A., R. Vanderlip and A. Casady (1973). "Performance of 2-dwarf and 3-dwarf grain sorghum hybrids harvested at various moisture contents." *Crop Science* **13**(2): 215-219.
- Yamaguchi, M., H. Fujimoto, K. Hirano, S. Araki-Nakamura, K. Ohmae-Shinohara, A. Fujii, M. Tsunashima, X. J. Song, Y. Ito, R. Nagae, J. Wu, H. Mizuno, J.-i. Yonemaru, T. Matsumoto, H. Kitano, M. Matsuoka, S. Kasuga and T. Sazuka (2016). "Sorghum Dw1, an agronomically important gene for lodging resistance, encodes a novel protein involved in cell proliferation." *Scientific Reports* **6**: 28366.
- Zhao, J., M. B. Mantilla Perez, J. Hu and M. G. Salas Fernandez (2016). "Genome-Wide Association Study for Nine Plant Architecture Traits in Sorghum." *The Plant Genome*.
- Zhao, Y. L., Y. Steinberger, M. Shi, L. P. Han and G. H. Xie (2012). "Changes in stem composition and harvested produce of sweet sorghum during the period from maturity to a sequence of delayed harvest dates." *biomass and bioenergy* **39**: 261-273.

- Murphy, R. L., R. R. Klein, D. T. Morishige, J. A. Brady, W. L. Rooney, F. R. Miller, D. V. Dugas, P. E. Klein and J. E. Mullet (2011). "Coincident light and clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum." *Proceedings of the National Academy of Sciences* **108**(39): 16469-16474.
- Murphy, R. L., D. T. Morishige, J. A. Brady, W. L. Rooney, S. Yang, P. E. Klein and J. E. Mullet (2014). "Ghd7 (Ma6) Represses Sorghum Flowering in Long Days: Ghd7 Alleles Enhance Biomass Accumulation and Grain Production." *The Plant Genome* **7**(2).
- Murray, S. C., A. Sharma, W. L. Rooney, P. E. Klein, J. E. Mullet, S. E. Mitchell and S. Kresovich (2008). "Genetic improvement of sorghum as a biofuel feedstock: I. QTL for stem sugar and grain nonstructural carbohydrates." *Crop Science* **48**(6): 2165-2179.
- Murray, S. C., W. L. Rooney, S. E. Mitchell, A. Sharma, P. E. Klein, J. E. Mullet and S. Kresovich (2008). "Genetic improvement of sorghum as a biofuel feedstock: II. QTL for stem and leaf structural carbohydrates." *Crop science* **48**(6): 2180-2193.
- Natoli, A., C. Gorni, F. Chegani, P. Ajmone Marsan, C. Colombi, C. Lorenzoni and A. Marocco (2002). "Identification of QTLs associated with sweet sorghum quality." *Maydica* **47**(3-4): 311-322.
- Olson, S. N., K. Ritter, W. Rooney, A. Kemanian, B. A. McCarl, Y. Zhang, S. Hall, D. Packer and J. Mullet (2012). "High biomass yield energy sorghum: developing a genetic model for C4 grass bioenergy crops." *Biofuels, Bioproducts and Biorefining* **6**(6): 640-655.
- Poornima, S., Geethalakshmi, V. and Leelamathi, M. (2008) 'Sowing dates and nitrogen levels on yield and juice quality of sweet sorghum', *Res. J. Agric. Biol. Sci*, **4**(6), pp. 651–654.
- Quinby, J. (1967). "The maturity genes of sorghum." *Advances in agronomy* **19**: 267-305.
- Quinby, J., J. Hesketh and R. Voigt (1973). "Influence of temperature and photoperiod on floral initiation and leaf number in sorghum." *Crop Science* **13**(2): 243-246.
- Quinby, J. and J. Martin (1954). "Sorghum improvement." *Adv. Agron* **6**: 305-359.
- Rao, S. S. *et al.* (2013) 'Sweet Sorghum Planting Effects on Stalk Yield and Sugar Quality in Semi-Arid Tropical Environment', *Agronomy Journal*, **105**(5), p. 1458. doi: 10.2134/agronj2013.0156.
- R development Core Team (2016). R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing.
- Ritter, K. B. *et al.* (2008) 'Identification of QTL for sugar-related traits in a sweet × grain sorghum (*Sorghum bicolor* L. Moench) recombinant inbred population', *Molecular Breeding*, **22**(3), pp. 367–384. doi: 10.1007/s11032-008-9182-6.
- Rocateli, A., R. Raper, K. Balkcom, F. Arriaga and D. Bransby (2012). "Biomass sorghum production and components under different irrigation/tillage systems for the southeastern US." *Industrial Crops and Products* **36**(1): 589-598.
- Rooney, W. and S. Aydin (1999). "Genetic control of a photoperiod-sensitive response in *Sorghum bicolor* (L.) Moench." *Crop Sci* **39**.
- Rooney, W. L. and S. Aydin (1999). "Genetic control of a photoperiod-sensitive response in *Sorghum bicolor* (L.) Moench." *Crop science* **39**(2): 397-400.
- Russo, V. M. and W. W. Fish (2012). "Biomass, extracted liquid yields, sugar content or seed yields of biofuel feedstocks as affected by fertilizer." *Industrial Crops and Products* **36**(1): 555-559.
- Sattler, S. E., N. A. Palmer, A. Saballos, A. M. Greene, Z. Xin, G. Sarath, W. Vermerris and J. F. Pedersen (2012). "Identification and characterization of four missense mutations in

## **V. Principaux résultats et discussion de l'article : « Genotype and sowing date effects on the production and composition of sorghum stems, disentangling the contributions of dwarfism, earliness and photoperiod sensitivity »**

### **V.1- Un réservoir de diversité génétique pour la composition de la biomasse : un atout vers le développement de variétés adaptées aux différentes voies de valorisation**

Une variabilité génétique conséquente a été observée au sein du panel de 35 génotypes qui a été analysé dans le cadre de cette étude. En comparaison des analyses menées précédemment cette analyse impliquant un panel comprenant des génotypes avec des sensibilités à la photopériode variables et des tailles d'entre-nœuds contrastantes a notamment permis d'étendre légèrement la gamme de variabilité observée pour les teneurs en lignines. Les génotypes présentant ces fortes teneurs en lignines correspondent principalement à des génotypes très sensibles à la photopériode de la race Guinea, ces derniers ayant rarement été intégrés aux panels analysés précédemment (Trouche *et al.*, 2014; Brenton *et al.*, 2016).

Cette importante variabilité repose sur de forts effets génétiques avec des héritabilités au sens large comprises entre 0.57 et 0.74 ce qui correspond aux valeurs déjà observées dans la bibliographie. Les coefficients de variation génétique relativement faibles pour les composantes des parois ( $\leq 10\%$  pour les teneurs en lignines, cellulose et hémicellulose) contrastent avec ceux plus élevés obtenus pour les caractères plus intégratifs de digestibilité de la matière sèche, organique ou des parois (entre 15 et 50 % avec des héritabilités du même ordre).

Les héritabilités et les coefficients de variation génétique obtenus confirment donc le potentiel de développement de variétés de sorgho présentant des compositions de la biomasse très différentes. Néanmoins, si la disponibilité de larges variabilités génétiques est vectrice de gains potentiels aux niveaux des différents caractères considérés individuellement, les corrélations entre ces derniers peuvent contraindre largement le développement de variétés devant présenter des combinaisons de caractères pertinentes.

### **V.2- Rendement et qualité(s) : des relations « génétiques » complexes à gérer en fonction des idéotypes ciblés par la sélection**

Dans le cadre de cette étude, seules les longueurs et les biomasses fraîches de tiges ont été mesurées, l'information relative au rendement en matière sèche par unité de surface n'était donc pas disponible. Néanmoins, il a été montré que la hauteur des tiges présentent des corrélations

positives avec le rendement en matière sèche (Trouche *et al.*, 2014 ; Brenton *et al.*, 2016) on peut donc espérer que les corrélations observées entre la hauteur des tiges et les autres caractères soient représentatives de celles qui seront obtenues avec le rendement en biomasse sèche (mais voir Trouche et al qui observent des corrélations différentes entre hauteur, rendement en biomasse sèche et les composantes de qualité des tiges...).

Comme mentionné précédemment, la qualité de la biomasse est un concept instable, cette qualité devant être définie en fonction des voies de valorisation de la biomasse. Jusqu'à présent les deux composantes principales qui ont été analysées sont la teneur en sucres solubles qui constitue la cible de sélection dans le contexte de production d'éthanol de première génération et la digestibilité de la biomasse qui est quant à elle la cible à optimiser dans le cadre de l'alimentation animale et de la production d'éthanol de seconde génération via la voie biochimique.

En termes de production en sucres solubles, nous avons observé comme dans la bibliographie des corrélations positives entre la hauteur des plantes et le Brix. Néanmoins lorsque la teneur en sucres solubles des tiges est considérée (en intégrant la teneur en jus des tiges et la concentration en sucre de ces jus (Brix), la corrélation avec la hauteur des tiges devient non significative ce qui contraste avec les valeurs reportées dans la bibliographie (corrélation de 0.69 observée par Burks *et al.*, (2015b)). Une des explications possibles à cette divergence est la composition des panels qui est très différente non seulement en termes de hauteur moyenne des génotypes mais aussi en termes de variabilité génétique analysée (des guinéas photopériodiques notamment dans le cas de notre étude). En outre cette absence de corrélation est en phase avec les résultats récents qui ont mis en évidence que la production de sucres solubles et la croissance en hauteur ne sont pas intimement liés (Shukla *et al.*, 2017) ce qui permet donc d'entrevoir le développement de variétés courtes avec des rendements en sucres élevés (par gramme de matière sèche).

Au niveau des carbohydrates structuraux, des corrélations négatives ont été mises en évidence entre croissance et digestibilité de la biomasse et des corrélations positives entre croissance et teneurs en lignines. Ces résultats sont d'une part en accord avec les résultats reportés dans la bibliographie et d'autre part en phase avec les effets très forts des classes de nanisme mis en évidence sur les différentes composantes de la biomasse structurale. Néanmoins il est important de noter que même au sein des génotypes les plus hauts des teneurs moyennes en lignines peuvent être identifiées.

L'analyse séparée des différentes classes de nanisme et de précocité a permis de mettre en évidence la complexité des relations entre croissance et composition. Si les effets des différentes classes de précocité sur les différentes composantes de la composition de la biomasse sont relativement faibles, les différentes classes de nanismes présentent en revanche des propriétés très différentes en termes de compositions en carbohydrates structuraux et en lignines. En outre, il est important de souligner que les structures des corrélations sont très différentes entre les différentes classes. A titre d'exemple, si la corrélation entre la hauteur et la teneur en sucres solubles de la tige est non significative lorsque la totalité des génotypes est considérée ( $r=0.084$ ,  $pvalue = 0.63$ ), celle-ci est positive pour les génotypes courts ( $r=0.77$ ,  $pvalue = 0.009$ ), et non significative pour les deux autres classes de nanisme ( $r_{dw2}=-0.38$  ( $pvalue = 0.18$ ),  $r_{dw3} = -0.16$  ( $pvalue= 0.6$ )). La corrélation entre la hauteur et la solubilité de la matière organique est quant à elle négative ( $r=-0.45$ ,  $pvalue =0.005$ ) lorsque tous les génotypes sont considérés, mais elle est non significative lorsque les différentes classes de nanisme sont considérées. Ces résultats sont à prendre en compte dans le contexte de la sélection de sorghos hybrides où des idéotypes différents sont sélectionnés des côtés mâles et femelles.

### **V.3- La date de semis impacte la qualité de la biomasse et le niveau de sensibilité au photopériodisme des génotypes affecte l'intensité cette réponse**

En combinaison avec les efforts de sélection, les méthodes culturales peuvent jouer un rôle clé dans l'adaptation des produits aux attentes des utilisateurs. Dans le contexte d'Afrique de l'Ouest, la date de semis est soumise à la variation des conditions environnementales annuelles et elle joue un rôle prépondérant dans le succès de la culture, des semis trop précoces ou trop tardifs pouvant s'accompagner d'une augmentation des contraintes biotiques et abiotiques au moment du remplissage ou de la maturation des grains. Dans le contexte des pays tempérés, l'évolution des itinéraires techniques incluant de nouvelles rotations de cultures (CIVE : Cultures Intermédiaires à Vocation Energétique) favorise le développement de cultures dérobées pour lesquelles des semis tardifs sont effectués.

Dans ces différents contextes, l'évaluation des effets des dates de semis sur la composition de la biomasse est donc primordiale. Dans le contexte de la production d'éthanol de première génération (basés sur l'extraction des sucres solubles des tiges), les semis tardifs ont été associés dans plusieurs cas à des réductions de rendement en sucres solubles (Burks *et al.*, 2013; Rao *et al.*, 2013; Almodares and Hoseini, 2016; Tovignan *et al.*, 2016). L'analyse de trois dates de semis a mis en évidence une réduction des teneurs en lignines et une augmentation des teneurs

en hémicellulose dans le cas de semis tardifs qui s'accompagne donc d'une augmentation de la solubilité de la biomasse. En d'autres termes, les semis tardifs permettent d'améliorer la qualité de la biomasse pour l'alimentation animale et la production énergétique via des voies biochimique (éthanol de seconde génération / méthane). Néanmoins cette amélioration de la qualité pour ces usages s'accompagne d'une réduction de la longueur des tiges (et donc du rendement final).

Si les génotypes présentent des réponses différentes aux décalages des semis, il est intéressant de noter que les différentes classes de sensibilité au photopériodisme présentent des réponses moyennes différentes. Les génotypes les moins sensibles à la photopériode sont ceux qui répondent le plus fortement aux décalages des semis tandis que les génotypes les plus photopériodiques présentent des évolutions très faibles de leur composition entre les dates de semis. Ces résultats ouvrent des questionnements sur la modulation des allocations de carbones entre les différentes classes de sensibilité à la photopériode :

Pourquoi des génotypes ayant des durées de cycle semis-floraison stables (génotypes non photopériodiques) ont leur composition très affectée par le décalage des semis ?

Comment des génotypes dont la durée du cycle semis floraison est très variable (fortement réduite dans le cas de semis tardifs) maintiennent une composition de la biomasse stable ?

#### **V.4- Les 35 génotypes analysés : 35 parents d'un dispositif de croisement multiparental devant permettre de disséquer le déterminisme génétique de la qualité de la biomasse**

Outre l'estimation de la variabilité génétique disponible sur un panel de génotype original (i.e. enrichi en génotypes d'Afrique de l'Ouest), l'analyse des effets du nanisme, de la précocité et de la sensibilité à la photopériode sur la qualité de la biomasse, ces analyses ont aussi été motivées par le statut des 35 génotypes analysés. En effet ces 35 génotypes ont été utilisés dans un plan de croisement multiparental ayant 2 objectifs principaux, d'une part le développement d'une population pertinente en termes d'analyses génétiques pour des caractères pertinents aux niveaux des pays en voies de développement mais aussi des pays tempérés (rendement en grains, adaptation à des stress biotiques et abiotiques, qualité du grain, qualité de la biomasse) et d'autre part le développement d'une ressource pertinente pour la sélection en Afrique de l'Ouest. Les analyses génétiques menées sur les aspects de qualité de la biomasse sont présentées dans le chapitre suivant.



## **Chapitre IV : Identification des bases génétiques de la qualité de la biomasse au sein d'un dispositif multiparental de type BackCrossNAM**

---

### **I. La qualité de la biomasse un caractère d'intérêt majeur pour des utilisations multiples**

La biomasse végétale est connue comme l'une des sources d'énergies renouvelables les plus prometteuses en termes d'abondance, de coût de production et de respect de l'environnement (Berndes, Hoogwijk and van den Broek, 2003; Antonopoulou *et al.*, 2008). En outre de nouvelles filières de valorisation non énergétiques de la biomasse émergent ouvrant de nouveaux marchés pour les productions de biomasse végétales (Parajuli *et al.*, 2015; Xiu and Shahbazi, 2015). Pour alimenter ces filières, trois sources de biomasse végétales peuvent être distinguées : le bois issu des forêts « naturelles », les résidus de l'agriculture et les plantes énergétiques dédiées.

Les graminées en C4 sont en mesure de contribuer significativement à la production de biomasse végétale dédiée aux nouvelles valorisations de la biomasse (van der Weijde *et al.*, 2013). Parmi ces dernières le sorgho présente de nombreux atouts (Mullet *et al.*, 2014). En effet, outre son intérêt pour l'alimentation animale, son rendement élevé en biomasse, son efficacité d'utilisation de l'azote (Gardner, Maranville and Pappozzi, 1994; Lama Tamang *et al.*, 2011), son adaptation à divers systèmes de culture et sa grande variabilité phénotypique pour sa composition de la biomasse et ses propriétés (Vandenbrink *et al.*, 2010; Trouche *et al.*, 2014; Brenton *et al.*, 2016 et Chapitre III de ce document) contribuent à en faire l'espèce cible de nombreux projets de transformation de la biomasse végétale. Néanmoins, si un fort rendement en biomasse par hectare, est une composante nécessaire dans les filières de valorisation, sa composition biochimique et ses propriétés de dégradabilité jouent aussi un rôle clé dans les utilisations envisageables (van der Weijde *et al.*, 2013).

### **II. Déterminisme génétique de la qualité de la biomasse**

La compréhension des mécanismes moléculaires impliqués dans le contrôle génétique de la qualité de la biomasse est un des facteurs importants pour sa transformation. Chez le sorgho, un

nombre relativement limité d'études (Tableau II-1) a été mené concernant la compréhension de l'architecture génétique de la qualité de la biomasse, et ce malgré l'existence d'une grande variabilité génétique (Chapitre III).

Comme mentionné précédemment la qualité de la biomasse dépend de deux grandes composantes : sa teneur en sucres solubles et la composition de ses parois cellulaires en termes de carbohydrates structuraux (cellulose, hémicellulose) et de lignines. La teneur en sucres solubles a été l'objet de plusieurs études en raison de son importance dans le cadre de l'utilisation de la biomasse pour la production d'éthanol de première génération (i.e. extraction des sucres solubles de la tige et leur fermentation en éthanol). En revanche la composition des parois cellulaires et plus globalement les propriétés de dégradabilité de la biomasse ligno cellulosique ont été relativement peu travaillées malgré leurs impacts dans les contextes de l'alimentation animale, de la production énergétique ou de biomatériaux.

### **II.1- La teneur en sucres solubles des tiges**

Le déterminisme génétique des teneurs en sucres solubles des tiges a été analysé dans le cadre de plusieurs études de cartographie génétique menées sur des populations biparentales (Murray, Sharma, et al., 2008; Ritter et al., 2008; Shiringani, Frisch and Friedt, 2010; Guan et al., 2011; Shiringani and Friedt, 2011; Qazi, Paranjpe and Bhargava, 2012). La majorité de ces études ont été menées sur des croisements entre des parents de types "Sorgho grain (de taille courte)" et "Sorgho sucré (de grande taille) » et ont donc mis en évidence de nombreuses colocalisations entre des QTL de hauteur, de floraison et de teneurs en sucres solubles. Ces résultats, cohérents avec les corrélations observées entre hauteur et teneur en sucres solubles (Brix), ne permettent donc pas d'identifier spécifiquement les régions chromosomiques contrôlant l'accumulation des sucres solubles dans les tiges. En outre, de fortes interactions génotypes x environnements ont été mises en évidence avec une forte instabilité des QTL détectés entre les différents environnements.

Deux études de génétique d'association ont ensuite été menées afin de tirer parti de la décroissance rapide du déséquilibre de liaison pour préciser les régions d'intérêt (Murray et al., 2009; Lv et al., 2013). Néanmoins ces deux études ont été effectuées sur des panels de génotypes comprenant là encore des mélanges de génotypes courts et hauts avec peu de génotypes non sucrés de taille haute. En outre, ces deux analyses ont été menées avec des effectifs faibles de génotypes (max =115) et de marqueurs (max = 369) ne permettant pas une analyse fine et robuste du déterminisme génétique de la teneur en sucres solubles. Sur la base

de ces deux études, une seule région présentant une association supportant les ajustements aux tests multiples a été identifiée sur le chromosome 1 à proximité d'un gène de glucose-6-phosphate isomerase (EC 5.3.1.9) (Murray et al., 2009). Plus récemment, deux études basées sur des panels de tailles plus importantes et développés spécifiquement dans le contexte de l'analyse de production de biomasse ont été réalisées (Burks et al., 2015; Brenton et al., 2016). L'étude menée par Burks et al (Burks et al., 2015) a mis en évidence l'effet du gène majeur D (Chrom 6, 51.8 Mb) sur la quantité de sucres extractibles des tiges. Ce gène D aussi appelée « Dry Midrib ou nervure sèche » contrôle les types de nervures qui sont des prédictors de la jutosité des tiges (les nervures blanches indiquant des tiges sèches tandis que les nervures vertes correspondent généralement à des tiges plus juteuses). En complément, 5 régions chromosomiques additionnelles ont été mises en évidence par Brenton et al 2016 (ces régions ne comprenant pas de gènes majeurs de floraison ou de nanisme) et des gènes potentiellement responsables des associations ont été détectés. Néanmoins une très large proportion de la variabilité génétique de la teneur en sucres solubles est encore inexplicée. A titre d'exemple bien que 49 % des géotypes du panel de Burks et al. 2015 présentent des nervures vertes, seules 24 % présentent l'allèle « favorable » pour le gène D.

En complément des analyses entre phénotypes et diversité nucléotidique, plusieurs analyses d'expression de gènes ont été menées pour tenter d'identifier les déterminants moléculaires impliqués dans l'accumulation des sucres solubles (Qazi, Paranjpe and Bhargava, 2012; Shakoor *et al.*, 2014; Bihmidine *et al.*, 2015; Yu *et al.*, 2015; McKinley *et al.*, 2016; Mizuno, Kasuga and Kawahigashi, 2016) mais jusqu'à présent peu de liens ont été faits entre les analyses du transcriptome et les études de génétique moléculaire. Il apparaît généralement peu de convergences entre les gènes différentiellement exprimés et les régions chromosomiques identifiées au travers des études QTL (en populations biparentales) et de génétique d'association.

## **II.2- Composition des parois cellulaires et digestibilités de la biomasse et des parois**

La qualité de la biomasse végétative sous forme de fourrages ou d'ensilages a fait l'objet de nombreuses études en raison de son importance dans le contexte de l'alimentation animale. Ces études ont porté principalement sur :

- l'évolution de la qualité en fonction des itinéraires techniques (date de récolte, tallage...),
- les qualités nutritionnelles des variétés arrivant sur le marché

En outre, des gènes majeurs impactant la teneur en lignine des parois cellulaires ont été identifiés chez le sorgho (Vermerris *et al.*, 2007; Saballos *et al.*, 2008, 2009; Vermerris, 2011; Saballos *et al.*, 2012; Sattler *et al.*, 2014) certains d'entre eux (bmr6, bmr12 et bmr19) ayant été rapidement mobilisés dans le contexte du développement de lignées à faible teneur en lignine pour les utilisations du fourrage ou de la biomasse totale en alimentation animale (Pedersen *et al.*, 2008). Trois des gènes impliqués dans le contrôle du phénotype « Nervure brune ou Brown midrib » ont été clonés et 5 autres sont encore à caractériser (Sattler *et al.*, 2014). Si ces gènes constituent de formidables outils de contrôle des teneurs en lignines dans la biomasse, leur utilisation en sélection a souvent été contrariée par des verses importantes des génotypes portant les allèles mutés. Il est aussi important de noter ici, que i) la variabilité génétique observée pour les teneurs en lignines (Murray, Rooney, *et al.*, 2008; Shiringani and Friedt, 2011; Trouche *et al.*, 2014) ne dépend pas uniquement de ces gènes majeurs et ii) que la digestibilité de la biomasse et plus largement ses propriétés structurales ne dépendent pas uniquement des teneurs en lignines mais aussi des proportions de celluloses, hémicelluloses et pectines dans les parois et de leurs interactions avec les lignines. Dans ce contexte, une meilleure compréhension des régions chromosomiques contrôlant la variabilité des différentes composantes de la biomasse et des propriétés de cette biomasse est nécessaire pour optimiser la sélection de variétés devant présenter des compositions de la biomasse spécifique.

Néanmoins, malgré l'intérêt du pilotage de la composition en carbohydrates structuraux et en lignines pour l'alimentation animale, la production de bioénergie de seconde génération ou de biomatériaux peu d'études ont été mises en œuvre pour analyser l'architecture génétique de ces composantes. Murray *et al.* (2008a), (2008b) ont été les premiers à étudier l'architecture génétique des teneurs en carbohydrates structuraux et des lignines au sein d'une population biparentale entre un sorgho de type grain (Btx623) et un sorgho de type sucré (Rio). Une forte instabilité des QTL entre les environnements a été observée. En outre, des déterminismes génétiques indépendants ont été mis en évidence pour les teneurs des carbohydrates structuraux et des lignines dans les feuilles et les tiges. Enfin comme dans le cas de l'analyse des sucres solubles de nombreuses colocalisations entre les QTL détectés pour les composantes des parois cellulaires, la croissance en hauteur et la floraison ont été mises en évidence. Shiringani and Friedt (2011) au travers de l'analyse d'une autre population issue d'un croisement entre un sorgho de type « grain » et un sorgho de type « sucré » sont arrivés aux mêmes conclusions : de fortes interactions Génotype x Environnement menant à une relative instabilité des QTL détectés dans les différents environnements et des colocalisations de QTL entre composition de

la biomasse, croissance et floraison. Néanmoins quelques régions communes ont tout de même été détectées entre ces deux études.

Une troisième analyse de cartographie génétique sur une population biparentale a été menée par Vandenbrink et al. (2013). Contrairement aux deux études précédentes, qui s'étaient focalisées sur des caractères descriptifs de la composition de la biomasse (NDF, ADF, ADL) et qui étaient ancrées sur des croisements entre des sorghos grains et des sorghos sucrés, cette analyse a porté sur le potentiel d'hydrolyse de la biomasse lignocellulosique (un caractère beaucoup plus intégratif) et sur l'indice de cristallinité de la cellulose (une des composantes de l'aptitude à l'hydrolyse de la biomasse lignocellulosique) au sein d'un croisement interspécifique entre un génotype de sorgho cultivé (*Sorghum bicolor ssp bicolor*) et une espèce sauvage pérenne à rhizome (*Sorghum propinquum*). Comme dans le cas de l'étude de Murray et al (2008) l'analyse simultanée des feuilles et des tiges a permis de mettre en évidence des déterminismes relativement indépendants entre les deux organes et aucun QTL commun n'a été détecté entre l'aptitude à la conversion et la cristallinité de la cellulose. Les colocalisations entre les études de Murray et al 2008, de Shiringani et Friedt, 2011 et de Vandenbrink et al (2013) n'ont pas été analysées précisément. Afin d'explorer une diversité génétique plus large et véritablement spécifique aux idéotypes « sorgho sucré – sorgho biomasse », un panel composé de 390 accessions comprenant 152 sorghos sucrés et 238 sorghos biomasse (génotype à forte production de biomasse mais avec une teneur en sucres solubles des tiges relativement faible) a été mis en place (Brenton *et al.*, 2016). L'analyse de génétique d'association effectuée pour la teneur de la biomasse en parois cellulaires (estimée au travers du caractère NDF, Neutral Detergent Fiber) a permis la détection de 5 régions chromosomiques indépendantes (ces 5 régions étant communes avec celle détectées pour la teneur en carbohydrates non structuraux sur le même panel, Cf. chapitre précédent). Les comparaisons avec les études menées dans des populations biparentales présentées précédemment n'ont pas été menées par manque de temps.

Comme pour le métabolisme des sucres solubles, plusieurs études ont eu pour ambition d'identifier les réseaux d'expression génique contrôlant les accumulations en carbohydrates structuraux et en lignines dans les tiges (Shakoor *et al.*, 2014; McKinley *et al.*, 2016; Rai *et al.*, 2016). Néanmoins, comme dans le cas des sucres solubles, peu de liens solides ont été mis en évidence entre les gènes différentiellement exprimés lors de l'évolution de la composition de la biomasse et les régions chromosomiques contribuant à la variabilité génétique des différentes composantes de la biomasse.

En conclusion, malgré l'importance du pilotage de la composition de la biomasse pour les applications classiques (alimentation animale basée sur le fourrage ou l'ensilage) et les applications émergentes (production d'énergies renouvelables et de biomatériaux) une faible compréhension du déterminisme génétique des composantes de la biomasse est actuellement disponible. En complément des gènes à effet majeurs (gènes bmr) qui induisent souvent des effets indésirables, quelques régions à effets relativement faibles ont été mises en évidence, mais une très large part de la variabilité génétique disponible est encore inexploitée.

Dans ce contexte, afin d'affiner notre compréhension du déterminisme génétique des composantes de la qualité de la biomasse, 29 populations issues d'un dispositif de croisement multiparental de type Backcross NAM ont été caractérisées durant 2 ans pour leur composition en carbohydrates structuraux, en lignines et leurs propriétés de digestibilité des tiges. L'ambition de cette étude était, en tirant parti des propriétés génétiques de ce plan de croisement, d'identifier les zones chromosomiques impliquées dans le déterminisme génétique de ces caractères. Plus précisément, l'objectif était ici de s'affranchir des limites des approches de cartographie en population biparentale (i.e. faible diversité génétique et déséquilibre de liaison à longue distance) et des populations d'association (faible puissance de détection dans le cas de fréquences alléliques très déséquilibrées) pour fournir un catalogue « générique » des régions chromosomiques d'intérêt pour le pilotage de la composition de la biomasse chez le sorgho.

### **III. Matériel et méthodes**

#### **III.1. Développement des populations BCNAM et dispositif expérimental**

Le développement des 29 populations BCNAM utilisées dans le cadre de cette étude a été décrit au sein du chapitre II (Partie IV). Le dispositif expérimental est le même que celui qui a été décrit dans le cadre de l'analyse du déterminisme génétique de la floraison et de la croissance en hauteur au sein du chapitre II.

#### **III.2. Caractérisation de la composition des tiges**

En complément des caractérisations agromorphologiques qui ont été présentées au sein du chapitre II, la qualité de la biomasse de l'ensemble des familles des 29 populations a été caractérisée. Le protocole utilisé pour ces caractérisations est décrit ci-dessous.



**a**



**b**



**c**

Figure IV-1 : Echantillons de tiges des familles BCNAM séchées à l'air libre (a) et sous-échantillons de 5 tiges / famille après broyage grossier (b) suivi d'un broyage plus fin avec un tamis de 1 mm (c).

### ***III.2.1. Préparation des échantillons de tiges***

Pour l'ensemble des populations BCNAM, les 5 plantes caractérisées en termes agromorphologiques dans chacune des familles BC1F4 ont été conservées pour analyser la composition biochimique des tiges. Les feuilles (y compris les gaines) les pédoncules et les panicules de chaque plante individuelle ont été enlevées. Ces lots de 5 tiges ont été séchés à l'air libre (Figure IV-1a) et à l'abri des précipitations en vue d'un pré-broyage grossier suivi d'un broyage fin. Une des raisons à l'application de cette technique réside d'une part dans le nombre élevé d'échantillons (>1200) qui ne permettait pas d'envisager un broyage direct juste après la récolte, et d'autre part dans le nombre trop réduit d'étuves disponibles (une seule).

Après un temps de séchage conséquent (i.e. de 3 à 4 mois), qui a permis d'atteindre un taux d'humidité de la matière sèche d'environ 10%, un pré-broyage grossier de chaque lot de 5 tiges par famille a été effectué à l'aide d'un broyeur à compost (VIKING GE335). Après le broyage, un échantillon de 200 grammes (Figure IV-1b) a été prélevé et le séchage a été finalisé à l'étuve (60 °C) dans des sacs en moustiquaire. Un second broyage plus fin a ensuite été effectué à l'aide d'un broyeur avec un tamis à 1 mm (SM100, Retsch, Haan, Germany). La poudre obtenue (Figure IV-1c) était ainsi compatible avec les analyses biochimiques présentées ci-dessous et l'utilisation de calibrations en spectrométrie proche infra-rouge qui ont été développées au sein de l'UMR SELMET du CIRAD.

### ***III.2.2 Caractérisation de la composition biochimique et la valeur alimentaire des tiges***

La caractérisation des tiges (composition biochimique et valeur alimentaire) provenant de l'ensemble des familles BCNAM a été effectuée grâce à l'utilisation de la spectroscopie en réflectance dans le proche infrarouge (SPIR ou NIRS : Near Infra Red Spectroscopy). Cette méthode analytique basée sur l'absorption des rayonnements infrarouges (longueur d'onde supérieure à celle du spectre visible) par la matière organique, est une technique rapide et non destructive.

La prédiction de la composition biochimique et de la valeur alimentaire des tiges grâce aux mesures indirectes issues de la SPIR comprend 3 étapes : l'acquisition de données de référence, le développement d'équations de calibration et enfin la prédiction.

En amont du développement des calibrations SPIR qui ont été utilisées pour effectuer les prédictions sur l'ensemble des échantillons, des mesures de références ont été effectuées sur un premier jeu d'échantillons. Ces échantillons de références correspondent d'une part à des



échantillons de tiges analysés dans le cadre de projets précédents et à d'autre part à des échantillons de la population BCNAM qui ont été sélectionnées sur la base de leurs caractéristiques spectrales. Les analyses de référence correspondant aux paramètres suivants ont été réalisées :

- Teneur en Matière minérale (MM)

La teneur en matière minérale (cendres) représente les résidus inorganiques restants après extraction complète de l'eau et de la matière organique. Elle est déterminée par calcination de l'échantillon à 550°C pendant 4 heures selon les normes de l'Association Française de Normalisation (AFNOR, 1977, AFNOR, 2009).

- Teneur en Matière azotée totale (MAT)

La teneur en protéines brutes ou matière azotée totale a été déterminée en utilisant la méthode standard de détermination de la teneur en azote (N) de Kjeldhal (Sáez-Plaza *et al.*, 2013). Cette méthode de mesure est largement utilisée pour la détermination des protéines dans le domaine de l'alimentation humaine et animale. La teneur en protéines brutes du produit est obtenue en multipliant la teneur en azote total (fraction organique et inorganique) par un coefficient de conversion de 6.25 (cas du fourrage). Cette teneur en exprimées en pourcentage de la matière sèche (%MS).

- Teneur en Cellulose brute de Weende (CBW)

La détermination de la teneur en cellulose brute ou teneur en fibre de l'échantillon a été réalisée conformément aux normes internationales (AFNOR, 2002). C'est une ancienne méthode d'analyse qui donne une estimation de la teneur en fibre globale de l'échantillon (cellulose, hémicellulose, lignines...). Il faut noter que lors du processus de mesure, une partie des hémicelluloses et des fibres solubles sont supprimées (portions non estimées). La teneur en CBW est exprimée en pourcentage de matière sèche (%MS).

- Teneurs en constituants pariétaux selon la méthode de Van Soest : NDF, ADF et ADL

Les fibres représentent l'une des principales composantes la biomasse végétale et de ce fait contribuent largement à l'alimentation des animaux qui est son utilisation la plus classique. Elles constituent la fraction de l'aliment la plus difficile à digérer.

Les teneurs des différentes fractions des fibres : les fibres insolubles dans les détergents neutres (NDF), les fibres insolubles dans les détergents acides (ADF), et la lignine (ADL), ont été

estimées selon la méthode de Van Soest et al. (Van Soest and Wine, 1967; Van Soest, Robertson and Lewis, 1991) conformément aux normes internationales AFNOR (AFNOR, 2006, 2008).

- La fraction de fibres insolubles dans les détergents neutres ou Neutral Detergent Fiber (NDF) comprend toutes les composantes des parois cellulaires (cellulose, hémicellulose et lignine) à l'exception de certaines pectines. Le NDF correspond à la fraction organique (sans les cendres) de l'aliment dont seulement une partie est digestible par les animaux.
- La fraction insoluble dans les détergents acides ou Acid Detergent Fiber (ADF), correspond à la fraction la moins digestible et est considérée par Van Soest comme étant à peu près équivalente à la cellulose brute et à la lignine. L'ADF constitue un bon indicateur de la capacité de l'animale à digérer le fourrage, plus la teneur en ADF augmente et plus la digestibilité du fourrage diminue.
- Enfin, la lignine ou Acid Detergent Ligin (ADL) représente la fraction du NDF totalement indigestible. Cette méthode de détermination de la teneur en lignine sous-estime la concentration réelle de la lignine. La lignine représente l'une des composantes majeures des parois cellulaires (après la cellulose et hémicellulose) et constitue le principal facteur influençant sa digestibilité. Par contre, d'un point de vue structural, la lignine confère force et rigidité aux parois cellulaires.

A partir de ces 3 paramètres, les teneurs en cellulose et en hémicellulose ont été calculées grâce aux formules suivantes :

- Teneur en hémicellulose Van Soest :  $Hemi = NDF - ADF$  (en % de MS)
- Teneur en cellulose Van Soest :  $cvs = ADF - ADL$  (en % de MS)

Ces différentes estimations (NDF, ADF, ADL, cellulose VS, hémicellulose VS) sont fortement corrélées en raison de leur mode de mesure, de calcul et leurs unités (% de MS). Afin de caractériser les teneurs relatives en cellulose, hémicellulose et lignine des fibres (NDF), les paramètres suivants ont donc été calculés :

- Teneur relative des fibres en cellulose :  
$$celSndf = \left[ \frac{ADF - ADL}{NDF} \right] * 100 \text{ (en \% NDF)}$$

- Teneur relative des fibres en hémicellulose :

$$hemiSndf = \left[ \frac{NDF-ADF}{NDF} \right] * 100 \text{ (en \% NDF)}$$

- Teneur relative des fibres en lignines :

$$adlSndf = \left[ \frac{ADL}{NDF} \right] * 100 \text{ (en \% NDF)}$$

- Teneur en cellulose brute de Weende ajustée à la teneur en parois :

$$cbwSndf = \frac{CBW}{NDF} * 100 \text{ (en \% NDF)}$$

Ce calcul permet de mettre en évidence les différences entre CBW et le NDF qui représentent normalement toutes deux quasiment le même caractère (quantités de parois cellulaires). Il est probable que ce rapport mette principalement en évidence les différences d'estimations des hémicelluloses entre ces deux paramètres.

La solubilité de la biomasse (ou digestibilité) a été déterminée selon la méthode de Aufrère et al (Aufrère *et al.*, 2007). Cette méthode enzymatique, basée sur la digestion par pepsine-cellulase permet de prédire la solubilité de la matière organique in vitro (SMO appelée In Vitro Organic Matter Digestibility (IVOMD) en anglais) et la solubilité de la matière sèche in vitro (SMS, appelé In Vitro Dry Matter Digestibility (IVDMD) en anglais).

A partir de ces valeurs de solubilité de la matière organique (SMO), et connaissant la teneur en matière minérale (MM) de l'échantillon, la digestibilité de la portion NDF (in-vitro NDF digestibility: IVNDFD) est calculée en faisant l'hypothèse que la partie non NDF de l'échantillon est complètement digestible :

$$IVNDFD = 100x(SMO - ((100 - MM) - NDF))/NDF$$

En amont de ce projet de thèse, une bibliothèque de données de référence (comportant environ 500 échantillons pour les différents paramètres) était déjà disponible (Tableau IV-1) au niveau de l'UMR SELMET (<https://umr-selmet.cirad.fr/>). En complément, afin de s'assurer de l'adéquation des calibrations au matériel végétal analysé dans le cadre de la thèse, de nouvelles analyses de références ont été effectuées sur une centaine d'échantillons issus des différentes populations de cette étude (ces derniers ayant été sélectionnés sur la base de leur signature spectrale en proche infrarouge). Par ailleurs, ces 100 échantillons de calibration sont les mêmes que ceux utilisés pour la prédiction des 35 géotypes parentaux (cf. chapitre III).

Tableau IV-1: Propriétés des modèles de prédiction pour les différents caractères analysés

Caractère <sup>A</sup>	Collection analysée				Calibration		Validation		
	N	Moy	SD	CV %	SEC	R <sup>2</sup> cal	SECV	R <sup>2</sup> val	RPD
MS	523	93.45	1.95	2.09	0.54	0.92	0.61	0.90	3.2
MM	517	4.41	2.65	60.14	0.55	0.96	0.68	0.94	3.9
MAT	518	3.73	2.21	59.3	0.43	0.96	0.5	0.95	4.4
CBW	515	37.74	6.84	18.14	1.02	0.98	1.17	0.97	5.9
NDF	455	68.77	10.9	15.85	1.36	0.98	1.57	0.98	6.9
ADF	461	41.66	8.58	20.61	1.24	0.98	1.41	0.97	6.1
ADL	463	6.46	2.56	39.58	0.56	0.95	0.66	0.94	3.9
SMS	467	38.48	12.33	32.04	1.69	0.98	1.91	0.98	6.5
SMO	463	36.47	12.16	33.34	1.74	0.98	1.9	0.98	6.4

<sup>A</sup>MM : Matière Minérale, MAT : Matière Azoté Totale, CBW : Cellulose Brute de Weende, NDF : Neutral Detergent Fiber, ADF : Acid Detergent Fiber, ADL : Acid Detergent Lignin, SMS : solubilité de la matière sèche, SMO : solubilité de la matière organique.

### ***III.2.3. Acquisition de spectres des échantillons de tiges***

Les acquisitions de spectres ont été réalisées à l'aide d'un spectromètre proche infrarouge (Tango Bruker Optics, Hong Kong Ltd) à transformée de Fourier. Le principe est basé sur l'absorption du rayonnement (infrarouge) par les molécules organiques de l'échantillon à différentes longueurs d'onde. Lors du processus, l'échantillon est éclairé à différentes longueurs d'onde, et à chacune de ces longueurs d'ondes, les molécules organiques contenues dans l'échantillon absorbent la lumière (les ondes infrarouges) en émettant un signal. Ce signal constitue donc le spectre de l'échantillon.

Au total, 4794 échantillons provenant des 29 populations BCNAM ont été caractérisés en SPIR. Chaque échantillon a été « scanné » à 2 nanomètres d'intervalle dans des longueurs d'ondes allant de 400 à 2500nm. Chaque échantillon a été scanné deux fois et les spectres obtenus ont été moyennés (lorsque la variabilité entre les spectres était trop importante, deux nouvelles prises de spectres ont été réalisées).

### ***III.2.4. Développement des calibrations***

Les équations de prédictions des paramètres analytiques ont été définies sur la base d'un modèle de régression linéaire. Ces modèles de calibration sont caractérisés par un écart-type résiduel de calibration (SEC : standard Error of Calibration) et un coefficient de détermination ( $R_{cal}^2$ ). Les précisions des modèles de prédictions ont ensuite été évaluées sur une collection de validation grâce aux paramètres suivants : l'écart-type résiduel de la validation croisée (SECV : standard error of cross validation), le coefficient de détermination sur la collection de validation ( $R^2_{val}$ ) et le rapport entre la variabilité de la population et la précision de la calibration (RPD : ratio performance/validation) (Tableau IV-1). Ces modèles de calibrations ont été utilisés pour prédire les valeurs des caractères sur les populations analysées. Même dans le cas où les échantillons ont été analysés avec les méthodes de référence, ce sont les données prédites qui ont été utilisées pour les analyses.

## **III.3. Analyse des données phénotypiques**

Les données de composition et de digestibilité de la biomasse, obtenues sur les essais des saisons normales de culture 2014 et 2015, ont subi une première étape de vérification des données (distribution et corrélations inter-caractères). Les calculs des statistiques de base (moyenne, écart-type et coefficient de variation phénotypique) ont été effectués sur chacune des variables obtenues dans les deux environnements de culture.

Les moyennes ajustées pour chaque famille BC1F4 ont été calculées sur les données brutes collectées sur les 1092 (essai 2014, 26 populations) et 1275 (2015, 25 populations) familles BC1F4 analysées dans le dispositif « global » en prenant en compte les dispositifs en blocs augmentés. Trois autres populations BCNAM ayant été caractérisées dans des dispositifs différents ont été analysées séparément. Pour la population BC06 un ajustement prenant en compte le dispositif en blocs augmentés a été effectué tandis que pour les deux autres populations (BC03 et BC010) les données ont été ajustées en prenant en compte les dispositifs en split plot au sein desquels elles étaient intégrées.

Les corrélations entre les deux années ont été calculées afin d'évaluer la stabilité des données entre les deux années.

Une première analyse de variance (modèle 1) a été réalisée sur chaque essai afin d'estimer les effets génétiques et résiduels sur chaque caractère en se basant sur les valeurs collectées sur les génotypes témoins qui sont répétés dans l'essai :

$$Y_{ij} = \mu + P_i + B_j + e_{ij} \quad (\text{Modèle 1})$$

Avec  $Y_{ij}$  : la valeur phénotypique mesurée sur le génotype  $i$  dans le bloc  $j$  ;

$\mu$  : la valeur moyenne de la population,

$P_i$  et  $G_i$  indiquent les effets de la variété témoin  $i$  et du génotype (famille)  $i'$  respectivement,

$B_j$  : effet du bloc  $j$ ,

$A_j$  : effet de l'année  $j$ , et

$e_{ij}$  : la résiduelle du modèle.

Sur la base des résultats d'Anova (1), la répétabilité des différents caractères a été estimée dans chaque environnement (année) à partir des valeurs obtenues sur les génotypes témoins selon la formule suivante :

$$\text{rep} = \frac{\sigma_G^2}{(\sigma_G^2 + \sigma_R^2)}$$

Avec  $\sigma_G^2 = 1/k (CM_g - CM_r)$  où  $k$  représente le nombre de répétitions de chaque génotype (témoin) dans l'essai ( $k=76$ ) pour les essais de 2015 et 2014 ;

$CM_g$  : correspond aux carrés moyens génotypiques et

$\sigma_R^2 = CM_r$  : indique le carré moyen résiduel

Une seconde analyse a été effectuée sur les 1007 familles communes aux deux années d'expérimentation, en utilisant le modèle linéaire mixte suivant :

$$Y_{ij} = \mu + A_j + G_i + e_{ij} \quad (\text{Modèle 2})$$

Où  $Y_{ij}$  est la valeur phénotypique mesurée sur la famille  $i$  dans l'année  $j$ ;

$\mu$  : est la moyenne générale de la population,

$A_j$  correspond à l'effet de l'année  $j$  déclaré en effet fixe,

$G_i$  : l'effet aléatoire de la famille  $i$  et

$e_{ij}$  : la résiduelle du modèle.

A partir des résultats du modèle (2), une estimation de la variabilité environnementale a été réalisée et les effets génétiques (Best linear unbiased predictors : BLUP) de chaque famille ont été extraits.

Sur la base des résultats du modèle (2), les coefficients de variation phénotypique (CVP%) et génotypique (CVG%) ont été calculés selon les formules ci-dessous :

$$CVp = \sqrt{\sigma_p^2 / \bar{x}} * 100$$

$$CVg = \sqrt{\sigma_g^2 / \bar{x}} * 100$$

Avec  $\sigma_p^2$  : la variance phénotypique  $\sigma_p^2 = \sigma_g^2 + CMe$  ;

Où  $\sigma_g^2$  : représente la variance génotypique :  $\sigma_g^2 = (CMg - CMe) / k$

Avec CMg et CMe : les carrés moyen génotypique et résiduel ; et

$k$  : le nombre de répétition de l'essai.

Enfin, connaissant les variances phénotypiques  $\sigma_p^2$  et génotypiques  $\sigma_g^2$ , l'héritabilité au sens large ( $h_{SL}^2$ ) a été calculée selon la formule suivante :

$$h_{SL}^2 = \sigma_g^2 / \sigma_p^2$$

Les corrélations intra-caractère (inter-années) et inter-caractères (intra-année) ont été calculées sur la base des valeurs ajustées obtenues sur chaque famille pour chaque année de culture.

Les données moyennes ajustées obtenues sur chaque essai pour chacune des variables, ainsi que les Blup obtenus pour chacune des 1007 familles communes aux deux essais ont été utilisées pour les analyses de génétique d'association.

Tableau IV-2: Variations phénotypiques observées sur les 2 essais mis en place à Sotuba au Mali et estimations des effets génétiques (Coefficient de variation génétique CVG% et héritabilité au sens large  $h_{bs}^2$ )

Caractères		Essai 2014			Essai 2015			Essais 2014 et 2015			
Abréviation	Description	Moy.	Ecart-type	Rep.	Moy.	Ecart-type	Rep.	Années	Fam	CVP% CVG%	$h_{bs}^2$
MAT	Matière Azotée Totale, %MS	3.89	1.45	0.94	3.73	2.01	0.94	*	***	45.34 21.48	0.37
CBW	Cellulose brute de Wende, %MS	31.74	4.81	0.85	34.45	5.24	0.97	***	***	15.32 8.77	0.49
NDF	Neutral Detergent Fiber, %MS	61.08	7.94	0.85	62.93	8.32	0.98	***	***	13.13 7.38	0.48
ADF	Acid Detergent Fiber, %MS	34.36	5.43	0.84	38.81	5.89	0.97	***	***	15.62 8.29	0.44
ADL	Acid Detergent Lignin, %MS	5.09	0.91	0.79	6.36	1.24	0.98	***	***	30.59 8.54	0.14
SMS	Solubilité de la matière sèche, %MS	47.19	7.98	0.81	40.44	8.75	0.97	***	***	18.99 10.45	0.46
SMO	Solubilité de la matière organique, %MO	46.27	8.43	0.83	38.46	8.95	0.97	***	***	20.66 11.79	0.49
cvs	Teneur en Cellulose Van Soest, %MS	29.29	4.76	0.86	32.45	4.8	0.97	***	***	15.3 8.29	0.45
hemi	Teneur en hémicellulose Van Soest, %MS	26.61	2.97	0.86	24.12	2.76	0.99	***	***	12.3 6.62	0.45
cbwSndf	Teneur en cellulose brute de Wende ajustée à la teneur en parois, %NDF	51.97	1.7	0.78	54.63	1.92	0.97	***	***	3.38 2.01	0.52
cvsSndf	Teneur en cellulose Van Soest ajustée à la teneur en parois, %NDF	47.83	1.69	0.83	51.48	1.49	0.97	***	***	3.36 1.11	0.2
hemiSndf	Teneur en hémicellulose Van Soest ajustée à la teneur en parois, %NDF	43.82	2.17	0.77	38.45	2.01	0.99	***	***	5.73 2.18	0.25
adlSndf	Teneur en lignines ajustée à la teneur en parois, %NDF	8.38	1.12	0.84	10.07	1.05	0.99	***	***	12.31 6	0.38
innndf	Digestibilité in vitro du NDF, %NDF	21.57	4.81	0.83	8.36	3.68	0.99	***	***	28.91 13.57	0.36



### **III.4- Identification des zones chromosomiques impliquées dans le contrôle génétique de la composition de la biomasse**

Suite au génotypage de l'ensemble des familles BC1F4 (Chapitre II), les associations entre la variabilité des polymorphismes mis en évidence et les données phénotypiques ont été détectées en utilisant le modèle linéaire mixte incorporé dans le package NAM développé par (Xavier et al., 2015) et implémenté dans le logiciel R. Un taux de faux positif de 1% correspondant à une valeur de  $-\log_{10}$  p-value de 2 a été utilisé comme valeur seuil pour déclarer les associations « significatives ». Ce seuil faible a été choisi afin de privilégier la découverte de régions intéressantes plutôt que l'élimination de faux positifs. La même stratégie « conservative » a été adoptée par Mace et al (2013). En effet, l'objectif de notre analyse n'était pas de valider les régions mais bien de les découvrir pour pouvoir les comparer à la bibliographie puis dans un second temps se focaliser sur les associations les plus stables (dans le temps et entre études). En outre, cette stratégie est cohérente avec le fait que les gènes causatifs ne sont pas forcément sous les pics d'association les plus forts comme cela a été montré par simulation (Cormier *et al.*, 2014). Les SNP proches physiquement (distance < à 50 Kb) et présentant des associations significatives avec les caractères ont été considérés comme définissant un QTL unique.

## **IV. Résultats**

### **IV.1. Variabilité des caractères et héritabilités**

Au niveau intra-essai, des répétabilités élevées ont été mises en évidence sur les 2 années d'expérimentation (de 0.77 à 0.99, Tableau IV-2) indiquant une faible variabilité expérimentale entre les blocs et des différences génétiques stables entre les génotypes témoins analysés. L'analyse simultanée des 2 années d'expérimentation (pour les 1007 familles communes aux 2 essais) a mis évidence des effets « familles » et « années » fortement significatifs ( $P < 0.0001$ ) pour tous les caractères. Les effets « année » importants qui ont été mis en évidence reflètent probablement les effets combinés du décalage des dates de semis (14 jours) et des différences de conditions environnementales (Figure IV-2) entre les 2 années.

Les effets génétiques fortement significatifs sont associés à des héritabilités au sens large de 0.14 (pour ADL) à 0.52 (pour cbwSndf) et des coefficients de variation génétique dans leurs majorités inférieures à 10 % (seule la teneur en matière azotée totale dépasse les 20 %).



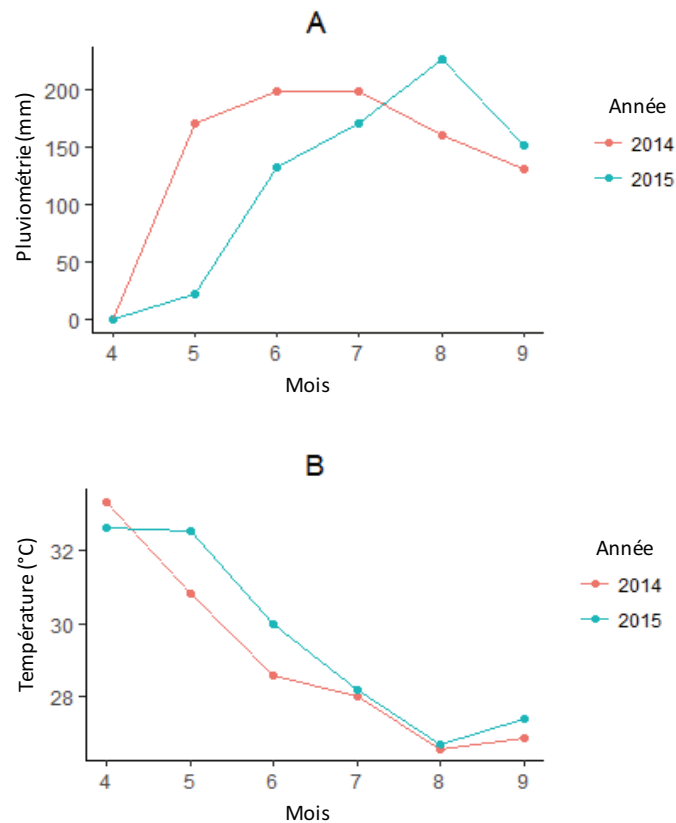


Figure IV-2 : Evolutions des précipitations et des températures pour les deux expérimentations mises en place sur le site de Sotuba au Mali. L'année 2015 est caractérisée par un déficit de précipitations important pour les mois de Juin et Juillet.

## IV.2. Corrélations entre les années et entre les caractères

Les mêmes structures de corrélations entre les caractères ont été mises en évidence pour les deux années d'expérimentation (Tableau IV-3). CBW, NDF ADF, CVS et Hemi sont très fortement corrélés entre elles ( $>0.8$ ), tandis qu'elles présentent des corrélations négatives ( $< -0.9$ ) avec les solubilités de la matière sèche (SMS), de la matière organique (SMO) et des parois cellulaires (IVNDFD). Le calcul des proportions des composantes de parois cellulaires met en évidence, une corrélation négative entre les teneurs en cellulose (cvsSndf) et en hémicellulose (hemiSndf) tandis que les teneurs en cellulose (cvsSndf) et en lignines (adlSndf) sont indépendantes. La digestibilité des parois, est quand à corrélée négativement aux teneurs en cellulose (cvsSndf) et en lignines (adlSndf) des parois et positivement à la teneur en hémicellulose (hemiSndf). Ces corrélations correspondent à celles publiées précédemment (Trouche *et al.*, 2014).

Si une convergence de structure globale des corrélations a été observée entre les deux années, en revanche pour l'ensemble des caractères analysés, les corrélations inter-annuelles sont faibles ( $r_{20104-2015} < 0.35$ ) (Tableau IV-3).

Tableau IV- 4 : Nombre d'associations détectées pour chaque caractère mesuré dans les différents environnements de culture au sein des 29 populations BCNAM.

Variables	Nombre d'associations présentant des LOD compris entre 2.64 et 8	Nombre d'associations présentant des LOD supérieurs à 8
MAT	8	2
CBW	8	0
NDF	6	0
ADF	6	0
ADL	6	8
cvs	5	0
hemi	14	0
cbwSndf	1	1
cvsSndf	6	0
hemiSndf	14	0
adlSndf	4	1
SMS	4	1
SMO	5	0
IVNDFD	6	1

Tableau IV-5 : Caractéristiques des 14 zones chromosomiques présentant des associations avec des valeurs de LOD supérieures à 8. Ces zones ont été détectées à minima sur la base des valeurs de blup inter-annuelles.

Variables	Chr	Pos (Mb) <sup>A</sup>	LOD	Association(s) déjà détectée(s)
MAT	1	1.92	10.51	
ADL	2	69.76	9.78	
ADL	3	6.86	8.83	
ADL	3	61.32	8.85	
ADL	3	70.30	12.09	
SMS	6	51.60	8.48	Dry Midrib » : Xu et al., 2000, Hart et al., 2001, Srinivas et al., 2009, Burks et al., 2015
cbwSndf	7	55.45	14.25	
IVNDFD	7	58.59	8.81	Hauteur : Rami et al., 1998, Parh, 2005, Brown et al., 2008, Morris et al., 2013, Higgins et al., 2014
adlSndf	7	58.68	8.35	
ADL	7	60.20	16.01	
ADL	8	49.99	45.01	
MAT	8	54.80	8.05	
ADL	9	49.46	11.76	
ADL	9	56.71	15.27	

<sup>A</sup> Position du SNP présentant la valeur de LOD la plus élevée. Cette position est basée sur la version 2.1 du génome du sorgho.

### **IV.3. Détection des zones chromosomiques impliquées dans le contrôle génétique de la composition biochimique des tiges et de leur digestibilité.**

Pour l'ensemble des caractères analysés, 394 associations présentant des LOD compris entre 3.26 et 45.01 ont été mises en évidence en considérant tous les environnements (201 zones détectées en 2014, 86 pour 2015 et 107 associations pour les Blup issus de l'analyse des deux années d'expérimentation). Ces associations correspondent à 378 zones chromosomiques uniques (Annexe 5). La figure IV-3 montre une représentation graphique « Manhattan plot » des différentes régions génomiques associées à la variation des caractères relatifs à la composition biochimique des tiges évaluée dans les environnements de culture 2014 et 2015.

Les 107 associations détectées sur les données BLup correspondent à 68 régions uniques (les SNP distants de moins de 50 kb ayant été fusionnées). Parmi ces dernières, 12 ont aussi été détectées sur une des deux années. Dans la majorité des cas, les valeurs de LOD observées pour les blup sont supérieures à celles observées pour les données annuelles. Dans la suite du document, seules les régions chromosomiques détectées au moins sur les données blup ont été considérées.

En fonction des caractères, des nombres variables de zones ont été mises en évidence (Tableau IV- 4). Les teneurs en hémicellulose dans la matière sèche (hemi) et dans les parois (hemiSndf) sont les caractères pour lesquels les plus grands nombres de zones ont été mis en évidence (14).

Avec des valeurs de LOD compris entre 3.5 et 6.95, seulement entre 0.01 et 0.11 % de variation phénotypique sont expliquées par ces deux caractères (Annexe 5).

Par ailleurs, aucune zone chromosomique à effet majeur n'a été détecté pour ces variables de teneurs en hémicellulose dans la matière sèche (hemi) et dans les parois (hemiSndf).

Quatorze régions chromosomiques présentant des valeurs de LOD supérieures ou égales à 8 ont été détectées (Tableau IV-5). Ces associations concernent les variables de teneurs en matière azotée totale (MAT : 2 QTL) ou teneur en protéines brutes, des teneurs en lignine des parois /fibres (ADL :8 QTL et adlSndf : 1 QTL), de teneur en cellulose brute par rapport aux fibres (cbwSndf : 1QTL), de solubilité de la matière sèche (SMS : 1 QTL) et de digestibilité des parois (IVNDFD : 1 QTL). Parmi ces zones, seulement deux sont contiguës (87 kb de distance).

La plus forte association significative a été obtenue pour la teneur en ADL (LOD= 45) sur le chromosome 8 à 49 Mb et explique 0.7 % de variation phénotypique associé à ce caractère.

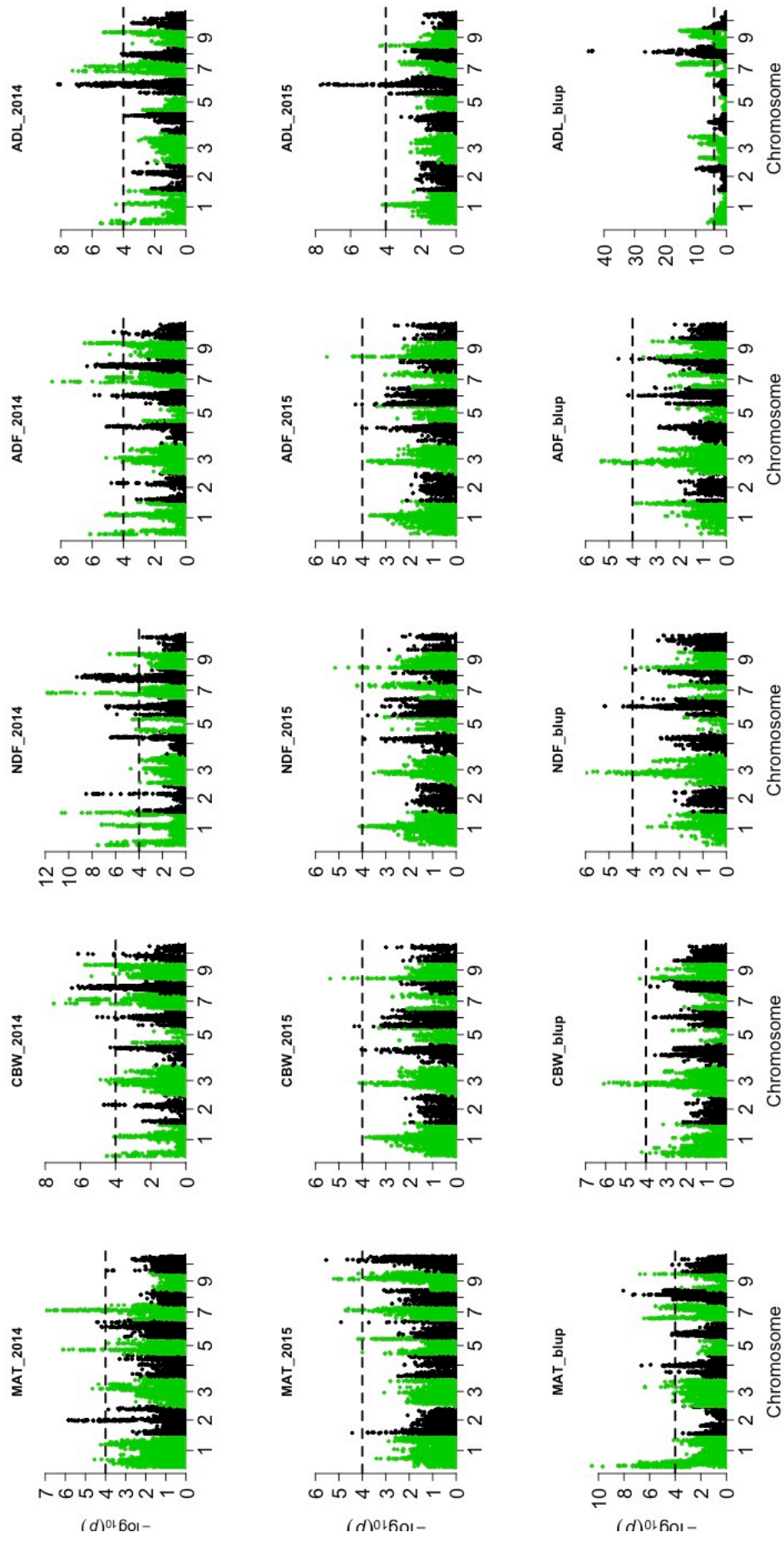


Figure IV-3 : Manhattan plot des résultats de cartographie d'association pour les variables de composition biochimique des tiges et de digestibilité évaluées dans les environnements de culture 2014 et 2015. Les résultats obtenus sur les valeurs de BLUP (2014 – 2015) sont aussi présentés. Les lignes horizontales en pointillés indiquent le seuil de significativité «  $\alpha=0.0001$  ».

Sur les 68 régions uniques mises en évidence, 44 concernent uniquement un caractère, tandis que les autres (24) correspondent à des zones où des associations entre des caractères différents sont mises en évidence.

La comparaison des zones impliquées dans le contrôle génétique de la composition de la biomasse et de celles détectées pour la durée du cycle et la longueur moyenne des entre-nœuds (Chapitre II) sur le même dispositif expérimental a permis de mettre en évidence 4 zones de co-localisation ;

- Chr 3, 51.3 Mb : cvsSndf et SFD (durée du cycle entre le semis et la feuille drapeau),
- Chr7 55.2 : cvs, NDF, cvsSndf et longueur moyenne des entrenœuds,
- Chr7 55.4 Mb : cbwSndf, Ivndfd et longueur moyenne des entrenœuds,
- Chr9 56.7 : MAT, ADL, Hemi, cvsSndf, hemiSndf et longueur moyenne des entrenœuds

Une distance minimale de 50 kb a été considérée pour déclarer une colocalisation. Il est intéressant de remarquer que les zones correspondant au chromosome 7 (entre 55.2 et 55.4 Mb) ont déjà été détecté par ailleurs pour leurs effets sur la croissance en hauteur (Rami et al., 1998, Parh, 2005, Brown et al., 2008, Morris et al., 2013, Higgins et al., 2014).

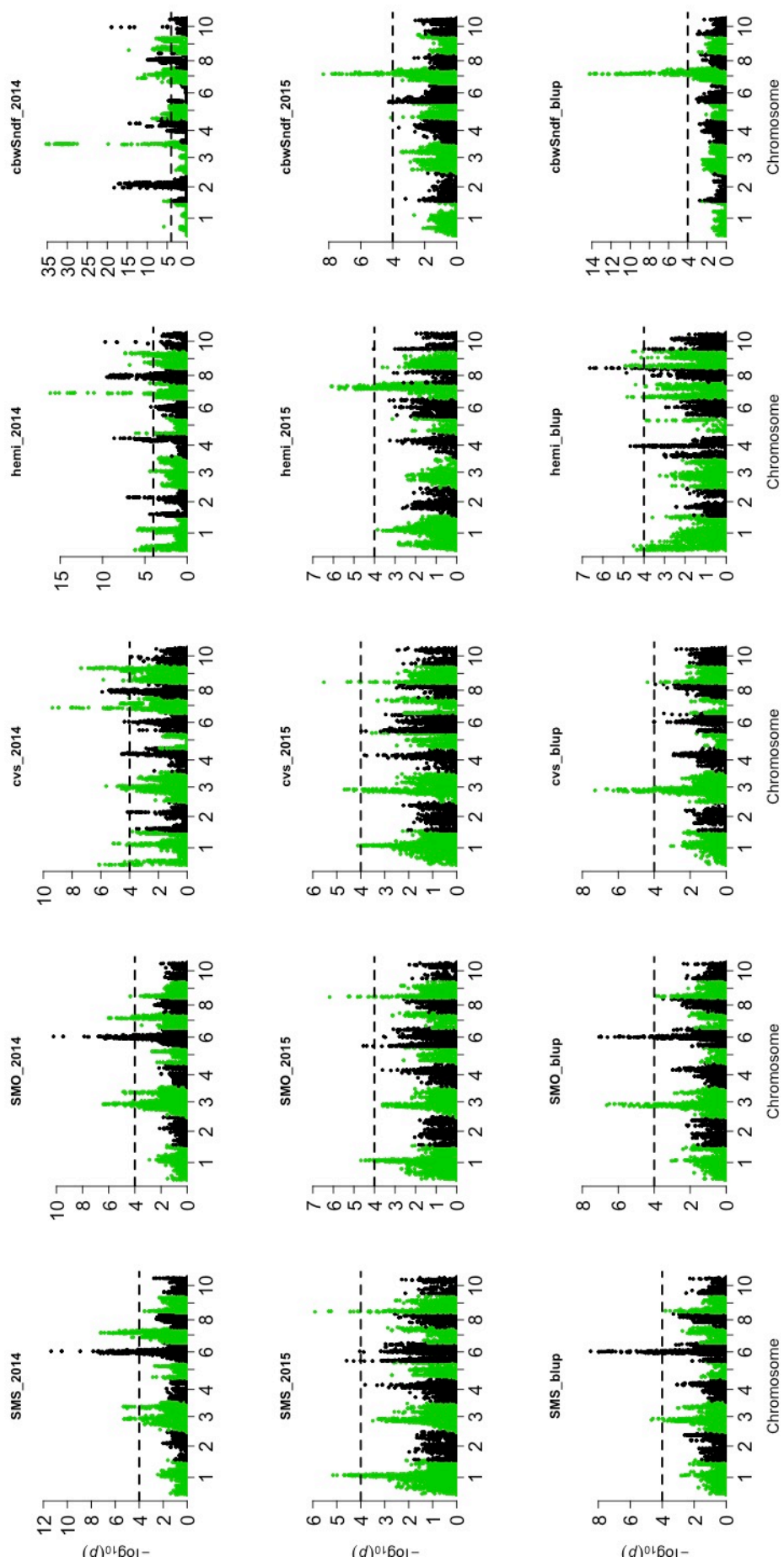


Figure IV-3 : suite



## V. Discussion : Une meilleure compréhension de l'architecture génétique de la qualité de la biomasse

La grande majorité des analyses ayant porté sur l'analyse de la composition de la biomasse végétative ont été focalisées sur la composante « sucres solubles » (Murray, Sharma, *et al.*, 2008; Ritter *et al.*, 2008; Shiringani, Frisch and Friedt, 2010; Shiringani and Friedt, 2011; Burks *et al.*, 2015). Ces études ont notamment mis en évidence, l'importance de la zone comprenant le gène D (Dry Midrib ou nervure sèche), située aux alentours de 51 Mb sur le chromosome 6, dans le contrôle génétique de ce caractère (Xu *et al.*, 2000; Hart *et al.*, 2001; Srinivas *et al.*, 2009). Cette même région a été mise en évidence au sein de la population BCNAM qui a été analysée dans le cadre de cette étude avec des associations fortes (LOD >8 sur les données Blup) avec la solubilité de la matière organique (SMO) qui est très liée aux proportions de sucres solubles et de carbohydrates pariétaux dans la biomasse. En outre, cette même région a aussi été détectée au sein d'un panel d'association de 175 génotypes non sensibles à la photopériode (David Pot, communication personnelle). Au sein de la région mise en évidence, 3 gènes présentent des fonctions compatibles avec une implication possible dans le contrôle génétique de la jutosité et de la teneur en sucres ont été identifiés :

- Sobic.006G143900 : est un impliqué dans le transport des sucres
- Sobic.006G145400 : est impliqué dans l'hydrolyse des O-glycosyl
- Sobic.006G145500 : présente une annotation fonctionnelle en lien avec le métabolisme des carbohydrates

Une analyse plus fine de la décroissance du déséquilibre de liaison dans la zone doit être menée afin de tenter d'affiner les zones d'intérêt. En outre, l'analyse précise des haplotypes présents au sein des parents du dispositif BCNAM ainsi que leur phénotypage plus précis pourra permettre de préciser l'implication des différents gènes.

Outre, cette région du gène D, Brenton *et al.* (2016), au travers d'une analyse de génétique d'association sur un panel composé de 390 accessions de génotypes à fortes productions de biomasse, ont mis en évidence 5 régions additionnelles impliquées dans le contrôle des teneurs en carbohydrates structuraux (NDF) et non structuraux (NFC). Ces cinq régions sont situées sur le chromosome 4 (entre 63.30 – 63.34 Mb) et sur le chromosome 6 (entre 43.20 et 49.78 Mb). Bien que la zone sur le chromosome 6 soit relativement proche de celle mises en évidence au sein de notre étude et des précédentes (zone du gène D), la décroissance du déséquilibre de liaison observée dans cette région indique que ce sont des régions indépendantes.

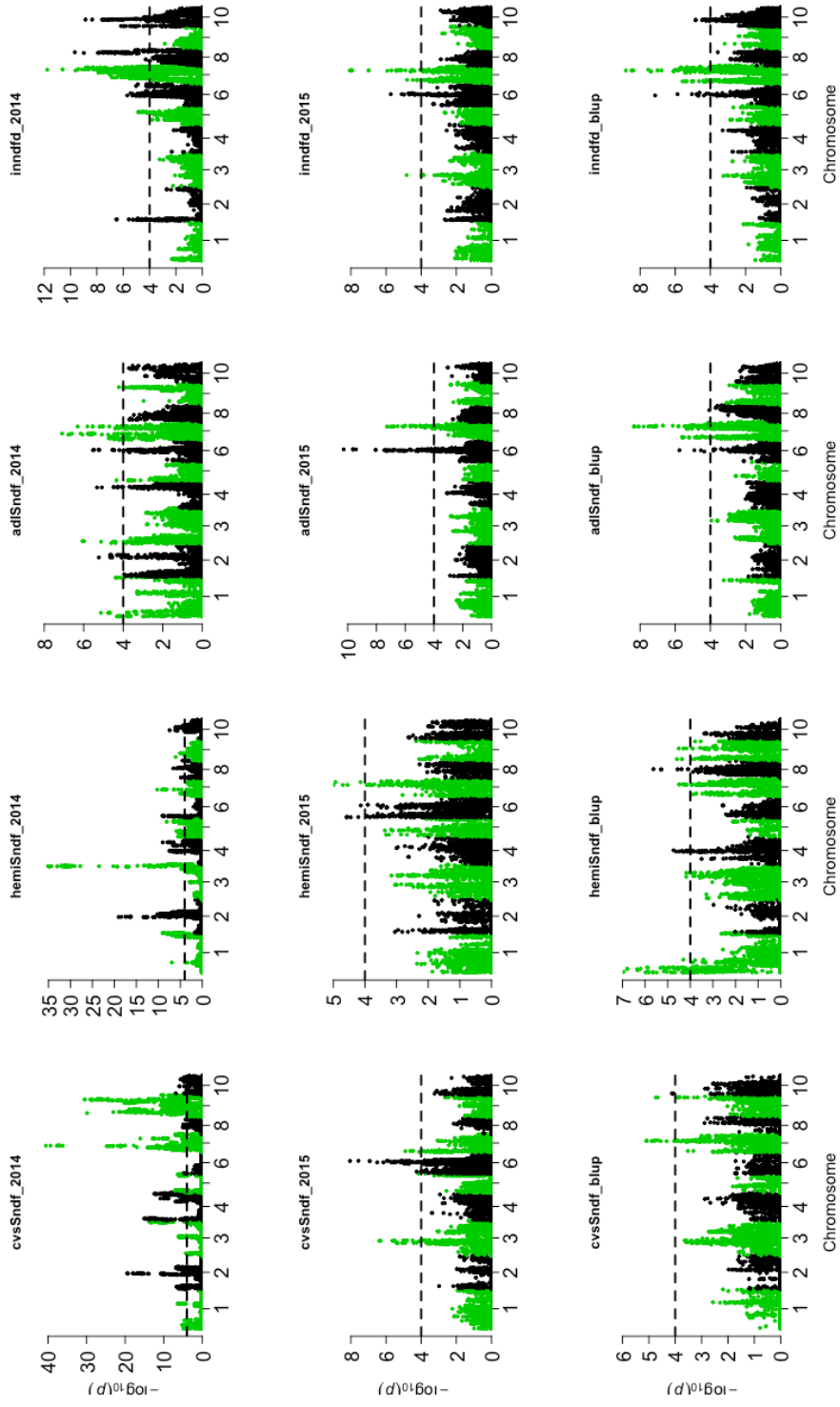


Figure IV-3 : suite

Sur les populations BCNAM aucune association (détectée sur les valeurs de Blup) n'a été mise en évidence à proximité de ces 5 régions détectées par Brenton et al (Brenton *et al.*, 2016).

En parallèle des études ciblant l'accumulation de sucres solubles dans les tiges, seules deux études ont exploré le déterminisme génétique des carbohydrates structuraux chez le sorgho. Murray et al en 2008 (Murray, Rooney, *et al.*, 2008) ont analysé le déterminisme génétique de ces composés dans les tiges et les feuilles d'une population biparentale issue du croisement entre un sorgho grain (Btx623) et un sorgho sucré (Rio). Ces derniers ont mis en évidence une déconnexion entre les déterminismes génétiques de l'accumulation de ces composés dans les tiges et les feuilles et de nombreuses colocalisations avec des zones impliquées dans le contrôle de la floraison, de la hauteur des plantes et le tallage. En outre, comme mentionné précédemment, la même équipe coordonnée par Stephen Kresovich a étendu la base de diversité génétique explorée en effectuant une étude de génétique d'association sur un panel de 390 génotypes sélectionnés pour leur aptitude à la production de biomasse (Brenton *et al.*, 2016). Néanmoins, dans le cadre de cette étude, seule la teneur globale en fibres totales (NDF) a été analysée et seules des zones communes avec celles détectées pour la teneur en sucres solubles (NDF) ont été mises en évidence (voir ci-dessus).

L'étude présentée ici, basée sur la caractérisation d'environ 1200 familles BC1F4, a permis la mise en évidence de 68 régions uniques impactant la composition biochimique de la tige et ses propriétés de digestibilité. Parmi celles-ci quatre co-localisent avec des régions impliquées dans le contrôle de la hauteur des tiges et la durée du cycle qui ont été mises en évidence au sein du chapitre II sur le même dispositif :

- Chr 3, 51.36 Mb : détection d'une association pour cvsSndf (LOD = 3.64) et colocalisation avec une association pour la durée du cycle semis floraison (LOD = 16.03),
- Chr 7, 55.19 – 55.20 – 55.45 Mb détection de deux associations pour cvsSndf (Chr7, 55.20, LOD = 12.145) et pour cbwSndf (Chr 7, 55.45 Mb, LOD = 14.25) qui colocalisent avec une association pour la longueur moyenne des entre-nœuds (LOD = 38.14). Cette région chromosomique avait déjà été mise en évidence par d'autres auteurs pour présenter des effets sur la hauteur des plantes (Rami et al., 1998, Parh, 2005, Brown et al., 2008, Morris et al., 2013, Higgins et al., 2014),

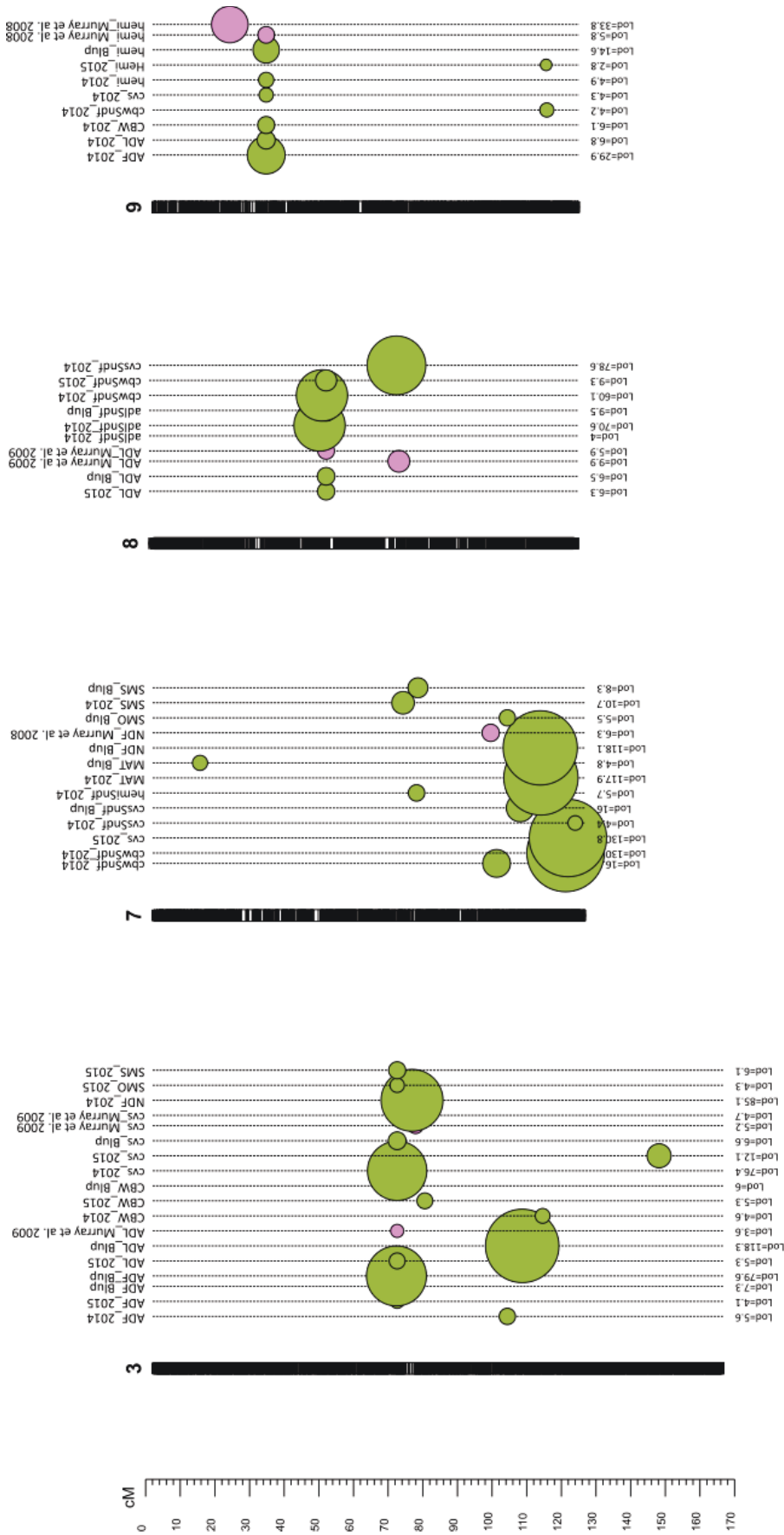


Figure IV-4 : Colocalisations entre les régions détectées au sein du dispositif BCNAM (en vert) et celles identifiées par Seth C Murray, Arun Sharma, et al. (2008) (en rose).

- Chr 9, 56.71 Mb : détection d'une association pour ADL (LOD = 15.27) qui colocalise avec une région impliquée dans le contrôle de la longueur moyenne des entre-nœuds (LOD =4.3).

Si il semble clair que les zones détectées sur le chromosome 3 et le chromosome 7 correspondent bien à des effets pléiotropiques de gènes de maturité (Chr3, 51.36 Mb) et de nanisme (Chr 7, 55.19 – 55.20 Mb) en raison des valeurs très élevées de LOD pour ces deux caractères, la valeur de LOD observée pour la teneur en lignine de la matière sèche sur le chromosome 9 qui est largement supérieure à celle observée pour la longueur moyenne des entre-nœuds nous amène à ne pas écarter trop rapidement cette zone.

De façon plus globale, le faible taux de colocalisation entre les zones associées avec les caractères de croissance, de cycle et de composition de la biomasse laissent penser que la plupart des régions détectées correspondent bien à des effets directs de gènes impliqués dans la biosynthèse des différents composés de la biomasse.

L'analyse des colocalisations entre les QTL détectés par Murray et al (2008) et les associations détectées sur la base des valeurs de blup des 2 expérimentations du dispositif BCNAM a mis en évidence une zone de colocalisation sur le chromosome 3 (Figure IV- 4). En effet, la zone localisée à 33 Mb ( $\approx$  72 cM) où a été détecté un QTL de teneur en ADL (Murray, Rooney, *et al.*, 2008) comprend aussi dans notre étude des associations avec CBW (LOD = 6.09), NDF (LOD = 5.97), ADF (LOD = 5.32), CVS (LOD = 7.29) et SMO (LOD = 6.64). Sur la base des valeurs de LOD et en prenant en compte les fortes corrélations existant entre ces caractères, il est probable que cette zone contrôle principalement la teneur en cellulose des tiges.

Néanmoins il est important de souligner ici, qu'en réduisant les critères de sélection des associations significatives (prise en considération des associations détectées uniquement sur une des deux années d'expérimentations) de plus nombreuses zones en commun entre les deux études existent. On peut notamment citer ici :

- Chr 7 : 63 Mb : QTL de NDF mis en évidence par Murray et al 208 et association pour la teneur en hémicellulose détectée au sein de l'essai 2014 (LOD =4.39),
- Chr8 : 18 Mb : QTL d'ADL mis en évidence par Murray et al 2008 et associations pour CBW, NDF (LOD=9.3), ADF (LOD = 6.32), ADL (LOD = 6.32), CVS (LOD=5.91) et Hemi (LOD = 9.48) détectées sur le dispositif BCNAM en 2014,

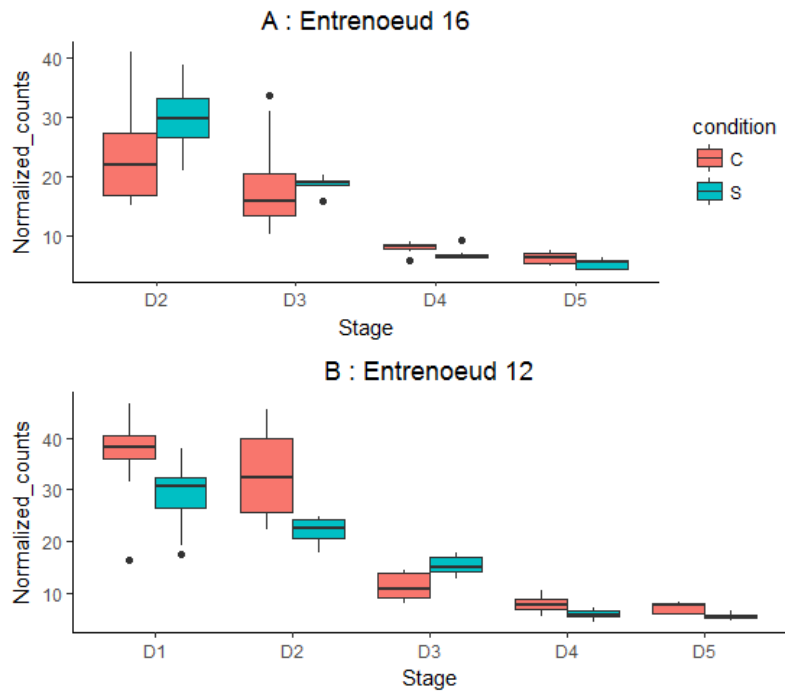


Figure IV-5 : Evolution du niveau d'expression du gène Sobic.003G391600 au cours du développement de deux entrenœuds (A : entrenœud 12, B : entrenœud 16) situés à deux niveaux différents sur la tige (base et milieu respectivement). Le stade D1 correspond au stade d'élongation de l'entrenœud 12 (le plus bas) et le stade D5 correspond au stade de maturité du grain (tige complètement mature). L'expression du gène a été analysée au sein de deux conditions d'alimentation hydrique (irriguée en rouge et alimentation hydrique limitée en bleu) pour un hybride commercial (Biomass140, Hercules de Semences de Provence) de type Biomasse durant une année de culture en conditions tempérées (INRA Mauguio, France).

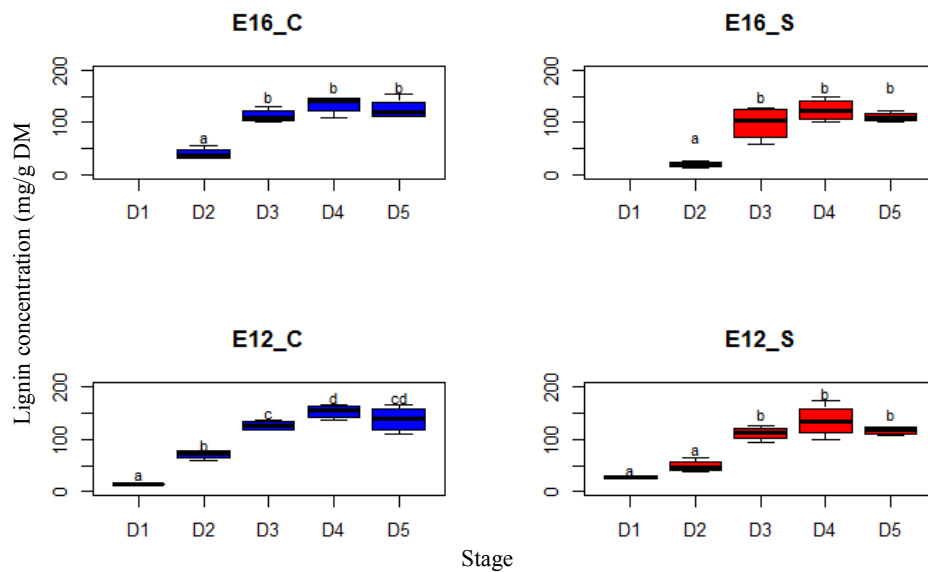


Figure IV-6 : Evolution de la teneur en lignines de la matière sèche au cours du développement de deux entrenœuds (12 et 16) soumis à deux conditions d'alimentation hydrique (C : condition contrôle, S : condition d'alimentation hydrique limitée).

- Chr 9 : 41 Mb : QTL d'hémicellulose détecté par Murray et al 2008 et associations pour NDF (LOD= 4.34), cwSndf (LOD=14.61), CVS (LOD= 5.83), cvsSndf (LOD=29.87), Hemi (LOD=6.8), HemiSndf (LOD=6.606) mises en évidence sur le dispositif BCNAM en 2014.

En ce qui concerne plus spécifiquement les 8 zones détectées avec des LOD supérieurs à 8 pour les teneurs en lignine de la matière sèche, aucune colocalisation avec les mutants bmr pré-identifiés n'a été mise en évidence (Saballos *et al.*, 2009, 2012; Sattler *et al.*, 2014). En revanche, un gène codant pour une protéine présentant une forte homologie avec une « 4-coumarate:coenzyme A ligase » (Sobic.003G391600) est localisé à proximité (3 kb) de la région du Chromosome 3 (70.3 Mb) présentant une association avec la lignine (LOD = 12.09). Ce gène présente une expression maximale dans les premiers stades du développement des entrenoeuds qui diminue ensuite pour être quasiment non détectable à partir du stade D4 (Figure IV-5, communication personnelle Angélique Berger). En étant encore bien loin d'une validation de l'implication fonctionnelle de ce gène dans le contrôle de l'accumulation des lignines, il est intéressant de remarquer que ce profil est en relatif accord avec l'expression importante de la teneur en lignines entre les stades D1 et D3 (Figure IV- 6).

Une analyse systématique des propriétés des gènes sous-jacents aux régions détectées sera nécessaire pour proposer des gènes candidats qui pourront ensuite être ciblés pour des études de validation fonctionnelles. Outre leurs profils d'expression au cours du développement de l'entrenœud, il sera notamment intéressant de regarder leur profil de diversité génétique (types de polymorphismes, nombre et fréquences des haplotypes).

## Chapitre V. Discussion générale et perspectives

---

La présente étude avait pour objectif d'améliorer notre compréhension des déterminants génétiques de la production et de qualité de la biomasse végétative chez le sorgho. La production de biomasse est sous la dépendance de 3 facteurs principaux : la durée du cycle (qui conditionne le nombre d'entre-nœuds), le nanisme (qui conditionne la taille individuelle des entre-nœuds) et le tallage (composante qui ne sera pas développée dans cette étude). La qualité de la biomasse végétative quant à elle, est influencée par la teneur en sucres solubles des tiges (c'est en effet la tige qui contribue le plus à la biomasse végétative du sorgho dans le cas des idéotypes moyens à hauts (>2m)) et par la composition et la structure des parois cellulaires (teneurs en lignines, cellulose et hémicellulose). Une meilleure compréhension de la mise en place de ces caractères et plus précisément de leur déterminisme génétique devrait permettre d'optimiser les stratégies de sélection visant la valorisation de la biomasse végétative.

Dans ce contexte, cette thèse avait pour objectifs :

- D'explorer les corrélations entre la production (en grande partie déterminée par la sensibilité à la photopériode et par le nanisme) et la qualité de la biomasse (sucres solubles, carbohydrates structuraux et lignines),
- D'affiner notre compréhension du déterminisme génétique de la floraison et de la croissance en hauteur (notamment pour des génotypes sensibles à la photopériode),
- Et d'améliorer notre compréhension du déterminisme génétique de la qualité de la biomasse.

### **I. Les génotypes sensibles à la photopériode : un pool peu exploré jusqu'à présent mais porteurs d'innovation variétale au Nord et aux Suds**

La plupart des programmes de sélection du sorgho ont été basés jusqu'à très récemment sur des génotypes rendus insensibles à la photopériode. Les deux objectifs de cette focalisation étaient : i) d'adapter les génotypes de sorgho aux climats tempérés et ii) de développer des variétés adaptées à des saisons des pluies courtes en milieu tropical. Cette élimination du photopériodisme par conversion des génotypes sensibles à la photopériode (notamment les



ressources génétiques Africaines) en géotypes insensibles a été possible grâce à des gènes majeurs dont les allèles récessifs inhibent la sensibilité à la photopériode (Ma1 - Ma6). La réduction de la taille des tiges a aussi été un objectif important des programmes d'améliorations afin d'optimiser les indices de récolte. Dans ce contexte, les gènes majeurs de nanisme mis en évidence chez le sorgho (dw1 -d w4) ont été intensément utilisés.

Néanmoins si le programme de conversion initié aux USA a été couronné de succès et a permis l'expansion de la culture du sorgho aux milieux tempérés (à des latitudes plus élevées), l'utilisation de variétés de tailles courtes et insensibles à la photopériode dans les conditions semis arides n'a pas apporté les résultats escomptés.

Récemment, un regain d'intérêt pour la sensibilité à la photopériode est apparu. En effet, ce caractère peut apporter des innovations significatives dans, au moins deux contextes :

- En premier lieu, l'augmentation de la sensibilité à la photopériode permet d'allonger la longueur des cycles et donc de maximiser la production de biomasse végétative. Cette dernière pouvant être valorisée au sein de différentes filières (alimentation animale (Nord et Sud), et énergie et biomatériaux (principalement au Nord pour l'instant)).
- La sensibilité à la photopériode permet une meilleure adaptation au milieu soudano sahélien caractérisé par une pluviométrie extrêmement variable. La sensibilité à la photopériode permet dans ces conditions, indépendamment de la date de semis, d'initier la phase de remplissage et la maturation des grains lorsque les conditions climatiques et biotiques ne sont pas trop adverses.

Au niveau de l'analyse de la qualité de la biomasse et de ses interactions avec la production, la grande majorité des études ont été basées sur des géotypes très insensibles à la photopériode. En outre ce sont souvent des populations issues de croisements entre des parents « courts et peu riches en sucres solubles » et des parents « hauts et riches en sucres solubles » qui ont été utilisés.

Dans ce contexte, les études présentées au sein de ce document ont largement mobilisé des géotypes sensibles à la photopériode et ont tenté de maximiser la variabilité disponible entre les parents au sein des différents idéotypes (large diversité au sein des géotypes de tailles courtes et de tailles plus hautes). Les analyses menées sur ce matériel végétal, original par rapport à ce qui a généralement été étudié dans la bibliographie, ont permis :

- D'affiner notre compréhension de la variabilité de la durée du cycle chez les génotypes sensibles à la photopériode (P118 ; BCNAM), ce qui avait été très peu exploré jusqu'ici
- De souligner la stabilité plus importante des génotypes sensibles en réponse à des modifications environnementales (Chapitre III : effet d'interaction PPS x Semis)

## **II. Une meilleure compréhension du déterminisme génétique de la croissance en hauteur, de la floraison et de la qualité de la biomasse**

A l'initiation de cette thèse en février 2014, peu d'informations étaient disponibles sur le contrôle génétique de la date de floraison au sein des génotypes sensibles à la photopériode (Mace, Hunt and Jordan, 2013), en outre au niveau de la qualité de la biomasse la majorité des études avaient été menées sur des populations biparentales ou des panels d'associations incluant des idéotypes très divers qui favorisaient la mise en évidence d'effet pléiotropiques de gènes de nanisme ou de maturité. Cette thèse nous a permis de :

- mettre en évidence un QTL majeur contrôlant la réaction à la photopériode au sein de génotypes sensibles à ce stimulus et de proposer un gène candidat sous-jacent. L'approche mise en œuvre combinant l'utilisation d'un modèle de culture et la mendélisation des paramètres analysés a permis d'analyser le déterminisme d'un caractère n'étant pas évident à analyser de façon directe. D'une part l'utilisation de 2 parents photopériodiques divergents relativement peu en termes de sensibilité à la photopériode a joué un rôle clé et d'autre part la décomposition du processus de floraison en ses paramètres unitaires au travers de l'utilisation d'un modèle écophysiological a donné accès à une composante inaccessible à un phénotypage direct de la longueur de cycle. Le QTL identifié impacte le contrôle de la photopériode critique en dessous de laquelle la plante ne réagit plus à la durée du jour. Ce résultat ouvre de nouvelles voies de sélection puisqu'il permet d'envisager de modifier la précocité des variétés sans éliminer la sensibilité à la photopériode,
- De valider des régions chromosomiques dont l'impact sur la hauteur des plantes avait déjà été mis en évidence (gènes majeurs...) mais aussi de détecter de nouvelles zones avec des effets conséquents tant sur la floraison que sur la composition biochimique de la biomasse et ses propriétés,

- De montrer que les relations entre les caractères varient significativement entre les classes de nanismes qui sont mobilisées (cf. chapitre III),
- De mettre en évidence de nouveaux QTL à effets forts (chromosomes 2, 3, 7, 8 et 9), mais également de valider l'importance de la région du gène D sur la teneur en sucres solubles des tiges,
- De proposer pour quelques régions des gènes candidats possibles qui pourront constituer les cibles de prochaines études de validation fonctionnelles.

### **III. Optimiser les analyses des résultats pour améliorer notre compréhension de l'architecture génétique des caractères**

#### **III.1- Des dispositifs de grandes tailles méritant une meilleure estimation des variabilités environnementales et des interactions GXE**

Bien que le dispositif expérimental en blocs augmentés utilisé pour l'évaluation des différentes familles soit pertinent, il présente des limites (évaluation sous optimale des variabilités environnementales en fonction de la structure de l'hétérogénéité présente). Dans ce contexte, il serait donc intéressant de s'orienter vers l'utilisation de modèles d'ajustement plus élaborés, prenant en compte la variabilité spatiale du terrain (variabilité intra annuelle), mais également de la variabilité environnementale interannuelle. Dans ce contexte, toujours dans l'objectif de mieux appréhender la variabilité environnementale subie par les génotypes et évaluer au mieux les effets d'interactions Génotype X Environnement qui ont été très mal pris en compte au sein des analyses qui ont été menées il sera pertinent de s'orienter vers une utilisation plus intensive des modèles de développement tels que Samara, CERES ou Ecomeristem qui ont été adaptés au sorgho (Ritchie and Alagarswamy, 1989b, Ritchie and Alagarswamy, 1991).

#### **III.2- Une amélioration possible et nécessaire des modèles d'analyses appliqués au dispositif BCNAM**

L'analyse des données phénotypiques collectées sur les différentes familles du dispositif BCNAM a été réalisée selon le modèle linéaire mixte développé par Yu et al. (2008) et implémenté dans le package NAM pour son utilisation avec des plans de croisement du type Nested Association Mapping. Néanmoins si cette stratégie prend bien en compte la structure familiale elle ne tire pas parti des connexions entre les parents et ne prend pas en compte la distribution des haplotypes entre les parents. Récemment, Giraud et al. (2014) et (Bouchet *et al.*, 2017) ont développés de nouvelles méthodes d'analyses spécifiques au dispositif NAM qui tirent vraiment partie des particularités de ces dispositifs.

#### **IV. Le dispositif BCNAM : accélérer et affiner notre compréhension du déterminisme génétique et moléculaire des caractères d'intérêt et optimiser l'efficacité de la sélection**

Contrairement aux dispositifs NAM au sein desquels ce sont des familles de lignées recombinantes qui sont développées, les plans de croisement BCNAM impliquent une étape complémentaire de rétrocroisement. Cette étape additionnelle offre des bénéfices évidents en termes d'analyses génétiques et de sélection qui ont été décrits au sein du Tableau II-2.

En termes de sélection, la possibilité d'utiliser l'analyse de liaison en complément des approches « globales » de génétique d'association facilite l'utilisation du matériel pour les programmes de sélection. En effet, de nombreux QTL spécifiques à un faible nombre de populations en sont pas forcément détectées sur les analyses globales (tout au moins celles qui ont été menées jusqu'à présent sans tirer au mieux parti des informations parentales). Dans ce contexte, les analyses de liaison menées au niveau intra-population permettent d'attribuer des valeurs génétiques aux différentes familles et d'initier ensuite des schémas de recombinaison inter-populations. Dans ce contexte, plusieurs populations de sélection ont été initiées au travers de croisements inter-populations dans le but, entre autres, de cumuler les allèles de précocité/tardiveté identifiés dans les différentes populations BCNAM. Les travaux présentés dans le cadre de cette thèse ne font qu'effleurer la richesse des informations exploitables sur le dispositif BCNAM qui a été développé. Outre la caractérisation de la floraison, de la croissance en hauteur et de la qualité des tiges qui a été présentée (en partie) dans ce document, les évaluations ont aussi porté sur des caractères de résistance et de tolérances à certaines maladies (anthracnose) et insectes (Cécidomyie) et sur le rendement et la qualité du grain qui sont des caractères cibles clés dans le contexte d'Afrique de l'Ouest (et plus généralement au niveau mondial).

En termes d'identification des gènes sous-jacents aux zones d'intérêt détectées, qui est une action primordiale dans le cadre des transferts d'informations entre populations et programmes de sélection, le chemin est encore long. En effet, comme cela a été initié pour les zones avec des impacts fortement significatifs sur les teneurs en lignines au sein du dispositif BCNAM, une caractérisation des gènes se trouvant dans ces zones est nécessaire. Cette caractérisation comprend différents volets :

- L'identification des haplotypes disponibles dans les populations et l'analyse de leur structure de diversité (nombre, fréquence, distance entre les haplotypes, proportion de mutations synonymes et non synonymes...),
- L'identification des polymorphismes disponibles sur des bases génétiques plus larges intégrant potentiellement des accessions sauvages,
- L'analyse des profils d'expression des gènes. Dans le cas de gènes candidats pour la biomasse, sont-ils exprimés dans les entre-nœuds, si oui à quel stade ?
- L'isolement des familles se rapprochant au mieux de lignées quasi-isogéniques pour tenter de préciser les effets génétiques des différents haplotypes...

En complément de ces approches sur la diversité naturelle, des approches d'analyses de mutants induits et d'ingénierie génétique sont aussi en cours de développement pour tendre, à terme, vers la constitution d'une bibliothèque de gènes impactant les caractères d'intérêt agronomiques qui constituera une ressource pertinente pour le développement de variétés à façon.

## Références bibliographiques

---

- ABDULAI, A. L., KOURESSY, M., VAKSMANN, M., ASCH, F., GIESE, M. & HOLGER, B. 2012. Latitude and Date of Sowing Influences Phenology of Photoperiod-Sensitive Sorghums. *Journal of Agronomy and Crop Science*, 198, 340-348.
- AFNOR 1977. Aliments des animaux. Dosage des cendres brutes. *Animal Feed Science and Technology*, 18-101, 2pp.
- AFNOR 2009. Aliments des animaux. Dosage de l'azote et calcul de la teneur en protéines brutes – Partie 2 : Méthode de digestion ne bloc et distillation à la vapeur. *Animal Feed Science and Technology*, 16pp.
- AINSWORTH, E. A. & ORT, D. R. 2010. How Do We Improve Crop Production in a Warming World? *Plant Physiology*, 154, 526-530.
- AITKEN, Y. 1974. *Flowering Time, Climate and Genotype: The Adaptation of Agricultural Species to Climate Through Flowering Responses*, Melbourne University Press.
- ALAGARSWAMY, G. & RITCHIE, J. T. Phasic Development in CERES-Sorghum Model. Hodges, T. (Ed.), *Predicting Crop Phenology*, CRC Press, Boca Raton, 1991. 143–152.
- ALAGARSWAMY, G., RITCHIE, J. & FLINT, B. Effect of high temperature on leaf appearance rates in maize, rice, sorghum, and pearl millet. 1986 Annual Meetings. New Orleans, Louisiana, ASA, 1986. 10.
- ALMODARES, A. & HADI, M. R. 2009. *Production of bioethanol from sweet sorghum: A review*.
- AMASINO, R. M. & MICHAELS, S. D. 2010. The timing of flowering. *Plant Physiology*, 154, 516-520.
- ANDREWS, D. J. 1973. Effects of date of sowing on photosensitive Nigerian sorghums. *Experimental Agriculture*, 9, 337-346.
- ARENDS, D., PRINS, P., JANSEN, R. C. & BROMAN, K. W. 2010. R/qlt: high-throughput multiple QTL mapping. *Bioinformatics*, 26, 2990-2992.
- ASHBURNER, M., BALL, C. A., BLAKE, J. A., BOTSTEIN, D., BUTLER, H., CHERRY, J. M., DAVIS, A. P., DOLINSKI, K., DWIGHT, S. S., EPPIG, J. T., HARRIS, M. A., HILL, D. P., ISSEL-TARVER, L., KASARSKIS, A., LEWIS, S., MATESE, J. C., RICHARDSON, J. E., RINGWALD, M., RUBIN, G. M. & SHERLOCK, G. 2000. Gene Ontology: tool for the unification of biology. *Nat Genet*, 25, 25-29.
- ATIS, I., KONUSKAN, O., DURU, M., GOZUBENLI, H. & YILMAZ, S. 2012. Effect of harvesting time on yield, composition and forage quality of some forage sorghum cultivars. *Int. J. Agric. Biol*, 14, 879-886.
- AYUB, M., NADEEM, M. A., TANVEER, A. & HUSNAIN, A. 2002. Effect of different levels of nitrogen and harvesting times on the growth, yield and quality of sorghum fodder. *Asian J. Plant Sci*, 4, 304-307.
- BANDILLO, N., RAGHAVAN, C., MUYCO, P. A., SEVILLA, M. A. L., LOBINA, I. T., DILLA-ERMITA, C. J., TUNG, C.-W., MCCOUCH, S., THOMSON, M. & MAULEON, R. 2013. Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. *Rice*, 6, 11.

- BANGBOL SANGMA, H. 2013. Genetic characterization of flowering time in sorghum.
- BANGBOL, S. H. 2013. *Genetic characterization of flowering time in sorghum*. Doctor of Philosophy Doctor of Philosophy, University of Queensland.
- BELAYACHI, L. & DELMAS, M. 1995. Sweet sorghum: A quality raw material for the manufacturing of chemical paper pulp. *Biomass and bioenergy*, 8, 411-417.
- BELLIARD, J. 1982. *analyse génétique et physiologique du système de contrôle photopériodique de la floraison chez le mil africain*. Doctorat d'état Doctorat d'état, Université de paris-Sud, Orsay.
- BENDIX, C., MARSHALL, C. M. & HARMON, F. G. 2015. Circadian Clock Genes Universally Control Key Agricultural Traits. *Molecular Plant*, 8, 1135-1152.
- BERENJI, J. & DAHLBERG, J. 2004. Perspectives of Sorghum in Europe. *Journal of Agronomy and Crop Science*, 190, 332-338.
- BIELUSZEWSKI, T., GALGANSKI, L., SURA, W., BIELUSZEWSKA, A., ABRAM, M., LUDWIKOW, A., ZIOLKOWSKI, P. A. & SADOWSKI, J. 2015. AtEAF1 is a potential platform protein for Arabidopsis NuA4 acetyltransferase complex. *BMC Plant Biology*, 15, 75.
- BLAIR, M. W., IRIARTE, G. & BEEBE, S. 2006. QTL analysis of yield traits in an advanced backcross population derived from a cultivated Andean× wild common bean (*Phaseolus vulgaris* L.) cross. *Theoretical and Applied Genetics*, 112, 1149-1163.
- BOIVIN, K., DEU, M., RAMI, J.-F., TROUCHE, G. & HAMON, P. 1999. Towards a saturated sorghum map using RFLP and AFLP markers. *TAG Theoretical and Applied Genetics*, 98, 320-328.
- BOSS, P. K., BASTOW, R. M., MYLNE, J. S. & DEAN, C. 2004. Multiple Pathways in the Decision to Flower: Enabling, Promoting, and Resetting. *The Plant Cell*, 16, S18-S31.
- BOUCHET, S., OLATOYE, M. O., MARLA, S. R., PERUMAL, R., TESSO, T., YU, J., TUINSTRA, M. & MORRIS, G. P. 2017. Increased Power To Dissect Adaptive Traits in Global Sorghum Diversity Using a Nested Association Mapping Population. *Genetics*, 206, 573-585.
- BRADY, J. A. 2006. *Sorghum Ma5 and Ma6 maturity genes*. Doctor of philosophy, Texas A&M University.
- BROMAN, K. W. & SEN, S. 2009. *A Guide to QTL Mapping with R/qtl*.
- BROMAN, K. W., WU, H., SEN, S. & CHURCHILL, G. A. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, 19, 889-890.
- BROWN, P., KLEIN, P., BORTIRI, E., ACHARYA, C., ROONEY, W. & KRESOVICH, S. 2006. Inheritance of inflorescence architecture in sorghum. *Theoretical and Applied Genetics*, 113, 931-942.
- BROWN, P. J., ROONEY, W. L., FRANKS, C. & KRESOVICH, S. 2008. Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes. *Genetics*, 180, 629-637.
- ÇAKMAKÇI, S., GÜNDÜZ, İ., TÜSÜZ, M. A., ÇEÇEN, S. & AYDINOĞLU, B. 1999. Effects of different harvesting times on yield and quality of sorghum silage (*Sorghum bicolor* L.). *Turkish Journal of Agriculture and Forestry*, 23, 603-612.
- CAMPBELL, I., CASADY, A. & CROOK, W. 1975. Effects of a Single Height Gene (DW3 of Sorghum on Certain Agronomic Characters. *Crop Science*, 15, 595-597.
- CARMI, A., AHARONI, Y., EDELSTEIN, M., UMIEL, N., HAGILADI, A., YOSEF, E., NIKBACHAT, M., ZENOU, A. & MIRON, J. 2006. Effects of irrigation and plant density on yield, composition and in vitro digestibility of a new forage sorghum variety, Tal, at two maturity stages. *Animal feed science and technology*, 131, 121-133.

- CARPITA, N. C. & MCCANN, M. C. 2008. Maize and sorghum: genetic resources for bioenergy grasses. *Trends in plant science*, 13, 415-420.
- CASADY, A. 1967. Effect of a single height gene (Dw3) of *Sorghum vulgare* Pers. on certain culm and leaf blade characteristics. *Crop Science*, 7, 595-598.
- CHANTEREAU, J., CRUZ, J.-F., RATNADASS, A. & TROUCHE, G. 2014. *Le sorgho*, Editions Quae.
- CHANTEREAU, J., DEU, M., GLASZMANN, J.-C., DEGREMONT, I., GONZALEZ DE LÉON, D. & HAMON, P. 1994. RFLP diversity in cultivated sorghum in relation to racial differentiation and heterosis in hybrids.
- CHANTEREAU, J., TROUCHE, G., RAMI, J., DEU, M., BARRO, C. & GRIVET, L. 2001a. RFLP mapping of QTLs for photoperiod response in tropical sorghum. *Euphytica*, 120, 183-194.
- CHILDS, K. L., MILLER, F. R., CORDONNIER-PRATT, M.-M., PRATT, L. H., MORGAN, P. W. & MULLET, J. E. 1997a. The sorghum photoperiod sensitivity gene, Ma3, encodes a phytochrome B. *Plant Physiology*, 113, 611-619.
- CLERGET, B., RATTUNDE, H. F. W., DAGNOKO, S. & CHANTEREAU, J. 2007. An easy way to assess photoperiod sensitivity in sorghum: Relationships of the vegetative-phase duration and photoperiod sensitivity. *SAT eJournal* [Online], 3.
- COCHEME, J. & FRANQUIN, P. 1967. Technical report on a study of the agroclimatology of the semiarid area south of the Sahara in West Africa= Rapport technique sur une etude d'agroclimatologie d l'afrique seche au sud du sahara en afrique occidentale.
- COVINGTON, M. F., PANDA, S., LIU, X. L., STRAYER, C. A., WAGNER, D. R. & KAY, S. A. 2001. ELF3 Modulates Resetting of the Circadian Clock in Arabidopsis. *The Plant Cell*, 13, 1305-1316.
- CRASTA, O., XU, W., ROSENOW, D., MULLET, J. & NGUYEN, H. 1999. Mapping of post-flowering drought resistance traits in grain sorghum: association between QTLs influencing premature senescence and maturity. *Molecular and General Genetics MGG*, 262, 579-588.
- CRAUFURD, P. Q., MAHALAKSHMI, V., BINDER, F. R., MUKURU, S. Z., CHANTEREAU, J., OMANGA, P. A., QI, A., ROBERTS, E. H., ELLIS, R. H., SUMMERFIELD, R. J. & HAMMER, G. L. 1999. Adaptation of sorghum: characterisation of genotypic flowering responses to temperature and photoperiod. *TAG Theoretical and Applied Genetics*, 99, 900-911.
- CRAUFURD, P. Q. & WHEELER, T. R. 2009. Climate change and the flowering time of annual crops. *J. Exp. Bot.*, 60, 2529-2539.
- CROSSA, J. 2012. I. 4 Screening Experimental Designs for Quantitative Trait Loci, Association Mapping, Genotype-by Environment Interaction, and Other Investigations. *Frontiers in Physiology*, 3, 156.
- CUEVAS, H. E., ZHOU, C., TANG, H., KHADKE, P. P., DAS, S., LIN, Y.-R., GE, Z., CLEMENTE, T., UPADHYAYA, H. D. & HASH, C. T. 2016. The evolution of photoperiod-insensitive flowering in sorghum, a genomic model for panicoid grasses. *Molecular biology and evolution*, 33, 2417-2428.
- CURTIS, D. L. 1968. The relation between yield and date of heading of Nigerian sorghums. *Experimental Agriculture*, 4, 93-101.
- DE WET, J. & HARLAN, J. 1971. The origin and domestication of *Sorghum bicolor*. *Economic Botany*, 25, 128-135.
- DELL'ACQUA, M., GATTI, D. M., PEA, G., CATTONARO, F., COPPENS, F., MAGRIS, G., HLAING, A. L., AUNG, H. H., NELISSEN, H. & BAUTE, J. 2015. Genetic



- properties of the MAGIC maize population: a new platform for high definition QTL mapping in *Zea mays*. *Genome biology*, 16, 167.
- DINGKUHN, M., SINGH, B., CLERGET, B., CHANTEREAU, J. & SULTAN, B. 2006. Past, present and future criteria to breed crops for water-limited environments in West Africa. *Agricultural Water Management*, 80, 241-261.
- DOGGETT, H. 1965. The development of cultivated sorghums. *Crop plant evolution*. Cambridge Univ. Press, London.
- DOGGETT, H. 1988. *Sorghum*, Longman Scientific & Technical.
- EJETA, G. & KNOLL, J. E. 2007. Marker-Assisted Selection in Sorghum. In: VARSHNEY, R. K. & TUBEROSA, R. (eds.) *Genomics-Assisted Crop Improvement: Vol 2: Genomics Applications in Crops*. Dordrecht: Springer Netherlands.
- EL MANNAI, Y., SHEHZAD, T. & OKUNO, K. 2011. Variation in flowering time in sorghum core collection and mapping of QTLs controlling flowering time by association analysis. *Genetic Resources and Crop Evolution*, 58, 983-989.
- ELSHIRE, R. J., GLAUBITZ, J. C., SUN, Q., POLAND, J. A., KAWAMOTO, K., BUCKLER, E. S. & MITCHELL, S. E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one*, 6, e19379.
- EMMAMBUX, M., STADING, M. & TAYLOR, J. 2004. Sorghum kafirin film property modification with hydrolysable and condensed tannins. *Journal of Cereal Science*, 40, 127-135.
- ENCISO, J., JIFON, J., RIBERA, L., ZAPATA, S. & GANJEGUNTE, G. 2015. Yield, water use efficiency and economic analysis of energy sorghum in South Texas. *Biomass and Bioenergy*, 81, 339-344.
- FAOSTAT 2014. United Nations Food and Agriculture Organization Statistical Database. Available Online at: <http://faostat.fao.org/>.
- FARRÉ, I. & FACI, J. M. 2006. Comparative response of maize (*Zea mays* L.) and sorghum (*Sorghum bicolor* L. Moench) to deficit irrigation in a Mediterranean environment. *Agricultural water management*, 83, 135-143.
- FAURE, S., TURNER, A., GRUSZKA, D., CHRISTODOULOU, V., J DAVIS, S., KORFF, M. & A LAURIE, D. 2012. Mutation at the circadian clock gene *EARLY MATURITY 8* adapts domesticated barley.
- FEDERER, W. T. 1961. Augmented designs with one-way elimination of heterogeneity. *Biometrics*, 17, 447-473.
- FELTUS, F., HART, G., SCHERTZ, K., CASA, A., KRESOVICH, S., ABRAHAM, S., KLEIN, P., BROWN, P. & PATERSON, A. 2006. Alignment of genetic maps and QTLs between inter-and intra-specific sorghum populations. *Theoretical and Applied Genetics*, 112, 1295.
- FERNANDEZ, M. G. S., BECRAFT, P. W., YIN, Y. & LÜBBERSTEDT, T. 2009. From dwarves to giants? Plant height manipulation for biomass yield. *Trends in plant science*, 14, 454-461.
- FOLLIARD, A., TRAORE, P. C. S., VAKSMANN, M. & KOURESSY, M. 2004. Modeling of sorghum response to photoperiod: a threshold-hyperbolic approach. *Field Crops Research*, 89, 59-70.
- FU, C., YANG, X. O., CHEN, X., CHEN, W., MA, Y., HU, J. & LI, S. 2009. OsEF3, a homologous gene of Arabidopsis ELF3, has pleiotropic effects in rice. *Plant Biology*, 11, 751-757.
- GARNER, W. & ALLARD, H. 1923. Further studies in photoperiodism: the response of the plant to relative length of day and night, US Government Printing Office.

- GAROFALO, P. & RINALDI, M. 2013. Water-use efficiency of irrigated biomass sorghum in a Mediterranean environment. *Spanish Journal of Agricultural Research*, 11, 1153-1169.
- GIRAUD, H., LEHERMEIER, C., BAUER, E., FALQUE, M., SEGURA, V., BAULAND, C., CAMISAN, C., CAMPO, L., MEYER, N. & RANC, N. 2014. Linkage disequilibrium with linkage analysis of multiline crosses reveals different multiallelic QTL for hybrid performance in the flint and dent heterotic groups of maize. *Genetics*, 198, 1717-1734.
- GLAMOCLIJA, D., JANKOVIC, S., RAKIC, S., MALETIC, R., IKANOVIC, J. & LAKIC, Z. 2011. Effects of nitrogen and harvesting time on chemical composition of biomass of Sudan grass, fodder sorghum, and their hybrid. *Turkish Journal of Agriculture and Forestry*, 35, 127-138.
- GLAUBITZ, J. C., CASSTEVENS, T. M., LU, F., HARRIMAN, J., ELSHIRE, R. J., SUN, Q. & BUCKLER, E. S. 2014. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PloS one*, 9, e90346.
- GRASSI, G., DULOBIDAS, D., CAPACCIOLI, S., GRASSI, A. 2008. Techno Economic Feasibility of Integrated Biorefineries Based on Sweet-sorghum Production. 2498–2500.
- GRENIER, C., BRAMEL-COX, P. J., NOIROT, M., RAO, K. P. & HAMON, P. 2000. Assessment of genetic diversity in three subsets constituted from the ICRISAT sorghum collection using random vs non-random sampling procedures A. Using morpho-agronomical and passport data. *Theoretical and Applied Genetics*, 101, 190-196.
- GUINDO, D. 2016. Analyse génétique de la qualité du grain de sorgho pour la fabrication du tô dans le cadre d'un schéma de sélection récurrente assistée par marqueurs. Montpellier SupAgro. *Montpellier SupAgro.*, 127.
- HARLAN, J. & DE WET, J. 1972. A simplified classification of cultivated sorghum. *Crop science*, 12, 172-176.
- HARRIS, K., SUBUDHI, P., BORRELL, A., JORDAN, D., ROSENOW, D., NGUYEN, H., KLEIN, P., KLEIN, R. & MULLET, J. 2006. Sorghum stay-green QTL individually reduce post-flowering drought-induced leaf senescence. *Journal of experimental botany*, 58, 327-338.
- HART, G., SCHERTZ, K., PENG, Y. & SYED, N. 2001. Genetic mapping of Sorghum bicolor (L.) Moench QTLs that control variation in tillering and other morphological characters. *TAG Theoretical and Applied Genetics*, 103, 1232-1242.
- HASH, C., RAJ, A. B., LINDUP, S., SHARMA, A., BENIWAL, C., FOLKERTSMA, R., MAHALAKSHMI, V., ZERBINI, E. & BLÜMMEL, M. 2003. Opportunities for marker-assisted selection (MAS) to improve the feed quality of crop residues in pearl millet and sorghum. *Field Crops Research*, 84, 79-88.
- HENDRIKS, A. & ZEEMAN, G. 2009. Pretreatments to enhance the digestibility of lignocellulosic biomass. *Bioresource technology*, 100, 10-18.
- HICKS, K. A., ALBERTSON, T. M. & WAGNER, D. R. 2001. EARLY FLOWERING3 Encodes a Novel Protein That Regulates Circadian Clock Function and Flowering in Arabidopsis. *The Plant Cell*, 13, 1281-1292.
- HICKS, K. A., MILLAR, A. J., CARRÉ, I. A., SOMERS, D. E., STRAUME, M., MEEKS-WAGNER, D. R. & KAY, S. A. 1996. Conditional Circadian Dysfunction of the *Arabidopsis* early-flowering 3 Mutant. *Science*, 274, 790-792.
- HIGGINS, R., THURBER, C., ASSARANURAK, I. & BROWN, P. 2014. Multiparental mapping of plant height and flowering time QTL in partially isogenic sorghum families. *G3: Genes| Genomes| Genetics*, 4, 1593-1602.

- HILLEY, J., TRUONG, S., OLSON, S., MORISHIGE, D. & MULLET, J. 2016. Identification of Dw1, a regulator of sorghum stem internode length. *PLoS one*, 11, e0151271.
- HIRANO, K., KAWAMURA, M., ARAKI-NAKAMURA, S., FUJIMOTO, H., OHMAE-SHINOHARA, K., YAMAGUCHI, M., FUJII, A., SASAKI, H., KASUGA, S. & SAZUKA, T. 2017. Sorghum DW1 positively regulates brassinosteroid signaling by inhibiting the nuclear localization of BRASSINOSTEROID INSENSITIVE 2. *Scientific Reports*, 7.
- HO, J., MCCOUCH, S. & SMITH, M. 2002. Improvement of hybrid yield by advanced backcross QTL analysis in elite maize. *TAG Theoretical and Applied Genetics*, 105, 440-448.
- HOLLAND, J. B. 2015. MAGIC maize: a new resource for plant genetics. *Genome biology*, 16, 163.
- HUANG, B. E., GEORGE, A. W., FORREST, K. L., KILIAN, A., HAYDEN, M. J., MORELL, M. K. & CAVANAGH, C. R. 2012. A multiparent advanced generation inter-cross population for genetic analysis in wheat. *Plant biotechnology journal*, 10, 826-839.
- HUANG, X., CÖSTER, H., GANAL, M. & RÖDER, M. 2003. Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics*, 106, 1379-1389.
- HUANG, X., PAULO, M.-J., BOER, M., EFFGEN, S., KEIZER, P., KOORNNEEF, M. & VAN EEUWIJK, F. A. 2011. Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant inbred line population. *Proceedings of the National Academy of Sciences*, 108, 4488-4493.
- IZAWA, T., OIKAWA, T., SUGIYAMA, N., TANISAKA, T., YANO, M. & SHIMAMOTO, K. 2002. Phytochrome mediates the external light signal to repress FT orthologs in photoperiodic flowering of rice. *Genes Dev*, 16.
- JANSEN, R. C. 1994. Controlling the type I and type II errors in mapping quantitative trait loci. *Genetics*, 138, 871-881.
- JIANG, W., ZHOU, H., BI, H., FROMM, M., YANG, B. & WEEKS, D. P. 2013. Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic acids research*, 41, e188-e188.
- JIMÉNEZ, L., LÓPEZ, F. & MARTÍNEZ, C. 1993. Paper from sorghum stalks. *Holzforchung-International Journal of the Biology, Chemistry, Physics and Technology of Wood*, 47, 529-533.
- JONES, C. A. & KINIRY, J. R. 1986. *CERES-Maize : A simulation model of maize growth and development*, Rexa A&M University Press, College Station, Texas.
- JORDAN, D., MACE, E., CRUICKSHANK, A., HUNT, C. & HENZELL, R. 2011. Exploring and exploiting genetic variation from unadapted sorghum germplasm in a breeding program. *Crop Science*, 51, 1444-1457.
- KAO, C.-H., ZENG, Z.-B. & TEASDALE, R. D. 1999. Multiple interval mapping for quantitative trait loci. *Genetics*, 152, 1203-1216.
- KARPER, R. 1932. A dominant mutation of frequent recurrence in sorghum. *The American Naturalist*, 66, 511-529.
- KASSAHUN, B., BINDER, F., HASH, C. & KURUVINASHETTI, M. 2010. Stay-green expression in early generation sorghum [*Sorghum bicolor* (L.) Moench] QTL introgression lines. *Euphytica*, 172, 351-362.
- KEBEDE, H., SUBUDHI, P., ROSENOW, D. & NGUYEN, H. 2001. Quantitative trait loci influencing drought tolerance in grain sorghum (*Sorghum bicolor* L. Moench). *TAG Theoretical and Applied Genetics*, 103, 266-276.

- KIM, J.-S. 2004. *Genomic analysis of sorghum by fluorescence in situ hybridization*. Texas A&M University.
- KIM, W.-Y., HICKS, K. A. & SOMERS, D. E. 2005. Independent Roles for *EARLY FLOWERING 3* and *ZEITLUPE* in the Control of Circadian Timing, Hypocotyl Length, and Flowering Time. *Plant Physiology*, 139, 1557-1569.
- KJELDAHL, J. 1883. A new method for the determination of nitrogen in organic matter. *Z. Anal. Chem*, 22, 366-382.
- KLEIN, R. R., MULLET, J. E., JORDAN, D. R., MILLER, F. R., ROONEY, W. L., MENZ, M. A., FRANKS, C. D. & KLEIN, P. E. 2008. The effect of tropical sorghum conversion and inbred development on genome diversity as revealed by high-resolution genotyping. *Crop science*, 48, S-12-S-26.
- KOURESSY, M., DINGKUHN, M., VAKSMANN, M., CLEMENT-VIDAL, A. & CHANTEREAU, J. 2008a. Potential contribution of dwarf and leaf longevity traits to yield improvement in photoperiod sensitive sorghum. *European Journal of Agronomy*, 28, 195-209.
- KOURESSY, M., DINGKUHN, M., VAKSMANN, M. & HEINEMANN, A. B. 2008b. Adaptation to diverse semi-arid environments of sorghum genotypes having different plant type and sensitivity to photoperiod. *Agricultural and Forest Meteorology*, 148, 357-371.
- KOURESSY, M., TRAORÉ, S., VAKSMANN, M., GRUM, M., MAIKANO, I., SOUMARÉ, M., TRAORÉ, P. S., BAZILE, D., DINGKUHN, M. & SIDIBÉ, A. 2008. Adaptation des sorghos du Mali à la variabilité climatique. *Cahiers Agricultures*, 17, 95-100.
- KOURESSY MAMOUTOU, V. M., TRAORÉ SEYDOU B, 2006. Le photopériodisme des sorghos africains, une réponse à la variabilité pluviométrique. in: AMMA CIRAD (Ed.), Monitoring and forecasting the African monsoon impacts on agriculture and vegetation (AMMA) Dakar, Sénégal. *Cirad - Agritrop*, 557-559.
- KOVER, P. X., VALDAR, W., TRAKALO, J., SCARCELLI, N., EHRENREICH, I. M., PURUGGANAN, M. D., DURRANT, C. & MOTT, R. 2009. A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS genetics*, 5, e1000551.
- LAFARGE, T. A. & HAMMER, G. L. 2002. Predicting plant leaf area production:: shoot assimilate accumulation and partitioning, and leaf area ratio, are stable for a wide range of sorghum population densities. *Field Crops Research*, 77, 137-151.
- LANDER, E. S. & BOTSTEIN, D. 1989. Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps. *Genetics*, 121, 185-199.
- LANDER, E. S., GREEN, P., ABRAHAMSON, J., BARLOW, A., DALY, M. J., LINCOLN, S. E. & NEWBURG, L. 1987. MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics*, 1, 174-181.
- LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Meth*, 9, 357-359.
- LAURIE, D. A., PRATCHETT, N., SNAPE, J. W. & BEZANT, J. H. 1995. RFLP mapping of five major genes and eight quantitative trait loci controlling flowering time in a winter × spring barley (*Hordeum vulgare* L.) cross. *Genome*, 38, 575-585.
- LEUNG, H., RAGHAVAN, C., ZHOU, B., OLIVA, R., CHOI, I. R., LACORTE, V., JUBAY, M. L., CRUZ, C. V., GREGORIO, G. & SINGH, R. K. 2015. Allele mining and enhanced genetic recombination for rice breeding. *Rice*, 8, 34.

- LI, X., LI, X., FRIDMAN, E., TESSO, T. T. & YU, J. 2015. Dissecting repulsion linkage in the dwarfing gene Dw3 region for sorghum plant height provides insights into heterosis. *Proceedings of the National Academy of Sciences*, 112, 11823-11828.
- LI, X.-F., LIU, Z.-X., LU, D.-B., LIU, Y.-Z., MAO, X.-X., LI, Z.-X. & LI, H.-J. 2013. Development and evaluation of multi-genotype varieties of rice derived from MAGIC lines. *Euphytica*, 192, 77-86.
- LIN, Y.-R., SCHERTZ, K. F. & PATERSON, A. H. 1995. Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population. *Genetics*, 141, 391-411.
- LIU, G., CAMPBELL, B. C. & GODWIN, I. D. 2014. Sorghum genetic transformation by particle bombardment. *Cereal Genomics: Methods and Protocols*, 219-234.
- LIU, G. & GODWIN, I. D. 2012. Highly efficient sorghum transformation. *Plant cell reports*, 31, 999-1007.
- LIU, X. L., COVINGTON, M. F., FANKHAUSER, C., CHORY, J. & WAGNER, D. R. 2001. ELF3 Encodes a Circadian Clock-Regulated Nuclear Protein That Functions in an Arabidopsis PHYB Signal Transduction Pathway. *The Plant Cell*, 13, 1293-1304.
- MACE, E., HUNT, C. & JORDAN, D. 2013. Supermodels: sorghum and maize provide mutual insight into the genetics of flowering time. *Theoretical and Applied Genetics*, 126, 1377-1395.
- MACE, E. & JORDAN, D. 2010. Location of major effect genes in sorghum (*Sorghum bicolor* (L.) Moench). *Theoretical and Applied Genetics*, 121, 1339-1356.
- MACE, E. & JORDAN, D. 2011. Integrating sorghum whole genome sequence information with a compendium of sorghum QTL studies reveals uneven distribution of QTL and of gene-rich regions with significant implications for crop improvement. *Theoretical and applied genetics*, 123, 169-191.
- MACE, E. S., RAMI, J.-F., BOUCHET, S., KLEIN, P. E., KLEIN, R. R., KILIAN, A., WENZL, P., XIA, L., HALLORAN, K. & JORDAN, D. R. 2009. A consensus genetic map of sorghum that integrates multiple component maps and high-throughput Diversity Array Technology (DArT) markers. *BMC Plant Biology*, 9, 13.
- MACKAY, I. J., BANSEPT-BASLER, P., BARBER, T., BENTLEY, A. R., COCKRAM, J., GOSMAN, N., GREENLAND, A. J., HORSNELL, R., HOWELLS, R. & O'SULLIVAN, D. M. 2014. An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: creation, properties, and validation. *G3: Genes, Genomes, Genetics*, 4, 1603-1610.
- MADHUSUDHANA, R. & PATIL, J. 2013. A major QTL for plant height is linked with bloom locus in sorghum [*Sorghum bicolor* (L.) Moench]. *Euphytica*, 191, 259-268.
- MAJOR, D., ROOD, S. & MILLER, F. 1990. Temperature and photoperiod effects mediated by the sorghum maturity genes. *Crop Science*, 30, 305-310.
- MAJOR, D. J. 1980. Photoperiod response characteristics controlling flowering of nine crop species. *Canadian Journal of Plant Science*, 60, 777-784.
- MAJOR, D. J., JOHNSON, D. R., TANNER, J. W. & ANDERSON, I. C. 1975. Effects of Daylength and Temperature on Soybean Development. *Crop Sci.*, 15, 174-179.
- MANICHAIKUL, A., MOON, J. Y., SEN, Ś., YANDELL, B. S. & BROMAN, K. W. 2009. A Model Selection Approach for the Identification of Quantitative Trait Loci in Experimental Crosses, Allowing Epistasis. *Genetics*, 181, 1077.
- MASTRORILLI, M., KATERJI, N., RANA, G. & STEDUTO, P. 1995. Sweet sorghum in Mediterranean climate: radiation use and biomass water use efficiencies. *Industrial Crops and Products*, 3, 253-260.

- MATHUR, S., UMAKANTH, A., TONAPI, V., SHARMA, R. & SHARMA, M. K. 2017. Sweet sorghum as biofuel feedstock: recent advances and available resources. *Biotechnology for Biofuels*, 10, 146.
- MATSUBARA, K., KONO, I., HORI, K., NONOUE, Y., ONO, N., SHOMURA, A., MIZUBAYASHI, T., YAMAMOTO, S., YAMANOUCHI, U., SHIRASAWA, K., NISHIO, T. & YANO, M. 2008. Novel QTLs for photoperiodic flowering revealed by using reciprocal backcross inbred lines from crosses between japonica rice cultivars. *Theoretical and Applied Genetics*, 117, 935-945.
- MATSUBARA, K., OGISO-TANAKA, E., HORI, K., EBANA, K., ANDO, T. & YANO, M. 2012. Natural Variation in Hd17, a Homolog of Arabidopsis ELF3 That is Involved in Rice Photoperiodic Flowering. *Plant and Cell Physiology*, 53, 709-716.
- MCCORMICK, R. F., TRUONG, S. K., SREEDASYAM, A., JENKINS, J., SHU, S., SIMS, D., KENNEDY, M., AMIREBRAHIMI, M., WEERS, B. D., MCKINLEY, B., MATTISON, A., MORISHIGE, D. T., GRIMWOOD, J., SCHMUTZ, J. & MULLET, J. E. 2017. The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J.*
- MCWATTERS, H. G., BASTOW, R. M., HALL, A. & MILLAR, A. J. 2000. The ELF3zeitnehmer regulates light signalling to the circadian clock. *Nature*, 408, 716-720.
- MENG, L., ZHAO, X., PONCE, K., YE, G. & LEUNG, H. 2016. QTL mapping for agronomic traits using multi-parent advanced generation inter-cross (MAGIC) populations derived from diverse elite indica rice lines. *Field Crops Research*, 189, 19-42.
- MESSINA, C. D., BOOTE, K. J., LÖFFLER, C., JONES, J. W. & VALLEJOS, C. E. 2006. Model-assisted genetic improvement of crops. In: WALLACH, D., MAKOWSKI, D. & JONES, J. W. (eds.) *working with Dynamic crop Models*. Amsterdam: Elsevier.
- MING, R., LIU, S.-C., LIN, Y.-R., DA SILVA, J., WILSON, W., BRAGA, D., VAN DEYNZE, A., WENSLAFF, T. F., WU, K. K., MOORE, P. H., BURNQUIST, W., SORRELLS, M. E., IRVINE, J. E. & PATERSON, A. H. 1998. Detailed Alignment of Saccharum and Sorghum Chromosomes: Comparative Organization of Closely Related Diploid and Polyploid Genomes. *Genetics*, 150, 1663-1682.
- MOHAMED, A., ALI, R., ELHASSAN, O., SULIMAN, E., MUGOYA, C., MASIGA, C. W., ELHUSIEN, A. & HASH, C. T. 2014. First products of DNA marker-assisted selection in sorghum released for cultivation by farmers in sub-saharan Africa. *Journal of Plant Science and Molecular Breeding*, 3.
- MONCADA, P., MARTINEZ, C., BORRERO, J., CHATEL, M., GAUCH JR, H., GUIMARAES, E., TOHME, J. & MCCOUCH, S. 2001. Quantitative trait loci for yield and yield components in an *Oryza sativa* × *Oryza rufipogon* BC2F2 population evaluated in an upland environment. *TAG Theoretical and Applied Genetics*, 102, 41-52.
- MORGAN, P. W. & FINLAYSON, S. A. 2000. Physiology and genetics of maturity and height. In: SMITH, C. W. & FREDERIKSEN, R. A. (eds.) *Sorghum: Origin, History, Technology, and Production*. New York: Wiley Series in Crop Science.
- MORRIS, G. P., RAMU, P., DESHPANDE, S. P., HASH, C. T., SHAH, T., UPADHYAYA, H. D., RIERA-LIZARAZU, O., BROWN, P. J., ACHARYA, C. B., MITCHELL, S. E., HARRIMAN, J., GLAUBITZ, J. C., BUCKLER, E. S. & KRESOVICH, S. 2013. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences*, 110, 453-458.
- MOURADOV, A., CREMER, F. & COUPLAND, G. 2002. Control of Flowering Time: Interacting Pathways as a Basis for Diversity. *Plant Cell*, 14, S111-130.

- MULLET, J., MORISHIGE, D., MCCORMICK, R., TRUONG, S., HILLEY, J., MCKINLEY, B., ANDERSON, R., OLSON, S. N. & ROONEY, W. 2014. Energy Sorghum—a genetic model for the design of C4 grass bioenergy crops. *Journal of experimental botany*, 65, 3479-3489.
- MULLET, J. E., ROONEY, W. L., KLEIN, P. E., MORISHIGE, D., MURPHY, R. & BRADY, J. A. 2016. *Discovery and utilization of sorghum genes (Ma5, Ma6)*. 13/633,771.
- MULTANI, D. S., BRIGGS, S. P., CHAMBERLIN, M. A., BLAKESLEE, J. J., MURPHY, A. S. & JOHAL, G. S. 2003. Loss of an MDR Transporter in Compact Stalks of Maize br2 and Sorghum dw3 Mutants. *Science*, 302, 81-84.
- MURPHY, R. L., KLEIN, R. R., MORISHIGE, D. T., BRADY, J. A., ROONEY, W. L., MILLER, F. R., DUGAS, D. V., KLEIN, P. E. & MULLET, J. E. 2011a. Coincident light and clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum. *Proceedings of the National Academy of Sciences*.
- MURPHY, R. L., MORISHIGE, D. T., BRADY, J. A., ROONEY, W. L., YANG, S., KLEIN, P. E. & MULLET, J. E. 2014. Ghd7 (Ma6) Represses Sorghum Flowering in Long Days: Ghd7 Alleles Enhance Biomass Accumulation and Grain Production. *Plant Gen.*, 7, 1-10.
- MURRAY, S. C., ROONEY, W. L., HAMBLIN, M. T., MITCHELL, S. E. & KRESOVICH, S. 2009. Sweet sorghum genetic diversity and association mapping for brix and height. *The Plant Genome*, 2, 48-62.
- MURRAY, S. C., ROONEY, W. L., MITCHELL, S. E., SHARMA, A., KLEIN, P. E., MULLET, J. E. & KRESOVICH, S. 2008. Genetic improvement of sorghum as a biofuel feedstock: II. QTL for stem and leaf structural carbohydrates. *Crop science*, 48, 2180-2193.
- OLLITRAULT, P. 1987. Évaluation génétique des Sorghos cultivés (*Sorghum bicolor* L. Moench) par l'analyse conjointe des diversités enzymatique et morphophysiologique: relations avec les sorghos sauvages. Paris 11.
- OLSON, S. N., RITTER, K., ROONEY, W., KEMANIAN, A., MCCARL, B. A., ZHANG, Y., HALL, S., PACKER, D. & MULLET, J. 2012. High biomass yield energy sorghum: developing a genetic model for C4 grass bioenergy crops. *Biofuels, Bioproducts and Biorefining*, 6, 640-655.
- PARH, D. K. 2005. DNA-based markers for ergot resistance in sorghum. PhD thesis. University of Queensland, School of Land and Food Sciences.
- PASCUAL, L., DESPLAT, N., HUANG, B. E., DESGROUX, A., BRUGUIER, L., BOUCHET, J. P., LE, Q. H., CHAUCHARD, B., VERSCHAVE, P. & CAUSSE, M. 2015. Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. *Plant biotechnology journal*, 13, 565-577.
- PATERSON, A., BOWERS, J. & CHAPMAN, B. 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 9903-9908.
- PATERSON, A. H., BOWERS, J. E., BRUGGMANN, R., DUBCHAK, I., GRIMWOOD, J., GUNDLACH, H., HABERER, G., HELLSTEN, U., MITROS, T. & POLIAKOV, A. 2009. The Sorghum bicolor genome and the diversification of grasses. *Nature*, 457, 551-556.
- PEREIRA, M. & LEE, M. 1995. Identification of genomic regions affecting plant height in sorghum and maize. *TAG Theoretical and Applied Genetics*, 90, 380-388.

- PILLEN, K., ZACHARIAS, A. & LÉON, J. 2004. Comparative AB-QTL analysis in barley using a single exotic donor of *Hordeum vulgare* ssp. *spontaneum*. *Theoretical and applied genetics*, 108, 1591-1601.
- PIRAUX, M., BULDGEN, A., STEYAERT, P. & DIENG, A. 1997. Intensification agricole en région sahélo-soudanienne. 1. Itinéraires techniques dans un contexte à risques. *Biotechnologie, Agronomie, Société et Environnement*, 1, 196-208.
- QUINBY, J. 1966. Fourth maturity gene locus in sorghum. *Crop Science*, 6, 516-518.
- QUINBY, J. 1967. The maturity genes of sorghum. *Advances in agronomy*, 19, 267-305.
- QUINBY, J., HESKETH, J. & VOIGT, R. 1973. Influence of temperature and photoperiod on floral initiation and leaf number in sorghum. *Crop Science*, 13, 243-246.
- QUINBY, J. & KARPER, R. 1945. Inheritance of three genes that influence time of floral initiation and maturity date in milo. *Journal of the American Society of Agronomy*.
- QUINBY, J. & MARTIN, J. 1954. Sorghum improvement. *Adv. Agron*, 6, 305-359.
- QUINBY, J. R. 1974. *Sorghum improvement and the genetics of growth*, College Station, M Univ. Press.
- R DEVELOPMENT CORE TEAM 2016. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
- 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). *TAG Theoretical and Applied Genetics*, 97, 605-616.
- REED, J. W., NAGPAL, P., BASTOW, R. M., SOLOMON, K. S., DOWSON-DAY, M. J., ELUMALAI, R. P. & MILLAR, A. J. 2000. Independent Action of ELF3 and phyB to Control Hypocotyl Elongation and Flowering Time. *Plant Physiology*, 122, 1149.
- RISTERUCCI, A. M., GRIVET, L., N'GORAN, J. A. K., PIERETTI, I., FLAMENT, M. H. & LANAUD, C. 2000. A high-density linkage map of *Theobroma cacao* L. *Theoretical and Applied Genetics*, 101, 948-955.
- RITCHIE, J. T. & ALAGARSWAMY, G. 1989a. Genetic Coefficients for CERES Models. In: VIRMANI, S. M. A. T., H L S AND ALAGARSWAMY (ed.) *Reserch bulletin n°12*. International Crops Research Institute for the Semi-Arid Tropics.
- RITCHIE, J. T. & ALAGARSWAMY, G. Simulation of Sorghum and Pearl Millet Phenology Modeling the Growth and Development of Sorghum and Pearl Millet, 1989b. Res.Bull. n°12, ICRISAT, 24-29.
- RITCHIE, J. T. & ALAGARSWAMY, G. Phasic Development in CERES-Sorghum Model. Hodges, T. (Ed.), *Predicting Crop Phenology*, CRC Press, Boca Raton, 1991. 143-152.
- RITTER, K. B., JORDAN, D. R., CHAPMAN, S. C., GODWIN, I. D., MACE, E. S. & MCINTYRE, C. L. 2008. Identification of QTL for sugar-related traits in a sweet× grain sorghum (*Sorghum bicolor* L. Moench) recombinant inbred population. *Molecular Breeding*, 22, 367-384.
- ROCATELI, A., RAPER, R., BALKCOM, K., ARRIAGA, F. & BRANSBY, D. 2012. Biomass sorghum production and components under different irrigation/tillage systems for the southeastern US. *Industrial Crops and Products*, 36, 589-598.
- ROONEY, W. L. & AYDIN, S. 1999. Genetic control of a photoperiod-sensitive response in *Sorghum bicolor* (L.) moench. *Crop Sci*, 39, 397-400.
- ROONEY, W. L., BLUMENTHAL, J., BEAN, B. & MULLET, J. E. 2007. Designing sorghum as a dedicated bioenergy feedstock. *Biofuels, Bioproducts and Biorefining*, 1, 147-157.



- RUSSO, V. M. & FISH, W. W. 2012. Biomass, extracted liquid yields, sugar content or seed yields of biofuel feedstocks as affected by fertilizer. *Industrial Crops and Products*, 36, 555-559.
- SAITO, H., OGISO-TANAKA, E., OKUMOTO, Y., YOSHITAKE, Y., IZUMI, H., YOKOO, T., MATSUBARA, K., HORI, K., YANO, M., INOUE, H. & TANISAKA, T. 2012. Ef7 Encodes an ELF3-like Protein and Promotes Rice Flowering by Negatively Regulating the Floral Repressor Gene Ghd7 under Both Short- and Long-Day Conditions. *Plant and Cell Physiology*, 53, 717-728.
- SANNEMANN, W. 2015. *Multi-parent advanced generation inter-cross in barley: high-resolution quantitative trait locus mapping for flowering time as a proof of concept*.
- SANON, M., HOOGENBOOM, G., TRAORÉ, S. B., SARR, B., GARCIA, A. G. Y., SOMÉ, L. & RONCOLI, C. 2014. Photoperiod sensitivity of local millet and sorghum varieties in West Africa. *NJAS - Wageningen Journal of Life Sciences*, 68, 29-39.
- SAPIN, P. 1983. Le sorgho et son amélioration. Synthèse Haute Volta 1961-1981.
- SARATH, G., MITCHELL, R. B., SATTLER, S. E., FUNNELL, D., PEDERSEN, J. F., GRAYBOSCH, R. A. & VOGEL, K. P. 2008. Opportunities and roadblocks in utilizing forages and small grains for liquid fuels. *Journal of industrial microbiology & biotechnology*, 35, 343-354.
- SATTLER, S. E., SABALLOS, A., XIN, Z., FUNNELL-HARRIS, D. L., VERMERRIS, W. & PEDERSEN, J. F. 2014. Characterization of novel Sorghum brown midrib mutants from an EMS-mutagenized population. *G3 (Bethesda)*, 4, 2115-24.
- SEPTININGSIH, E., PRASETIYONO, J., LUBIS, E., TAI, T., TJUBARYAT, T., MOELJOPAWIRO, S. & MCCOUCH, S. 2003. Identification of quantitative trait loci for yield and yield components in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. *Theoretical and applied genetics*, 107, 1419-1432.
- SHIRINGANI, A. L., FRISCH, M. & FRIEDT, W. 2010. Genetic mapping of QTLs for sugar-related traits in a RIL population of Sorghum bicolor L. Moench. *Theoretical and Applied Genetics*, 121, 323-336.
- SIMPSON, G. G., GENDALL, A. R. & DEAN, C. 1999. When to switch to flowering. *Annual review of cell and developmental biology*, 15, 519-550.
- SNOWDON, J. 1936. The cultivated races of sorghum. *The cultivated races of sorghum*.
- SRINIVAS, G., SATISH, K., MADHUSUDHANA, R., REDDY, R. N., MOHAN, S. M. & SEETHARAMA, N. 2009. Identification of quantitative trait loci for agronomically important traits and their association with genic-microsatellite markers in sorghum. *Theoretical and Applied Genetics*, 118, 1439-1454.
- STEPHENS, J., MILLER, F. & ROSENOW, D. 1967. Conversion of alien sorghums to early combine genotypes. *Crop Science*, 7, 396-396..
- STOOP, W. A., PATTANAYAK, C. M., MATLON, P. J. & ROOT, W. R. 1982. *A strategy to raise the productivity of subsistence farming systems in the West African semi-arid tropics*, International crops research institute for the semi-arid tropics (ICRISAT).
- SWAMINATHAN, M. S. 2006. An Evergreen Revolution. *Crop Sci*, 46, 2293-2303.
- TANKSLEY, S. D. & NELSON, J. C. 1996a. Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theor Appl Genet*, 92, 191-203.
- TANKSLEY, S., GRANDILLO, S., FULTON, T., ZAMIR, D., ESHED, Y., PETIARD, V., LOPEZ, J. & BECK-BUNN, T. 1996b. Advanced backcross QTL analysis in a cross

- between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. *TAG Theoretical and Applied Genetics*, 92, 213-224.
- TANKSLEY, S. D., MEDINA-FILHO, H. & RICK, C. M. 1982. Use of naturally-occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. *Heredity*, 49, 11-25.
- TARUMOTO, I., YANASE, M., IWAHARA, Y., KUZUMI, Y., MORIKAWA, T. & KASUGA, S. 2003. Inheritance of a Thermo-sensitivity Gene Controlling Flower Initiation in Sorghum. *Breeding Science* 53, 353-357.
- THOMAS, B. & VINCE-PRUE, D. 1996. *Photoperiodism in plants*, San Diego, U.S.A, Academic Press, 428 pp.
- THOMSON, M., TAI, T., MCCLUNG, A., LAI, X., HINGA, M., LOBOS, K., XU, Y., MARTINEZ, C. & MCCOUCH, S. 2003. Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. *Theoretical and applied genetics*, 107, 479-493.
- THURBER, C. S., MA, J. M., HIGGINS, R. H. & BROWN, P. J. 2013. Retrospective genomic analysis of sorghum adaptation to temperate-zone grain production. *Genome biology*, 14, R68.
- TRAORE, S. B., REYNIERS, F.-N., VAKSMANN, M., BATHER, K., SIDIBE, A., YOROTE, A., YATTARA, K. & KOURESSY, M. 2001. Adaptation à la sécheresse des écotypes locaux de sorghos du Mali. *Science et changements planétaires/Sécheresse*, 11, 227-37.
- TROUCHE, G. V., MICHEL; KOURESSY, MAMOUTOU; CHANTEREAU, JACQUES; BARRO, CLARISSE; RAMI,JEAN-FRANÇOIS 1998. Déterminisme génétique mendélien de la sensibilité à la photopériode des sorghos Ouest-africains *Cirad - Agritrop*, 225-246.
- UPADHYAYA, H. D., WANG, Y.-H., GOWDA, C. & SHARMA, S. 2013. Association mapping of maturity and plant height using SNP markers with the sorghum mini core collection. *Theoretical and applied genetics*, 126, 2003-2015.
- UPADHYAYA, H. D., WANG, Y.-H., SHARMA, S. & SINGH, S. 2012. Association mapping of height and maturity across five environments using the sorghum mini core collection. *Genome*, 55, 471-479.
- VAKSMANN, M., CHANTEREAU, J., BAHMANI, I., AG HAMADA, M., CHARTIER, M. & BONHOMME, R. 1998. Influence of night temperature on photoperiod response of a west african guinea sorghum landrace. In: RATNADASS A. (ÉD.), C. J. É., GIGOU J. (ÉD.). (ed.) *Amélioration du sorgho et de sa culture en Afrique de l'Ouest et du Centre. Atelier de Restitution du Programme Conjoint sur le Sorgho*. Bamako, Mali: CIRAD.
- VAKSMANN, M., TRAORÉ, S. & NIANGADO, O. 1996. Le photopériodisme des sorghos africains. *Agriculture et développement*, 9, 13-18.
- VAN DER WEIJDE, T., KAMEI, C. L. A., TORRES, A. F., VERMERRIS, W., DOLSTRA, O., VISSER, R. G. & TRINDADE, L. M. 2013. The potential of C4 grasses for cellulosic biofuel production. *Frontiers in Plant Science*, 4.
- VANDENBRINK, J. P., GOFF, V., JIN, H., KONG, W., PATERSON, A. H. & FELTUS, F. A. 2013. Identification of bioconversion quantitative trait loci in the interspecific cross *Sorghum bicolor* × *Sorghum propinquum*. *Theoretical and applied genetics*, 126, 2367-2380.
- VERMERRIS, W. 2011. Survey of genomics approaches to improve bioenergy traits in maize, sorghum and sugarcane. *Journal of Integrative Plant Biology*, 53, 105-119.

- VERMERRIS, W., SABALLOS, A., EJETA, G., MOSIER, N. S., LADISCH, M. R. & CARPITA, N. C. 2007. Molecular breeding to enhance ethanol production from corn and sorghum stover. *Crop Science*, 47, S-142-S-153.
- VO, L. T., GIRONES, J., BELOLI, C., CHUPIN, L., DI GIUSEPPE, E., VIDAL, A. C., SOUTIRAS, A., POT, D., BASTIANELLI, D. & BONNAL, L. 2017. Processing and properties of sorghum stem fragment-polyethylene composites. *Industrial Crops and Products*, 107, 386-398.
- VON KORFF, M., WANG, H., LEON, J. & PILLEN, K. 2005. AB-QTL analysis in spring barley. I. Detection of resistance genes against powdery mildew, leaf rust and scald introgressed from wild barley. *Theoretical and Applied Genetics*, 111, 583-590.
- VON KORFF, M., WANG, H., LÉON, J. & PILLEN, K. 2006. AB-QTL analysis in spring barley: II. Detection of favourable exotic alleles for agronomic traits introgressed from wild barley (*H. vulgare* ssp. *spontaneum*). *Theoretical and Applied Genetics*, 112, 1221-1231.
- WINDSCHEFFEL, J. A., VANDERLIP, R. & CASADY, A. 1973. Performance of 2-dwarf and 3-dwarf grain sorghum hybrids harvested at various moisture contents. *Crop Science*, 13, 215-219.
- WOLABU, T. W., ZHANG, F., NIU, L., KALVE, S., BHATNAGAR-MATHUR, P., MUSZYNSKI, M. G. & TADEGE, M. 2016. Three FLOWERING LOCUS T-like genes function as potential florigens and mediate photoperiod response in sorghum. *New Phytologist*, 210, 946-959.
- XAVIER, A., XU, S., MUIR, W. M. & RAINEY, K. M. 2015. NAM: association studies in multiple populations. *Bioinformatics*, 31, 3862-3864.
- XIAO, J., LI, J., GRANDILLO, S., AHN, S. N., YUAN, L., TANKSLEY, S. D. & MCCOUCH, S. R. 1998. Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. *Genetics*, 150, 899-909.
- XIN, Z., WANG, M. L., BARKLEY, N. A., BUROW, G., FRANKS, C., PEDERSON, G. & BURKE, J. 2008. Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population. *BMC Plant Biology*, 8, 103.
- XU, Y. & CROUCH, J. H. 2008. Marker-Assisted Selection in Plant Breeding: From Publications to Practice All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher. *Crop Science*, 48, 391-407.
- YAMAGUCHI, M., FUJIMOTO, H., HIRANO, K., ARAKI-NAKAMURA, S., OHMAE-SHINOHARA, K., FUJII, A., TSUNASHIMA, M., SONG, X. J., ITO, Y., NAGAE, R., WU, J., MIZUNO, H., YONEMARU, J.-I., MATSUMOTO, T., KITANO, H., MATSUOKA, M., KASUGA, S. & SAZUKA, T. 2016. Sorghum Dw1, an agronomically important gene for lodging resistance, encodes a novel protein involved in cell proliferation. *Scientific Reports*, 6, 28366.
- YANG, S., MURPHY, R. L., MORISHIGE, D. T., KLEIN, P. E., ROONEY, W. L. & E., M. J. 2014. Sorghum Phytochrome B Inhibits Flowering in Long Days by Activating Expression of SbPRR37 and SbGHD7, Repressors of SbEHD1, SbCN8 and SbCN12. *PLoS ONE*, 9, 1-12.

- YANG, S., WEERS, B. D., MORISHIGE, D. T. & MULLET, J. E. 2014b. CONSTANS is a photoperiod regulated activator of flowering in sorghum. *BMC Plant Biology*, 14, 1-15.
- YANG, Y., PENG, Q., CHEN, G., LI, X. & WU, C. 2013. OsELF3 Is Involved in Circadian Clock Regulation for Promoting Flowering under Long-Day Conditions in Rice. *Molecular Plant*, 6, 202-215.
- YAPI, A. M., KERGNA, A. O., DEBRAH, S. K., SIDIBE, A. & SANOGO, O. 2000. *Analysis of the economic impact of sorghum and millet research in Mali*, International Crops Research Institute for the Semi-Arid Tropics.
- YOSHIDA, R., FEKIH, R., FUJIWARA, S., ODA, A., MIYATA, K., TOMOZOE, Y., NAKAGAWA, M., NIINUMA, K., HAYASHI, K., EZURA, H., COUPLAND, G. & MIZOGUCHI, T. 2009. Possible role of EARLY FLOWERING 3 (ELF3) in clock-dependent floral regulation by SHORT VEGETATIVE PHASE (SVP) in *Arabidopsis thaliana*. *New Phytologist*, 182, 838-850.
- YU, J., HOLLAND, J. B., MCMULLEN, M. D. & BUCKLER, E. S. 2008. Genetic Design and Statistical Power of Nested Association Mapping in Maize. *Genetics*, 178, 539-551.
- YU, J., PRESSOIR, G., BRIGGS, W. H., VROH BI, I., YAMASAKI, M., DOEBLEY, J. F., MCMULLEN, M. D., GAUT, B. S., NIELSEN, D. M., HOLLAND, J. B., KRESOVICH, S. & BUCKLER, E. S. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet*, 38, 203-208.
- YUAN, Q., SAITO, H., OKUMOTO, Y., INOUE, H., NISHIDA, H., TSUKIYAMA, T., TERAISHI, M. & TANISAKA, T. 2009. Identification of a novel gene ef7 conferring an extremely long basic vegetative growth phase in rice. *Theoretical and Applied Genetics*, 119, 675-684.
- ZAKHRABEKOVA, S., GOUGH, S. P., BRAUMANN, I., MÜLLER, A. H., LUNDQVIST, J., AHMANN, K., DOCKTER, C., MATYSZCZAK, I., KUROWSKA, M., DRUKA, A., WAUGH, R., GRANER, A., STEIN, N., STEUERNAGEL, B., LUNDQVIST, U. & HANSSON, M. 2012. Induced mutations in circadian clock regulator Mat-a facilitated short-season adaptation and range extension in cultivated barley. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 4326-4331.
- ZENG, N. 2003. Drought in the Sahel. *Science*, 302, 999-1000.
- ZENG, Z.-B. 1993. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proceedings of the National Academy of Sciences*, 90, 10972-10976.
- ZHAO, J., HUANG, X., OUYANG, X., CHEN, W., DU, A., ZHU, L., WANG, S., DENG, X. W. & LI, S. 2012a. OsELF3-1, an Ortholog of *Arabidopsis* EARLY FLOWERING 3, Regulates Rice Circadian Rhythm and Photoperiodic Flowering. *PLOS ONE*, 7, e43705.
- ZHAO, Y. L., STEINBERGER, Y., SHI, M., HAN, L. P. & XIE, G. H. 2012b. Changes in stem composition and harvested produce of sweet sorghum during the period from maturity to a sequence of delayed harvest dates. *biomass and bioenergy*, 39, 261-273.
- ZOU, G., ZHAI, G., FENG, Q., YAN, S., WANG, A., ZHAO, Q., SHAO, J., ZHANG, Z., ZOU, J., HAN, B. & TAO, Y. 2012. Identification of QTLs for eight agronomically important traits using an ultra-high-density map based on SNPs generated from high-throughput sequencing in sorghum under contrasting photoperiods. *Journal of Experimental Botany*, 63, 5451-5462.

Annexe 1 : Caractéristiques des QTL détectés sur la population BC06 pour les variables phénologiques et agromorphologiques : durée du cycle (SFD), de longueur des tiges (LTIG) et ses composantes (nombre total d'entre-nœud allongés « NBETN » et longueur moyenne d'entre-nœuds « LgMoyETN »), du nombre de feuilles vertes à maturité (NBFV), des longueurs du pédoncule (LPED) et de la panicule (LPAN) pour les semis de juin, juillet (2014 et 5014) et de septembre (2015).

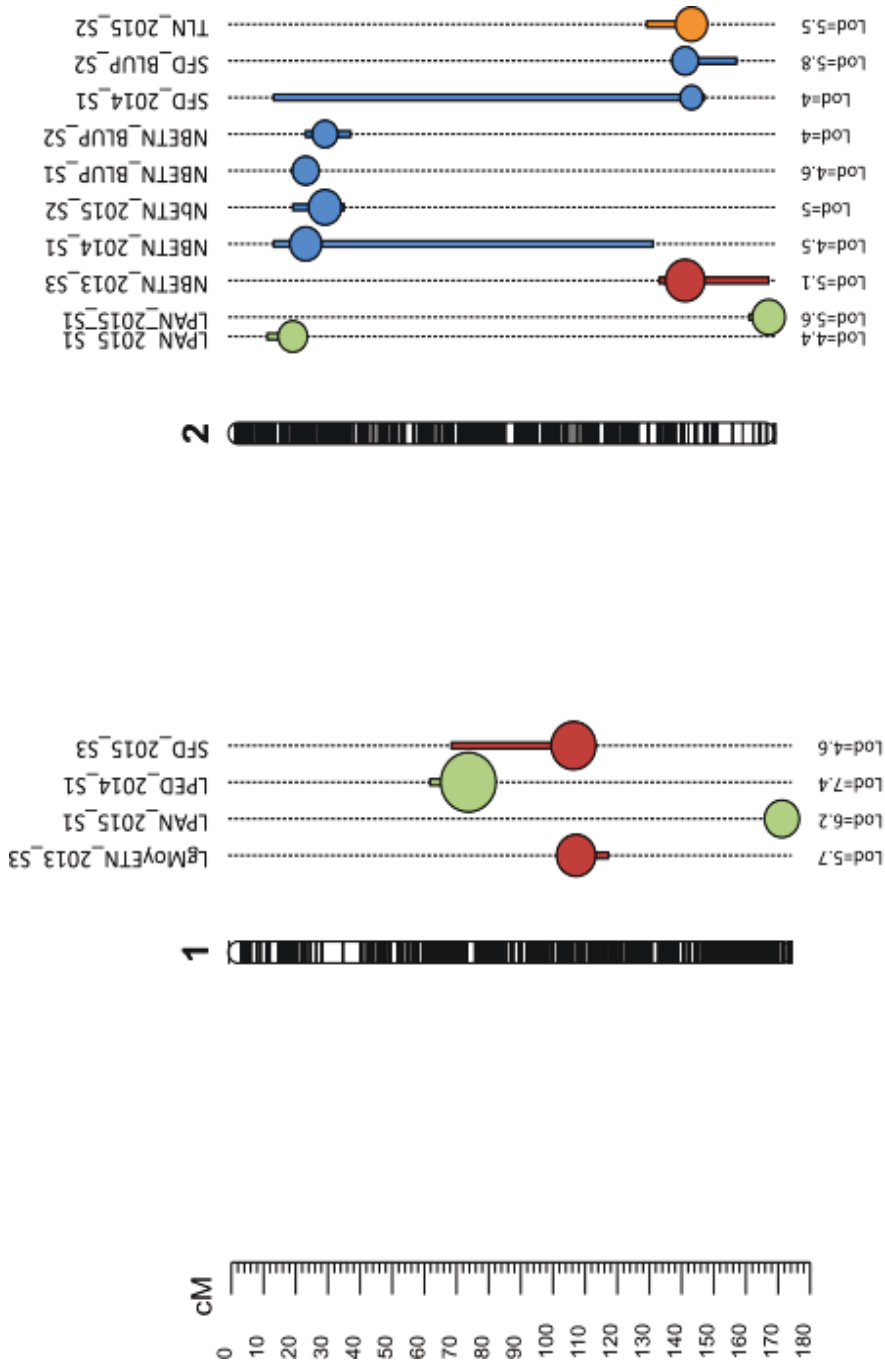
QTL_ID	Variables	Années / Semis	Marqueurs proche	Chr.	LOD	R2 (%)	Binf.	Pos.	Bsup.	Add.
1	SFD	2015_S3	S1_54235531	1	4.65	13.91	69.38	107.38	113.38	3.37
2	SFD	2014_S1	S2_70490684	2	3.95	3.64	14.10	144.10	148.10	-0.87
2	SFD	BLUP_S2	S2_69911173	2	5.78	4.90	138.10	142.10	158.10	-1.27
3	SFD	2014_S2	S3_18977016	3	16.59	31.73	62.85	64.85	82.85	6.24
3	SFD	2015_S2	S3_49026599	3	10.44	17.52	60.85	72.85	82.85	3.33
4	SFD	2015_S1	S3_50900251	3	11.67	14.41	70.85	76.85	80.85	4.81
4	SFD	BLUP_S1	S3_50900251	3	14.27	15.21	70.85	76.85	82.85	4.58
4	SFD	BLUP_S2	S3_50900251	3	23.26	26.72	70.85	76.85	82.85	4.08
4	SFD	2014_S1	S3_52498920	3	14.32	15.72	78.85	80.85	84.85	5.35
5	SFD	2014_S2	S4_61559398	4	3.96	5.77	111.48	113.48	119.48	2.77
5	SFD	2015_S1	S4_62517561	4	4.40	4.81	115.48	117.48	135.48	2.55
5	SFD	BLUP_S1	S4_62517561	4	4.08	3.66	115.48	117.48	135.48	2.19
5	SFD	BLUP_S2	S4_62517561	4	7.36	6.41	113.48	117.48	119.48	1.93
6	SFD	2013_S3	S6_367564	6	5.45	14.56	1.33	1.33	5.33	-3.29
6	SFD	2014_S1	S6_367564	6	33.64	52.42	1.33	1.33	3.33	-9.34
6	SFD	2014_S2	S6_928884	6	9.51	15.57	1.33	3.33	5.33	-3.95
6	SFD	2015_S1	S6_367564	6	29.17	49.21	1.33	1.33	3.33	-7.61
6	SFD	2015_S2	S6_367564	6	19.04	37.13	1.33	1.33	3.33	-4.29
6	SFD	BLUP_S1	S6_367564	6	34.85	54.02	1.33	1.33	3.33	-7.89
6	SFD	BLUP_S2	S6_367564	6	28.05	35.19	1.33	1.33	3.33	-4.02
7	SFD	2014_S2	S8_52875123	8	4.91	7.29	103.38	111.38	115.38	2.74
7	SFD	BLUP_S2	S8_51997979	8	4.02	3.32	103.38	105.38	121.38	1.48
8	SFD	2013_S3	S9_49515157	9	4.02	10.49	52.72	72.72	84.72	2.61
9	SFD	BLUP_S2	S10_11549421	10	6.46	5.54	48.59	58.59	64.59	-1.28
10	Kp	2014_Kp	S3_52498920	3	12.43	25.80	78.85	80.85	84.85	6.08
10	Kp	2015_Kp	S3_50900251	3	11.64	14.46	70.85	76.85	80.85	5.27
10	Kp	Blup_Kp	S3_52498920	3	19.87	27.20	78.85	80.85	82.85	5.49
11	Kp	2015_Kp	S4_62517561	4	4.36	4.79	115.48	117.48	135.48	2.79
11	Kp	Blup_Kp	S4_62148417	4	5.16	5.01	111.48	115.48	119.48	2.50
12	Kp	2014_Kp	S6_367564	6	14.03	30.21	1.33	1.33	3.33	-7.78
12	Kp	2015_Kp	S6_367564	6	29.02	49.28	1.33	1.33	3.33	-8.33
12	Kp	Blup_Kp	S6_367564	6	21.65	30.95	1.33	1.33	3.33	-6.51
13	Kp	Blup_Kp	S7_3965458	7	5.06	4.91	18.71	24.71	32.71	1.14
14	Kp	Blup_Kp	S8_52875123	8	6.01	5.96	103.38	111.38	113.38	2.35
15	LTIG	2014_S1	S3_6268239	3	5.15	7.50	14.85	26.85	160.85	12.07
15	LTIG	2015_S1	S3_6268239	3	5.01	5.77	18.85	26.85	40.85	10.02
15	LTIG	2015_S2	S3_6268239	3	5.66	7.18	18.85	26.85	40.85	9.47
15	LTIG	BLUP_S1	S3_6268239	3	5.60	6.03	14.85	26.85	32.85	9.49
16	LTIG	BLUP_S2	S3_15942198	3	5.46	13.07	60.85	62.85	64.85	0.00
17	LTIG	2015_S2	S3_50900251	3	5.34	6.74	60.85	76.85	90.85	12.77

## Annexe 1 : suite

QTL_ID	Variables	Années / Semis	Marqueurs proche	Chr.	LOD	R2 (%)	Binf.	Pos.	Bsup.	Add.
17	LTIG	2015_S1	S3_54721296	3	5.79	6.76	84.85	86.85	112.85	16.52
18	LTIG	BLUP_S1	S3_58744591	3	4.95	5.27	84.85	110.85	120.85	10.57
19	LTIG	2015_S1	S6_367564	6	6.56	7.77	1.33	1.33	5.33	-14.58
19	LTIG	BLUP_S1	S6_367564	6	4.89	5.20	1.33	1.33	5.33	-11.39
20	LTIG	2014_S1	S7_54915750	7	16.90	30.10	72.71	74.71	78.71	-30.82
20	LTIG	2015_S1	S7_54915750	7	18.79	27.44	72.71	74.71	78.71	-27.11
20	LTIG	2015_S2	S7_55156400	7	15.77	23.80	72.71	76.71	80.71	-21.05
20	LTIG	BLUP_S1	S7_54915750	7	23.05	33.59	72.71	74.71	78.71	-29.12
21	LTIG	2014_S1	S7_58358808	7	24.11	48.87	94.71	96.71	98.71	42.48
21	LTIG	2015_S1	S7_58358808	7	25.91	43.02	94.71	96.71	98.71	37.22
21	LTIG	2015_S2	S7_58358808	7	23.88	41.66	94.71	96.71	98.71	29.99
21	LTIG	BLUP_S1	S7_58358808	7	31.74	54.31	94.71	96.71	98.71	40.56
21	LTIG	BLUP_S2	S7_58761329	7	7.56	18.76	94.71	98.71	100.71	0.00
22	NBETN	2014_S1	S2_4490785	2	4.47	6.95	14.10	24.10	132.10	0.66
22	NbETN	2015_S2	S2_5492926	2	4.99	7.52	20.10	30.10	36.10	0.45
22	NBETN	BLUP_S1	S2_4490785	2	4.63	4.97	20.10	24.10	26.10	0.40
22	NBETN	BLUP_S2	S2_5492926	2	4.01	4.49	24.10	30.10	38.10	0.26
23	NBETN	2013_S3	S2_69911173	2	5.14	10.70	134.10	142.10	168.10	-0.34
24	NBETN	2014_S1	S3_50414812	3	4.82	7.53	58.85	74.85	88.85	0.95
24	NBETN	2014_S2	S3_50900251	3	7.59	14.70	70.85	76.85	84.85	1.37
24	NBETN	2015_S1	S3_20979230	3	5.19	6.22	60.85	66.85	86.85	0.55
24	NbETN	2015_S2	S3_50900251	3	10.51	17.40	60.85	76.85	86.85	1.01
24	NBETN	BLUP_S1	S3_20979230	3	6.46	7.15	60.85	66.85	80.85	0.71
24	NBETN	BLUP_S2	S3_50900251	3	13.93	18.48	60.85	76.85	82.85	0.89
25	NBETN	2013_S3	S6_928884	6	9.48	21.22	1.33	3.33	5.33	-0.86
25	NBETN	2014_S1	S6_367564	6	18.78	37.23	1.33	1.33	3.33	-1.84
25	NBETN	2014_S2	S6_367564	6	11.80	25.00	1.33	1.33	3.33	-1.80
25	NBETN	2015_S1	S6_367564	6	24.78	41.74	1.33	1.33	3.33	-1.46
25	NbETN	2015_S2	S6_928884	6	16.85	31.14	1.33	3.33	5.33	-1.05
25	NBETN	2015_S3	S6_367564	6	4.19	11.82	1.33	1.33	7.33	-0.68
25	NBETN	BLUP_S1	S6_367564	6	28.01	45.22	1.33	1.33	3.33	-1.40
25	NBETN	BLUP_S2	S6_367564	6	20.39	30.31	1.33	1.33	3.33	-1.01
26	NBETN	2013_S3	S6_48088280	6	5.72	12.03	61.33	65.33	67.33	-0.57
26	NBETN	2014_S2	S6_51238248	6	4.66	8.48	71.33	75.33	93.33	-0.90
26	NBETN	2015_S1	S6_51238248	6	4.53	5.36	69.33	75.33	77.33	-0.53
26	NBETN	BLUP_S2	S6_51238248	6	5.46	6.27	71.33	75.33	93.33	-0.47
27	NBETN	2015_S1	S7_58358808	7	5.36	6.45	80.71	96.71	100.71	0.48
27	NBETN	BLUP_S1	S7_58358808	7	5.22	5.65	90.71	96.71	100.71	0.35
28	NBETN	2015_S3	S8_52308546	8	4.36	12.33	35.38	107.38	123.38	0.63
29	LgMoyETN	2013_S3	S1_54414622	1	5.72	10.64	104.23	108.23	118.23	-0.61
30	LgMoyETN	2014_S1	S3_4612127	3	4.97	6.08	14.85	20.85	32.85	0.73
31	LgMoyETN	2014_S1	S6_367564	6	4.09	4.92	1.33	1.33	3.33	0.76
31	LgMoyETN	2014_S2	S6_367564	6	7.60	14.10	1.33	1.33	3.33	1.64
31	LgMoyETN	2015_S2	S6_928884	6	3.97	4.59	1.33	3.33	7.33	0.63
31	LgMoyETN	BLUP_S2	S6_367564	6	6.12	7.39	1.33	1.33	5.33	0.76

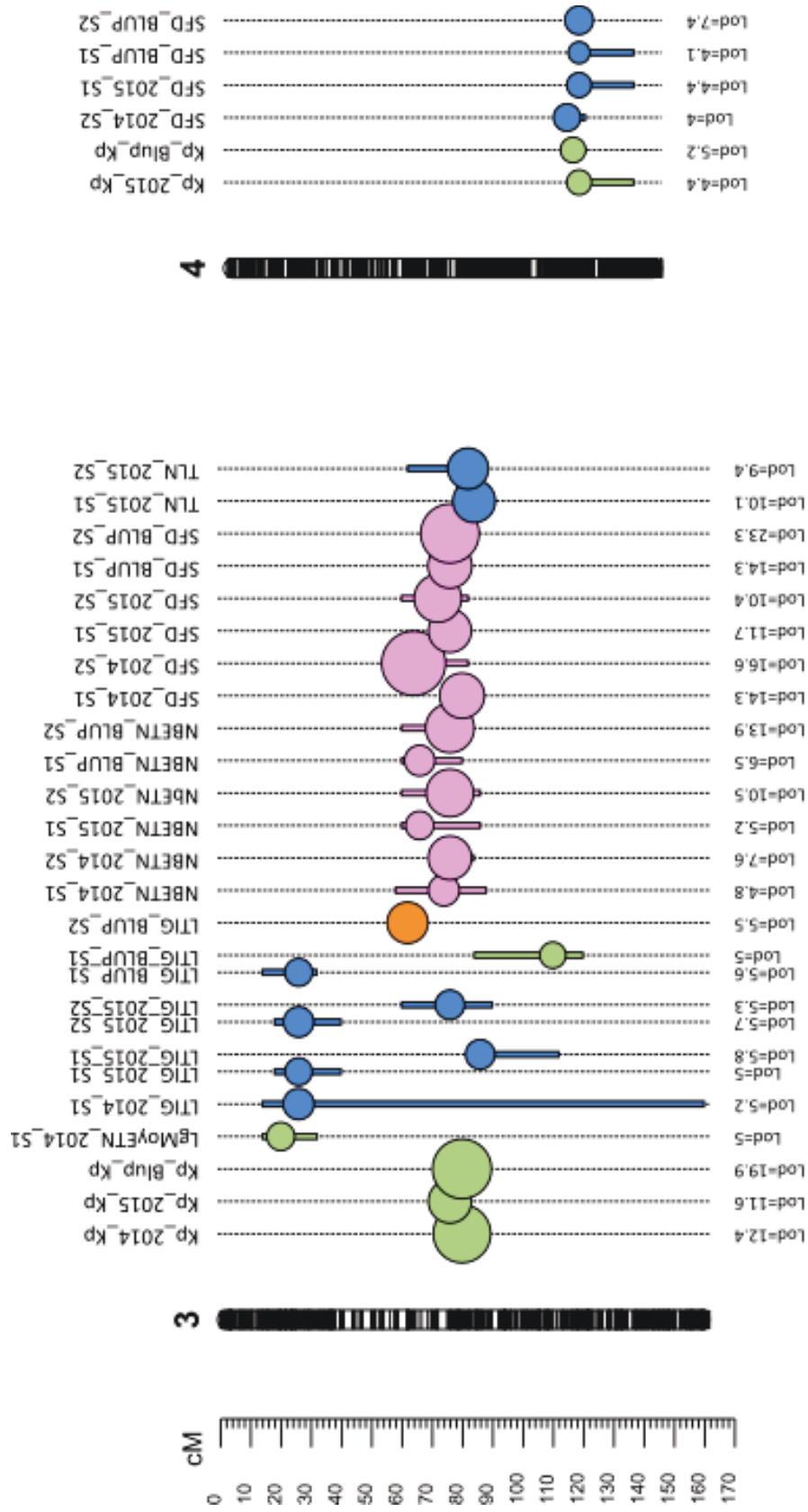
## Annexe 1 : Suite

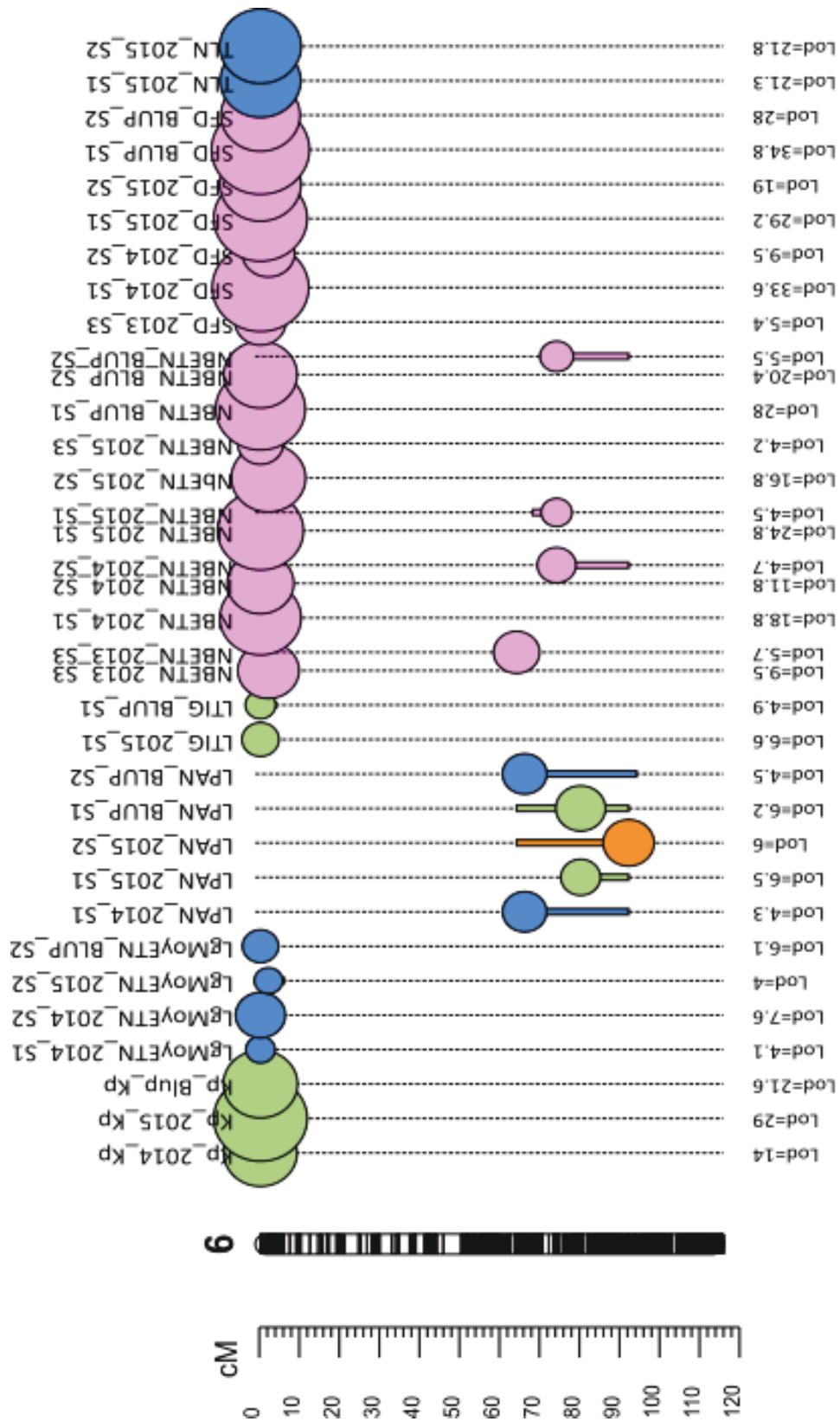
QTL_ID	Variables	Années / Semis	Marqueurs proche	Chr.	LOD	R2 (%)	Binf.	Pos.	Bsup.	Add.
32	LgMoyETN	2015_S3	S7_19441215	7	7.97	19.90	52.71	56.71	82.71	-1.12
33	LgMoyETN	2013_S3	S7_55934608	7	9.50	18.80	72.71	80.71	84.71	-1.75
33	LgMoyETN	2014_S1	S7_54915750	7	21.64	35.28	72.71	74.71	78.71	-2.07
33	LgMoyETN	2014_S2	S7_55629954	7	12.02	24.53	72.71	78.71	82.71	-2.02
33	LgMoyETN	2015_S1	S7_54915750	7	23.39	44.86	72.71	74.71	78.71	-1.73
33	LgMoyETN	2015_S2	S7_55156400	7	21.93	34.66	72.71	76.71	80.71	-1.90
33	LgMoyETN	BLUP_S1	S7_54915750	7	24.02	45.48	72.71	74.71	78.71	-1.77
33	LgMoyETN	BLUP_S2	S7_55156400	7	21.56	34.01	72.71	76.71	80.71	-1.63
34	LgMoyETN	2013_S3	S7_58358808	7	16.42	36.62	94.71	96.71	98.71	2.69
34	LgMoyETN	2014_S1	S7_58358808	7	23.31	39.17	94.71	96.71	98.71	2.44
34	LgMoyETN	2014_S2	S7_58358808	7	14.61	31.57	94.71	96.71	100.71	2.35
34	LgMoyETN	2015_S1	S7_58358808	7	26.22	52.94	94.71	96.71	98.71	2.00
34	LgMoyETN	2015_S2	S7_58358808	7	29.39	53.30	94.71	96.71	98.71	2.48
34	LgMoyETN	2015_S3	S7_58358808	7	12.72	34.67	92.71	96.71	98.71	1.82
34	LgMoyETN	BLUP_S1	S7_58358808	7	26.99	53.97	94.71	96.71	98.71	2.07
34	LgMoyETN	BLUP_S2	S7_58358808	7	27.33	47.88	94.71	96.71	98.71	2.03
35	NBFV	2014_S1	S1_1418359	1	12.17	32.42	1.33	3.33	5.33	1.58
36	NBTFD	2015_S2	S2_70490684	2	5.48	7.09	130.10	144.10	148.10	-0.22
37	NBTFD	2015_S1	S3_54118580	3	10.06	14.20	80.85	84.85	88.85	1.08
37	NBTFD	2015_S2	S3_53285858	3	9.42	13.01	62.85	82.85	84.85	0.88
38	NBTFD	2015_S1	S6_367564	6	21.30	36.57	1.33	1.33	3.33	-1.70
38	NBTFD	2015_S2	S6_367564	6	21.79	37.42	1.33	1.33	5.33	-1.31
39	NBTFD	2015_S1	S7_51352218	7	6.54	8.71	58.71	62.71	96.71	0.54
40	NBTFD	2015_S2	S10_60180434	10	5.13	6.59	122.59	130.59	132.59	0.46
41	LPED	2014_S1	S1_32788768	1	7.41	21.24	62.71	74.71	80.71	-5.07
41	LPED	2015_S1	S7_54915750	7	9.23	21.42	72.71	74.71	82.71	-3.79
41	LPED	2015_S2	S7_55156400	7	6.15	16.20	70.71	76.71	82.71	-3.06
41	LPED	BLUP_S1	S7_54915750	7	8.31	20.66	70.71	74.71	80.71	-3.48
42	LPED	2015_S1	S9_53317323	9	4.29	9.18	88.72	98.72	112.72	-2.49
43	LPED	2015_S2	S9_53869426	9	5.68	14.86	88.72	100.72	112.72	-3.00
44	LPED	2015_S1	S10_55721757	10	3.96	8.42	100.59	104.59	132.59	2.34
45	LPED	BLUP_S1	S10_55721757	10	4.17	9.69	100.59	104.59	122.59	2.33
46	LPAN	2015_S1	S1_71561031	1	6.17	8.42	168.23	172.23	174.23	-1.20
47	LPAN	2015_S1	S2_3765980	2	4.42	5.86	12.10	20.10	22.10	-0.98
48	LPAN	2015_S1	S2_76231279	2	5.55	7.49	162.10	168.10	168.10	-0.62
49	LPAN	2014_S1	S6_48667936	6	4.30	11.08	65.33	67.33	93.33	-1.43
49	LPAN	BLUP_S2	S6_48667936	6	4.46	11.17	65.33	67.33	95.33	-0.70
50	LPAN	2015_S1	S6_52549852	6	6.50	8.93	77.33	81.33	93.33	-1.16
50	LPAN	BLUP_S1	S6_52549852	6	6.17	13.95	65.33	81.33	93.33	-1.02
51	LPAN	2015_S2	S6_55282785	6	5.97	14.70	65.33	93.33	99.33	-1.71
52	LPAN	2014_S1	S7_56386292	7	6.52	17.41	60.71	82.71	86.71	-1.80
52	LPAN	2015_S1	S7_54915750	7	11.95	17.98	70.71	74.71	80.71	-1.55
52	LPAN	2015_S2	S7_56386292	7	6.95	17.43	78.71	82.71	88.71	-1.47
52	LPAN	BLUP_S1	S7_56386292	7	10.59	25.80	64.71	82.71	84.71	-1.22
52	LPAN	BLUP_S2	S7_56386292	7	7.59	20.02	78.71	82.71	88.71	-1.09
53	LPAN	2015_S1	S8_49498524	8	5.25	7.05	83.38	89.38	97.38	-1.09

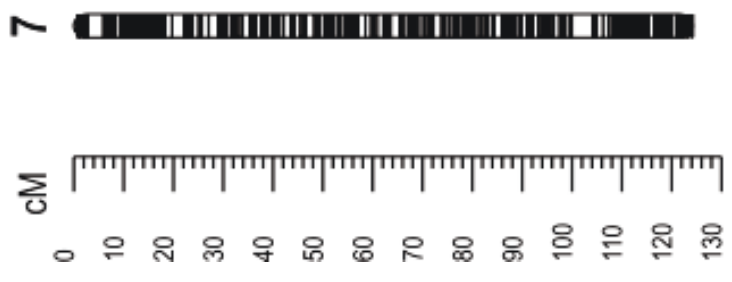
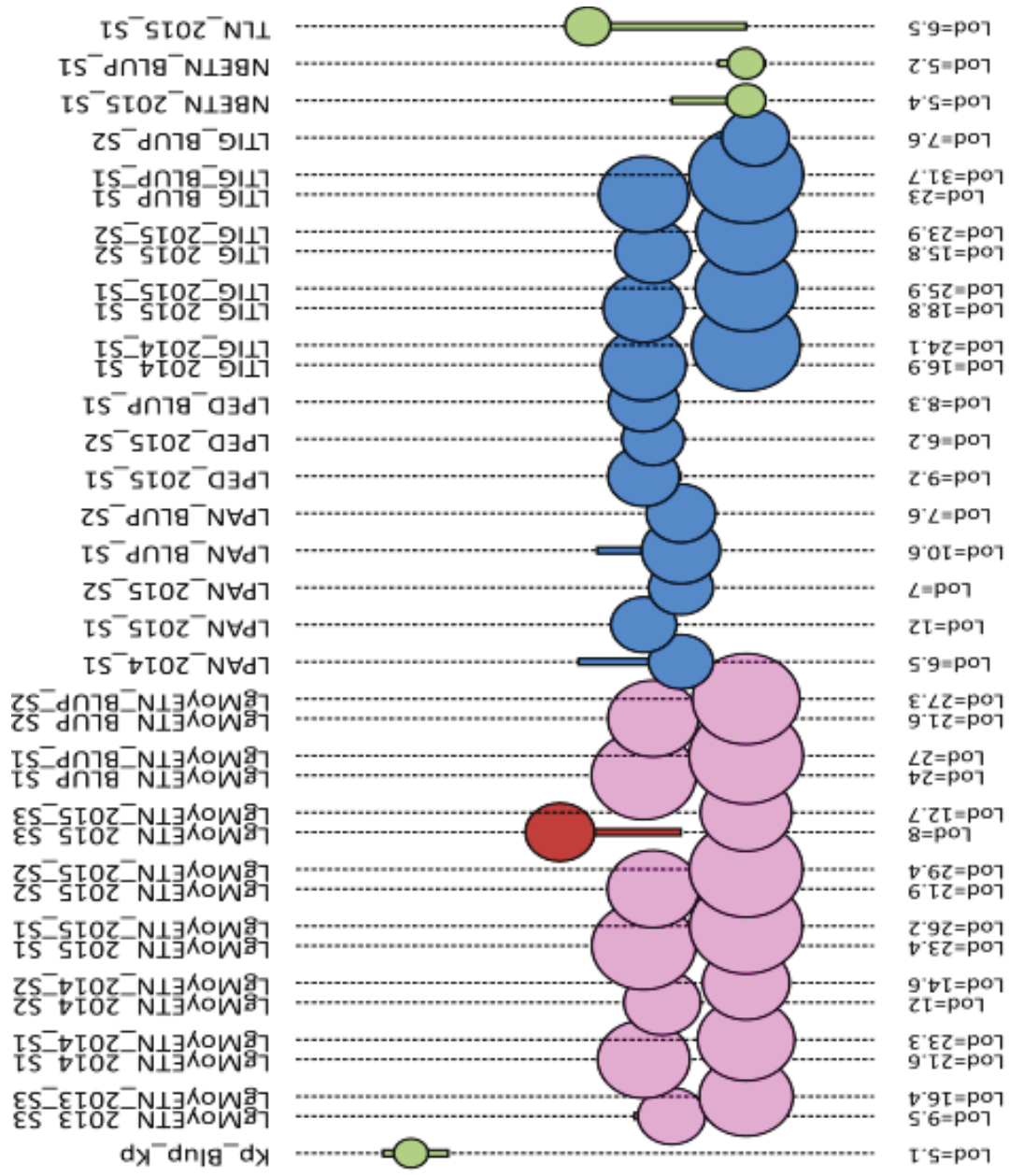


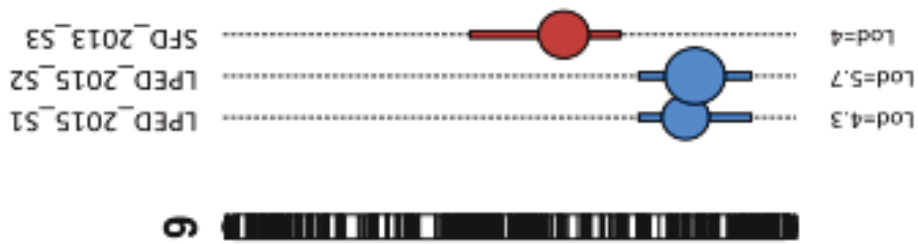
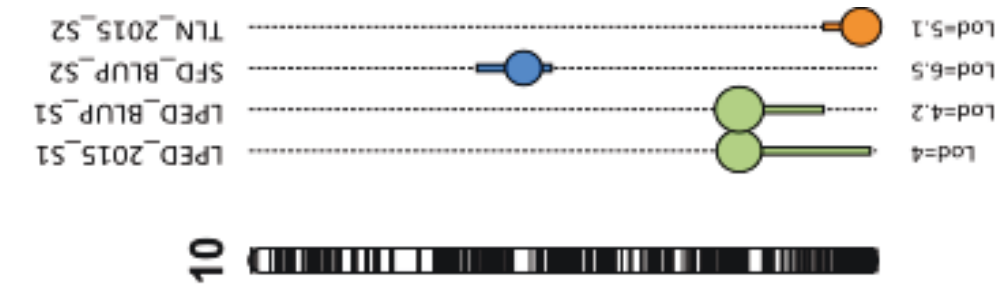
Annexe2 : QTL détectés dans les 6 environnements pour l'ensemble des variables analysées. La stabilité des QTL à travers des différents environnements/semis est illustrée par des couleurs. La couleur verte indique les QTL spécifiques au semis le plus précoce (S1 : mois de juin), les QTL propres au semis le plus tardif (S2 : mois de Juillet) sont représentés par la couleur orange et les QTL spécifiques au semis de contre saison (S3 = mois de septembre) sont représentés en marron. La couleur bleue indique les QTL communs aux semis de juin et juillet et en rose les QTL communs aux trois semis.



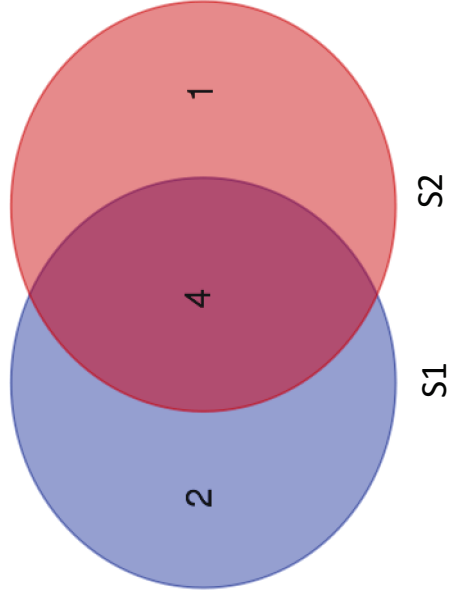




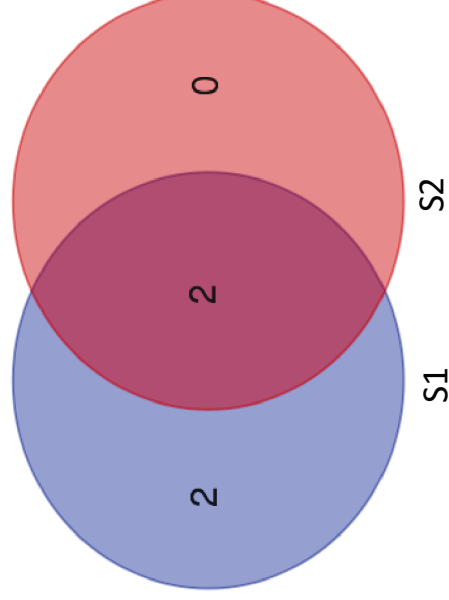




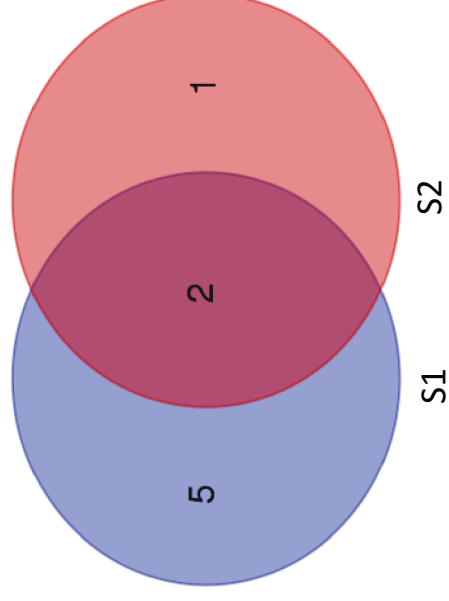
LTIG: 7 QTL



LPED: 4 QTL



LPAN: 8 QTL



Annexe 3 : Stabilité des QTL entre les 3 dates de semis pour les caractères de longueur des tiges (LTIG), des longueurs du pédoncule (LPED) et de la panicule (LPAN) représentée au travers de diagrammes de Venn. Ces QTL sont identifiés uniquement dans les semis de saison normale de culture (Juin : S1 et Juillet : S2). Le semis S1 présente le plus grand nombre de QTL spécifique (47 % (9/19)), tandis que seulement 42% des QTL sont détectés simultanément pour les semis S1 et S2 (8/19). L'ensemble des QTL détectés sur les deux années d'essais a été considéré pour effectuer cette analyse, les QTL correspondant à chaque date de semis ont donc pu être détectés sur une des deux années, sur les deux années simultanément ou sur les valeurs de blup calculées à partir des 2 années.

Annexe 4 : Régions chromosomiques associées à la variation des 9 caractères évalués dans les différents environnements de saison normale de culture 2014 et 2015 et de l'analyse Blup (2014-2015) au sein des 29 populations BCNAM.

QTI_ID	Trait	ANNEES	Chr.	Marqueurs proche	Position (cM)	Lod	R2 (%)
1	SFD	2015	1	S1_55361265	114.48	3.35	2.14
2	SFD	2015	1	S1_68253253	156.39	2.75	2.32
3	SFD	2015	2	S2_3342314	17.38	3.61	2.26
4	SFD	2015	3	S3_20758317	66.00	6.24	3.47
5	SFD	2015	3	S3_51364624	77.95	16.03	3.15
6	SFD	2015	4	S4_48644617	63.03	5.07	3.33
7	SFD	2015	6	S6_618387	2.11	39.91	4.67
8	SFD	2015	6	S6_38136509	34.40	13.3	6.45
9	SFD	2015	7	S7_9039671	53.66	2.79	1.92
10	SFD	2015	10	S10_9471985	50.79	4.02	2.80
11	LTIG	2015	1	S1_14835673	46.54	2.41	43.31
12	LTIG	Blup	1	S1_5612973	15.60	2.52	11.68
13	LTIG	2015	1	S1_7282991	18.52	2.4	42.47
14	LTIG	2015	2	S2_14846850	67.19	2.46	51.37
15	LTIG	2015	3	S3_53793797	84.30	3.55	34.07
16	LTIG	2014	4	S4_50888448	71.72	2.76	39.24
17	LTIG	2015	4	S4_62985615	119.83	2.46	29.32
18	LTIG	2015	5	S5_58133066	109.28	2.61	5.95
19	LTIG	Blup	5	S5_58626654	112.61	2.5	15.06
20	LTIG	Blup	6	S6_40782541	37.79	2.76	16.99
21	LTIG	2014	6	S6_42780401	45.26	2.95	34.17
22	LTIG	2015	6	S6_45123546	54.03	4.14	70.35
23	LTIG	2015	6	S6_618387	2.11	5	16.96
24	LTIG	2015	7	S7_58487671	96.96	22.78	112.95
25	LTIG	2014	7	S7_14725942	55.47	4.06	57.21
26	LTIG	2014	7	S7_55452838	77.95	36.6	191.56
26	LTIG	Blup	7	S7_55452838	77.95	32.46	125.47
27	LTIG	Blup	7	S7_9891139	54.53	4.23	27.62
28	LTIG	2015	9	S9_2448875	15.52	3.19	46.00
29	LTIG	2014	9	S9_51302927	87.16	4.23	70.86
30	LTIG	2014	9	S9_56411124	110.56	5.1	77.57
30	LTIG	Blup	9	S9_56411124	110.56	4.75	52.35
31	LTIG	2015	9	S9_56905564	112.11	5.73	88.03
32	LTIG	Blup	9	S9_7424023	49.20	2.39	19.10
33	LTIG	2015	10	S10_56988108	112.86	2.35	30.80
34	LTIG	Blup	10	S10_640329	3.32	3	34.89
35	NBETN	2014	1	S1_12904805	35.73	3.61	0.25
36	NBETN	Blup	1	S1_1622402	4.70	2.98	0.02
37	NBETN	Blup	1	S1_45064818	80.88	2.31	0.01
38	NBETN	Blup	2	S2_447031	1.60	2.63	0.02
39	NBETN	2014	2	S2_73872122	155.77	2.62	0.15
40	NBETN	Blup	2	S2_75017814	161.09	2.29	0.04
41	NBETN	2015	2	S2_77149208	171.83	3.71	0.24
42	NBETN	2014	3	S3_51364624	77.95	14.12	0.22
42	NBETN	2015	3	S3_51364624	77.95	11.23	0.28
42	NBETN	Blup	3	S3_51364624	77.95	13.38	0.10
43	NBETN	Blup	3	S3_63606943	127.20	3.72	0.04
44	NBETN	2015	3	S3_66628361	134.65	4.95	0.32
45	NBETN	2014	3	S3_70592324	147.32	2.76	0.06
46	NBETN	Blup	3	S3_70952800	149.11	2.69	0.02
47	NBETN	Blup	4	S4_52233153	79.84	2.43	0.04
48	NBETN	2015	4	S4_53063567	83.06	2.55	0.14
49	NBETN	2015	4	S4_8199967	33.46	2.72	0.15
50	NBETN	2014	5	S5_9409738	48.92	2.79	0.16
51	NBETN	2014	6	S6_40052515	35.49	8.83	0.34
52	NBETN	2015	6	S6_40756181	37.69	10.54	0.43
52	NBETN	Blup	6	S6_40756181	37.69	13.35	0.13
53	NBETN	2014	6	S6_52444694	81.08	3.94	0.22
54	NBETN	Blup	6	S6_59602334	109.71	3	0.04
55	NBETN	2015	6	S6_61506430	115.89	2.57	0.12
56	NBETN	2014	6	S6_618387	2.11	10.69	0.12
56	NBETN	2015	6	S6_618387	2.11	21.72	0.28
56	NBETN	Blup	6	S6_618387	2.11	12.57	0.10
57	NBETN	2015	7	S7_58241033	95.22	3.68	0.26
58	NBETN	Blup	7	S7_58340702	96.13	2.37	0.03
59	NBETN	2014	7	S7_9846377	54.52	2.68	0.14
60	NBETN	2015	8	S8_41887563	62.56	2.51	0.15
60	NBETN	Blup	8	S8_41887563	62.56	2.3	0.04

## Annexe 4 : suite

QTI ID	Trait	ANNEES	Chr.	Merqueurs proche	Position (cM)	Lod	R2 (%)
61	NBETN	2014	8	S8_7209173	46.49	2.41	0.09
62	NBETN	2014	9	S9_57515603	114.04	2.35	0.12
63	NBETN	Blup	9	S9_53584846	99.85	2.27	0.03
64	NBETN	2015	9	S9_58323448	118.30	2.6	0.08
65	NBETN	2015	10	S10_16074841	65.66	3.18	0.19
65	NBETN	Blup	10	S10_16074841	65.66	3.37	0.06
66	NBETN	2014	10	S10_51584052	82.30	3.09	0.17
67	NBETN	Blup	10	S10_57101890	113.71	2.37	0.01
68	NBETN	2015	10	S10_60072772	129.82	3.37	0.22
69	LgMoyETN	2014	1	S1_16569662	53.11	2.51	0.11
70	LgMoyETN	Blup	1	S1_5479654	15.37	3.19	0.19
71	LgMoyETN	2015	1	S1_56568966	122.74	2.34	0.13
72	LgMoyETN	2015	1	S1_8571404	19.80	3.07	0.18
73	LgMoyETN	2015	2	S2_16423267	70.30	2.3	0.13
74	LgMoyETN	2014	3	S3_14333071	61.67	3.84	0.37
75	LgMoyETN	2015	3	S3_50440497	75.87	2.53	0.06
76	LgMoyETN	2014	4	S4_63352444	121.81	2.44	0.15
77	LgMoyETN	2015	5	S5_1041007	3.74	2.37	0.08
78	LgMoyETN	2015	5	S5_58133066	109.28	3.49	0.03
79	LgMoyETN	2015	6	S6_43211385	46.88	2.96	0.25
80	LgMoyETN	2015	6	S6_33718749	31.87	3.03	0.22
81	LgMoyETN	Blup	6	S6_48859331	67.93	2.49	0.07
82	LgMoyETN	2015	7	S7_14650483	55.46	3.79	0.12
83	LgMoyETN	2014	7	S7_55196102	76.63	38.14	1.38
83	LgMoyETN	2015	7	S7_55452838	77.95	23.82	0.31
83	LgMoyETN	Blup	7	S7_55452838	77.95	34.58	0.59
84	LgMoyETN	2014	7	S7_58487675	96.96	21.22	0.59
84	LgMoyETN	2015	7	S7_58487675	96.96	22.85	0.39
85	LgMoyETN	Blup	7	S7_9825208	54.52	5.03	0.17
86	LgMoyETN	2015	8	S8_513629	2.8	2.62	0.15
86	LgMoyETN	Blup	8	S8_513629	2.8	2.77	0.16
87	LgMoyETN	2015	9	S9_56905564	112.11	6.22	0.35
88	LgMoyETN	2015	9	S9_3547302	26.84	2.26	0.08
89	LgMoyETN	Blup	9	S9_4367789	34.55	2.48	0.1
90	LgMoyETN	2014	9	S9_56662476	111.35	4.3	0.33
91	LgMoyETN	Blup	9	S9_56878071	112.02	5.92	0.27
92	LgMoyETN	2014	9	S9_6994412	47.75	2.59	0.07
93	LPED	2015	1	S1_26861739	71.6	2.93	1.53
94	LPED	2015	1	S1_8494658	19.72	3.12	1.52
95	LPED	2015	2	S2_1196618	5.7	2.47	0.96
96	LPED	2015	2	S2_73415379	153.95	2.32	1.05
97	LPED	2015	4	S4_55115484	91.02	3.25	1.84
98	LPED	2015	4	S4_57153739	98.27	3.48	1.99
99	LPED	2015	6	S6_3537961	21.14	3.91	2.1
100	LPED	2015	6	S6_51026409	74.36	3.2	0.75
101	LPED	2015	7	S7_55452838	77.95	26	4.54
102	LPED	2015	7	S7_63918540	123.33	2.36	2.01
103	LPED	2015	7	S7_9296826	54.13	4.32	1.86
104	LPED	2015	8	S8_40263917	61.81	2.37	1.65
105	LPED	2015	9	S9_52533282	94.07	3.71	1.71
106	LPED	2015	10	S10_60585013	132.55	2.84	1.28
107	LPAN	2015	1	S1_2638288	7.76	3.08	0.91
108	LPAN	2015	1	S1_8793588	20.02	2.67	0.44
109	LPAN	2015	2	S2_12747904	63.05	2.62	0.16
110	LPAN	2015	2	S2_61874199	117.28	2.35	0.48
111	LPAN	2015	3	S3_5106030	22.42	2.28	0.24
112	LPAN	2015	4	S4_25926739	60.12	3.02	0.58
113	LPAN	2015	4	S4_66826922	140.57	3.29	0.23
114	LPAN	2015	5	S5_52917004	81.2	2.29	0.26
115	LPAN	2015	6	S6_15071243	27.79	2.87	0.38
116	LPAN	2015	6	S6_46967910	60.93	3.49	0.58
117	LPAN	2015	6	S6_53955033	87.79	2.82	0.51
118	LPAN	2015	6	S6_60806170	113.78	2.97	0.29
119	LPAN	2015	7	S7_55293252	77.13	6.25	0.4
120	LPAN	2015	7	S7_64104190	124.3	2.28	0.45
121	LPAN	2015	8	S8_6761511	46.28	2.6	0.17
122	LPAN	2015	9	S9_133594	0.5	2.72	0.22
123	LPAN	2015	9	S9_57055532	112.58	2.57	0.73
124	LPAN	2015	10	S10_7738899	42.86	4.73	1.24

Annexe 5 : Zones chromosomiques impliquées dans le contrôle génétique de la composition biochimique des tiges et de leur digestibilité pour l'ensemble des familles relatives aux 29 populations BCNAM dans les deux années de cultures.

QTL ID	Variables	ANNEES	Marqueurs proche	Chr.	Position (cM)	Lod	R2 (%)
1	ADF	2014	S1_854618	1	2.48	6.12	1.51
2	ADF	2014	S1_4758561	1	14.11	4.49	1.29
3	ADF	2015	S1_52517294	1	99.42	3.68	1.39
4	ADF	2014	S1_53208947	1	102.26	5.14	0.78
5	ADF	Blup	S1_70243740	1	167.23	3.87	0.10
6	ADF	2014	S2_65658115	2	127.94	4.77	2.06
7	ADF	2014	S3_33491528	3	70.56	4.08	0.63
7	ADF	Blup	S3_33491528	3	70.56	5.32	0.10
8	ADF	2015	S3_50502868	3	76.01	3.77	0.55
9	ADF	Blup	S3_51671464	3	78.64	5.33	0.10
10	ADF	2014	S3_57304052	3	102.45	5.1	0.86
11	ADF	2014	S3_67094061	3	135.85	3.58	0.61
12	ADF	2015	S4_60092067	4	108.57	4.01	0.37
13	ADF	2014	S4_61510567	4	113.15	5.07	1.73
14	ADF	2015	S5_60170063	5	118.87	3.38	0.58
15	ADF	2015	S6_838601	6	2.91	4.29	0.35
16	ADF	Blup	S6_51482108	6	76.52	4.18	0.04
17	ADF	2014	S6_51632199	6	77.24	5.85	0.97
18	ADF	2014	S7_9846377	7	54.52	8.56	2.73
19	ADF	2014	S7_55960363	7	80.55	5.07	0.72
20	ADF	2014	S8_17181814	8	51.11	6.32	1.96
21	ADF	Blup	S8_53962857	8	117.57	4.6	0.07
22	ADF	2015	S9_1065779	9	5.44	5.51	1.03
22	ADF	Blup	S9_1065779	9	5.44	3.53	0.10
23	ADF	2014	S9_4126488	9	32.81	4.88	1.39
24	ADF	2014	S9_54859583	9	104.65	6.47	2.49
25	ADF	2014	S10_17105459	10	66.24	4.62	0.91
26	ADL	2014	S1_854618	1	2.48	5.45	0.05
27	ADL	Blup	S1_1011133	1	2.88	6.02	0.00
28	ADL	2014	S1_4758561	1	14.11	4.74	0.05
29	ADL	2015	S1_52517294	1	99.42	4.19	0.06
30	ADL	2014	S1_53208947	1	102.26	4.47	0.03
31	ADL	2014	S1_71978188	1	173.89	3.67	0.04
32	ADL	Blup	S2_69755303	2	140.36	9.78	0.01
33	ADL	Blup	S3_6856142	3	29.19	8.83	0.01
34	ADL	Blup	S3_61320574	3	121.90	8.85	0.02
35	ADL	Blup	S3_70300999	3	146.16	12.09	0.02
36	ADL	Blup	S4_7322069	4	28.89	3.92	0.00
37	ADL	Blup	S4_52951489	4	82.63	5.76	0.00
38	ADL	2014	S4_61431589	4	112.86	3.88	0.06
39	ADL	2015	S6_838601	6	2.91	3.79	0.01
40	ADL	2015	S6_50568091	6	72.18	7.71	0.03
41	ADL	Blup	S6_51019555	6	74.33	3.85	0.00
42	ADL	2014	S6_51632199	6	77.24	8.19	0.05
43	ADL	Blup	S7_1639986	7	9.98	6.44	0.01
44	ADL	2014	S7_9846377	7	54.52	7.25	0.09
45	ADL	2014	S7_56542008	7	83.48	6.43	0.02
46	ADL	Blup	S7_60203696	7	106.10	16.01	0.02
47	ADL	2014	S8_17181814	8	51.11	4.02	0.03
48	ADL	2014	S8_42679489	8	63.31	4.1	0.01
49	ADL	Blup	S8_49994978	8	92.32	45.01	0.7
50	ADL	2015	S9_1065779	9	5.44	4.33	0.04
51	ADL	Blup	S9_49461301	9	72.22	11.76	0.02
52	ADL	2014	S9_54560810	9	103.53	5.27	0.07
53	ADL	Blup	S9_56711089	9	111.50	15.27	0.02
54	ADL	Blup	S10_761863	10	3.95	7.36	0.01
55	ADL	2014	S10_9587060	10	51.30	3.45	0.02
56	adlSndf	2014	S1_63880	1	0.17	4.32	0.07
57	adlSndf	2014	S1_4758561	1	14.11	5.13	0.08
58	adlSndf	2014	S1_71970811	1	173.87	4.39	0.08
59	adlSndf	2014	S2_62788783	2	121.38	5.23	0.03
60	adlSndf	2014	S2_65539838	2	127.61	4.66	0.16
61	adlSndf	2014	S3_2834180	3	12.74	6.05	0.09
62	adlSndf	Blup	S3_61365764	3	122.00	3.94	0.01
63	adlSndf	2014	S4_61364754	4	112.61	5.32	0.12
64	adlSndf	2014	S5_481709	5	1.65	4.34	0.04
65	adlSndf	2014	S6_51193516	6	75.15	5.53	0.13



## Annexe 5 : suite

QTL ID	Variables	ANNEES	Marqueurs proche	Chr.	Position (cM)	Lod	R2 (%)
66	adlSndf	2015	S6_51901469	6	78.52	10.28	0.04
67	adlSndf	2014	S7_2148270	7	13.21	4.74	0.06
68	adlSndf	Blup	S7_3174860	7	19.63	5.59	0.01
69	adlSndf	2015	S7_5331205	7	37.73	3.64	0.01
70	adlSndf	2014	S7_9846377	7	54.52	7.09	0.15
71	adlSndf	2015	S7_58487675	7	96.96	7.25	0.03
72	adlSndf	2014	S7_58593229	7	97.55	6.3	0.07
73	adlSndf	Blup	S7_58680185	7	98.04	8.35	0.01
74	adlSndf	Blup	S7_59449779	7	102.35	5.48	0.02
75	adlSndf	Blup	S8_49985279	8	92.26	4.06	0.01
76	adlSndf	2014	S9_55114325	9	105.63	4.22	0.10
77	adlSndf	2015	S9_57381937	9	113.62	3.85	0.01
78	CBW	2014	S1_217818	1	0.69	4.49	0.89
79	CBW	2014	S1_4758561	1	14.11	3.58	0.79
80	CBW	Blup	S1_6395952	1	16.97	4.21	0.17
81	CBW	Blup	S1_12905189	1	35.73	3.59	0.10
82	CBW	2015	S1_52517294	1	99.42	3.84	1.07
83	CBW	2014	S1_53642880	1	104.07	4.08	0.99
84	CBW	2014	S2_65658115	2	127.94	4.68	1.50
85	CBW	2014	S3_33491528	3	70.56	4.31	0.57
85	CBW	Blup	S3_33491528	3	70.56	6.09	0.11
86	CBW	2015	S3_50502868	3	76.01	4.12	0.44
87	CBW	2014	S3_57304052	3	102.45	4.86	0.60
88	CBW	Blup	S4_51875609	4	78.09	3.55	0.17
89	CBW	2015	S4_60092067	4	108.57	4.03	0.29
90	CBW	2014	S4_62792639	4	118.79	4.29	0.63
91	CBW	2015	S6_838601	6	2.91	4.34	0.27
92	CBW	Blup	S6_51019555	6	74.33	3.61	0.04
93	CBW	2014	S6_51632199	6	77.24	5.05	0.50
94	CBW	2014	S7_9721308	7	54.50	7.52	1.01
95	CBW	2014	S7_55960369	7	80.55	6.61	0.61
96	CBW	2014	S8_17181814	8	51.11	6.48	1.42
97	CBW	Blup	S8_42922042	8	63.70	3.78	0.05
98	CBW	2014	S8_44450024	8	66.13	6.13	1.35
99	CBW	2015	S9_1065779	9	5.44	5.37	0.71
99	CBW	Blup	S9_1065779	9	5.44	4.3	0.09
100	CBW	2014	S9_10240092	9	52.74	4.13	0.43
101	CBW	Blup	S9_49461317	9	72.22	3.42	0.23
102	CBW	2014	S9_54879230	9	104.72	5.75	1.46
103	CBW	2014	S10_17105459	10	66.24	6.13	0.80
104	cbwSndf	2014	S1_12968630	1	36.08	5.84	1.33
105	cbwSndf	2014	S1_73187662	1	177.98	5.85	0.27
106	cbwSndf	2014	S2_58535259	2	95.47	18.28	4.49
107	cbwSndf	2014	S2_63875154	2	124.54	17.14	3.83
108	cbwSndf	2015	S3_62430451	3	124.47	3.45	0.06
109	cbwSndf	2014	S3_72386632	3	156.38	35.29	6.95
110	cbwSndf	2014	S4_53063567	4	83.06	5.01	0.74
111	cbwSndf	2015	S4_57271277	4	98.73	3.62	0.04
112	cbwSndf	2014	S4_59030008	4	106.05	10.08	1.00
113	cbwSndf	2014	S4_62792639	4	118.79	14.44	0.59
114	cbwSndf	2014	S5_639198	5	2.24	8.72	0.89
115	cbwSndf	2015	S5_1579656	5	5.75	4.11	0.04
116	cbwSndf	2014	S5_7084773	5	41.44	6.28	0.27
117	cbwSndf	2015	S6_256309	6	0.80	4.25	0.06
118	cbwSndf	2015	S6_15090432	6	27.80	3.68	0.13
119	cbwSndf	2014	S7_2103077	7	12.93	6.27	0.81
120	cbwSndf	2014	S7_9846377	7	54.52	12.41	1.16
121	cbwSndf	2014	S7_54351609	7	72.30	10.7	0.31
122	cbwSndf	2015	S7_55182396	7	76.56	8.34	0.10
123	cbwSndf	2015	S7_55410965	7	77.73	6.63	0.10
123	cbwSndf	Blup	S7_55452838	7	77.95	14.25	0.03
124	cbwSndf	Blup	S7_59894055	7	104.57	6.48	0.02
125	cbwSndf	2014	S8_46183403	8	72.08	9.92	0.84
126	cbwSndf	2014	S8_55221495	8	123.30	7.18	0.29
127	cbwSndf	2014	S9_4126488	9	32.81	14.61	1.02
128	cbwSndf	2014	S9_54871441	9	104.69	7.71	1.13
129	cbwSndf	2014	S10_17105459	10	66.24	18.92	0.66

## Annexe 5 : suite

QTL ID	Variables	ANNEES	Marqueurs proche	Chr.	Position (cM)	Lod	R2 (%)
130	cvs	2014	S1_217818	1	0.69	6.12	1.15
131	cvs	2014	S1_4758561	1	14.11	4.34	0.97
132	cvs	2015	S1_52517294	1	99.42	4.12	1.02
133	cvs	2014	S1_55369811	1	114.54	5.1	0.84
134	cvs	2014	S1_71977041	1	173.89	3.81	0.98
135	cvs	2014	S2_2670693	2	13.49	3.5	0.53
136	cvs	2014	S2_65658115	2	127.94	4.14	1.31
137	cvs	2015	S3_29496218	3	70.45	4.28	0.37
138	cvs	Blup	S3_33491528	3	70.56	7.29	0.08
139	cvs	2015	S3_50502868	3	76.01	4.68	0.45
140	cvs	2014	S3_57304052	3	102.45	5.63	0.77
141	cvs	2014	S3_67094061	3	135.85	3.38	0.45
142	cvs	2015	S4_60092067	4	108.57	3.84	0.22
143	cvs	2014	S4_61197863	4	111.99	4.54	1.17
144	cvs	2015	S5_60170063	5	118.87	3.53	0.39
145	cvs	2015	S6_838601	6	2.91	3.98	0.22
146	cvs	Blup	S6_51482108	6	76.52	4.01	0.03
147	cvs	2014	S6_51632199	6	77.24	4.36	0.50
148	cvs	Blup	S6_61771295	6	116.38	3.46	0.03
149	cvs	2014	S7_9846377	7	54.52	9.37	2.17
150	cvs	2014	S7_55208247	7	76.69	4.63	0.64
151	cvs	2015	S7_60660307	7	108.36	3.29	0.26
152	cvs	2014	S8_17181814	8	51.11	5.91	1.44
153	cvs	Blup	S8_53962857	8	117.57	3.94	0.03
154	cvs	2015	S9_1065779	9	5.44	5.54	0.67
154	cvs	Blup	S9_1065779	9	5.44	4.38	0.08
155	cvs	2014	S9_4126488	9	32.81	5.83	1.18
156	cvs	2014	S9_54871441	9	104.69	7.37	2.04
157	cvs	2014	S10_17105459	10	66.24	3.87	0.60
158	cvsSndf	2014	S1_55222321	1	113.53	6.32	0.35
159	cvsSndf	2014	S2_2291378	2	11.25	6.7	0.37
160	cvsSndf	2014	S2_57857971	2	91.14	19.36	2.60
161	cvsSndf	2014	S2_64624739	2	125.62	6.82	2.24
162	cvsSndf	2015	S3_19141114	3	64.91	5.4	0.07
163	cvsSndf	Blup	S3_33491528	3	70.56	3.6	0.00
164	cvsSndf	2015	S3_50502868	3	76.01	6.38	0.07
165	cvsSndf	Blup	S3_51364624	3	77.95	3.64	0.01
166	cvsSndf	2014	S3_56730697	3	98.68	6.02	0.35
167	cvsSndf	2014	S3_72386632	3	156.38	13.64	3.32
168	cvsSndf	2014	S4_2324630	4	8.76	15.04	1.75
169	cvsSndf	2015	S4_12151571	4	53.58	3.39	0.07
170	cvsSndf	2014	S4_59030008	4	106.05	6.89	0.89
171	cvsSndf	2014	S4_62742484	4	118.52	10.87	0.97
172	cvsSndf	2014	S4_66826922	4	140.57	12.41	1.40
173	cvsSndf	2014	S5_3172402	5	16.93	7.06	0.75
174	cvsSndf	2014	S5_61714300	5	123.59	6.22	0.39
175	cvsSndf	2015	S5_62104033	5	124.30	4.16	0.12
176	cvsSndf	2015	S6_3279540	6	18.90	4.26	0.14
177	cvsSndf	2015	S6_52444694	6	81.08	8.05	0.09
178	cvsSndf	2015	S7_1142296	7	6.81	4.9	0.04
179	cvsSndf	Blup	S7_1240265	7	7.44	3.52	0.01
180	cvsSndf	2014	S7_9846377	7	54.52	40.58	3.23
181	cvsSndf	Blup	S7_55208247	7	76.69	5.09	0.01
182	cvsSndf	2014	S7_58908922	7	99.32	16.09	2.55
183	cvsSndf	2014	S9_4126488	9	32.81	29.87	1.98
184	cvsSndf	2014	S9_54871441	9	104.69	30.49	3.95
185	cvsSndf	Blup	S9_56711089	9	111.50	4.74	0.01
186	cvsSndf	2014	S9_59171325	9	121.97	10.65	1.31
187	cvsSndf	Blup	S10_2096418	10	10.89	4.13	0.01
188	cvsSndf	2014	S10_2274895	10	11.87	6.84	0.79
189	cvsSndf	2014	S10_17105459	10	66.24	5.94	0.39
190	hemi	2014	S1_205605	1	0.65	6.13	0.43
190	hemi	Blup	S1_205605	1	0.65	4.34	0.05
191	hemi	2014	S1_4758561	1	14.11	5.37	0.46
192	hemi	Blup	S1_6395952	1	16.97	4.5	0.06
193	hemi	Blup	S1_12613668	1	34.09	3.61	0.04
194	hemi	2014	S1_53562386	1	103.71	5.85	0.41

## Annexe 5 : suite

QTL ID	Variables	ANNEES	Marqueurs proche	Chr.	Position (cM)	Lod	R2 (%)
195	hemi	2015	S1_53949905	1	105.78	3.85	0.32
196	hemi	Blup	S1_71291859	1	171.37	3.59	0.03
197	hemi	2014	S2_2291378	2	11.25	4.4	0.27
198	hemi	2014	S2_65658115	2	127.94	7.03	0.80
199	hemi	2014	S3_57503267	3	103.76	4.37	0.24
200	hemi	Blup	S4_51869581	4	78.05	4.66	0.06
201	hemi	2014	S4_62792639	4	118.79	8.63	0.46
202	hemi	2014	S5_403252	5	1.36	6.06	0.39
203	hemi	Blup	S5_55739226	5	97.58	3.88	0.03
204	hemi	2014	S6_51632199	6	77.24	4.28	0.12
205	hemi	Blup	S7_2226391	7	13.70	4.81	0.04
206	hemi	2014	S7_9721308	7	54.50	16.18	0.76
207	hemi	2015	S7_55120635	7	76.24	5.73	0.15
208	hemi	2015	S7_59892688	7	104.56	6.08	0.17
209	hemi	Blup	S7_62299584	7	116.30	4.41	0.11
210	hemi	2014	S7_63676581	7	122.01	4.39	0.24
211	hemi	2014	S8_17181814	8	51.11	9.48	0.75
212	hemi	Blup	S8_49985279	8	92.26	4.85	0.06
213	hemi	Blup	S8_55222869	8	123.31	6.63	0.05
214	hemi	2015	S9_1203433	9	6.23	3.6	0.17
215	hemi	Blup	S9_2785305	9	18.98	4.9	0.07
216	hemi	2014	S9_4126488	9	32.81	6.8	0.47
217	hemi	Blup	S9_49461300	9	72.22	4.13	0.09
218	hemi	2014	S9_55364012	9	106.67	7.33	0.46
219	hemi	Blup	S9_56711089	9	111.50	4.54	0.10
220	hemi	2015	S10_643358	10	3.34	4.04	0.10
221	hemi	Blup	S10_761863	10	3.95	3.99	0.02
222	hemi	2014	S10_17105459	10	66.24	9.68	0.47
223	hemiSndf	Blup	S1_911933	1	2.63	6.95	0.04
224	hemiSndf	Blup	S1_4758561	1	14.11	6.79	0.02
225	hemiSndf	2014	S1_12968630	1	36.08	6.91	1.57
226	hemiSndf	Blup	S1_18169118	1	58.60	4.31	0.02
227	hemiSndf	2014	S1_72567090	1	175.88	9.05	0.40
228	hemiSndf	2014	S2_58535259	2	95.47	19.03	4.95
229	hemiSndf	2014	S2_61951410	2	117.77	10.23	1.99
230	hemiSndf	Blup	S3_61320574	3	121.90	4.18	0.03
230	hemiSndf	Blup	S3_61320574	3	132.92	4.18	0.03
231	hemiSndf	2014	S3_72386632	3	156.38	35.08	7.39
232	hemiSndf	Blup	S4_7322069	4	28.89	3.5	0.01
233	hemiSndf	2014	S4_53063567	4	83.06	7.51	1.00
234	hemiSndf	Blup	S4_53384679	4	84.31	4.74	0.01
235	hemiSndf	2014	S4_59030008	4	106.05	6.65	0.55
236	hemiSndf	2014	S4_62792639	4	118.79	9	0.42
237	hemiSndf	2015	S5_2201209	5	9.73	3.33	0.05
238	hemiSndf	2014	S5_8397100	5	46.00	5.89	0.26
239	hemiSndf	2015	S5_9657528	5	49.75	3.37	0.10
240	hemiSndf	2014	S5_55445320	5	96.07	8.09	0.38
241	hemiSndf	2015	S6_256309	6	0.80	4.58	0.07
242	hemiSndf	2014	S6_2419500	6	11.23	9.08	0.37
243	hemiSndf	2015	S6_51940421	6	78.70	4.13	0.17
244	hemiSndf	Blup	S7_1979500	7	12.15	3.87	0.01
245	hemiSndf	2014	S7_2989217	7	18.48	5.01	0.62
246	hemiSndf	2014	S7_9846377	7	54.52	10.24	0.73
247	hemiSndf	2015	S7_55182396	7	76.56	4.99	0.06
248	hemiSndf	Blup	S7_55960361	7	80.55	4.49	0.01
249	hemiSndf	2015	S7_59812781	7	104.16	4.93	0.10
250	hemiSndf	2014	S8_792266	8	4.45	4.85	0.33
251	hemiSndf	Blup	S8_42922042	8	63.70	5.63	0.01
252	hemiSndf	2014	S8_45736514	8	69.86	6.57	0.63
253	hemiSndf	2014	S8_55244689	8	123.44	4.5	0.26
254	hemiSndf	Blup	S9_1837024	9	10.61	4.59	0.01
255	hemiSndf	2014	S9_4126488	9	32.81	6.06	0.43
256	hemiSndf	2014	S9_15340014	9	56.10	5.04	0.22
257	hemiSndf	Blup	S9_49461301	9	72.22	4.51	0.03
258	hemiSndf	Blup	S9_56711089	9	111.50	3.79	0.02
259	hemiSndf	Blup	S10_7742027	10	42.88	3.38	0.01
260	hemiSndf	2014	S10_17105459	10	66.24	7.42	0.34
261	hemiSndf	2014	S10_49920321	10	77.10	6.04	1.03

## Annexe 5 : suite

QTL_ID	Variables	ANNEES	Marqueurs proche	Chr.	Position (cM)	Lod	R2 (%)
262	inndfd	2014	S2_1252143	2	5.95	6.5	2.06
263	inndfd	Blup	S3_14961414	3	62.10	3.26	0.05
264	inndfd	2015	S3_16628947	3	63.22	4.83	0.34
265	inndfd	Blup	S4_62328451	4	116.29	3.3	0.02
266	inndfd	2014	S5_6151462	5	36.17	3.79	1.24
267	inndfd	2014	S5_49019811	5	72.86	4.82	1.93
268	inndfd	2015	S6_1806235	6	6.42	3.31	0.17
269	inndfd	2014	S6_49320060	6	69.44	5.69	1.06
269	inndfd	Blup	S6_49320060	6	69.44	7.14	0.02
270	inndfd	2015	S6_51026409	6	74.36	5.73	0.27
271	inndfd	2014	S6_59721507	6	110.11	5.02	1.70
272	inndfd	Blup	S7_2919632	7	18.05	5.58	0.10
273	inndfd	2015	S7_5117558	7	36.14	4.8	0.53
274	inndfd	2014	S7_9721308	7	54.50	6.84	2.29
275	inndfd	2014	S7_55452838	7	77.95	7.88	2.31
276	inndfd	2015	S7_58487675	7	96.96	8.05	0.38
277	inndfd	Blup	S7_58593229	7	97.55	8.81	0.04
278	inndfd	2014	S7_59894055	7	104.57	11.76	3.03
279	inndfd	2014	S8_13480611	8	49.40	4.29	1.38
280	inndfd	2014	S8_51774965	8	103.41	9.67	2.88
280	inndfd	Blup	S8_51774965	8	103.41	3.58	0.06
281	inndfd	2014	S10_932575	10	4.84	6.17	2.43
282	inndfd	Blup	S10_10421454	10	54.99	4.84	0.09
283	inndfd	2014	S10_11155900	10	58.24	8.86	2.81
284	MAT	Blup	S1_1921019	1	5.60	10.51	0.03
285	MAT	2014	S1_11786191	1	29.28	4.55	0.11
286	MAT	2015	S1_13314928	1	38.02	3.4	0.09
287	MAT	2014	S1_57758029	1	128.41	4.23	0.10
288	MAT	2015	S2_1106921	2	5.19	4.43	0.10
289	MAT	2014	S2_59725258	2	102.57	5.83	0.12
290	MAT	2014	S2_72784392	2	151.79	3.86	0.08
291	MAT	2014	S3_59145302	3	112.61	4.65	0.14
292	MAT	Blup	S3_61153174	3	121.51	6.36	0.02
293	MAT	2014	S3_65402245	3	131.51	4.35	0.08
294	MAT	Blup	S4_8220598	4	33.57	4.7	0.01
295	MAT	Blup	S4_52951489	4	82.63	6.63	0.01
296	MAT	2015	S5_1973052	5	8.26	4	0.06
297	MAT	2014	S5_6649360	5	38.98	6.12	0.12
298	MAT	2014	S5_49327487	5	73.78	3.85	0.07
299	MAT	2015	S5_60113323	5	118.66	4.18	0.11
300	MAT	2014	S6_53273664	6	84.32	4.21	0.07
301	MAT	2015	S6_59043460	6	107.89	4.91	0.09
301	MAT	2014	S6_59059786	6	107.94	4.43	0.10
302	MAT	Blup	S7_1979500	7	12.15	6.43	0.02
303	MAT	2014	S7_55196102	7	76.63	6.94	0.04
304	MAT	2015	S7_55667891	7	79.05	4.72	0.04
305	MAT	Blup	S7_62379810	7	116.68	5.53	0.03
306	MAT	2014	S7_63320124	7	119.71	3.7	0.07
307	MAT	Blup	S8_49985279	8	92.26	7.3	0.03
308	MAT	Blup	S8_54801975	8	120.82	8.05	0.01
309	MAT	2015	S9_51021466	9	84.64	5.21	0.13
310	MAT	Blup	S9_56711089	9	111.50	6.75	0.01
311	MAT	2015	S9_57438983	9	113.80	4.17	0.18
312	MAT	2014	S10_2388340	10	12.75	3.9	0.07
313	MAT	Blup	S10_17104636	10	66.24	4.23	0.01
314	MAT	2015	S10_54387299	10	97.72	5.54	0.15
315	MAT	2015	S10_60084853	10	129.89	3.65	0.07
316	NDF	2014	S1_854618	1	2.48	7.5	4.27
317	NDF	2014	S1_4758561	1	14.11	6.39	4.27
318	NDF	Blup	S1_51820309	1	96.56	3.35	0.38
319	NDF	2015	S1_52709577	1	100.21	4.12	2.97
320	NDF	2014	S1_55369811	1	114.54	7.17	3.58
321	NDF	2014	S1_71977041	1	173.89	10.58	7.39
322	NDF	2014	S1_73317415	1	178.42	4.63	3.41
323	NDF	2014	S2_65658115	2	127.94	8.51	9.25
324	NDF	2015	S3_29496218	3	70.45	3.51	0.87
325	NDF	Blup	S3_33491528	3	70.56	5.97	0.22
326	NDF	2014	S3_57304052	3	102.45	4.71	1.68

## Annexe 5 : suite

QTL_ID	Variables	ANNEES	Marqueurs proche	Chr.	Position (cM)	Lod	R2 (%)
327	NDF	2014	S3_68853986	3	140.38	3.93	3.75
328	NDF	2015	S4_60092067	4	108.57	3.92	0.71
329	NDF	2014	S4_61197863	4	111.99	6.4	5.18
330	NDF	2014	S5_3314616	5	18.06	4.26	2.44
331	NDF	2014	S5_55456927	5	96.13	4.57	3.35
332	NDF	2015	S6_838601	6	2.91	3.76	0.61
333	NDF	2014	S6_2545524	6	12.35	5.89	3.72
334	NDF	Blup	S6_51482108	6	76.52	5.19	0.10
335	NDF	2014	S6_51632199	6	77.24	6.82	3.32
336	NDF	Blup	S6_61771295	6	116.38	3.58	0.10
337	NDF	2014	S7_9846377	7	54.52	11.82	8.67
338	NDF	2014	S7_55208247	7	76.69	3.97	2.20
339	NDF	2015	S7_60660246	7	108.36	4.22	0.93
340	NDF	2014	S8_17181814	8	51.11	9.3	6.90
341	NDF	Blup	S8_53962857	8	117.57	3.88	0.11
342	NDF	2015	S9_1065779	9	5.44	5.16	1.96
342	NDF	Blup	S9_1065779	9	5.44	4.3	0.27
343	NDF	2014	S9_4126488	9	32.81	4.34	3.22
344	NDF	2014	S9_54871441	9	104.69	6.49	6.32
345	NDF	2014	S10_56095271	10	106.76	3.68	1.84
346	SMO	2015	S1_52517294	1	99.42	4.66	3.69
347	SMO	2015	S3_19679186	3	65.28	3.59	1.04
347	SMO	Blup	S3_20979230	3	66.15	5.7	0.32
348	SMO	Blup	S3_33491528	3	70.56	6.64	0.35
349	SMO	2014	S3_50502868	3	76.01	6.41	3.29
350	SMO	2015	S3_19679186	3	76.01	3.59	1.04
351	SMO	2014	S3_66599752	3	134.58	4.73	1.39
352	SMO	2014	S3_68858653	3	140.40	4.88	2.63
353	SMO	2015	S4_60092067	4	108.57	3.61	0.66
354	SMO	2015	S6_914672	6	3.19	4.54	0.87
355	SMO	Blup	S6_51026409	6	74.36	7.06	0.17
356	SMO	2015	S6_51596069	6	77.06	3.55	0.77
356	SMO	2014	S6_51632199	6	77.24	10.22	2.22
357	SMO	2014	S7_55960369	7	80.55	6	2.13
358	SMO	2014	S7_59696578	7	103.59	4.19	1.32
359	SMO	Blup	S8_53962857	8	117.57	3.51	0.12
360	SMO	2015	S9_1065779	9	5.44	6.18	2.57
361	SMO	2014	S9_1848781	9	10.70	4.35	1.81
362	SMO	Blup	S9_2163743	9	13.21	3.91	0.18
363	SMS	2015	S1_52517294	1	99.42	5.16	3.85
364	SMS	2015	S3_19679186	3	65.28	3.5	0.91
365	SMS	Blup	S3_20979230	3	66.15	3.89	0.18
366	SMS	Blup	S3_33491528	3	70.56	4.67	0.20
367	SMS	2014	S3_50440461	3	75.87	5.25	2.55
368	SMS	2014	S3_68858653	3	140.40	5.31	2.89
369	SMS	2015	S4_60092067	4	108.57	3.82	0.68
370	SMS	2015	S6_914672	6	3.19	4.59	0.85
371	SMS	2015	S6_51596069	6	77.06	4.04	0.87
371	SMS	Blup	S6_51596069	6	77.06	8.48	0.20
371	SMS	2014	S6_51632199	6	77.24	11.35	2.44
372	SMS	2014	S7_9721308	7	54.50	4.61	2.13
373	SMS	2014	S7_55960369	7	80.55	7.2	2.56
374	SMS	2014	S7_59696578	7	103.59	5.67	2.02
375	SMS	Blup	S8_53962857	8	117.57	3.3	0.11
376	SMS	2015	S9_1065779	9	5.44	5.91	2.36
377	SMS	2014	S9_1848781	9	10.70	3.56	1.57
378	SMS	Blup	S9_2163743	9	13.21	3.8	0.12

## **Analyse des déterminants génétiques contrôlant la production et la composition de la tige chez le sorgho. Intégration des approches bi-et multiparentales**

Le sorgho est une des espèces de céréales les plus importantes au niveau mondial en termes d'alimentation humaine et animale. En outre, de nouvelles voies de valorisations liées à son haut potentiel de production de biomasse émergent (énergie, biomatériaux) avec des possibilités d'applications aux niveaux des pays du Nord et du Sud. L'identification des facteurs génétiques contrôlant la variabilité de la production et de la qualité de la biomasse est un des vecteurs de développement de variétés adaptées aux différentes utilisations. Actuellement, seule une vision partielle de l'architecture génétique de la production et de la qualité de la biomasse est disponible, pénalisant l'efficacité des programmes de sélection. Les principales raisons qui peuvent être évoquées sont : l'utilisation quasi exclusive de populations à base génétique très étroite (population biparentales correspondant à des idéotypes peu divers) et un manque de compréhension des interactions existantes entre la qualité de la biomasse et la phénologie. Dans ce contexte, cette thèse s'est articulée autour de 3 objectifs : i) apporter une meilleure compréhension du déterminisme génétique de la durée du cycle « semis - floraison » et de la hauteur des plantes dans le contexte de variétés sensibles à la photopériode, lesquelles ont été relativement peu travaillées jusqu'à présent, ii) analyser les interactions entre la qualité de la biomasse, la croissance en hauteur, la durée du cycle et la sensibilité à la photopériode et iii) affiner notre compréhension des mécanismes moléculaires impliqués dans la variation de la qualité de la biomasse. Pour mener à bien ces analyses, deux populations de cartographie biparentales, un panel de diversité correspondant aux 35 parents d'un schéma de croisement multiparental (Backcross Nested Association Mapping) ainsi que leurs descendances (29 populations totalisant environ 1200 BC1F4) ont été caractérisés en termes de croissance, de phénologie et de qualité de la biomasse. L'analyse du déterminisme génétique de la phénologie couplée à une approche de modélisation écophysiological a permis la mise en évidence d'une région chromosomique contrôlant une large part de la variabilité de la photopériode critique. Le gène *ELF3* pour lequel plusieurs évidences fonctionnelles ont été mises en évidence chez d'autres espèces, est un candidat potentiel pour expliquer cet effet. Outre une large convergence entre les zones détectées pour la phénologie au sein de cette thèse et les études précédentes, une forte instabilité des zones détectées en fonction des conditions environnementales (dates de semis, année) a été mise en évidence. L'analyse des relations entre la qualité de la biomasse, la croissance en hauteur, la précocité et la sensibilité à la photopériode a mis en évidence des effets importants du nanisme et de la date de semis sur la composition des tiges et des relations entre ses composants. En outre, des différences de réactions à la variabilité des dates de semis ont été observées en fonction de la sensibilité à la photopériode des génotypes considérés. Enfin, la caractérisation de la composition biochimique des tiges (teneur en cellulose, en hémicellulose et en lignine des parois) au sein des 29 populations BC1F4, a mis en évidence de nombreuses zones chromosomiques pour lesquelles des gènes potentiellement causatifs ont été proposés. Ces analyses ont permis d'alimenter un répertoire de régions chromosomiques contrôlant les caractères cibles de la sélection. Les potentialités de ces résultats et des populations développées (bi- et multiparentales) dans le contexte de la sélection ainsi que de la validation des gènes sous-jacents à ces zones sont discutées.

### **Genetic determinism of stem biomass production and composition in sorghum. Integrating bi- and multi-parental designs**

Sorghum ranks fifth in terms of grain production at the worldwide level. It contributes to food security in the developing countries and provides a significant share of the animal caloric intake. In addition, sorghum is expected to contribute significantly to the emerging energy and biomaterial value chains. In this context, identification of the genetic factors controlling biomass production and quality is of crucial interest to develop varieties fitting the expectations of the different uses. Currently only a partial vision of the genetic architecture of biomass production and quality is available, hampering optimal breeding efficiencies. This can be explained by a limited use of the genetic diversity pool to dissect the genetic control of the traits of interest (mainly restricted to non-photoperiod sensitive varieties until now) and a limited understanding of the relationships between biomass production (largely driven by phenology variability) and quality. In this context, this PhD aimed at i) providing a better understanding of the genetic bases of flowering and height variabilities using photoperiod sensitive genetic material ii) disentangling the interactions between biomass quality, height, cycle duration and photoperiod sensitivity and iii) refining our understanding of the molecular mechanisms involved in the variation in biomass quality. To tackle these challenges, two biparental QTL mapping populations, a diversity panel including 35 parents of a Backcross Nested Association Mapping design and a subset of their progenies (29 populations encompassing a total of 1200 BC1F4) were characterized for their phenology, growth and biomass composition. The analysis of the genetic determinism of phenology coupled with an ecophysiological modeling approach, revealed a chromosomal region controlling a large part of the variability (60%) of critical photoperiod (a component of cycle duration). The gene *ELF3*, for which functional evidence of impact on photoperiod sensitivity is available in other species can be proposed as a functional driver of the variability induced by this region. Although a significant proportion of the genomic regions detected in this study have already been revealed in previous studies, several new regions have been discovered, underlying the interest of multiparental based designs. Phenology analyses over different sowing dates highlighted the complexity of the genetic determinism of this trait through the detection of genomic regions with variable effects. In depth analyses of biomass quality relationships with height, precocity and photoperiod sensitivity underlined strong dwarfism and sowing dates effects on the components of biomass quality and their relationships. In addition, different behaviors of photoperiod sensitive and insensitive genotypes to delayed sowing dates have been observed. Finally, characterization of stem biochemical composition (cellulose, hemicellulose and lignin content of the cell walls) on the 29 BC1F4 populations, revealed numerous chromosomal regions with variable effects. Putative candidate genes from these regions of interest have been proposed. These analyses enriched a library of regions impacting the breeding targets linked to biomass production and quality. The potentials of these results and of the populations developed towards optimization of breeding efficiency and gene function validation have been discussed.