TAKING SORGHUM TO NEW HEIGHTS: IDENTIFICATION OF GENES CONTROLLING HEIGHT VARIATION IN SORGHUM

A Dissertation

by

JOSIE LYNN HILLEY

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	John Mullet
Committee Members,	Mary Bryk
	Clare Gill
	William Rooney
Intercollegiate Faculty Chair,	Dorothy Shippen

May 2017

Major Subject: Genetics

ABSTRACT

Sorghum is an important cereal crop worldwide though it is particularly important in semi-arid regions. It is grown for many uses including food, feed, forage, sugar, and bioenergy. In its native Africa, sorghum is 3-4 meters in height. However, in the U.S. shorter plants were selected for grain production to reduce lodging and to facilitate mechanical harvesting. In the 1950s, researchers determined that this variation in height was controlled by four major genes they termed the *dwarfing* (*Dw1-Dw4*) genes. In 2003, *Dw3* was identified as an ABCB efflux transporter of the plant hormone auxin. The locations of *Dw1* and *Dw2* have also been determined though the underlying genes remain to be elucidated. *Dw1* was found to be on chromosome 9 at ~57 Mbp and *Dw2* is located at ~42 Mbp on chromosome 6. The location of *Dw4* has not been definitively determined though locations of ~6 Mbp on chromosome 6 and ~67 Mbp on chromosome 4 have both been suggested.

In the work described in this dissertation, I determined that the gene that underlies *Dw1* is Sobic.009G229800, a highly conserved gene of unknown function. Furthermore, *Dw1* is found to interact with a QTL on chromosome 7. *Dw2* was determined to be Sobic.006G067700 a kinase whose closest homolog in Arabidopsis is KCBP INTERACTING PROTEIN KINASE (KIPK). KIPK is a member of the AGC protein kinase family subgroup AGCVIII, which includes several kinases involved in the regulation of auxin transport. Lastly, I attempted to locate *Dw4* through crosses with two different broomcorns. Surprisingly, no QTL matching the description of *Dw4* was

found. Overall this work increased our understanding of the genetic control of height in sorghum, as well as revealing some exciting possible new regulators of growth.

DEDICATION

To my mother, Nancy, and in memory of my father, Robert, who always supported and encouraged me.

ACKNOWLEDGEMENTS

I would like to thank my committee chair Dr. John Mullet, and my committee members, Dr. Mary Bryk, Dr. Clare Gill, and Dr. William Rooney, for their assistance and encouragement throughout the course of this research. Thanks also to my lab mates in the Mullet lab for all their help and support, as well as their willingness to celebrate the highs and commiserate the lows through the years. Special thanks to Dr. Daryl Morishige for all of his assistance and aid in troubleshooting. I would also like to thank Dr. Patricia Klein for her contributions toward genotyping many of my populations. Finally, thanks to my fellow Interdisciplinary Program of Genetics graduate students for their encouragement.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a dissertation committee consisting of Professor Mullet and Professor Bryk of the Department of Biochemistry and Biophysics, Professor Clare Gill of the Department of Animal Science, and Professor William Rooney of the Department of Soil and Crop Sciences.

The QTL mapping and fine mapping of Chapter II was aided by Dr. Patricia Klein of the Department of Horticulture and Dr. Daryl Morishige, Dr. Sara Olson, Ryan McCormick, Sandra Truong, and Susan Hall from the Department of Biochemistry and Biophysics.

The QTL mapping and fine mapping in Chapter III was aided by Dr. Daryl Morishige, Dr. Brock Weers, Ashley Mattison, Ryan McCormick, Sandra Truong, Susan Hall, and Robin Jakubik of the Department of Biochemistry and Biophysics.

The QTL mapping in Chapter IV was aided by Dr. Patricia Klein of the Department of Horticulture and Ryan McCormick and Susan Hall of the Department of Biochemistry and Biophysics.

All other work conducted for the dissertation was completed by the student independently.

This work was made possible by the Perry Adkisson Chair in Agricultural Biology, the USDA-DOE Feedstock Genomics Program under Grant Number NIFA No. 2010-67009-21507, the Great Lakes Bioenergy Research Center under Grant Number DE-FC02-07ER64494, and Ceres, Inc.

Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the various funding sources.

NOMENCLATURE

SAM	Shoot Apical Meristem
GA	Gibberellic Acid, also known as Gibberellin
BR	Brassinosteroid
DG	Digital Genotyping
MQM	Multiple QTL Mapping
HIF	Heterogeneous Inbred Family
RIL	Recombinant Inbred Line
CAPS	Cleaved Amplified Polymorphic Sequence
SYM	Standard Yellow Milo
DYM	Dwarf Yellow Milo
DDYM	Double Dwarf Yellow Milo
KCBP	Kinesin-like Calmodulin Binding Protein
KIPK	KCBP Interacting Protein Kinase
USDA ARS GRIN	U.S. Dept. of Agriculture Agricultural Research Service Germplasm Resources Information Network

TABLE OF CONTENTS

Page

Page

Page

LIST OF FIGURES

Page

Page

LIST OF TABLES

Page

Page

CHAPTER I

INTRODUCTION

Sorghum is the fifth most widely grown cereal crop worldwide (www.fao.org). It is particularly important in semi-arid environments, as sorghum is drought tolerant. It is grown for many uses including grain, feed, forage, sugar, and bioenergy. Sorghum has a lot of phenotypic variety with preferences based on the end use. When the grain is harvested, shorter plants are favored to reduce lodging and to facilitate mechanical harvesting. When the stem is the product, such as stem sugar in sweet sorghum, forage, and bioenergy from biomass, a larger stem increases the yield. Indeed, in bioenergy sorghum 83% of the shoot biomass was from the stem [1].

Sorghum is a C4 grass. It diverged from maize ~12 million years ago and from rice ~50 million years ago [2]. Sorghum is a diploid with a relatively small genome of ~730 Mbp and 34,000 genes. The genome is split into ten chromosomes. The first sorghum genome sequence and annotation was released in 2009 [3]. The genome and the gene annotation have been revised twice since then. These characteristics can make sorghum an appealing model organism for C4 grasses with more complex genomes.

Like other plants, the above ground tissue in sorghum is called the shoot. The shoot is divided into repeating units called phytomers. In grasses, the phytomer consists of a node, an internode, and a leaf coming from the node at the base of the internode (Fig 1). The leaf is divided into the leaf sheath, which surrounds the internode and the leaf blade that grows out from the stem. The final leaf is referred to as the flag leaf. The

final internode is called the peduncle which terminates in a collection of many flowers known as the panicle or head (Fig 1).

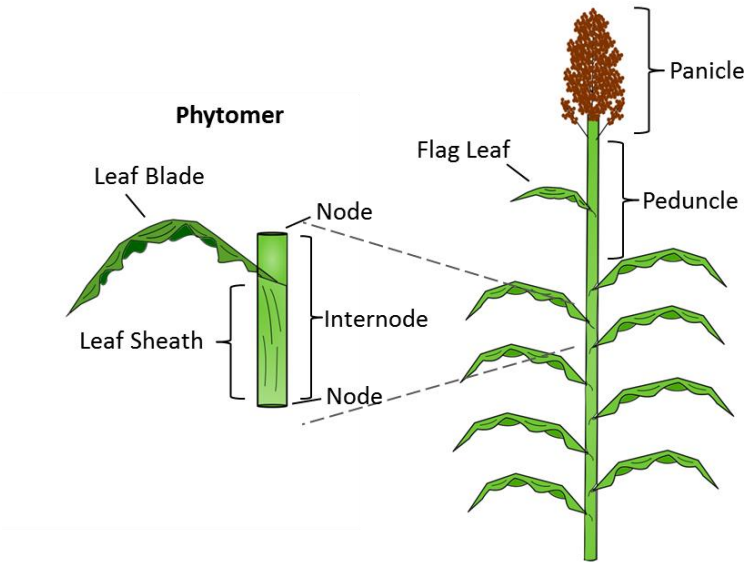


Fig 1. Diagram of the sorghum stem. The sorghum shoot, or above ground tissue, is made of repeating units called phytomers (enlarged). A phytomer is made of the node, internode, leaf sheath, and leaf blade (all labeled). The final internode is called the peduncle and the leaf that surrounds the peduncle is the flag leaf (both labeled on whole shoot). The stem terminates in a panicle, also known as the head, which contains the flowers and, later, seeds.

HISTORY OF SORGHUM

Sorghum bicolor (L) Moench is native to Africa where it is has many uses ranging from food to building material [4]. In this environment, it is generally tall (3-4 m) and photoperiod sensitive and flowers when the day length decreases toward 12

hours of daylight. Modern, domesticated sorghum is grouped into five races based on the morphology of the panicle, the spikelet, and the seed [5]. The most primitive of the five races is bicolor which is thought to come from central Africa. The guinea race is from western Africa and has many characteristics adaptive to higher rainfall. The race caudatum is from central Africa and is one of the more important races agronomically. Kafir is thought to have originated in southern Africa. From Africa, sorghum was taken to India and possibly taken back to Africa. Durra is thought to have come from eastern Africa and/or India. In addition to the five races, intermediates from hybrids of the races also exist for each combination [5].

Sorghum in the United States

Systematic sorghum introduction into the United States from Africa and Europe occurred in the mid-1800s. The earliest lines that are documented are sweet sorghum lines. The first lines that were grown for grain that contributed to future breeding stock are Milo Maize and Guinea Kafir [6]. Sorghum had an appeal as it was drought tolerant and could still produce a decent yield when maize would fail.

These early lines were tall and late flowering. Shorter plants were selected to reduce lodging and enable mechanical harvesting, while earlier flowering plants were able to produce grain in the temperate climate of the U.S. Seed color was also an important trait in early selections. Standard Yellow Milo (SYM) is thought to have originated from the early lines but with mutations that cause it to flower earlier. A mutation resulting in shorter plants was found in SYM resulting in Dwarf Yellow Milo

(DYM). A second mutation resulting in even shorter plants was found in DYM resulting in Double Dwarf Yellow Milo (DDYM). Meanwhile, a mutation for shorter height was also found in kafirs resulting in Texas Blackhull Kafir [6,7]. Hegari was introduced in the early 1900s, after the milos and kafirs.

Originally sorghum was grown as an inbred crop. The USDA released many of the early sorghum lines that were commonly grown in the southern plains. Gains in yield were limited during this era. From the 1940s hybrid vigor was greatly increasing yields in maize. Sorghum breeders were interested in breeding hybrids as well. However, in maize, the female and male flowers are physically separate on the same plant and so it is easy to produce a female plant by simply removing the male flower. But in sorghum the flowers are both in the same spikelet, so producing a female plant is more difficult and costly. This problem was overcome through cytoplasmic male sterility. This system makes use of cytoplasmically inherited genes, i.e. genes in the mitochondria or chloroplast, which cause male sterility but do not affect female fertility. Nuclear genes can be used to restore fertility and so produce seed in a field of a single hybrid. In the commonly used scheme for hybrid breeding, the male sterile lines (A-lines) are maintained by B-lines that are identical to the A-line except that they have normal cytoplasm. To produce the hybrid the A-line is crossed with a restorer (R-line) that restores male fertility. Once this system was developed, hybrid sorghum quickly took over commercial production for both grain and forage sorghum [4].

The limited number of introductions to the U.S. resulted in a strong bottleneck. The USDA does maintain a large, diverse collection of lines and landraces; however,

many of these are tall and late flowering. The Sorghum Conversion Program began in the early 1960s to introduce diversity from tropical landraces. To make the germplasm more accessible to breeders a breeding scheme was designed to introduce genes for short height and early flowering into the landraces. These landraces were crossed with an inbred line from the U.S. that was fixed for short height and early flowering, namely BTx406. The lines were then selected for short height and early flowering and repeatedly backcrossed to the exotic parent [8,9]. Many of these converted lines have subsequently proven important in commercial breeding programs [4].

STEM GROWTH IN PLANTS

Plants grow through both cell division and expansion. Cell division occurs in the meristematic tissues, which is a group of actively dividing, pluripotent cells. There are two principal meristems in plants: the shoot apical meristem (SAM) and the root apical meristem (RAM), located at the tips of shoots and roots, respectively. There are also intercalary meristems, which are found between differentiated tissues, e.g. at the base of each internode in grasses.

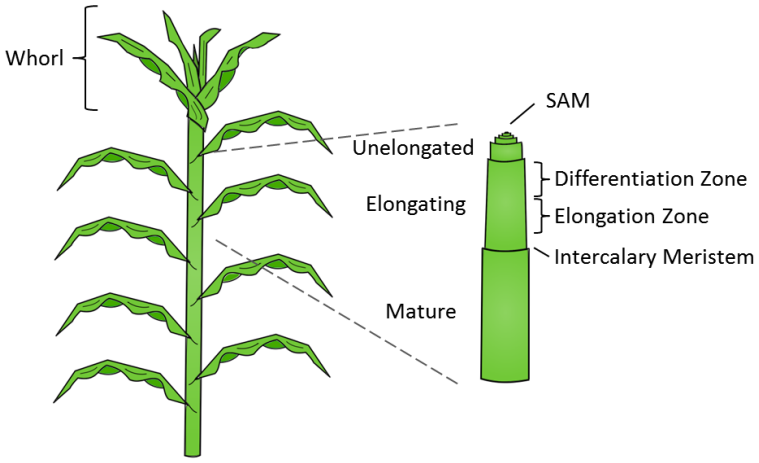


Fig 2. Diagram of the growing sorghum stem. The growing sorghum stem elongates mostly one internode at a time with the leaf sheath protecting the growing internode. As a result, the growing stem is hidden in the leaves (whorl) towards the top of the plant. The enlargement shows the top of the stem that has been stripped of leaves. The unelongated internodes are at the apex of the stem. Below them is the elongating internode and farther down the mature internodes (all labeled). The location of the intercalary meristem, elongation zone, and differentiation zone are labeled on the elongating internode.

As described above, in grasses, the stem is divided into internodes that are divided by nodes and surrounded by leaf sheaths. As each node and internode is produced, an intercalary meristem is established at the base of the internode from which the internode cells will be generated. However, in young plants, the internodes undergo limited elongation. At this stage, the grass plant is mostly leaves. As the leaves are established first throughout plant growth, a cluster of leaves called the whorl is formed at the top of the plant (Fig 2). The whorl surrounds and so protects the SAM and the unelongated internodes at the apex of the stem. When it comes time for the stem to

grow, the internodes elongate from the base with the leaf sheath protecting the more delicate, growing stem. The elongation is caused by cell division in the intercalary meristem followed by some of the daughter cells being pushed out of the meristem and elongating. When the cells reach their final size the cell wall is strengthened and elongation stops. Once the appropriate number of cells is produced, the intercalary meristem becomes dormant. As the elongation of one internode slows, the internode above it begins to elongate more rapidly. Thus there is one internode where most elongation is happening at any one time and internode growth moves sequentially up the stem (Fig 2). The peduncle elongates last and pushes the panicle above the leaves shortly before flowering.

Elongation of Plant Cells

Plant cells are surrounded by both a plasma membrane and a cell wall. The space outside of the plasma membrane, including the cell wall, which can transport solutes, is called the apoplast. The plant cell wall is originally laid down as the primary cell wall which consists of cellulose, hemicellulose, pectins, and structural proteins [10]. The cell walls of grasses differ from that of dicots primarily by different types of hemicelluloses and the concentration of pectins [11]. Cellulose is composed of glucans that are organized into microfibrils that give the wall structure. The orientation of the microfibrils determines the direction of expansion when the cell is growing. The orientation and location for deposition of microfibrils is, in turn, determined by the organization of microtubules in the nearby cytoplasm [12].

Cell expansion is accomplished through the action of turgor pressure exerting force on the cell wall and by modifying cell wall extensibility. Water moves into cells that are growing due to their low water potential compared to the apoplast. During long term irreversible growth, the cell wall is loosened reducing turgor pressure within the cell. The reduction in turgor pressure causes water absorption and an increase in cell volume. Growth is caused by these two simultaneous processes [13]. When the cell wall extensibility decreases, eliminating the difference in turgor pressure, growth stops.

Cell wall loosening has multiple causes. According to the acid growth hypothesis, cell wall loosening is due, at least in part, to a decrease in pH from H^+ pumps in the plasma membrane that increase the concentration of protons between the membrane and wall [14,15]. Proteins, including expansins and xyloglucan endotransglucosylase/hydrolase (XTH), aid in the loosening of the cell walls [10]. Expansins, XTH, and H^+ pumps are often the downstream targets of growth regulators such as the phytohormones auxin, gibberellin, and brassinosteroid [16] (see below). After the cell reaches its final size, it may produce a thicker secondary cell wall internally to the primary wall. The secondary cell wall is composed of cellulose and hemicellulose, like the primary wall, but includes the phenolic lignin which provides extra strength [17].

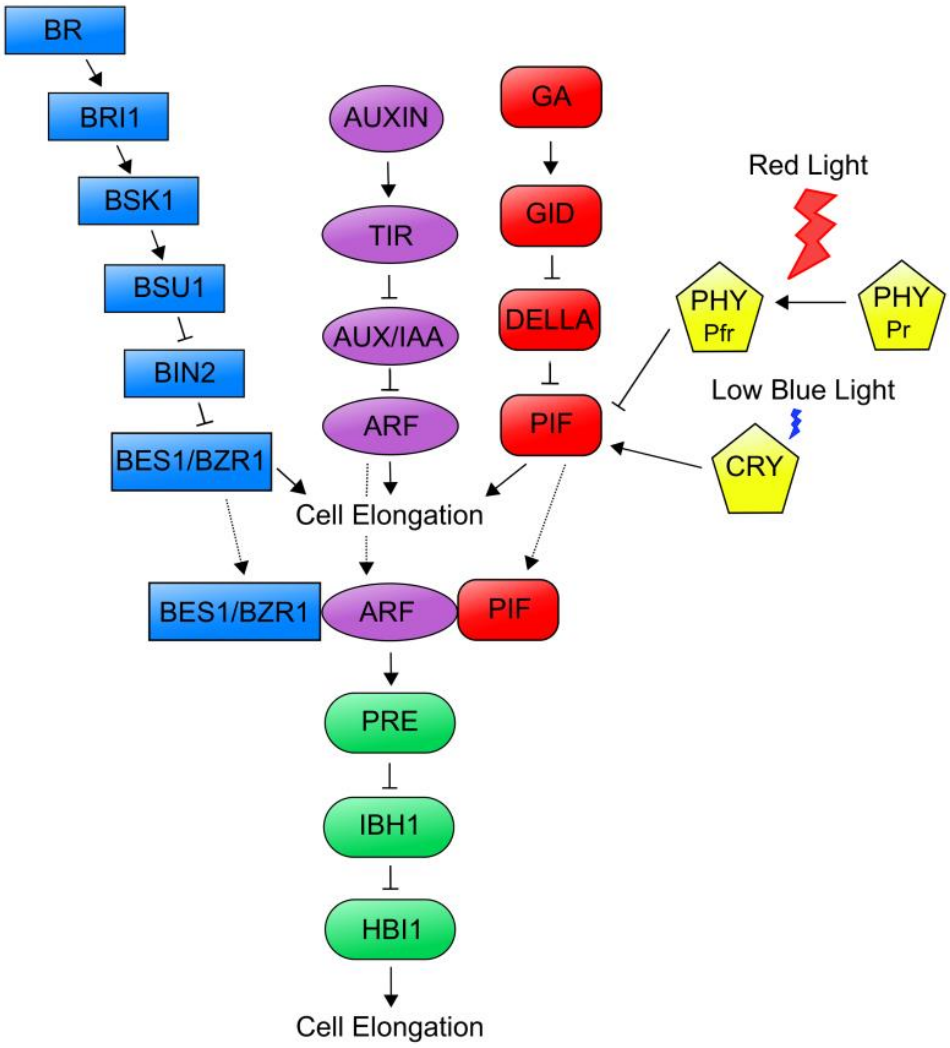


Fig 3. Overview of the regulation of cell elongation. The canonical signaling pathways for auxin, GA, and BR are shown. Phytochrome and cryptochrome sense the light environment and incorporate it into the overall pathway. These signals feed into the transcription factors ARF, BZR, and PIF, whose targets are partially overlapping. In addition, cell elongation is regulated by the triple HLH module. This module includes PRE1, which is downstream of ARF, BZR, and PIF, as well as IBH1 and HBI1.

REGULATION OF GROWTH

As plants are sessile, they have very flexible growth forms to cope with their environments. This flexibility is regulated by hormones. While plants have many different hormones, those that have been shown to have the greatest effect on growth are auxin, gibberellin, and brassinosteriod, with ethylene and cytokinin also contributing to growth regulation. Indeed, mutations in genes involved in these hormones' metabolism, signaling, or transport have been found to underlie dwarf or semi-dwarf lines that are important in breeding. All of these hormones have multiple functions that vary with cell and tissue type and concentration. While the core signaling components have been discovered in recent years for many hormones (summarized in Fig 3), questions remain about how the hormones achieve all their various functions. Additional or alternative signaling components possibly exist that have yet to be discovered or described. It should also be noted, that most work on the regulation of growth has been done using the dicot Arabidopsis, which has a rosette habit and so the roots or hypocotyl, referring to the seedling stem, are generally used to study growth.

Auxin

Auxin is an important hormone that has been implicated in many different aspects of development, including growth of stems and roots, phototropism, apical dominance, phyllotaxy, vascular differentiation, and lateral root formation. The biologically active form of auxin is indole-3-acetic acid (IAA) which is produced in all tissues in small amounts, but primarily in the meristems. It is made through the

tryptamine pathway (TAM) which involves the YUCCA enzymes and the indole-3-pyruvic acid (IPA) pathway via Trp aminotransferase (TAA) [18].

The current model for auxin signaling is that in the absence of auxin, the AUX/IAA family of proteins acts as repressors of the AUXIN RESPONSE FACTORS (ARFs) class of transcription factors with the aid of the corepressor TOPLESS (TPL) [19,20]. When auxin enters the cell, it interacts with TRANSPORT INHIBITOR RESISTANT1/AUXIN F-BOX BINDING (TIR/AFB) which is a component of the SCF ubiquitin E3 ligase. Auxin functions as a molecular glue for SCF^{TIR} and the AUX/IAA repressors which are subsequently ubiquinated [21–23]. The ubiquinated AUX/IAA proteins are then targeted for degradation via the 26S proteasome. This releases the ARF proteins to function as transcription factors.

However, not all auxin responses are thought to be a result of this signal cascade, including the increase in pH around the cell wall from the H^+ pumps that is a key component of acid growth [24]. Another protein that has been shown to bind auxin is AUXIN BINDING PROTEIN 1 (ABP1) which has been found to activate the H^+ pumps in the presence of auxin [25]. ABP1 has also been shown to be involved in auxins ability to reorganize microtubules from a transverse to longitudinal arrangement allowing for cell expansion [26]. However, a recent paper has called these findings into question as the researchers' two null ABP1 mutants did not have any noticeable difference in phenotype from wildtype [27].

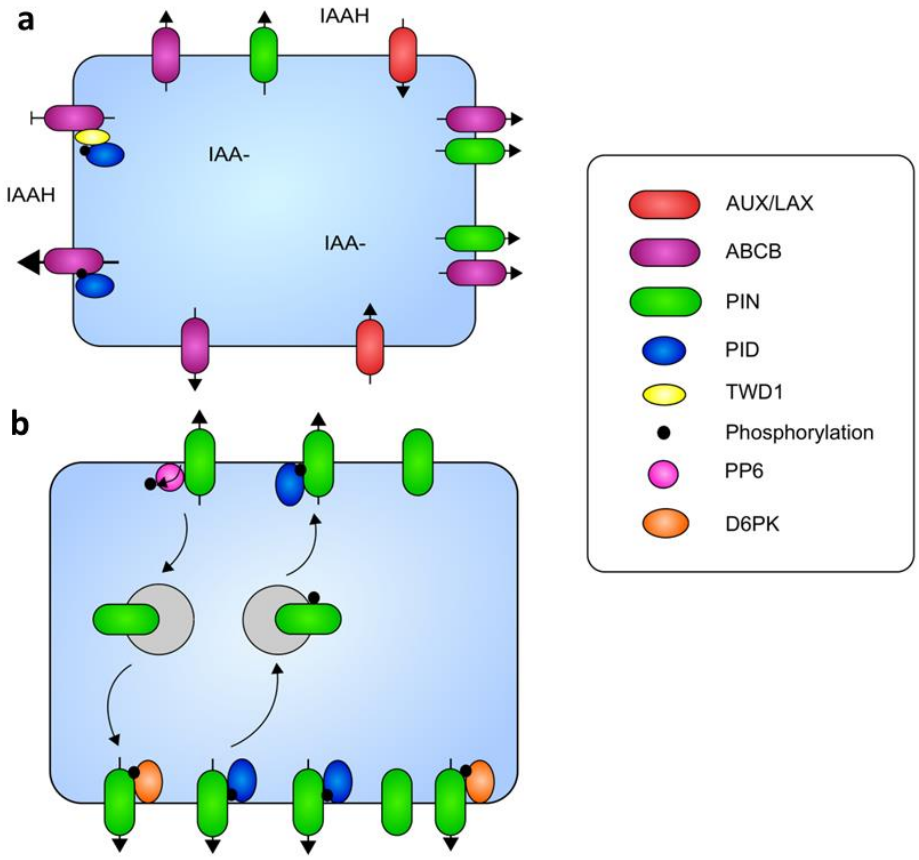


Fig 4. Cell to cell transport of auxin. (**a**) Due to the H^+ pumps, the apoplast has a lower pH than the cytoplasm causing auxin to be protonated (IAAH) in the apoplast. IAAH can diffuse through the plasma membrane. Auxin can also be transported across the membrane via AUX/LAX influx transporters. In the cell, auxin is ionized to IAA- and so cannot diffuse through the membrane; therefore, the primary control of polar auxin transport is via the efflux transporters, especially the PINs. PINs are distributed in a polar manner; whereas, the other efflux transporters, ABCBs, are primarily nonpolar. (**b**) The distribution of PINs on the plasma membrane is highly regulated. PINs are initially nonpolar in distribution but are constitutively recycled between the plasma membrane and the endosomes. Phosphorylation and dephosphorylation by PID and PP6, respectively, control the side of the cell that PIN is incorporated into, with phosphorylation favoring the apical side and dephosphorylation favoring the basal side of the cell. Additionally, phosphorylation of PIN by PID and/or D6PK activates it. PID also phosphorylates ABCB1 to activate it when TWD1 is not present. When TWD1 is present, it prevents the phosphorylation of ABCB1 thus inhibiting ABCB1 auxin transport (**a**).

Auxin is transported throughout the plant via the vascular system, diffusion, and cell-to-cell transporters. The H^+ pumps maintain the pH of the apoplast at ~5.5. As the pKa of auxin is 4.75, this results in a portion of auxin being in the protonated form which can easily diffuse through the plasma membrane. Once in the neutral cytoplasm auxin is deprotonated and so cannot diffuse through the membrane. In addition to diffusion, auxin can be pumped into the cell via the AUXIN RESISTANT1/LIKE AUX1 (AUX1/LAX) symporters [28]. PIN-FORMED (PIN) and P-glycoprotein/ATP-binding cassette subfamiliy B (ABCB) transporters [29,30] are involved in the export of auxin from the cell. PIN transporters have a polar distribution, thus contributing to the directional flow of auxin [31]. On the other hand, the ABCB transporters are generally randomly distributed and thus primarily influence the amount and distribution of auxin [32]. However, ABCB stabilizes the PIN proteins and, when associated with PIN, ABCB transporters may have a polar distribution [33,34] (Fig 4A).

As the proper polar transport of auxin is so important, the proteins involved in transport are carefully regulated. PIN has been shown to be regulated by several AGC VIII protein kinases, including PINOID (PID), WAG1 and 2, and four D6 PROTEIN KINASEs (D6PKs). AGC kinases are named after the cAMP-dependent protein kinase A, cGMP-dependent protein kinase G, and phospholipid-dependent protein kinase C which, in animals, are involved in the signaling cascades of secondary messengers, such as cAMP, cGMP, Ca2+, and phospholipids [35–37]. PID and WAG1 and 2 all phosphorylate PIN, thereby activating PIN and regulating the continuous recycling of PIN proteins at the plasma membrane [38,39]. While PID induces a switch in PIN

distribution from the basal side of the cell to the apical side, dephosphorylation of PIN by the PP6 phosphatase holoenzyme results in the opposite distribution [40]. The D6PKs also phosphorylate and activate PIN, but the residues phosphorylated and the function differs from that of PID and WAG1 and 2 [41,42]. Furthermore, PID enhances the efflux function of ABCB when it is just the two proteins. However, in the presence of a third protein, TWISTED DWARF1 (TWD1), it inhibits ABCB function [43] (summarized in Figure 4).

Gibberellin

Another major hormone involved in plant growth is gibberellin (GA). GA regulates plant height, seed germination, and pollen development. Bioactive GAs are synthesized through several steps; the last two of which are oxidations catalyzed by GA20ox and GA3ox. Additionally, bioactive GAs are deactivated by GA2ox [44]. The GA signaling pathway is similar to that of auxin with GA-INSENSITIVE DWARF1 (GID1) serving as a receptor for GA [45]. Once it binds GA, it can also bind the DELLA proteins that are subsequently ubiquitinated and targeted for degradation [46,47]. The DELLA family of proteins is named for a conserved amino acid sequence that is in the domain that binds GID1 and includes GA-INSENSITIVE (GAI) in Arabidopsis and SLENDOR RICE 1 (SLR1) in rice [48,49]. DELLAs are repressors of the various PHTYOCHROME INTERACTING FACTORS (PIF) transcription factors [50,51].

Brassinosteroid

A third hormone involved in plant growth is brassinosteroid (BR). BR regulates cell division and elongation, leaf senescence, and stress responses. BR binds the membrane bound protein, BR INSENSITIVE 1 (BRI1) which is a leucine-rich repeat (LRR) receptor kinase [52]. BRI1 is associated with the kinase BRI1-ASSOCIATED RECEPTOR KINASE1 (BAK1). When in the presence of BR these two kinases phosphorylate each other. When activated, BRI1 phosphorylates two kinases, BR-SIGNALING KINASE1 (BSK1) and CONSTITUTIVE DIFFERENTIAL GROWTH1 (CDG1) which in turn phosphorylate BRI1-SUPPRESSOR1 (BSU1) phosphatase [53,54]. BSU1 dephosphorylates and inactivates BRASSINOSTEROID INSENSTIVE2 (BIN2) [55,56]. This prevents BIN2 from phosphorylating the transcription factors, BRASSINAZOLE RESISTANT1 (BZR1) and BRI1-EMS-SUPPRESSOR1 (BES1) [57–59]. Phosphorylated BZR1/BES1 is retained in the cytoplasm by members of the 14-3-3 family [60]. In the absence of active BIN2, BZR1/BES1 are dephosphorylated by PROTEIN PHOSPHATASE 2A (PP2A) and can enter the nucleus and regulate gene transcription [61–63].

Ethylene

Ethylene is a gaseous hormone involved in fruit ripening and seedling development in addition to repressing hypocotyl stem growth. The signaling pathway of ethylene starts with receptors that are in the ER membrane. There are several related receptors that, when not bound to ethylene, activate the kinase CTR1 that, in turn,

phosphorylates EIN2. EIN2 is a unique protein that is also bound to the membrane except the C-terminal end which when phosphorylated remains attached and so the ethylene response is not active. In the presence of ethylene, the receptors do not phosphorylate CTR1. Thus EIN2 is not phosphorylated and the C-terminal end is cleaved and moves to the nucleus [64]. Once in the nucleus the various ethylene responsive transcription factors (ERFs) are activated [65].

Cytokinin

Cytokinin is involved in regulating cell division and has been implicated in meristem development and maintenance, vasculature development, lateral root formation, and nodule formation. The core cytokinin signaling pathway involves several hybrid histidine kinases that function as receptors. The phosphorelay system continues through several phosphotransfer proteins that when phosphorylated enter the nucleus to activate the cytokinin responsive transcription factors [66]. Interestingly, cytokinin often interacts with auxin in regulating its various functions.

Light Sensing in Growth Regulation

Another important influence on plant height is the intensity and quality of light. The shade avoidance syndrome (SAS) is where the plant increases stem growth to seek out more intense or higher quality light sources resulting in a longer stem with a smaller diameter. This response depends on the light sensing proteins phytochromes and

cryptochromes. However, plants with a strong SAS response are more likely to lodge when planted at high densities so breeders have selected for a weaker response.

While the five different phytochromes are involved in many light sensing functions, the best elucidated pathway is PhyB detection of the red light to far red light ratio that the plant uses to sense shading from nearby plants. PhyB is made as the Pr form that is found in the cytosol and absorbs red light. When the Pr form of PhyB absorbs red light, PhyB switches to its Pfr form that absorbs far-red light. The Pfr form can enter the nucleus where it interacts with PIFs, stimulating the proteins' degradation [67]. In addition to its light sensing function, phytochrome has recently been shown to function as a thermosensor with warmer temperatures favoring the inactive Pr form [68,69].

Two cryptochromes sense blue light and are involved in the SAS, supplementing phytochrome. In the shade, or low blue light, cry interacts with PIF4 and 5 enabling each of these proteins to bind DNA [70]. On the other hand, in high light and higher temperatures cry represses PIF4 [71]. In addition to PIFs, cryptochrome and phytochrome induce SAS through auxin, with BR also required for full response in low blue light [72,73].

Blue light is also detected by the phototropins. Phot1 and 2 are AGC VIII kinases that have two LOV (light, oxygen, voltage) domains in their N-terminal end. Both phototropins are involved in phototropism, chloroplast movement, and leaf expansion. In the absence of blue light phot1 is dephosphorylated and localized to the plasma membrane. Upon blue light exposure, phot1 is thought to autophosphorylate and

then interact with and dephosphorylate NON-PHOTOTROPIC HYPOCOTYL 3 (NPH3) [74]. NPH3 is part of the CULLIN RING E3 ligase complex and interaction with phot1 results in the ubiquitation and internalization or degradation of phot1 [75]. Phot1 also interacts with and phosphorylates ABCB19 thus deactivating ABCB19 in the presence of blue light [76].

Integration of Various Signals in Cell Elongation

Final plant size is determined through both the number of cells and the elongation of those cells. Cell elongation is determined through integration of plant metabolism (sucrose, nitrogen status), plant hormones and light signals along with temperature, aging, and pathogen stress. A major intersection of these signals is at the level of transcription factors involved in auxin, GA, and BR signaling, namely ARF6, PIF4, and BZR1/BES1. These transcription factors have both overlapping and specific targets [77]. The repressor protein DELLA has also been shown to regulate ARF6 and BZR1, in addition to the PIFs [78,79]. Downstream of the transcription factors is a series of helix-loop-helix and basic helix-loop-helix factors. PACLOBUTRAZOLE RESISTANT (PRE1) is a positive regulator of growth while IL1 BINDING bHLH PROTEIN1 (IBH1) inhibits a group of bHLH factors, including HOMOLOG OF BEE2 INTERACTING WITH IBH1 (HBI1), that positively regulates growth [80–82].

In addition to hormones and the pathways described above, plants also have many small peptides that may function as hormones and hundreds of receptor like kinases. For example, the small peptide RALF is a ligand for the receptor-like kinase

FERONIA. FERONIA is located in the plasma membrane with one end imbedded in the cell wall. RALF binding FERONIA results in the H^+ pumps being phosphorylated and inhibited thereby inhibiting cell expansion [83].

Regulation of the Cell Cycle

The regulation of the cell cycle is well conserved across eukaryotes. The cell cycle is divided into four phases, G1-S-G2-M, with gatekeepers controlling the transition between phases. As in animals, in plants these gatekeepers are cyclins and cyclin dependent kinases (CDKs). The transition from G1 to S is regulated through RETINOBLASTOMA-RELATED (RBR) and the E2F transcription factors. RBR binds E2F preventing it from binding DNA. The appropriate cyclin/CDK pair (CYCD/CDKA) phosphorylates RBR causing it to dissociate from E2F. Cytokinin and auxin are thought to induce the expression of at least some of the CYCDs [84]. The transition from G2 to M is regulated, in part, by MYB transcription factors [85]. Interestingly, plant cells more often undergo endoreduplication than animal cells. There is some evidence that cytokinin is involved in regulating whether a cell goes through mitosis or through the endocycle.

Regulation of the SAM

The SAM is divided into four sections: the central zone (CZ), the organizing center (OC), the rib zone (RZ), and the peripheral zone (PZ). The CZ is located at the apex of the stem and is where the slowly dividing stem cells are located. The PZ

surrounds the CZ and consists of more rapidly dividing cells. The leaf primordia form from the PZ and flank the SAM. The OC is below the CZ and the RZ, which consists of flattened cells that are starting to differentiate, is below that. The stem cell population is maintained through a negative feedback loop of WUSCHEL (WUS), CLAVATA3 (CLV3), and CLV1/2 [86]. WUS is produced in the OC and moves to the CZ. There it promotes stem cell identity, as well as inducing the expression of CLV3. CLV3 is the ligand for the leucine-rich repeat (LRR) receptor like kinase CLV1 and LRR receptor like CLV2 which interacts with the pseudokinase CORYNE and RECEPTOR-LIKE PROTEIN KINASE 2 (RPK2)/TOADSTOOL2 (TOAD2). CLV1 is expressed in the OC and the nearest layers of the CZ. CLV3 binding represses the expression of WUS [87,88].

WUS also represses the expression of ARABIDOPSIS RESPONSE REGULATOR (ARR) 7 and 15. ARR7/15, in turn, inhibits cytokinin signal transduction; however, cytokinin positively regulates ARR7/15 and WUS. ARR7/15 are also repressed by auxin via the ARF, MONOPTEROS (MP) [89]. KNOTTED1 HOMEOBOX (KNOX) represses differentiation throughout the SAM. KNOX increases cytokinin biosynthesis and so enhances WUS expression. On the other hand, KNOX represses synthesis and induces degradation of GA, which promotes differentiation [90,91]. Primordial organs, such as leaves, form to the outside of the PZ. The establishment of organ primordia is regulated primarily by gradients in auxin concentration with cytokinin and the cytokinin inhibitor ARABIDOPSIS HISTIDINE PHOSPHOTRANSFER PROTEIN (AHP5) adding robustness to the system [92].

Between the SAM and the organ primordia a boundary region is established by BR and auxin [93,94].

GENETICS OF HEIGHT IN CEREAL CROPS

In the 1960s, shorter wheat and rice cultivars were developed which enabled the plants to support higher grain yields without lodging. These cultivars, along with modern agriculture practices, were used to increase the food supply in several developing countries, including Mexico and India. This was the so-called "Green Revolution." For his work in this project, Norman Borlaug was awarded the Noble Prize and credited with saving a billion lives.

Subsequent studies in wheat and rice have shown that the "Green Revolution" genes are involved in gibberellin signaling and biosynthesis. As plant hormones can have pleiotropic effects, it was important that the genes breeders utilized to control height would not also negatively affect yield. The wheat Rht locus is the wheat version of the DELLA protein, GAI [95]. Interestingly, the two Rht mutants (Rht-B1b and Rht-D1b which are in the B and D genomes, respectively, of hexaploid wheat) have deletions in the N-terminal region that prevent the mutants from binding GID and so are constitutive repressors of the PIFs hence the semi-dwarf phenotype [95]. In rice, the dwarfing phenotype is due to the GA biosynthesis protein GA20ox2. The mutation did not have a negative effect on yield because GA20ox1 is principally expressed in the floral tissue, while GA20ox2 is expressed in the stem and leaves [96–98].

Additional examples of important mutations in height regulation include the "uzu" gene in barley and two different groups of ERFs in rice. In barley, the "uzu" gene has been used to produce semidwarfs. It is a weak allele of the BR receptor, BRI1 [99]. While rice is grown in partially flooded environments, complete flooding can severely reduce the yield. There are two different mechanisms of coping with flooding stress: in deepwater rice the plant grows extra tall to get above the water level and submergence tolerant rice which stops growth to conserve resources. Both of these involve ethylene and GA. In deepwater rice, the ERFs SNORKEL1 and 2, enhance growth through increasing the levels of bioactive GA [100]. In submergence tolerant rice, the ERF Sub1A upregulates the DELLA homolog SLR1 and the closely related, though missing the DELLA domain, SLRL1 which repress GA signaling [101,102].

Genetics of Height in Sorghum

Sorghum exhibits a great range of height from <1 meter in height to 3-4 meters. These differences are due to the length of the internodes, the number of internodes, which is strongly influenced by flowering time, and the rate of phytomer production. Fig 5 shows some of the variation in height and internode length in sorghum. The height of the plant typically grown varies with the end product. Grain sorghum is generally shorter, so it can produce a large head without lodging. On the other hand, sorghum grown for stem sugar, forage, or biomass is generally taller.

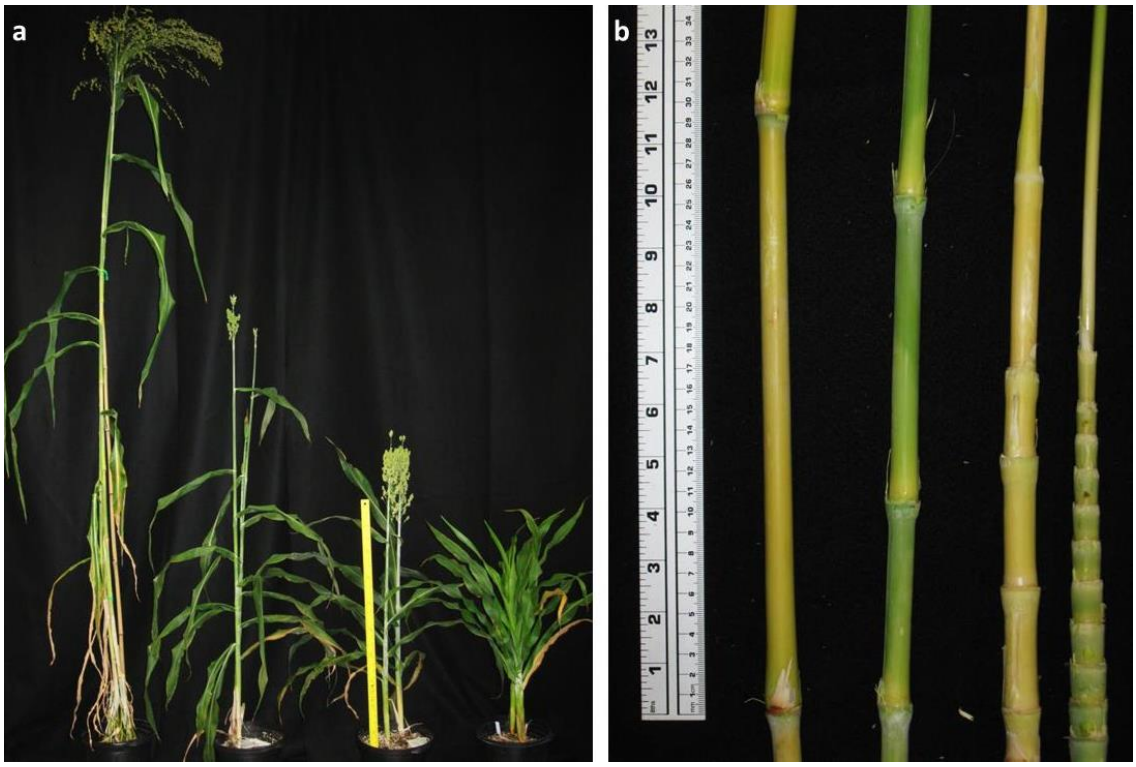


Fig 5. Height (a) and internode length (b) variation in sorghum. (**a**) Photograph of a representative plant from each of the four main height classes (from left to right): 1-dwarf (Standard Broomcorn), 2-dwarf (Texas Blackhull Kafir), 3-dwarf (BTx623), and 4-dwarf (BTx642). 1-dwarf plants have the dominant allele that increases height at three of the four *Dw* genes, while 2-dwarf are dominant at two of the four *Dw* loci, and so forth. Yellow meter stick shown for reference. (**b**) Photograph of stem internodes from the same four plants (from left to right) that have had the leaves removed.

In the 1930s through 1950s, several studies were conducted to determine the genes that control height in sorghum. Sieglinger [103] used broomcorns for his studies and based on various crosses determined there were two genes that affect height. In 1954, Quinby and Karper [104] used many different varieties of sorghum to determine that there were four genes that control height through the length of the internodes. They named these genes *Dw1-Dw4*. For their work they measured plants from the base of the

plant to the flag leaf to determine height. *Dw1* and *Dw2* were the two genes that they had found segregating in the milos. The recessive allele of *Dw3* was found in the kafirs. The dominant allele at *Dw4* was found only in the broomcorns.

At each locus, the dominant allele increases the length of the plants and each gene displays incomplete dominance. The genes interact in an additive fashion, though the more genes that were dominant the less affect an additional gene that is dominant has on the total height. While the genes were originally described as primarily affecting height [104], pleiotropic effects have been described for *Dw2* and *Dw3*. *Dw2* has been shown to also affect panicle length, yield, seed weight and leaf area [105,106]. *Dw3* has been shown to affect seeds per panicle and seed weight, tiller number, panicle size, and leaf angle [106–108].

Of the four *Dw* genes, only one, *Dw3*, a gene located at 59.8 Mbp on chromosome 7, had been cloned prior to the research described in this dissertation [109]. The gene was determined to be a homolog of the Arabidopsis ABCB1 auxin efflux transporter. This work was based on the maize mutant *brachytic2* (*br2*) which has severely shortened lower internodes with the upper internodes being less affected by the mutation [109]. This is a more severe phenotype than what is found in Arabidopsis. Since Arabidopsis is a dicot, it does not have the node/internode structure of the grasses, like maize and sorghum, and so does not have intercalary meristems in growing internodes. In Arabidopsis, ABCB1 and ABCB19 have very similar functions where they export auxin from the SAM and into the vascular system. In maize, in addition to both proteins exporting auxin from the SAM, ABCB1 is involved in export from the

intercalary meristem but ABCB19 is not [109,110]. Interestingly, the sorghum *Dw3* mutant is less severe than the maize mutant with the internodes affected fairly evenly throughout the stem. Also, *Dw3* is an unstable mutation, with a reversion rate of ~1:600 due to the large 882 bp tandem repeat insertion in the final exon of the mutant form of the gene. In revertants the repeat was lost due to uneven crossing over in the region [109].

Dw1 was mapped to ~57 Mbp on chromosome 9 in multiple populations [111–114]. Morris et al. [112] suggested that *Dw1* is a GA2ox. However, a more recent study [115] found that the mutants in the gibberellin signaling and biosynthesis pathway were bent in addition to being short. *Dw1* recessive plants do not exhibit bending. Furthermore, Ordonio et al. [115] sequenced the gene suggested by Morris et al. [112] in *Dw1* dominant and recessive lines and did not find any sequence difference. Additionally, GA2ox deactivates the bioactive gibberellins. Thus a null or knockdown mutant would result in a taller plant as seen in the pea SLENDER mutants as opposed to the shorter mutant seen in sorghum [116]. Meanwhile an overexpression mutant would be dominant instead of recessive. Thus the GA2ox is not a good candidate for *Dw1*.

Dw2 has been mapped to chromosome 6 at ~42 Mbp. Morris et al [112] suggested that a histone deacetylase underlies *Dw2*. Interestingly, *Ma1*, a mutant that affects the time to flowering that has been important in sorghum adaptation to temperate environments, is at ~40 Mbp on chromosome 6. Both of these loci were selected for in the Sorghum Conversion Program and so there is limited diversity of chromosome 6 for the conversion lines. Additionally, Higgins et al. [114] found that the location of the

most significantly associated SNPs varied from ~42 to ~44 Mbp in the various populations they mapped in. They suggested that the ~42Mbp location was due to *Ma1* causing synthetic associations between it and *Dw2*, which they thought was most likely located at ~44 Mbp [114].

Dw4 has not been conclusively mapped. Morris et al. [112] suggest that a QTL found on the opposite end of chromosome 6 to *Dw2* at ~6.6 Mbp is *Dw4*. On the other hand, Li et al. [117] found a QTL on chromosome 4 at ~66.7 Mbp that they suggest is *Dw4*.

Recently, Li et al. [117] found a QTL that affects height that does not match up with any of the known *Dw* loci. This locus is on chromosome 7 at ~54 Mbp, which is close to *Dw3*. The authors speculated that because of the location, the alleles at this locus and *Dw3* were dragged along with each other during breeding. Also, they found that this new locus affected all of the expanded internodes and peduncle whereas *Dw3* does not affect the peduncle length [117].

OVERVIEW OF THIS STUDY

The goal of this project was to increase our understanding of the regulation of height in sorghum. Height has long been an important trait for sorghum improvement with shorter plants being favored for grain production to reduce lodging and taller plants favored for sugar and biomass production. In this study QTL mapping and map-based cloning were used to identify *Dw1* and *Dw2*. An attempt to QTL map *Dw4* was made but a location for *Dw4* was not identified.

Dw1 was QTL mapped in an F_2 population derived from Hegari (*Dw1dw2Dw3dw4*) and 80M (*dw1dw2Dw3dw4*), thus the population should be segregating for only *Dw1*. However, mapping revealed that the population was actually segregating for *Dw1*, *Dw2,* a QTL on chromosome 7 (*Dw07_56*), and a fourth QTL on chromosome 1 (*Dw01_62*). Through Multiple QTL Mapping (MQM) *Dw1* was shown to interact with *Dw07_56*. *Dw1* was fine mapped in F_3 and F_4 Heterogeneous Inbred Families (HIFs) narrowing the region containing *Dw1* to 33 kb. All seven of the genes in the region were sequenced in the two parents and in Standard Yellow Milo (SYM) and Dwarf Yellow Milo (DYM), dominant and recessive at *Dw1* respectively but otherwise isogenic. The only gene in the region with a polymorphism in the coding region between the parents or between SYM and DYM is Sobic.009G229800, a highly conserved gene of unknown function.

Dw2 was initially mapped in a RIL population derived from BTx623 (*dw1Dw2dw3dw4*) and IS3620c (*dw1dw2Dw3dw4*), and so the population should be segregating for *Dw2* and *Dw3*. In addition to total length, the length of each internode was mapped, which revealed some interesting trends across development. *Dw2* and *Dw3* had similar additive effects for the first internode below the peduncle. The effect of *Dw2* decreased slightly down the stem for the first five internodes, after which it was not significant. *Dw3* had twice the additive effect of *Dw2* for the next few internodes (internodes 2-5) and then declined. *Dw2* was fine mapped in a RIL population derived from BTx642 and Tx7000, which is segregating for *Dw2* only. *Dw2* was determined to

encode a kinase in the AGCVIII family whose closest homolog in Arabidopsis is KCBP INTERACTING PROTEIN KINASE (KIPK).

Several crosses were made in an attempt to determine the location of *Dw4*. According to a previous study, the dominant allele of *Dw4* increases height and is found only in broomcorns. Additionally, the broomcorns were reported to be recessive at *dw3*. The broomcorns used in this study were Standard Broomcorn (SB) and Acme Broomcorn (AB). SB was crossed with SC170, BTx623, and Hegari, while AB was crossed with BTx623. No QTL corresponding to *Dw4* was found in the four populations. The AB x BTx623 population was segregating for a QTL on chromosome 4 that was previously suggested to be *Dw4*; however, the AB allele decreased length though AB should have the allele that increases length. Additionally, both broomcorns were found to be dominant at *Dw3* not recessive as previously described.

This study increases our knowledge of the genetic control of height variation in sorghum. First, the genes the underlie *Dw1* and *Dw2* were identified. Second, additional QTL were identified, one of which interacts with the *Dw1*. In addition, several additional QTL with small effects were identified. The results of this study demonstrate that the genetic basis of variation in height in sorghum is complex and modified by alleles of *Dw1, Dw2, Dw3* and numerous other QTL.

CHAPTER II

IDENTIFICATION OF *Dw1*, A REGULATOR OF SORGHUM STEM INTERNODE LENGTH*

INTRODUCTION

Sorghum is the fifth most widely cultivated cereal crop worldwide. This C4 grass is grown for grain, feed, forage, sugar, and biofuels. Sorghum diverged from a common ancestor with maize ~12 MYA and rice ~50 MYA [2]. It is native to Africa and parts of India and Australia with most African landraces growing to 3-4 meters in height before harvest. When grown in the U.S., many sorghum accessions from Africa produce tall, late flowering plants. However, after its initial introduction to the U.S., breeders found naturally occurring shorter genotypes that were subsequently used to breed short grain sorghum varieties to reduce stalk lodging. Sorghum genotypes with longer stems are grown for forage, sugar, and biomass to increase yield. Energy sorghum hybrids are 3-4 meters in height with long internodes and biomass yield ranging from 15-40 Mg/ha depending on genotype and environment [118–120]. Stem biomass of a first generation energy sorghum hybrid accounted for ~80% of harvested shoot biomass [1]. Therefore, a more complete understanding of the genetic and biochemical basis of stem growth could identify ways to increase the stem biomass yield of bioenergy sorghum.

Plant height is affected by the length of each internode, the rate of internode production, and the duration of vegetative growth. The latter influences height because production of internodes stops at floral induction even though internode elongation continues until anthesis. In the 1950s, Quinby and Karper [104] identified four loci, *Dw1-Dw4*, that control height by modifying internode length. Recessive alleles at the four loci reduce internode length [104]. Pleiotropic effects of *Dw2* and *Dw3* have been reported and include panicle length, seed weight, and leaf area for the former [105,106] and seed weight, panicle size, tiller number, and leaf angle for the latter [106,108,121]. However, pleiotropic effects have not been described for *Dw1* or *Dw4*. Additionally, QTL for height, including *Dw3* and a QTL on chromosome 9, have been found to co-localize with QTL for stem and total biomass [122].

The gene corresponding to *Dw3* was cloned by Multani et al. [109] and determined to encode an ABCB1 auxin efflux transporter. Further analysis showed that the maize homolog, *br2*, transports auxin from intercalary meristems located at the base of a stem internode into the elongating internode [110]. QTL corresponding to *Dw1* and *Dw2* have been identified, but the underlying genes are unknown. *Dw1* was mapped to the distal end of SBI-09 [111] and *Dw2* to SBI-06 adjacent to *Ma1* [9]. Recently, a QTL for stem length was identified on SBI-07 located near *Dw3* in a RIL population from a cross of Tx430 and P898012 [117].

The Green Revolution dwarfing genes in rice and wheat reduce gibberellin induced stem elongation producing semi-dwarf varieties with reduced lodging. In rice, semi-dwarf genotypes were found to encode a less active version of gibberellin 20

oxidase, an enzyme involved in GA synthesis [96]. In wheat, dwarf varieties contain alleles of a gene encoding a DELLA protein that is involved in gibberellin (GA) signaling [95]. Because of this, several researchers have suggested that *Dw1* encodes a gibberellin 2 oxidase that is located in the genomic region near SNPs associated with this height locus on SBI-09 [112–114]. However, recent work showed that gibberellin mutants in sorghum have bent stems, which are not observed in genotypes recessive for the sorghum dwarfing genes. Furthermore, there were no sequence variants in the GA2 oxidase coding region located on SBI-09 near *Dw1* between genotypes that were *Dw1* and *dw1* [115].

In this study, the gene corresponding to *Dw1* was map-based cloned using an F_2 population and HIFs derived from Hegari and 80M. *Dw1* encodes a protein of unknown function that is highly conserved in plants. In the process of identifying *Dw1*, a QTL that modulates internode length was identified on SBI-01 and a QTL on SBI-07 corresponding to one recently identified by Li et al. [117] was found to interact with *Dw1*.

METHODS

QTL Mapping of Stem Traits in Hegari x 80M

A map-based cloning approach was used to identify the gene corresponding to *Dw1*. A population segregating for *Dw1* was constructed by crossing Hegari, which is *Dw1dw2Dw3dw4* according to Quinby and Karper [104], and 80M (*dw1dw2Dw3dw4*) [123]. The F_1 plants were selfed and the F_2 population (n=218) was planted in April

2011 and grown in a greenhouse in long days (14 hours light, 10 hours dark), three plants per 3.8 gallon pot in soil that was a mixture of vermiculite (Sun Gro Horticulture) and Belk Clay soil (2:1) obtained from the Texas A&M University Field Station west of College Station, Texas. Osmocote Classic 13-13-13 (Scotts) was mixed into the soil and plants were subsequently fertilized every two weeks with Peters General Purpose 20-20-20 (JR Peters, Inc.). Plants were phenotyped for days to flowering, total stem fresh and dry weight, total stem length, and length and diameter of each internode at grain maturity for early flowering plants and after 190 days of growth for late flowering genotypes. The length of expanded internodes was measured for all plants in the population with the first expanded internode being labeled as number 5. DNA was extracted from leaf tissue using the FastDNA Spin Kit (MP Biomedicals). Each plant was genotyped using Digital Genotyping [124], using the enzyme FseI for digesting the genomic DNA. The Illumina GAII was used for sequencing and the reads were mapped onto the *Sorghum bicolor* genome v1.0 (Phytozome v6).

A genetic map for this population was constructed using MapMaker [125], with the Kosambi function. QTL analysis was performed in QTL Cartographer [126] using Composite Interval Mapping with a walk speed of 1.0cM and forward and backward model selection. The threshold was set using 1000 permutations and α=0.05. QTL mapping was performed with the entire population, early flowering plants only (n=85), and late flowering plants only (n=118). To look for possible gene interactions multiple-QTL analysis was used. A single QTL analysis using the EM algorithm initially identified four primary additive QTL which were used to seed model selection. The

method of Manichaikul et al. [127] was employed for model selection as implemented in R/qtl [128] for multiple-QTL analysis. Computational resources on the WSGI cluster at Texas A&M were used to calculate the penalties for main effects, heavy interactions, and light interactions. These penalties were calculated from 24,000 permutations for the average internode length to find a significance level of 5% in the context of a two-dimensional, two-genome scan.

Fine Mapping of *Dw1*

To refine the location of *Dw1*, plants were selected from early flowering lines that were segregating for *Dw1*, but fixed for the other loci controlling internode length. These plants (n=6) were selfed to create Heterogeneous Inbred Families (HIFs) [129]. For each family, the F_3 plants (n=75 for each HIF) were planted in December 2011 and grown in the greenhouse as with the F_2 population, phenotyped as described above, and genotyped using Digital Genotyping. The phenotypes were used to classify plants as dominant, heterozygous, or recessive at *Dw1*. The phenotype data were then correlated with genotype data spanning *Dw1*. The region encoding *Dw1* was further refined using F_4 HIFs derived from F_3 plants that were heterozygous at *Dw1*. The plants were planted in June 2013 and grown in the greenhouse as with the previous generations, except in Sunshine MVP soil (Sun Gro Horticulture). At grain maturity the plants were phenotyped for stem and internode length (n=78 for each HIF). The population was screened for individuals with breakpoints in the delimited *Dw1* region using two CAPS (Cleaved Amplified Polymorphic Sequence) markers, except for Family 2 which was

genotyped using Digital Genotyping because one of the CAPS markers was fixed in that family. The CAPS markers are described in Table A1. Restriction enzyme digests were performed using the manufacturer's recommended temperature for each enzyme (New England Biolabs) and incubations of at least 2 hours. All PCR amplification was done with Phusion (New England Biolabs). The breakpoints were refined using SNPs that were genotyped through Sanger sequencing using Big-Dye Terminator cycle sequencing kit v3.1 (Invitrogen) (Table A1).

Sequencing of Candidate Genes

All of the genes in the region encoding *Dw1* delimited by fine mapping were sequenced in the parental genotypes used for *Dw1* mapping as well as Standard Yellow Milo (*Dw1Dw2Dw3dw4*) and Dwarf Yellow Milo (*dw1Dw2Dw3dw4*) [104] by Sanger sequencing. The yellow milos are nearly isogenic except at *Dw1* [6,104]. The primers used to amplify and sequence genes in the delimited *Dw1* region are listed in Table A2. A polymorphism in Sobic.009G229800 that distinguished 80M and Hegari created a stop codon and truncated protein in 80M (*dw1*).

cDNA Sequencing and qRT-PCR

RNA was collected from stem tissue for cDNA sequencing and to characterize the expression of *Dw1* (Sobic.009G229800). The two parents (n=3 for each) were planted in the greenhouse in August 2013, and after 42 days of growth, stem tissue was collected from plants in the mid-morning. Plants were cut at soil level and leaves and

leaf sheaths were quickly stripped from the stem. Internodes that were in the process of elongating were located and divided into an upper portion of the internode that had stopped elongating, a mid-lower region containing cells that are in the process of elongation, and the base of the internode containing the intercalary meristem. A fully expanded internode was also harvested. The tissue was ground in liquid nitrogen and the RNA extracted using a Direct-zol RNA kit (Zymo Research) with TRI-Reagent (Molecular Research Center). The RNA was quantified on the Nanodrop spectrophotometer. RNA quality was confirmed by visualizing final samples with the BioAnalyzer (Agilent Technologies). Two technical replicates of cDNA and a no reverse transcriptase control were made using SuperScript III primed with both random hexamers and oligo (dT) at a ratio of 9:1 from 1µg of RNA.

Sobic.009G229800 cDNA from elongating stem tissue from each parental genotype was Sanger sequenced. The primers used to sequence the cDNA are listed in Table A3. Gene expression was analyzed using qRT-PCR on the 7900HT Fast Real-Time PCR System (Applied Biosystems) running SDS v2.3 software. *Dw1* was amplified in the presence of SYBR green using the following conditions: hold at 95°C for 10 mins, 40 cycles of 95°C 15 sec. and 60°C for 1 min. Primer efficiencies were determined based on a standard curve from a serial dilution of five 10-fold dilutions of PCR product for each parent. Primer specificity was checked using a dissociation curve and running PCR products on a gel. The primers used for *Dw1* amplification were: 5'-TACGCTAAAGATGGCACAAGTC-3' and 5'-TCCTTTGAACACGTCCAAGC-3'. The data was analyzed according to the comparative Ct (ΔΔCt) method [130] using the

18S ribosomal RNA to normalize the expression values and the sample from the 80M mature tissue as the calibrator. 18S ribosomal RNA reactions were performed with the TaqMan rRNA primers and probe (Applied Biosystems) and TaqMan MasterMix. Three technical replicates of qPCR were performed for each sample. The three biological replicates were averaged and the standard error of the mean calculated.

Protein Sequence Analysis

To gain insight into the function of *Dw1*, the protein sequence translated from the Hegari cDNA sequence was compared to other plants, using BLAST in Phytozome v.10 and to the NCBI database using NCBI BLAST. A sequence comparison of the protein's homologs in maize, rice, and Arabidopsis was generated in Jalview [131] using T-Coffee [132] with default settings. A phylogenetic tree of several protein homologs was constructed with MEGA6 [133] using MUSCLE [134,135] to align the sequences and Maximum Likelihood to construct the tree. Protein function and structure was examined using several web-based programs: PSIPRED-MEMSAT-SVM [136,137], PSIPRED-DISOPRED [138], PONDR [139] , and FoldIndex [140] using default settings for each program.

RESULTS

QTL Mapping of Stem Traits

The Hegari (*Dw1*) x 80M (*dw1*) F_2 population segregated for flowering time and height. Four QTL were identified that modulate the average length of internodes 5-10

(Fig 6, Table 1). A QTL corresponding to *Dw1* was identified on SBI-09 with a peak at ~56.6 Mbp on *Sorghum bicolor* genome v2 (Phytozome v10). This QTL explained ~22% of the trait variance observed. The *Dw1* allele in Hegari increased the lengths of all expanded internodes compared to plants containing the *dw1* allele present in 80M (Fig 7). A second QTL for internode length was located on SBI-06 at ~42.6 Mbp that aligned with *Dw2* [9]. A previously reported QTL for internode length was identified on SBI-01 at ~54.7 Mbp (*Dw01_54.7*) that explained ~5% of the variance [141,142]. A QTL on SBI-07 at ~55.1 Mbp (*Dw07_55.1*) that was recently described by Li et al [117] explained 19% of the variance. The QTL on SBI-07 (*Dw07_55.1*) was 3 Mbp from the *ABCB1* gene corresponding to *Dw3* (58.6 Mbp). No QTL aligned with *ABCB1* as expected because both parental genotypes are *Dw3*.

Table 1. QTL for Average Internode Length Identified in the Entire Population of Hegari x 80M F_2. For the additive effect, a positive number indicates that the 80M allele increases length while a negative number indicates the Hegari allele increases length.

QTL	Chr	Peak (cM)	Peak LOD	Peak (Mbp)	Additive	Dominance	R^2	*Dw* locus
1	1	104.2	5.53	54.7	12.5848	-5.5165	0.0503	*Dw01_54.7*
2	6	46.5	15	42.6	-22.8162	4.1926	0.1358	*Dw2*
3	7	62.4	44.37	55.1	39.2763	22.2605	0.1945	*Dw07_55.1*
4	9	112.2	21.8	56.6	-27.3763	6.4375	0.2186	*Dw1*

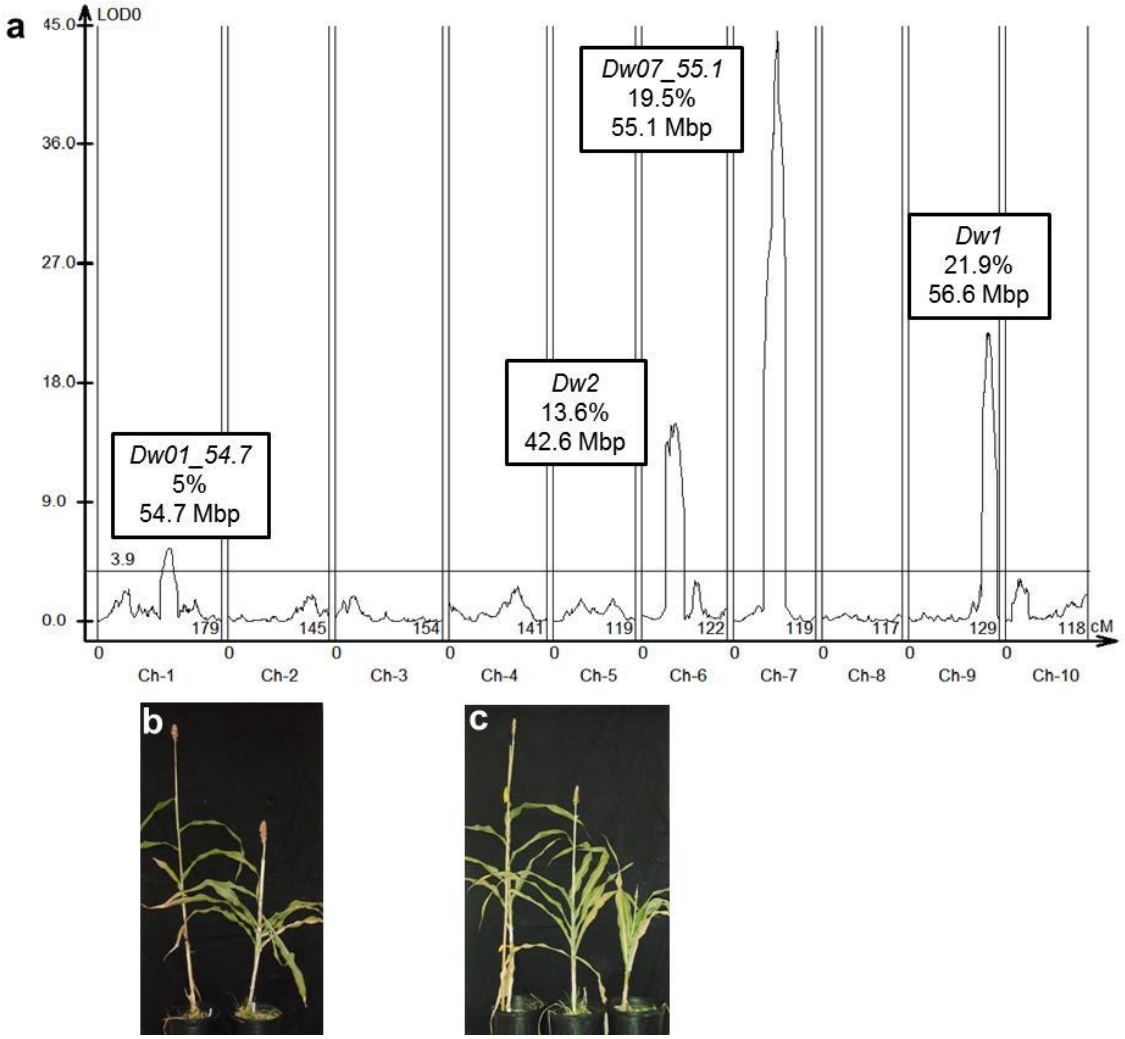


Fig 6. Stem internode length QTL identified in a population from Hegari x 80M. F_2 plants from a cross of Hegari and 80M (n=218) were grown in the greenhouse and the length of each internode was measured. The average internode length was used to map QTL. (**a**) The resulting graph shows four QTL, including *Dw1* and *Dw2*. The x-axis is the genetic map and the y-axis is the LOD score. The boxes above each trait identify the *Dw* loci, if any, the percentage of the variation explained by the QTL, and the location of the peak LOD value. (**b**) Photograph of Hegari (left) and 80M. (**c**) Photograph of F_5 plants that are *Dw1Dw1* (left), *Dw1dw1* (center), and *dw1dw1* (right) in otherwise uniform genetic backgrounds at the other loci that affect internode length.

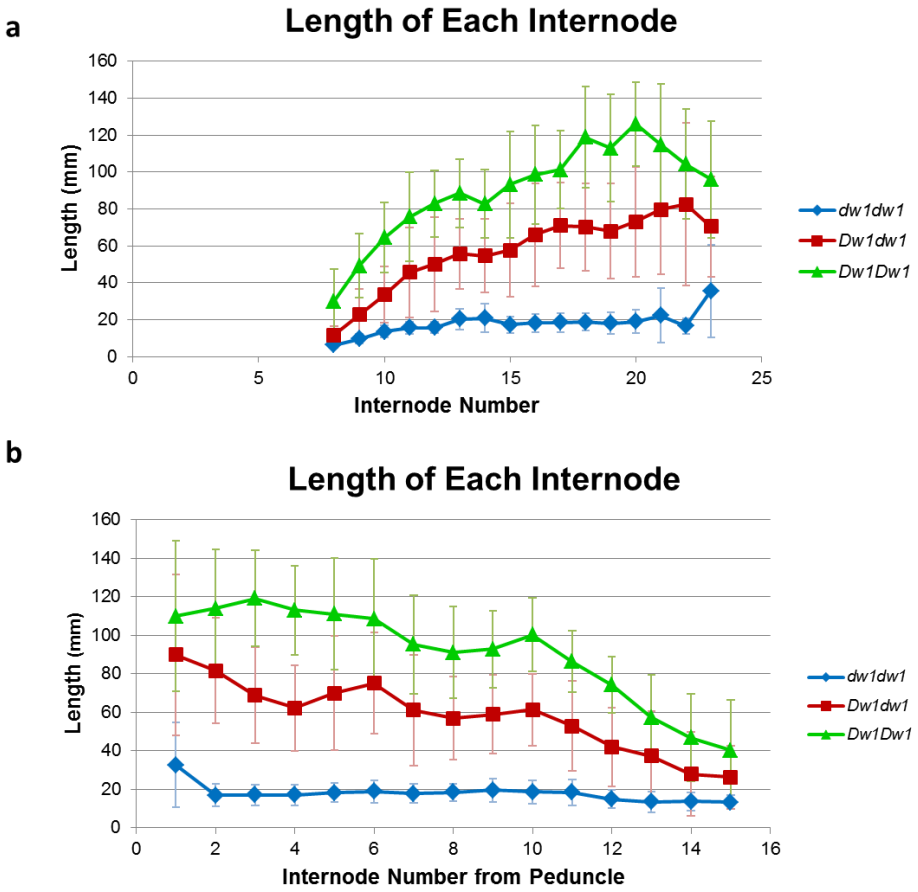


Fig 7. Internode length versus internode number for a HIF. The average internode length for each internode was calculated for each genotype at *Dw1* for one of the F_3 HIFs (n=75). In (**a**) the internodes are numbered from the bottom of the stem, whereas in (**b**) they are numbered from the peduncle.

QTL mapping was also performed using data on fresh and dry weight per internode, fresh or dry weight per unit stem length, and diameter of internode 7 (Table 2). Alleles of *Dw1* contributed to variation for internode fresh weight and dry weight.

Table 2. *Dw1* QTL for Each Trait for Hegari x 80M F_2. For the additive effect, a positive number indicates that the 80M allele increases length or weight while a negative number indicates the Hegari allele increases length or weight.

Trait	# of QTL	QTL at Dw1?	Peak (Mbp)	Peak LOD	Additive	Dominance	R^2
Average Internode Length	4	Yes	56.64	21.8	-27.3763	6.4375	0.2186
Length of Internode 5	4	Yes	56.64	7.67	-18.1324	8.5803	0.1144
Length of Internode 7	4	Yes	56.47	16.05	-29.4216	10.1362	0.2158
Length of Internode 10	4	Yes	57.07	13.09	-29.6994	11.5228	0.2388
Total Length	4	Yes	57.07	18.91	-46.3611	23.7702	0.3695
Stem Fresh Weight	3	Yes	57.07	9.91	-69.9536	23.7784	0.1869
Stem Dry Weight	2	Yes	57.07	8.04	-15.8367	7.1609	0.1614
Diameter	2	No					
Stem Fresh Weight/Stem Length	3	No					

Analysis of Epistasis

Potential interactions among the four QTL modulating internode length were investigated using multiple-QTL mapping in R/qtl [127]. The best model (y ~ *Dw01_54.7 + Dw2 + Dw07_55.1+ Dw1 + Dw10_3.2+ Dw07_55.1*:*Dw1*) had a pLOD of 50.1 and included five QTL and an interaction between two of the QTL (*Dw1* and *Dw07_55.1*, Table 3). The analysis showed an interaction between *Dw1* and *Dw07_55.1* such that allelic variation in *Dw1* has minimal impact on internode length in the presence of the 80M allele at *Dw07_55.1* which increased internode length (Fig 8). In addition, the 80M allele of *Dw07_55.1* increased internode length in *Dw1Dw1*, *Dw1dw1*, and *dw1dw1* backgrounds, although to a greater extent in genotypes that were *dw1dw1*. These results indicate that *Dw1* and *Dw07_55.1* independently activate the same downstream regulator of internode elongation, or act through different pathways to stimulate internode growth.

Table 3. QTL for Average Internode Length Identified Using MQM in R/qtl.

QTL	Chr	Peak (cM)	LOD	Peak (Mbp)	Additive	Dominance	Percent Variation	*Dw* locus
1	1	97.1	7.31	54.67	-12.66	-3.956	3.693	*Dw01_54.7*
2	6	41.1	20.274	42.64	23.531	2.016	11.849	*Dw2*
3	7	58.8	50.968	55.15	-39.248	22.749	43.127	*Dw07_55.1*
4	9	107.2	31.628	57.07	26.329	7.254	21.11	*Dw1*
5	10	19.3	4.883	3.17	8.839	-8.392	2.403	*Dw10_3.2*

QTL	LOD	Percent Variation	Add:Add	Add:Dom	Dom:Add	Dom:Dom	*Dw* locus
3:4	5.593	2.773	12.773	5.845	-11.038	-4.486	*Dw07_55.1:Dw1*

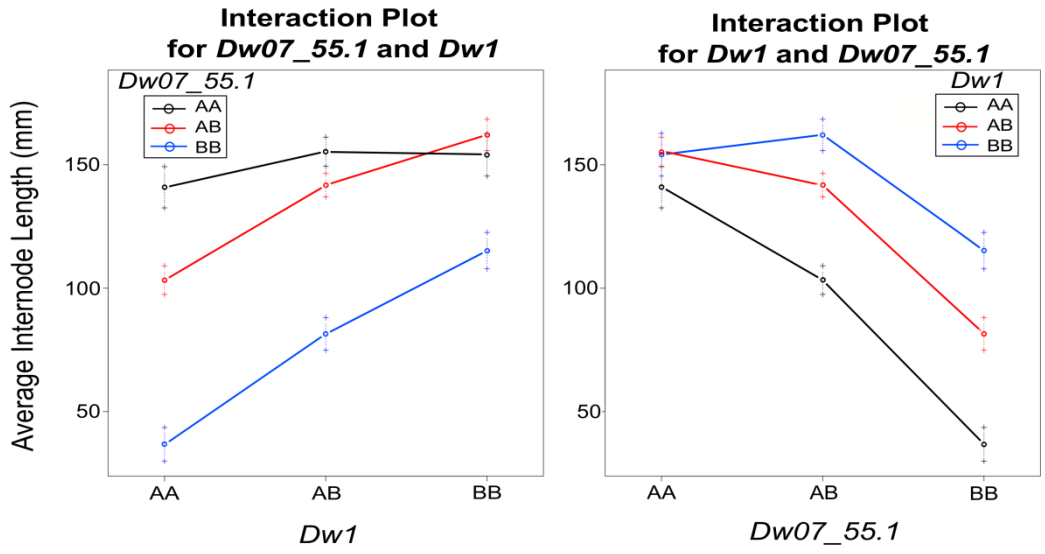


Fig 8. Interaction plots from MQM mapping in R/qtl. The interaction plots show the interaction between *Dw1* and the locus on chromosome 7 (*Dw07_55.1*) in the Hegari x 80M F_2. The A allele is 80M and the B allele is Hegari. Phenotypes distinguishing *Dw1* from *dw1* are greater when the *Dw07_55.1* locus on LG-07 is BB (fixed Hegari).

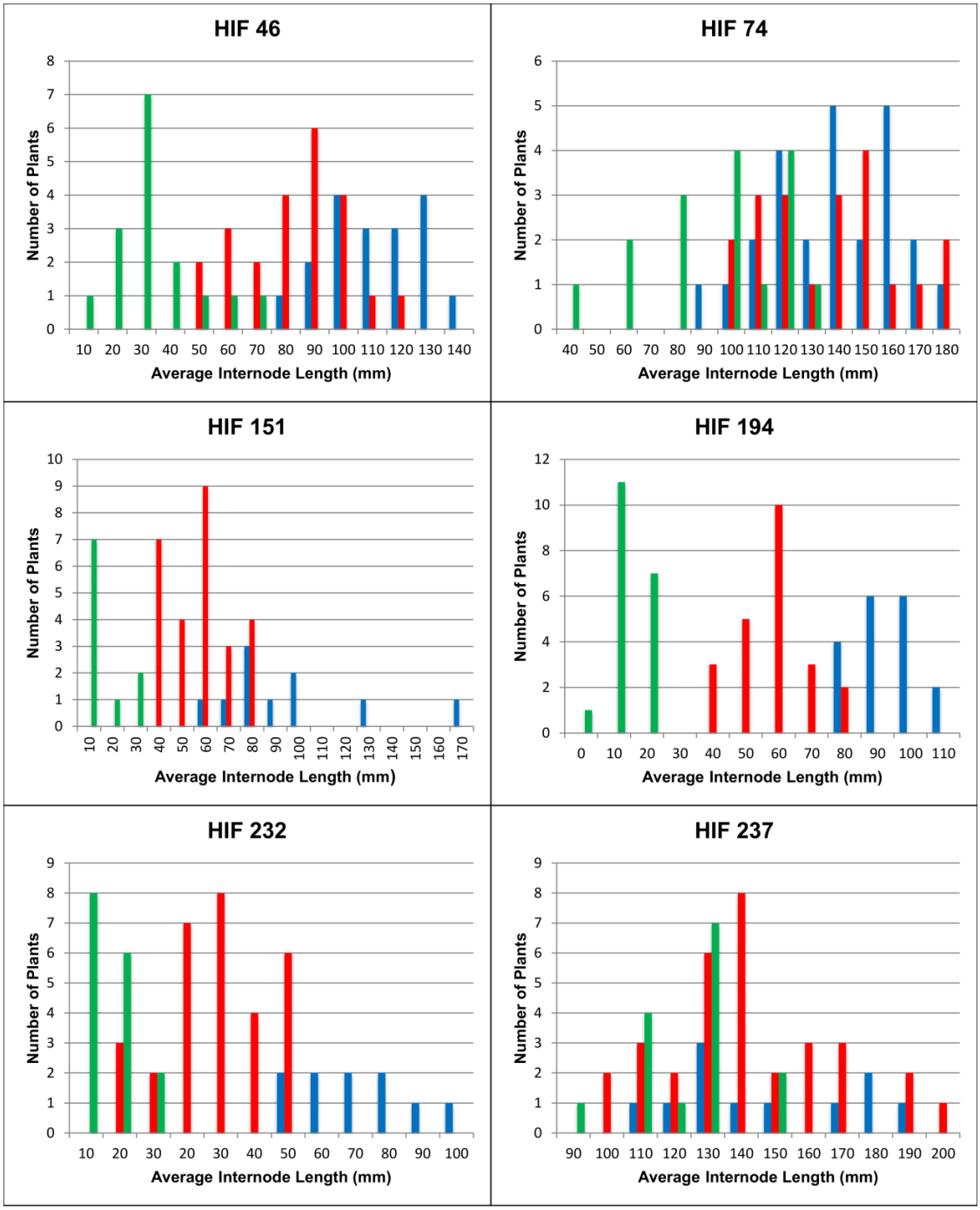


Fig 9. Histograms of the average internode length for each Hegari x 80M F_3 HIF. For each HIF, the lines that had recombination break points in the region of *Dw1* were removed and the remainder of the plants grouped into *Dw1Dw1* (blue), *Dw1dw1* (red), and *dw1dw1* (green) and plotted in a histogram. Note that HIFs 74 and 237 have the 80M allele at *Dw7_55.1* while the others have the Hegari allele.

Fine Mapping *Dw1*

Dw1 was fine mapped by constructing HIFs from seed of F_2 plants of the QTL mapping population that were heterozygous for *Dw1* and homozygous at the other QTL that affect internode length. HIFs derived from F_2 plants homozygous for the Hegari allele at *Dw07_55.1* were most useful for fine mapping *Dw1*. Histograms of the average internode length for each HIF are shown in Fig 9. Breakpoint analysis of the first set of HIFs narrowed the region encoding *Dw1* to 313 kb. The location of breakpoints in a few key lines was further refined using Digital Genotyping based on the restriction enzyme NgoMIV [124]. This information delimited the *Dw1* locus to 230 kb, a region encoding 35 genes as annotated in the v1.4 gene set (Phytozome v.9). A further round of fine mapping was carried out using five HIFs derived from F_3 plants heterozygous for *Dw1dw1*. These plants were screened for recombinants with CAPS markers and six plants were identified with recombination breakpoints in the delimited *Dw1* region. Phenotyping and identification of breakpoints by sequencing SNPs delimited *Dw1* to a region that spanned 33 kb and encoded seven genes as annotated in v2.1 (Phytozome v.10) (Table 4). Markers used for fine mapping and the location of the delimited *Dw1* locus are shown in Fig 10. Information about the seven putative genes in the delimited *Dw1* locus is provided in Table 4. Four of the genes were annotated with a function: an E3-ubiquitin ligase involved in syntaxin degradation, Photosystem I reaction center subunit VI, PRONE-Rop nucleoide exchanger, and a serine/threonine kinase. There were also three genes annotated as having unknown functions.

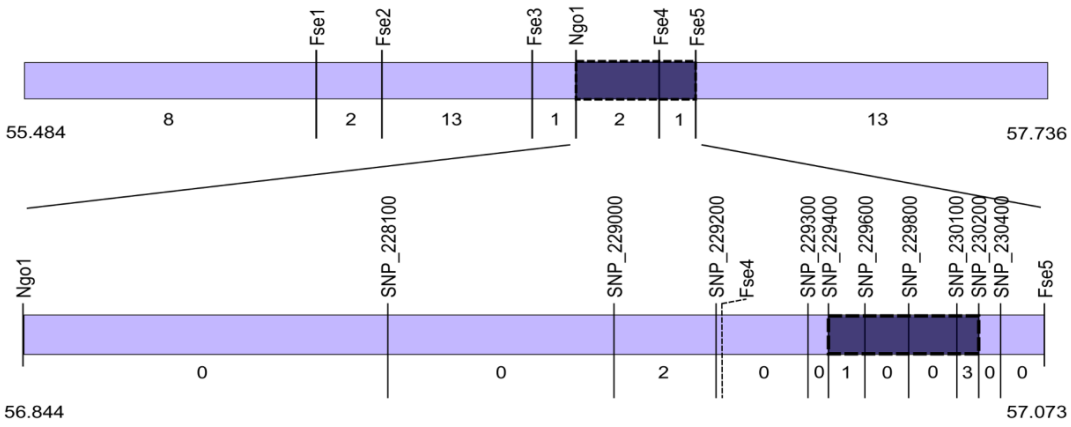


Fig 10. A schematic of the region of SBI-09 encoding *Dw1*. The top bar shows the *Dw1* locus delimited by QTL mapping in the F_2. The region was refined in the F_3 population (n=75 for each of six families) using the DG markers labeled in the diagram. The numbers below the bar are the number of recombinants (both bars). Note that all members of one of the families (237) had a breakpoint in between Fse5 and the end of the region shown. The lower bar represents the delimited *Dw1* locus defined by mapping in the F_3 generation with SNP markers labeled. Dark purple shows the location of *Dw1* based on fine mapping. SNP markers are named with the last six digits of the gene name of the gene the SNP is in or near. Fse4 is included for perspective though it was not scored in the F_4.

Table 4. Genes in the Delimited *Dw1* Locus.

Gene Name	Probable Function	Location
Sobic.009G229500	Unknown	57,026,900 - 57,027,289
Sobic.009G229600	E3 ubiquitin ligase/syntaxin degradation	57,027,335 - 57,036,566
Sobic.009G229700	Photosystem I reaction center, subunit VI	57,036,793 - 57,037,995
Sobic.009G229800	Unknown	57,042,620 - 57,045,133
Sobic.009G229900	PRONE-Rop nucleotide (guanine) exchanger	57,046,394 - 57,049,526
Sobic.009G230000	Unknown	57,050,065 - 57,051,463
Sobic.009G230100	Serine/threonine kinase	57,051,814 - 57,055,008

Identification of Polymorphisms in the Delimited *Dw1* locus

All seven genes located in the fine mapped *Dw1* locus were sequenced in Hegari and 80M (Table 5). No sequence variants were found in Sobic.009G229700 or Sobic.009G229900. The only sequence variants in Sobic.009G229600 and Sobic.009G230100 were located in introns and/or the 5'UTR. Of the genes annotated with an unknown function, Sobic.009G229500 had no sequence variants while Sobic.009G230000 had two INDELs in the 5'UTR and a SNP in the first exon that resulted in a synonymous mutation. Sobic.009G229800 was the only gene in the delimited *Dw1* locus that had a polymorphism distinguishing the parental genotypes that resulted in a change in amino acid sequence (Table 5). Hegari (*Dw1*) encoded a full-length protein, whereas the sequence in 80M (*dw1*) (and BTx623 (*dw1*)) contained an A > T mutation that caused a Lys199 > stop codon change in the second exon of Sobic.009G229800 (Fig 11B).

All seven of the genes in the delimited region were also sequenced in Standard Yellow Milo (*Dw1*) and Dwarf Yellow Milo (*dw1*). Quinby [123] noted that *dw1* was originally identified in the Standard Yellow Milo (*Dw1, Dw2, Dw3*) background [143]. The shorter version of Yellow Milo containing *dw1* was named Dwarf Yellow Milo. Therefore, the sequences of Standard Yellow Milo and Dwarf Yellow Milo are expected to vary only at *Dw1*. Sequence analysis revealed only one polymorphism in the delimited *Dw1* region that distinguished the two milo lines: the A > T SNP in Sobic.009G229800 that caused a premature stop codon. For all the other

polymorphisms found between Hegari and 80M in the region, Standard Yellow Milo and Dwarf Yellow Milo had the same allele as 80M.

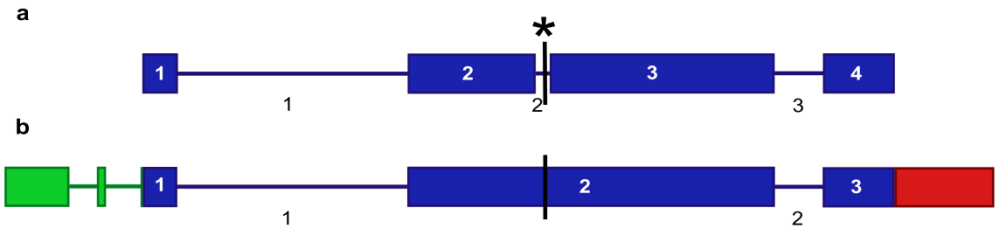


Fig 11. Gene annotation models of *Dw1* (Sobic.009G229800). (**a**) Gene model from *Sorghum bicolor* Genome v2.1 (Phytozome). (**b**) Gene model based on cDNA sequence analysis. Boxes (blue) represent exons and lines are introns. Regions colored green represent the 5'UTR and those colored red the 3'UTR. Exons are numbered within boxes and introns are numbered in black. The asterisk/vertical line marks the location of the Lys199 > stop codon mutation that distinguishes *Dw1* from *dw1*.

Table 5. Polymorphisms Distinguishing 80M and Hegari in Genes in the Delimited *Dw1* Locus.

Gene	#	Type	Polymorphism	Location	Region
Sobic.009G229500	None				
Sobic.009G229600	1	SNP	C > T	2660	Intron
	2	INDEL	- > A	6597	Intron
Sobic.009G229700	None				
Sobic.009G229800	1	INDEL	A > -	-707	5' UTR
	2	SNP	A > T; K > Stop	1350	Exon
Sobic.009G229900	None				
Sobic.009G230000	1	INDEL	- > CAGGCAGG	-64	5'UTR
	2	INDEL	- > ACGACG	-25	5'UTR
	3	SNP	G > T; L > L	126	Exon
Sobic.009G230100	1	INDEL	T > -	-397	5' UTR
	2	SNP	A > T	537	Intron
	3	INDEL	A > -	1841	Intron

The gene-model for Sobic.009G229800 in v2.1 (Phytozome v10) included a very short intron (intron 2) (Fig 11A). However, cDNA sequence analysis of Sobic.009G229800, and RNA-seq analysis (see below), failed to provide evidence for intron 2. Instead, cDNA sequences from Hegari (*Dw1*) contain a continuous coding region that spanned intron 2 of the v2.1 gene-model. Gene-models of homologs of Sobic.009G229800 in other plant species (e.g. maize, rice, and Arabidopsis) also lack intron 2 and show continuous reading frames across this region. The cDNA sequence also clarified splicing in the 5'UTR (Fig 11, regions in green). Based on this analysis, we propose the revised annotation of Sobic.009G229800 shown in Fig 11B that contains three exons and conclude that the polymorphism that distinguishes Hegari and 80M generates a truncated protein lacking most of exon 2 and all of exon 3 (mutation marked by an asterisk in Fig 11) presumably resulting in a loss of function.

The intron/exon structures of the other genes in the delimited *Dw1* locus were identical to homologs in maize and/or rice (Table 6). Furthermore, the RNA-seq data for v3.1 (Phytozome v11) is consistent with the annotations of the other genes in the delimited *Dw1* locus and the updated annotation of Sobic.009G229800 that lacks intron 2 (Fig 11B).

Table 6. Maize and Rice Homologs of the Seven Genes in the Delimited *Dw1* Region.

Sorghum	Maize	Rice
Sobic.009G229500	GRMZM2G405706	LOC_Os05g48610
Sobic.009G229600	N/A	LOC_Os05g48620
Sobic.009G229700	GRMZM2G451224	LOC_Os05g48630
Sobic.009G229800	GRMZM2G079832; GRMZM2G060467	LOC_Os01g01390
Sobic.009G229900	GRMZM2G359664; GRMZM2G377615	LOC_Os05g48640
Sobic.009G230000	GRMZM2G377613	LOC_Os05g48650
Sobic.009G230100	GRMZM2G079583	LOC_Os05g48660

Sobic.009G229800 was sequenced in other genotypes of sorghum previously identified as *Dw1* or *dw1* (Tables 7 and 8). Genotypes previously designated as *Dw1* encoded full-length proteins similar to Hegari. Numerous grain sorghum-breeding lines with shorter internodes were generated from the Dwarf Yellow Milo source of *dw1*. Therefore, it is not surprising that all of the lines designated *dw1* have the same recessive allele as Dwarf Yellow Milo. Sobic.009G229800 sequences from Rio and Early White Milo (both *Dw1*) contain several additional polymorphisms (Table 8). SIFT [144] analysis of a non-synonymous coding mutation found in Rio and Early White Milo (A425S) predicted that this change in *Dw1* would not affect function.

Table 7. Sequence Variants in Exons of Sobic.009G229800 in Diverse Sorghum Genotypes. Location is from the start codon.

Number	**2**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**22**
Polymorphism	A > T	C > T	G > A	G > A	C > A	T > C	T > A	T > G	T > C
Location (bp)	1350	1127	1259	1583	1586	1667	1733	2028	2316
Exon	2	2	2	2	2	2	2	2	3
Type		Syn	Syn	Syn	Syn	Syn	Syn	Nonsyn	Syn
Change in Protein	K > Stop	F > F	P > P	S > S	P > P	T > T	P > P	S > A	N > N
SIFT	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.36= tolerated	N/A

Table 8. Distribution of *Dw1* Coding Sequence Variants in Sorghum Genotypes. The polymorphism number corresponds to the number in Table 7.

Line	***Dw1* Genotype**	**Polymorphism Number**								
		2	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**22**
Hegari	*Dw1*	A	T	A	A	A	C	A	G	C
80M	*dw1*	T	T	A	A	A	C	A	G	C
Standard Yellow Milo	*Dw1*	A	T	A	A	A	C	A	G	C
Dwarf Yellow Milo	*dw1*	T	T	A	A	A	C	A	G	C
Double Dwarf Yellow Milo	*dw1*	T	T	A	A	A	C	A	G	C
BTx623	*dw1*	T	T	A	A	A	C	A	G	C
BTx406	*dw1*	T	T	A	A	A	C	A	G	C
SC170	*dw1*	T	T	A	A	A	C	A	G	C
R.07007	*dw1*	T	T	A	A	A	C	A	G	C
IS3620c	*dw1*	T	T	A	A	A	C	A	G	C
Rio	*Dw1*	A	C	G	G	C	T	T	T	T
M35-1	*Dw1*	A	T	A	A	A	C	A	G	C
Texas Blackhull Kafir	*Dw1*	A	T	A	A	A	C	A	G	C
Spur Feterita	*Dw1*	A	T	A	A	A	C	A	G	C
Early White Milo	*Dw1*	A	C	G	G	C	T	A	T	T

Expression of *Dw1* in Stem Tissue

Sobic.009G229800 was expressed in fully elongated internodes and elongating internodes (Fig 12). The highest levels of expression were observed in the lower portion

of the elongating internode. *Dw1* mRNA levels were ~3-fold higher in stems of Hegari compared to 80M.

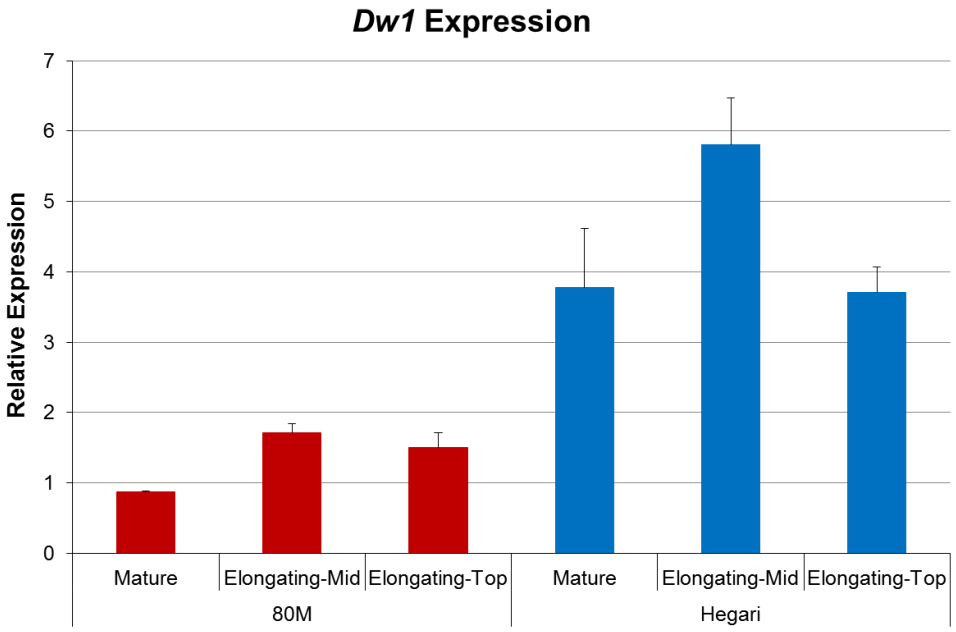


Fig 12. Relative expression of *Dw1* in stem internodes. RNA was extracted from a full length internode (Mature), the lower half of an elongating internode, and the upper half of an elongating internode for each parental genotype (n=3 each). Relative expression was determined by qRT-PCR using the ΔΔCt method with 18S rRNA as the normalizer and the sample from 80M mature tissue as the calibrator.

Protein Sequence Analysis

Sobic.009G229800 is currently annotated as having an unknown function. BLAST analysis showed that homologous genes/proteins are present in maize, rice, and Arabidopsis among other plants. Fig 13 shows the sequence alignment of Sobic.009G229800 and maize, rice, and Arabidopsis homologs. A phylogenetic tree of

select homologs has two distinct groups corresponding to the monocots and dicots (Fig 14). The Arabidopsis homolog of *Dw1* is annotated as associated with the plasma membrane based on experimental evidence [145] and located in the nucleus based on prediction (TAIR). PSIPRED-MEMSAT-SVM predicts that the sorghum Dw1 protein contains a single transmembrane/pore-lining domain from residues 263-278. Interestingly, these residues are missing in the Arabidopsis homolog (Fig 13). PSIPRED-DISOPRED, PONDR, and FoldIndex all predicted a high degree of disorder in the protein (Table 9).

Table 9. Summary of Protein Function Searches.

Program	**Program Description**	**Annotations**
BLAST-Arabidopsis homolog (TAIR)	finds homologs of subject	involved in: biological_process; located in: nucleus (predicted), plasma membrane (experimental); closest paralog: AT5G52430 hydroxyproline-rich glycoprotein
NCBI-conserved domain	database search of domains and proteins	large tegument protein
PSIPRED-MEMSAT-SVM	membrane helix prediction	pore-lining/transmembrane residues 263-278
PSIPRED-DISOPRED	predicts disorder based on homologs	highly disordered; possibly protein binding
PONDR	predicts disordered regions	~52% disordered; two long regions of disorder
FoldIndex	predicts disordered regions	~44% disordered

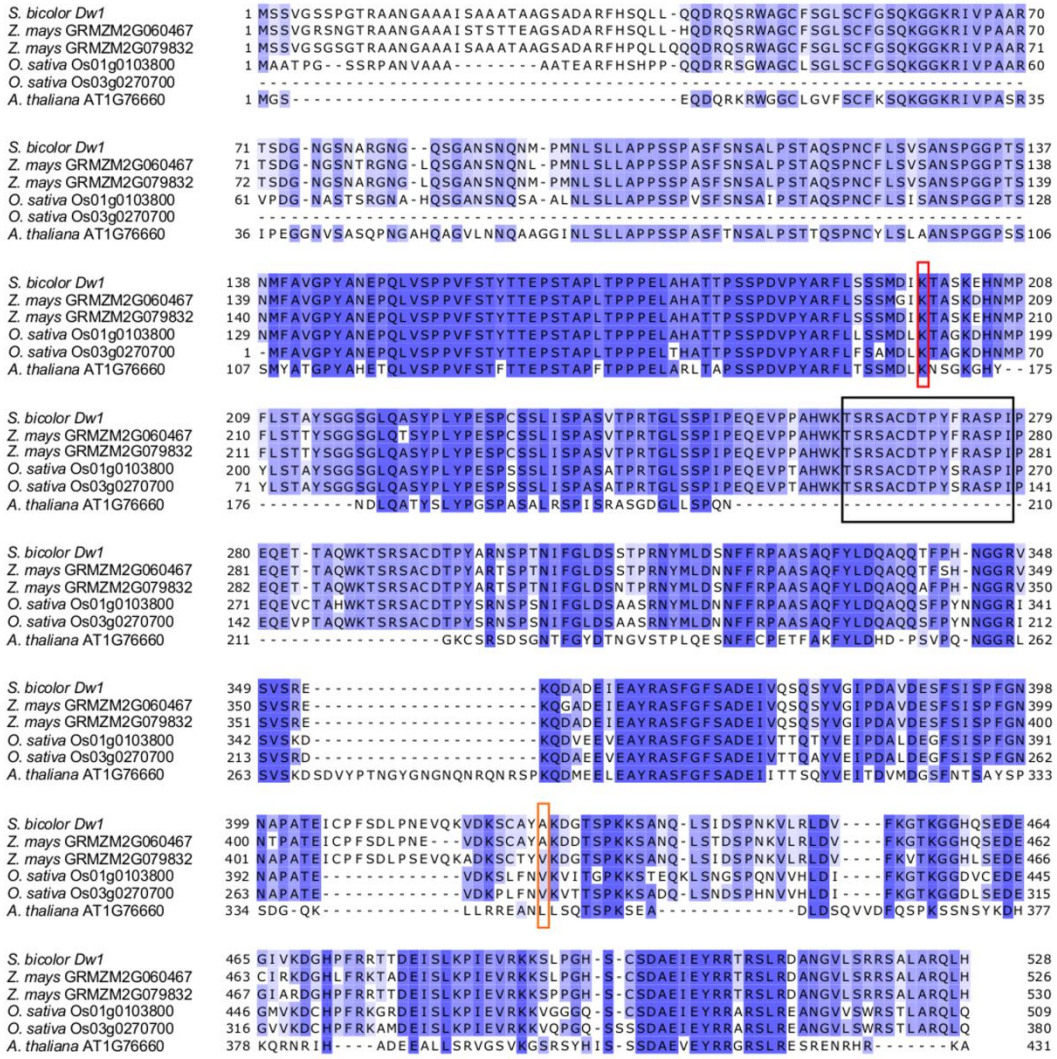


Fig 13. Protein alignment of *Dw1* and select homologs. Alignment of *Dw1* with the two maize homologs, the two rice homologs, and the Arabidopsis homolog compiled in Jalview using the T-Coffee function (dark blue color indicates higher percent identity). The red rectangle marks the functional polymorphism that distinguishes Hegari (*Dw1*) and 80M (*dw1*). The orange rectangle marks a polymorphism present in Rio and Early White Milo not found in the other sequenced lines. The black box is the possible transmembrane domain predicted by PSIPRED-MEMSAT-SVM.

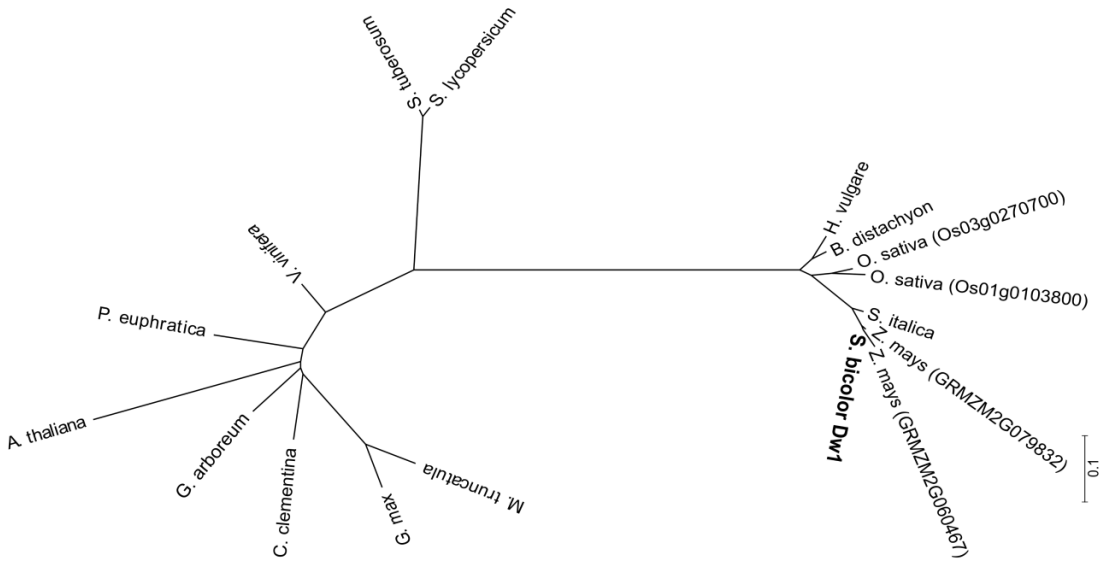


Fig 14. A phylogenetic tree of a diverse selection of *Dw1* homologs. Tree was constructed in MEGA6 using Maximum Likelihood. Sorghum *Dw1* is in bold letters.

Maize homologs of Sobic.009G229800 located on chromosomes 6 and 8 are syntenic to sorghum chromosome 9. Genes flanking *ZmDw1* on maize chromosome 8 show collinearity with the region on SBI-09 encoding *Dw1*. On the other hand, the *OsDw1* homologs are located on rice chromosomes 1 and 3 while sorghum chromosome 9 is syntenic to rice chromosome 5. This suggests that *Dw1* moved to its position on SBI-09 after separation from rice and before separation from maize.

DISCUSSION

In this study, *Dw1* was identified using a F_2 population and HIFs derived from Hegari (*Dw1*) and 80M (*dw1*). *Dw1* was identified as Sobic.009G229800 a gene of

unknown function that is highly conserved in plants. The recessive *dw1* allele corresponds to a loss of function mutation that creates a stop codon in the middle of the protein encoded by Sobic.009G229800. The recessive *dw1* allele identified in 80M was present in Dwarf Yellow Milo (*dw1*) and Double Dwarf Yellow Milo (*dw1,dw2*) but not in Standard Yellow Milo (*Dw1*) consistent with reports that short plants containing *dw1* originated as a spontaneous mutation in Standard Yellow Milo [104,123]. 80M and the other maturity standards (i.e., 100M, 90M, 80M, 60M) were derived from a cross of Early White Milo (*Dw1*) and Double Dwarf Yellow Milo (*dw1, dw2*) and progeny recessive for *dw1* and *dw2* were selected so that the maturity standards have similar internode lengths (*dw1dw2Dw3dw4*) [123].

The Dwarf Yellow Milo *dw1* allele is present in BTx623, an elite seed parent, and in other genotypes used for grain sorghum breeding in the U.S. (i.e., BTx406, SC170, R07007). The *dw1* allele described in this study is present in many grain sorghum lines because BTx406 (*dw1*) was used to convert tall late flowering sorghum accessions to short early flowering genotypes useful for grain sorghum breeding in the U.S. [9]. This also explains why Brown et al. [111] mapped a QTL for height (*Sb_HT9.1*) corresponding to allelic variation at the *Dw1* locus in a panel of grain genotypes many of which included BTx406 in their pedigrees. Markers most tightly linked to *Sb_HT9.1* identified a region of SBI-09 from 57.14-57.21, the same region we found that encodes *Dw1*. This region includes Sobic.009G229800; however, this gene was initially annotated in Phytozome as two genes (v1.4 gene set). Subsequently, Sobic.009G229800 was annotated with an intron spanning the portion of the coding

region that contains the causative mutation (v2.1). Two additional mapping studies identified the same region of SBI-09 as encoding *Dw1* [112,114]. Both studies suggested that mutations in a GA2 oxidase (GA2ox5) could be responsible for variation in height caused by *Dw1*. However, subsequent sequence analysis of *GA2ox5* from genotypes that were *Dw1Dw1* and *dw1dw1* did not show sequence variants consistent with the identification of this gene as *Dw1* [115]. Moreover, mutations causing reduced GA levels in sorghum result in short internodes but also abnormal culm bending, a phenotype not observed in *dw1dw1* sorghum genotypes [115].

Dw1 (Sobic.009G229800) is present in maize, rice, other grasses, and dicots such as Arabidopsis. Several large INDELS distinguish the proteins in grasses and Arabidopsis. Homologs of Sobic.009G229800 in maize are collinear with *Dw1* in sorghum; however, homologs in rice are not located on the homeologous chromosome suggesting that this gene moved to its current location in sorghum after separation of these grasses. The closest homolog in Arabidopsis is annotated as a plasma membrane protein, a localization that was verified experimentally [145]. The Arabidopsis protein was also annotated with a nuclear location. Analysis of the sorghum protein identified a stretch of amino acids (263-278) that could be associated with the lining of a transmembrane pore. The protein was also predicted to have highly disordered protein domains. Research clarifying the localization and biochemical function of the protein encoded by Sobic.009G229800 will be needed to understand how *Dw1* regulates the length of stem internodes.

Quinby and Karper [146] showed that alleles of *Dw1* do not affect leaf size, only internode lengths. The restriction of *Dw1* action to stems is useful because *dw1dw1* can be used to reduce internode length without affecting leaf morphology or canopy development. Furthermore, a QTL corresponding to *Dw1* was also found to modulate the weight of the stem but not weight per unit length of stem. Thus, *Dw1* increases length and weight of internodes. Heterozygous *Dw1dw1* progeny derived from Hegari x 80M had internode lengths that were intermediate compared to plants that were *dw1dw1* and *Dw1Dw1* (Fig 8), indicating gene dosage alters the gene's action on internode growth. *Dw1* was expressed in stem internodes, with ~3-fold higher expression in Hegari (*Dw1*) compared to 80M (*dw1*). Higher expression in Hegari could be due to feedback from *Dw1* resulting from greater growth of the internode, or due to differences in Hegari/80M genetic background.

This research was undertaken to further our understanding of genetic factors influencing internode elongation and stem length in sorghum with a focus on *Dw1*. QTL analysis of an F_2 population derived from Hegari and 80M used for fine mapping *Dw1* identified QTL that modulate stem internode length aligned with *Dw1*, *Dw2*, a minor QTL on SBI-01 (*Dw01_54.7*) and a QTL on SBI-07 approximately 3 Mbp from *Dw3* (*Dw07_55.1*) (16). Interactions between *Dw07_55.1* and *Dw1* were detected and plants homozygous for the *Dw07_55.1* allele from 80M had long internodes and showed attenuated influence of *Dw1* alleles in this background. *Dw3* is an ABCB1 efflux auxin transporter that has homologs in many other plants. However, the phenotypic effect of mutation of ABCB1 is attenuated in dicots like Arabidopsis where auxin is exported

from apical meristems via two different ABCB transporters: ABCB1 and ABCB19 [30]. In grasses, auxin is exported from the apical meristem and intercalary meristems of the stem. ABCB1 in maize is the only ABCB transporter in the intercalary meristem leading to more severe stem internode length phenotypes when this gene is mutated. Interestingly, in maize the ABCB1 mutant causes severe shortening of the lower internodes while the upper internodes are essentially normal in length [110]. In contrast, *dw1dw1* caused a reduction in the length of all internodes (Fig 7). The current study and prior studies showed that recessive *dw1* alleles decrease internode length/plant height in *Dw3* backgrounds (Standard Yellow Milo, Dwarf Yellow Milo) as well as in plants that are homozygous for *dw3* (Texas Blackhull Kafir (*Dw1Dw2dw3*) vs Martin (*dw1Dw2dw3*) [104]. This result suggests that *Dw1* action is not dependent on *Dw3*, although *Dw3* alleles may modulate the extent of *Dw1* action on internode elongation. As noted above, *Dw1* is not a GA2 oxidase as previously suggested and recessive alleles do not result in stem bending. However, it is possible that *Dw1* mediates signaling by hormones (GA, auxin, brassinosteroids, strigolactone, ethylene), photoreceptors (phytochromes, PIFs), or other factors that modulate internode growth. Ongoing research is focused on characterizing the molecular basis of *Dw1* action.

CHAPTER III

SORGHUM *Dw2* ENCODES A PROTEIN KINASE REGULATOR OF STEM INTERNODE LENGTH

INTRODUCTION

Sorghum is the fifth most widely grown cereal crop worldwide (faostat.fao.org). Its drought and heat tolerance make this crop especially important in semi-arid regions. Sorghum is a C4 grass with a diverse germplasm that has been selected for many uses including production of grain, forage, sugar, and biomass for bioenergy. In its native Africa, sorghum grows 4-5 meters tall and many genotypes are photoperiod sensitive, resulting in delayed flowering in long day environments. Upon introduction to temperate locations, photoperiod insensitive varieties that flower early were selected for production of grain [123]. Additionally, shorter grain varieties were selected to reduce lodging and to aid mechanical harvesting. In contrast, sorghum genotypes with longer stems and delayed flowering enhance biomass and sugar production [118,147]. In sweet sorghum, stem length is associated with higher sugar yield because stems accumulate high levels of sucrose post floral initiation [122,147,148]. In energy sorghum, 83% of the shoot biomass accumulates in the stem [1]. Therefore, increasing our knowledge of stem growth will aid the improvement of sorghum hybrids for bioenergy production.

Plant height is determined primarily by the length and number of stem internodes. The number of internodes produced by a plant is a consequence of growth duration and the rate of internode production. Quinby and Karper [104] identified four

loci (*Dw1*-*Dw4*) that control internode length by measuring the height of the stem from the ground to the flag leaf. At each *Dw* locus the dominant allele increased internode length. Recessive alleles of *Dw1* and *Dw2* were identified in Milo lines, while recessive alleles of *Dw3* were identified in Kafir backgrounds, and dominant alleles at *Dw4* were only found in broomcorns [104]. *Dw2* was shown to have pleiotropic effects on panicle length, seed weight, and leaf area [105,106]. In addition to internode length, *Dw3* influences grain yield, tiller number [121], and leaf angle [108].

Dw3 was the first dwarfing gene to be cloned in sorghum [109]. *Dw3* encodes a homolog of the maize *Br2* gene and is an ATP-binding cassette type B1 (ABCB1) auxin efflux transporter. This is in contrast to dwarfing or semi-dwarfing genes in other important crops, such as rice and wheat, which have mutations in genes involved in the gibberellin pathway [95,96]. *Dw1* was mapped to a region on chromosome 9 between 56.8-57.1 Mb [111]. The gene corresponding to *Dw1* was recently identified as Sobic.009G229800 by map-based cloning [149,150]. This gene regulates internode cell proliferation [150] and encodes a putative membrane protein not previously assigned a function [149]. The recessive *dw1* allele in Dwarf Yellow Milo (DYM), first identified by Quinby and Karper [104], contains a stop codon in exon 2 that results in protein truncation [149]. The *dw1* allele originating from Dwarf Yellow Milo has been used extensively in grain sorghum breeding programs.

Dw2 has also been used extensively in grain sorghum breeding programs to reduce plant height. *Dw2* is linked to *Ma1*, an important flowering time gene that confers photoperiod sensitivity [123]. *Ma1* is located on chromosome 6 at ~40.3 Mb

and encodes PRR37 [151]. *Dw2* was previously mapped to a location near *Ma1* at ~42 Mb in several QTL mapping studies [9,112,114,124] and suggested to be a histone deacetylase (Sobic.006G067600) based on GWAS analysis[112]. Recessive alleles of *Ma1* and the dwarfing genes were used in the Sorghum Conversion Program to convert tall late flowering landraces from Africa into short, early flowering genotypes that are useful for grain sorghum breeding. The landraces were crossed to BTx406 (*dw1dw2dw3dw4*) to introduce one or more of the recessive alleles at the *Dw* loci into landrace backgrounds [9]. Recent analysis of the sorghum conversion lines has shown that large portions of chromosome 6 have been introgressed from BTx406 into landrace accessions during conversion and that the peak of introgression frequency aligned with *Dw2* [152].

In the current study, *Dw2* was map-based cloned using two RIL populations: BTx623 (*dw1Dw2dw3dw4*) x IS3620c (*dw1dw2Dw3dw4*) and BTx642 (*dw1dw2dw3dw4*) x Tx7000 (*dw1Dw2dw3dw4*). *Dw2* was identified as a protein kinase whose closest homolog in Arabidopsis is the kinesin-like calmodulin-binding protein (KCBP)-interacting protein kinase (KIPK), a member of the AGCVIII subfamily that also includes PINOID (PID) and PHOTOTROPIN1 and 2 (PHOT1 and 2).

METHODS

Phenotypic Analysis of DYM and DDYM Stems

The progenitor genotypes Dwarf Yellow Milo (DYM; *Dw2*) and Double Dwarf Yellow Milo (DDYM; *dw2*) [104] were grown to examine the internode length

phenotypes caused by the two *Dw2* alleles. For each genotype, three plants were individually grown in 3.8-gallon pots (Custom2000) containing MetroMix MVP (Sun Gro Horticulture) with supplemental fertilizer (Peters 20-20-20) in the greenhouse during the summer. At grain maturity, the plants were harvested and the total stem length and length of each internode were measured.

QTL Mapping of *Dw2* in a RIL Population Derived from BTx623 and IS3620c

The BTx623 x IS3620c RIL population was used for mapping *Dw2* [153]. Seed for the population was obtained from the USDA-ARS Plant Genetic Resources Conservation Unit (Griffin, GA). BTx623 is *dw1Dw2dw3dw4* and IS3620c is *dw1dw2Dw3dw4* [124,154]; therefore, the population segregated for both *Dw2* and *Dw3*. The population (n=380) was grown in the greenhouse in the summer of 2013 with natural day lengths. Three plants of each RIL were grown per pot, one pot per line in the same manner as DYM and DDYM. Plants were harvested at grain maturity. For each plant, the total length of the plant (base of the plant to the base of the panicle) and the length of each internode and peduncle were measured. Internodes were numbered from the peduncle. Plants differed for flowering time, with earlier flowering lines producing fewer elongated internodes. As a consequence, the length of the 6^{th}, 7^{th}, and 8^{th} internodes below the peduncle had smaller sample sizes (n=375, n=356 and n=296, respectively). Genotyping and genetic map construction (n=398) were performed as described in Truong et al [155] except the DG marker sequences were mapped to version 3 of the sorghum reference genome assembly (*Sorghum bicolor* v3.1 DOE-JGI,

http://phytozome.jgi.doe.gov/), using BWA [156], and INDEL realignment and joint variant calling were performed with the GATK using the naive pipeline of the RIG workflow [157–160]. QTL mapping was performed in R/qtl using interval mapping (IM) with 1000 permutations and an α=0.05 [128]. Both the genetic map and QTL mapping were performed as an F_7 instead of a RIL due to excess heterozygosity.

MQM was performed using the same phenotypes, except peduncle length, and genotypes that were used for IM, except the genetic map was thinned to obtain a marker set with at least 1cM spacing between markers. Also, measurements of the length of each internode, average internode length, and total internode length were normalized using Empirical Quantile Normal Transformation prior to QTL mapping with R/qtl [127,128,161]. Penalties (main effect, heavy interaction, and light interaction) for all normalized phenotypes were calculated from 25,000 permutations of two-dimensional genome scans using the TIGGS-HPC cluster at Texas A&M; penalties calculated were negligibly different between phenotypes (i.e. same to the tenths place). Significant QTL identified from an initial IM analysis (alpha=0.05, main effect LOD = 3.2) were used to seed multiple-QTL model selection analysis (maximum number of QTL in a model was restricted to 7; main effect LOD = 3.2, heavy interaction LOD = 4.3, light interaction LOD = 1.9) [127,128]. The best scoring multiple-QTL model from model selection of each phenotype was then merged into a composite multiple-QTL model. The composite multiple-QTL model was generated by merging all overlapping 2-LOD intervals into one QTL and designating the position of the MLOD (maximum LOD) marker as the

QTL position [162], where loci with an epistatic interaction were merged independently of strictly additive loci.

QTL Mapping of *Dw2* in a RIL Population Derived from BTx642 and Tx7000

BTx642 is *dw1dw2dw3dw4* [163] and Tx7000 is *dw1Dw2dw3dw4* [123]; therefore, the population derived from a cross of these genotypes will segregate for alleles of *Dw2*. The BTx642 x Tx7000 RIL population (n=89) was grown in the field in the spring and summer of 2009. It was planted in a Norwood silty clay loam (fine-silty, mixed (calcareous), thermic Typic Udifluvent) in duplicate in a randomized block design at the Texas A&M Research Farm located near Snook, TX on 03/04/2009. The blocks were arrayed in 20 rows 4.6 m long and spaced 76 cm apart with two buffer rows on each end of the block. Each block was offset from the next by approximately 1.5 m. The plants emerged on 08/04/2009 and were thinned to a within-row spacing of 10 cm at 16 days after emergence (DAE). The average daily maximum temperature was 33.3°C and the average daily minimum temperature was 21.1°C. The population received 24.9 cm of natural rainfall during the growing season with supplemental flood irrigation as needed. The population was harvested on 23/06/2009 (76 DAE), approximately at anthesis for the population. Three plants of each RIL and parental lines from each of two replicates were harvested. For QTL mapping, the average of the two replications was used. Plants were phenotyped for total height, which was measured from the base of the plant to the top of the panicle.

DNA was extracted from leaf tissue harvested from each RIL and processed using ZR Plant/Seed DNA MiniPrep (Zymo Research). Digital Genotyping (DG) was performed as previously described [124] using the enzyme NgoMIV to digest genomic DNA. Reads were mapped to the reference genome and variants were processed as described for the BTx623 x IS3620c RIL population. The genetic map was constructed using R/qtl (n=93) after removing any markers that did not define a recombination breakpoint. QTL mapping was also performed in R/qtl using IM with 1000 permutations and an α=0.05 [128].

Fine Mapping of *Dw2*

The BTx642 x Tx7000 RIL population was used for fine mapping *Dw2*. Lines that had recombination breakpoints in or near *Dw2* were used to delimit the locus to the extent possible using additional DG genotypes and SNPs identified by Sanger sequencing genes in the region. Primers used for Sanger sequencing are listed in Table A4. All PCR amplification was done with Phusion© High-Fidelity DNA Polymerase (New England BioLabs, Inc.) using the standard conditions. The PCR product was gel purified using QIAquick Gel Extraction Kit (Qiagen) and prepared for capillary sequencing with BigDye© Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) using standard reaction conditions. Sequencing was performed with the ABI 3130xl Genetic Analyzer (Applied Biosystems) and the results were analyzed with Sequencher v4.8 (Gene Codes Corp.).

RILs with recombination breakpoints in the delimited *Dw2* region were grown to confirm stem and internode length phenotypes. Two pots containing two plants from each RIL were grown in two different greenhouses for a total of eight plants per RIL; otherwise the RILs were grown in the same manner as DYM and DDYM. At anthesis, the plants were harvested and the total length of the stem (measured from the base of the plant to the base of the panicle) and the length of each internode were recorded.

Sequencing of Genes in the Genomic Region Spanning *Dw2*

Once the region encoding *Dw2* was delimited to the extent possible with available genetic resources, the genes in this region were sequenced to search for functional mutations that distinguish DYM (*Dw2*) from DDYM (*dw2*). The genes in the *Dw2* locus were identified using the sorghum reference genome version 3.1 gene set (*Sorghum bicolor* v3.1 DOE-JGI, http://phytozome.jgi.doe.gov/). The primers for sequencing the genes are listed in Table A5 and capillary sequencing was performed as with fine mapping SNPs. DDYM was identified as a short plant in a field of DYM and alleles of *Dw2* differentiate the two genotypes [6]. For Sobic.006G067600 only the exons were sequenced, for all other genes, the entire gene was sequenced. Sobic.006G067700 was further sequenced in the other important breeding lines to examine the distribution and extent of allelic variation in *Dw2*.

Whole Genome Sequencing

Whole genome sequencing was used to identify polymorphisms that distinguish the parents of the two populations used to map *Dw2*. Tx7000 and BTx642 seeds were obtained from Dr. W.L. Rooney (Dept of Soil and Crop Sciences, TAMU). IS3620c seed (PI 659986 MAP) was obtained from the USDA-ARS Plant Genetic Resources Conservation Unit (Griffin, GA). Seeds were soaked in 20% bleach for 20 minutes and washed extensively in distilled water for one hour. Seeds were germinated on water-saturated germination paper in a growth chamber (14 hr light; 30° C/10 hr dark; 24° C). Genomic DNA was isolated from 8-day old root tissue using a FastPrep DNA Extraction kit and FastPrep24 Instrument (MP Biomedicals LLC, Solon, OH, USA), according to the manufacturer's specifications. DNA template (350 bp average insert size) was prepared using a TruSeq® DNA PCR-Free LT Kit, according to the manufacturer's directions. Paired-end sequencing (125 x 125 bases) was performed on an Illumina HiSeq2500. Sequence reads were mapped to version 3 of the sorghum reference genome assembly (*Sorghum bicolor* v3.1 DOE-JGI, http://phytozome.jgi.doe.gov/), using BWA v0.7.12 [156]. Base quality score recalibration, INDEL realignment, duplicate removal, joint variant calling, and variant quality score recalibration were performed using GATK v3.3 with the RIG workflow [157–160]. Whole genome sequence of Tx7000, BTx6424, and IS3620c are available at the Sequence Read Archive (www.ncbi.nlm.nih.gov/sra).

Protein Sequence Analysis

Each of the AGCVIII proteins in Arabidopsis was aligned with the sorghum genome using BLAST and the best hits were recorded. The resulting sorghum AGCVIII protein family was used to make a phylogenetic tree in MEGA6 [133]. The sequences were aligned using the MUSCLE algorithm [134,135]. The tree was estimated using maximum likelihood with the substitution model developed by Le & Gascuel [164] and the Gamma distribution. To estimate the reliability of the branches, 1000 boostraps were performed. Protein alignments were performed in Jalview v2.0 [131] using the TCoffee algorithm [132] with defaults.

RESULTS

Comparison of DYM and DDYM Internode Lengths

The recessive *dw2* allele present in Double Dwarf Yellow Milo (DDYM), the original source of *dw2*, arose as a mutation in Dwarf Yellow Milo (DYM) [6,9]. DYM and DDYM are both photoperiod sensitive (*Ma1*) [6] and when grown in long days, these genotypes showed delayed flowering relative to photoperiod insensitive plants and produced ~25 elongated internodes prior to anthesis. Comparison of DYM and DDYM stem internode lengths at grain maturity showed that the recessive allele of *dw2* in DDYM caused a reduction in the length of nearly every elongated internode compared to the corresponding internodes in DYM (Fig 15). The *dw2* allele found in DDYM was used extensively in U.S. grain sorghum breeding programs and the Sorghum Conversion

Program [9] to reduce the length of stems of sorghum genotypes such as IS3620c and BTx642 that were used in this study to clone *Dw2*.

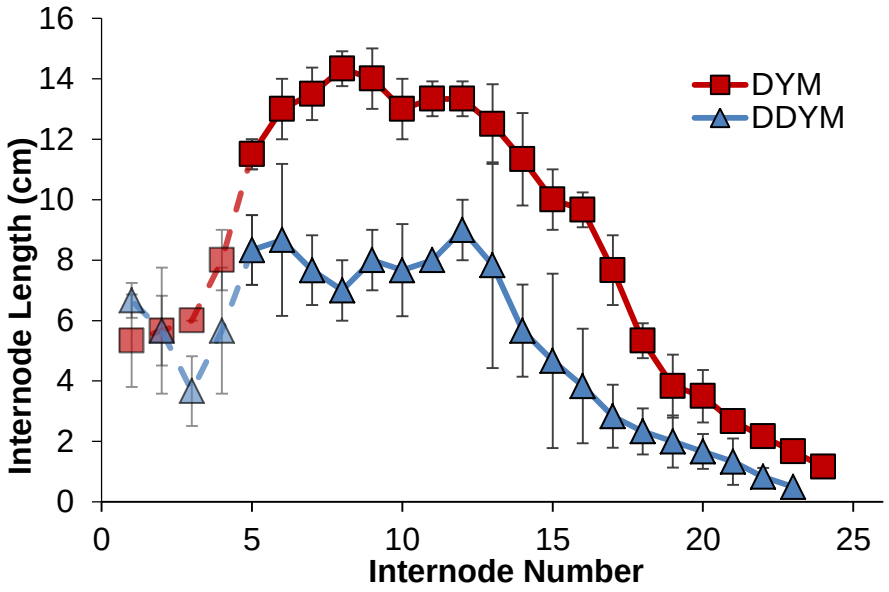


Fig 15. Internode lengths of the Yellow Milos. DYM and DDYM (n=3 per line) were grown in the greenhouse in the summer. The head of DYM died from stress, so the length of the first few internodes may not be representative of normal growth of DYM and so are represented with a dotted line and lightened points. DDYM did flower and produced seed. At grain maturity of DDYM, the plants were measured for the length of each internode, numbered from the peduncle, with the average and standard deviation shown.

QTL Mapping Using a RIL Population Derived from a Cross of BTx623 x IS3620c

QTL for total stem length, average internode length, the length of each internode numbered from the peduncle, and the length of the peduncle were mapped using the BTx623 x IS3620c RIL population (Fig 16, Fig 17, Table 10). As expected the

population segregated for *Dw2* on chromosome 6 (~42.7 Mb) and for *Dw3* on chromosome 7 (~59.8 Mb) and these loci affected both total stem length and internode length. An additional QTL (*Dw03_67.5*) at ~67.5 Mb on chromosome 3 affected total stem length (Fig 16). The influence of *Dw2* and *Dw3* on the length of the eight internodes was analysed to determine if the action of these genes varies with development (Table 10). *Dw3* affected the length of all eight internodes measured. *Dw2* influenced the length of the first five internodes but had minimal impact on the length of internodes 7-8. There is an additional QTL on chromosome 6 (48.6 Mb, *Dw06_48.6*) near *Dw2* segregating for the length of the sixth internode below the peduncle. However, the peaks for the fifth and sixth internode are broad and the 2-LOD interval for the peak on chromosome 6 for both internodes includes both *Dw2* and *Dw06_48.6* (Table 10, Fig 17). The additive effect of *Dw2* and *Dw3* on internode length varied with internode number (Fig 18). The additive effect for *Dw2* was highest for the internode immediately below the peduncle. The additive effect of *Dw3* on the length of the same internode was similar to that of *Dw2*. However, *Dw3* influenced the length of internodes formed earlier in development more than *Dw2*. The additive effect of *Dw3* decreased from the sixth to eighth internodes (Fig 18). QTL for peduncle length did not align with *Dw2* or *Dw3* (Fig 17).

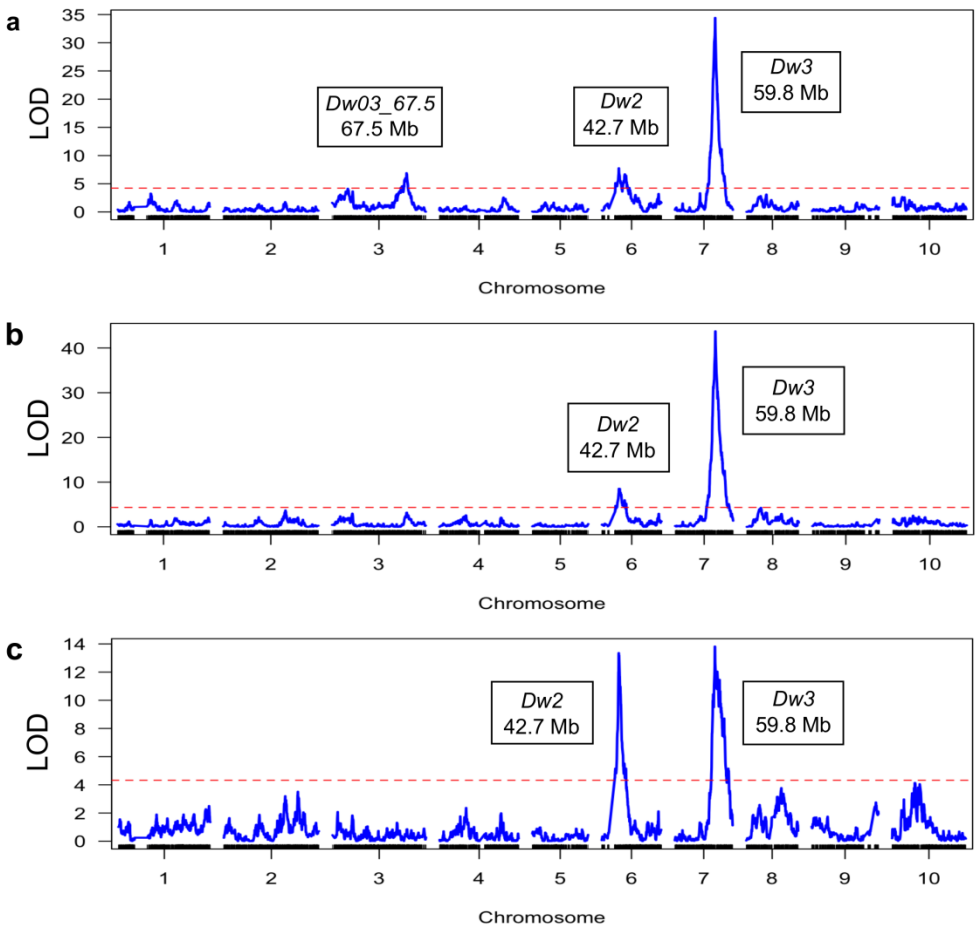


Fig 16. QTL identified using the BTx623 x IS3620c RIL population. The RIL population was grown in the greenhouse and genotyped using DG**.** Stem length (**a**) was measured from the base of the plant to the base of the panicle. Genetic map generation and QTL mapping were performed in R/qtl using interval mapping (IM). The x-axis is the markers along the chromosomes and the y-axis is the LOD score. The significant QTL peaks are labeled with the *Dw* locus and location (Mb). Stem length (**a**), average internode length (**b**), and the length of the first internode below the peduncle (**c**) are shown.

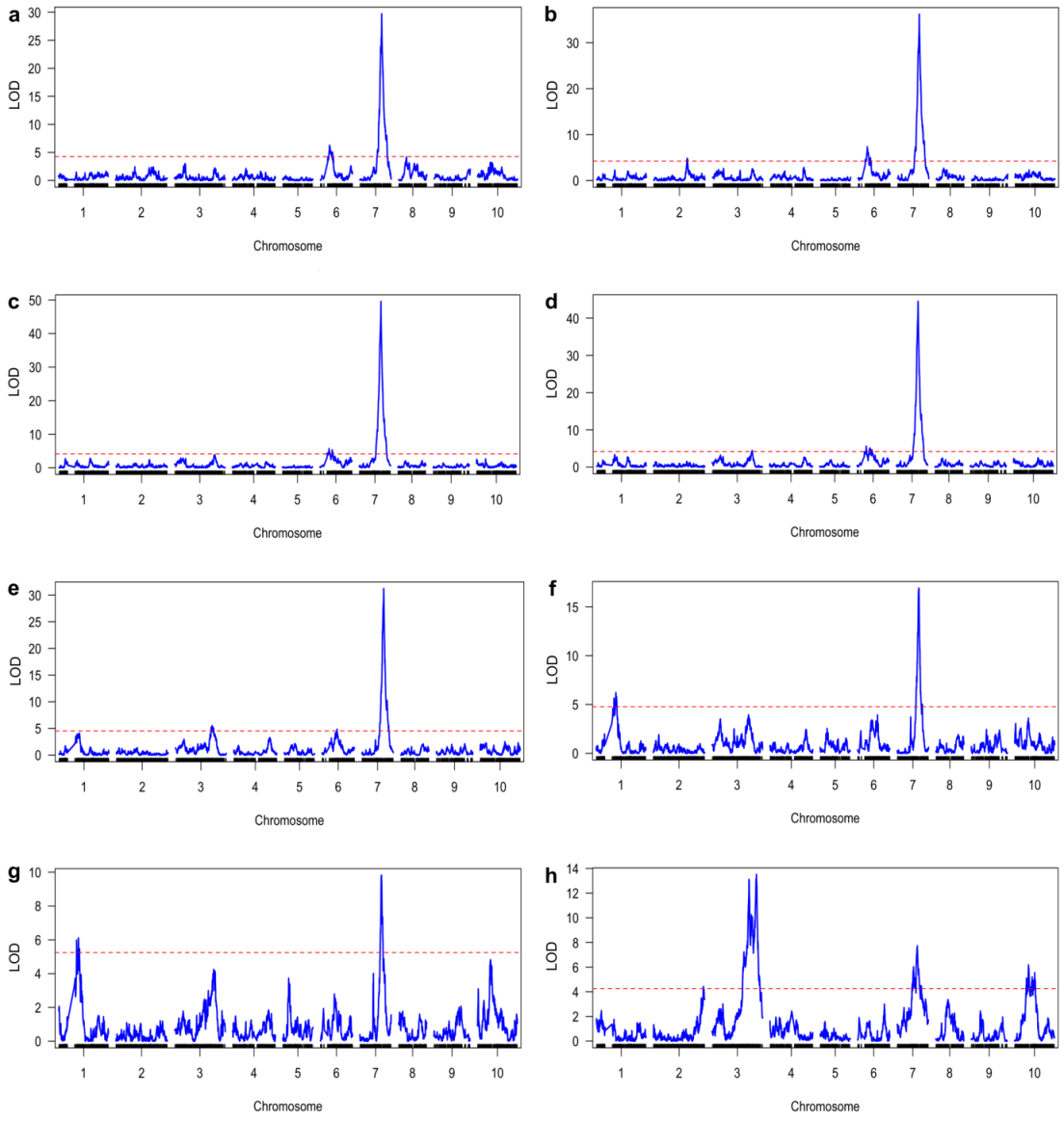


Fig 17. Internode length QTL identified using BTx623 x IS3620c RILs. The RIL population was grown in the greenhouse and genotyped using DG. QTL mapping was performed in R/qtl using IM. For each graph, the markers from the genetic map are listed on the x-axis and the LOD score on the y-axis. Each graph is for a different internode starting with the second internode below the peduncle (**a**) and ending with the eighth internode below the peduncle (**g**). The last graph is the peduncle (**h**).

Table 10. QTL Segregating for Stem Traits in the BTx623 x IS3620c Population. Internodes are numbered from the peduncle. Chromosome is abbreviated as "Chr". A positive additive effect indicates that the IS3620c allele increases length.

Trait	Chr	Peak (bp)	Peak (cM)	LOD	2-LOD Interval		Additive Effect
					Start	Stop	
Total Length	3	67,503,832	136.10	6.89	65,530,485	68,206,300	-12.48
	6	42,691,080	31.14	7.76	42,355,109	46,697,460	-13.29
	7	59,830,285	73.54	34.43	59,654,592	59,867,828	26.00
Average Internode Length	6	42,691,080	31.14	8.52	42,355,109	44,831,591	-20.25
	7	59,830,285	73.54	43.75	59,654,592	59,847,033	41.05
Length Peduncle	2	76,607,596	169.46	4.42	74,943,883	77,320,040	15.37
	3	70,750,399	150.22	13.51	62,718,371	71,404,420	-39.20
	7	59,086,124	68.41	7.73	55,545,487	59,785,398	-31.25
	10	7,100,563	47.03	6.17	5,639,508	48,402,197	-26.97
Length Internode 1	6	42,691,080	31.14	13.35	42,355,109	43,632,616	-25.62
	7	59,785,398	73.44	13.82	59,533,447	60,458,272	25.77
Length Internode 2	6	42,691,080	31.14	6.27	41,934,840	45,943,225	-20.81
	7	59,785,398	73.44	29.72	59,654,592	59,991,087	41.65
Length Internode 3	2	64,347,846	113.94	4.83	63,835,432	64,886,659	9.50
	6	42,691,080	31.14	7.36	42,051,620	45,706,034	-22.53
	7	59,830,285	73.54	36.21	59,631,468	59,847,033	45.81
Length Internode 4	6	42,691,080	31.14	5.71	38,080,498	46,697,460	-19.04
	7	59,830,285	73.54	49.53	59,654,592	59,847,033	49.70
Length Internode 5	3	67,503,832	136.10	4.45	3,482,238	68,957,430	-15.62
	6	42,691,080	31.14	5.60	39,022,638	49,672,003	-16.98
	7	59,830,285	73.54	44.48	59,654,592	59,847,033	44.36
Length Internode 6	3	62,683,672	123.74	5.50	60,818,299	66,423,271	-15.63
	6	48,641,758	50.58	4.83	42,551,078	50,220,562	-13.67
	7	59,830,285	73.54	31.19	59,654,592	59,991,087	34.42
Length Internode 7	1	56,402,777	66.50	6.20	20,256,774	58,060,819	15.54
	7	59,785,398	73.44	16.91	59,481,526	59,991,087	23.64
Length Internode 8	1	56,499,134	66.61	6.10	24,523,367	58,177,975	14.12
	7	59,628,954	72.90	9.82	59,277,216	59,991,087	17.18

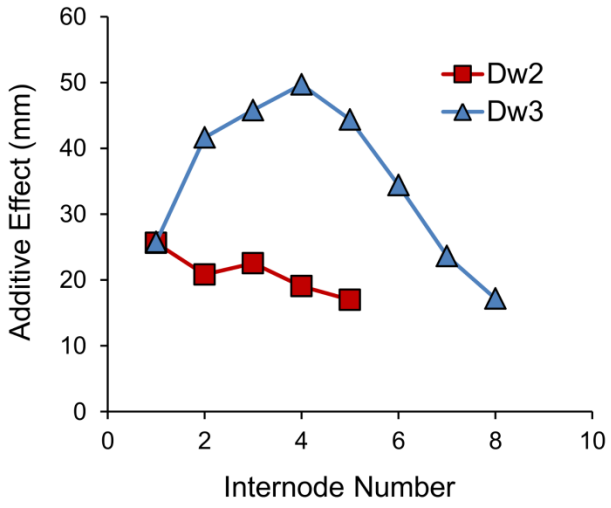


Fig 18. Additive effects of *Dw2* and *Dw3* on the length of each internode (BTx623 x IS3620c RIL). The RIL population was grown in the greenhouse and the length of each internode (numbered from the peduncle) was measured. Additive effects were determined as part of QTL mapping performed in R/qtl using IM. The BTx623 allele of *Dw2* increases internode length, whereas the IS3620c allele of *Dw3* increases internode length.

There was no strong statistical evidence of a genetic interaction between *Dw2* and any of the other loci from the multiple-QTL mapping (MQM) analysis (Table 11). For the best model for each phenotype, the only phenotype that included interactions in the model with the highest LOD is the length of internode 7. There are two interactions in this model, one between a QTL on chromosome 5 and *Dw3* (chromosome 7 at 59.8 Mb) and another between a QTL on chromosome 1 and a QTL close to *Dw3* (10.7 cM from *Dw3* at 61.2 Mb) (Table 11). The composite multiple-QTL model included both interactions and revealed interesting trends between the internode length traits with

internodes further from the peduncle having better support for the two interactions (Table 12, Fig 19). Additionally, composite model analysis clarified the effects of the two QTL on chromosome 6. *Dw2* affects the length of internodes 1-6, but starting at internode 4 and continuing through internode 7 the QTL at ~49 Mb on chromosome 6 also affected internode length (Table 12, Fig 19).

Table 11. Summary of the Best Model from MQM of Individual Phenotypes. Chromosome is shortened to "Chr."

Trait	QTL	Chr	Peak LOD	Peak (cM)	Peak (bp)	Start	Stop	Interactions (with Number; LOD)
Total Length	1	3	6.4	29.78	4309508	3057129	6163945	
	2	3	10.73	136.63	67760473	64467623	68260513	
	3	6	14.19	30.67	42508419	41934840	43596665	
	4	7	33.46	73.76	59847033	59247435	59991087	
	5	10	3.5	61.19	9626445	7791830	52293650	
Average Internode Length	1	2	3.18	120.91	66477452	61525510	69131669	
	2	3	5.35	136.63	67760473	65905794	72466480	
	3	6	18.33	31.71	42785280	41934840	43596665	
	4	7	42.44	73.76	59847033	59504276	59991087	
	5	8	4.03	28.3	3669596	2194037	53066186	
	6	10	7.39	60.06	9375593	8197931	11829372	
Internode 1 Length	1	6	15.59	31.71	42785280	42085051	43596665	
	2	7	16.56	73.76	59847033	59504276	61227548	
	3	10	4.22	38.41	5551100	4709177	51917685	
Internode 2 Length	1	6	9.72	29.36	42085051	39890464	43596665	
	2	7	25.15	73.76	59847033	59504276	59991087	
	3	10	5.96	60.06	9375593	6931729	54111672	
Internode 3 Length	1	3	4.57	135.53	67047035	64467623	69688959	
	2	6	14.55	29.36	42085051	41934840	43596665	
	3	7	30.09	71.36	59504276	59051589	59991087	
	4	10	4.71	60.06	9375593	6931729	51917685	

Table 11. Continued.

Trait	QTL	Chr	Peak LOD	Peak (cM)	Peak (bp)	Start	Stop	Interactions (with Number; LOD)
Internode 4 Length	1	3	7.11	135.53	67047035	65905794	69174377	
	2	6	11.29	29.36	42085051	39890464	43596665	
	3	6	3.37	64.34	51939240	49361779	53985519	
	4	7	3.98	69.75	59247435	58919267	59847033	
	5	7	5.7	73.76	59847033	59504276	59991087	
	6	10	2.94	69.97	47978440	1707726	55494359	
Internode 5 Length	1	3	7.08	135.53	67047035	63288516	68260513	
	2	6	10.02	29.36	42085051	39890464	45706034	
	3	7	36.41	71.36	59504276	59247435	59847033	
Internode 6 Length	1	1	5.21	56.53	21177180	8262098	57277940	
	2	3	8.95	120.25	61770650	60898775	63875751	
	3	6	6.54	29.36	42085051	1659623	50325848	
	4	7	26.94	71.36	59504276	59051589	59847033	
	5	8	3.88	75.53	59711692	58297740	61022028	
Internode 7 Length	1	1	10.32	66.73	56518269	54249162	58058247	6; 3.530
	2	3	6.54	120.25	61770650	60175252	63875751	
	3	5	7.29	27.92	5337348	3504889	6175733	5; 2.713
	4	6	8.07	51.47	48844243	46697460	50423659	
	5	7	13.6	71.36	59504276	59051589	59991087	3; 2.713
	6	7	3.9	82.07	61227548	60996573	63995754	1; 3.530
Internode 8 Length	1	1	6.29	65.71	55803782	8262098	58058247	
	2	7	6.04	73.76	59847033	59247435	60577582	
	3	8	4.46	75.53	59711692	57593772	62528965	

Table 12. Summary of the Best Model for Each Trait Based on Composite MQM. Includes the two interactions, notated with an "&". Chromosome is shortened to "Chr" and internode is shortened to "Int".

			LOD									
QTL	**Chr**	**Location (Mbp)**	**Total Length**	**Average Int Length**	**Length Int 1**	**Length Int 2**	**Length Int 3**	**Length Int 4**	**Length Int 5**	**Length Int 6**	**Length Int 7**	**Length Int 8**
1	1	56.52	1.73**	0.29	0.75	0.21	0.06	0.07	0.73	2.49***	4.33***	5.69***
2	2	66.48	0.32	2.7***	1.31*	2.74***	1.49*	1.2*	0.22	0.35	0.35	0.04
3	3	4.31	6.27***	2.68***	0.99*	0.92*	1.9**	2.37**	2.79***	1.75**	1.66**	0.02
4	3	67.76	10.97***	5.65***	1.41*	2.24**	3.99***	6.24***	6.16***	4.55***	1.39*	1.83**
5	5	5.34	1.11*	0.09	0.03	0.08	0.02	0.02	1.34*	2.84***	2.83***	1.21*
6	6	42.79	15.91***	20.25***	17.73***	11.82***	15.09***	13.06***	10.42***	5.17***	1.78**	0.07
7	7	59.85	18.61***	20.5***	5.34***	13.38***	16.02***	24.05***	20.58***	11.21***	10.2***	3.64***
8	7	61.23	0.01	0.98*	1.2*	0.05	0.03	0.05	0.01	0	1.55**	0.09
9	8	3.67	2.7***	4.32***	0.73	3.18***	2.65***	1.31*	1.1*	0.14	0	0.04
10	8	59.71	1.15*	0.03	0.6	0	0.12	0.45	2**	3.35***	2.08**	3.44***
11	10	9.38	4.15***	7.97***	4.33***	7.01***	5.26***	2.01**	1.11*	0.07	0.2	1.17*
1&8	n.a.	n.a.	0.01	0.2	0.19	0.07	0.03	0.95*	0.63	1.86**	1.66**	5.22***
5&7	n.a.	n.a.	0.04	0.05	0.16	0.89*	0.3	1.32*	0.51	3.8***	2.49***	4.27***

* P < 0.05, ** P < 0.01, *** P < 0.001.

Fig 19. Multiple QTL mapping (MQM) in BTx623 x IS3620c RILs. For all, MQM was performed on the same genotype and phenotype data as IM. MQM was performed in R/qtl. IM was used to seed multiple QTL model selection for each trait. The best model for each trait was combined to form the composite multiple-QTL model. This model consists of 11 QTL and two epistatic interactions. (**a**) and (**b**) a graph of the LOD score for the epistatic interaction for each internode length trait . The following eight graphs are the phenotype (y-axis) for each combination of genotype (x-axis and series) for each internode length (1-8). (**a**) is the interaction between a QTL on chromosome 1 and a QTL near *Dw3* while (**b**) is the interaction between a QTL on chromosome 5 and *Dw3*. (**c**) A heat map of the LOD value for each DG marker for each trait based on composite MQM analysis.

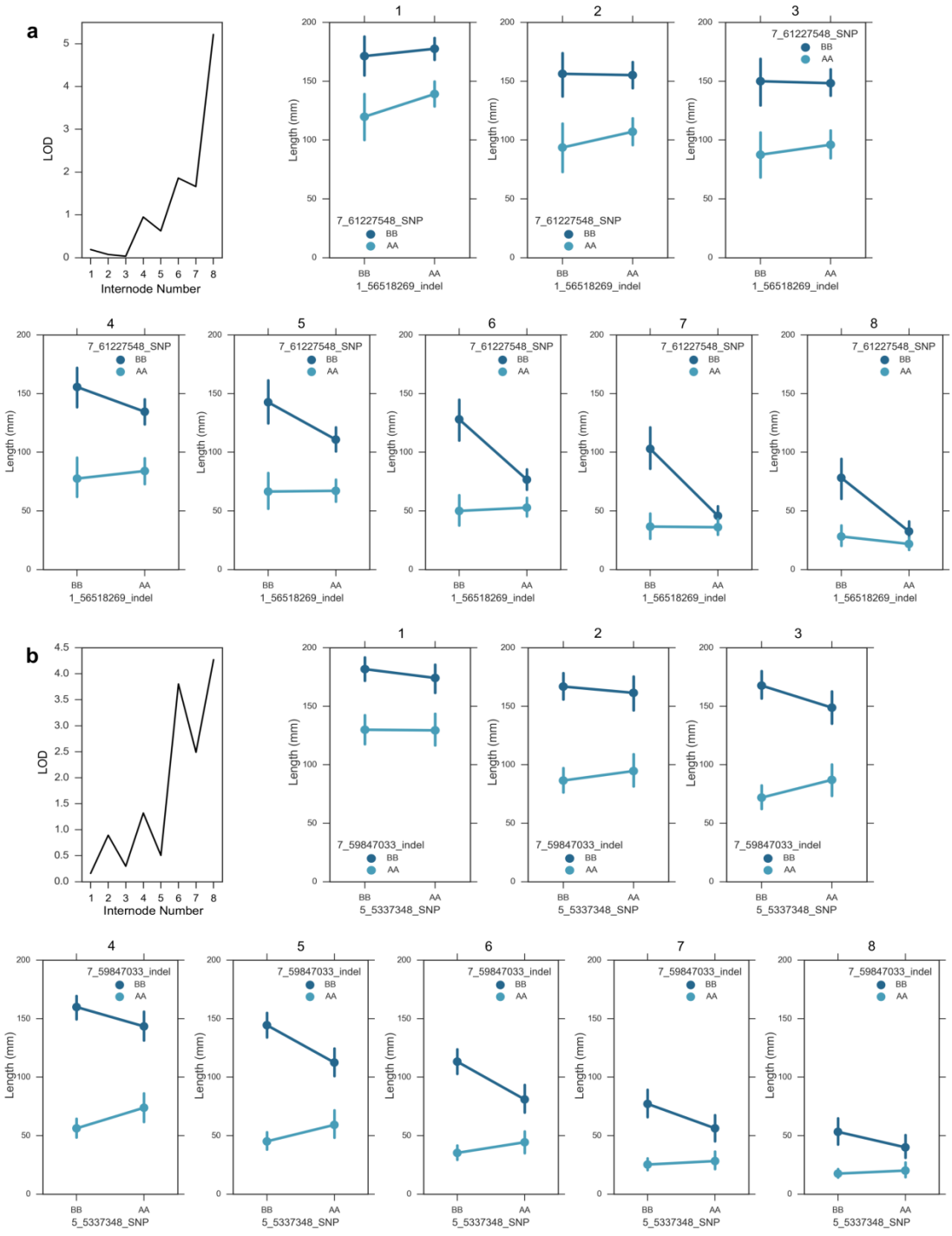
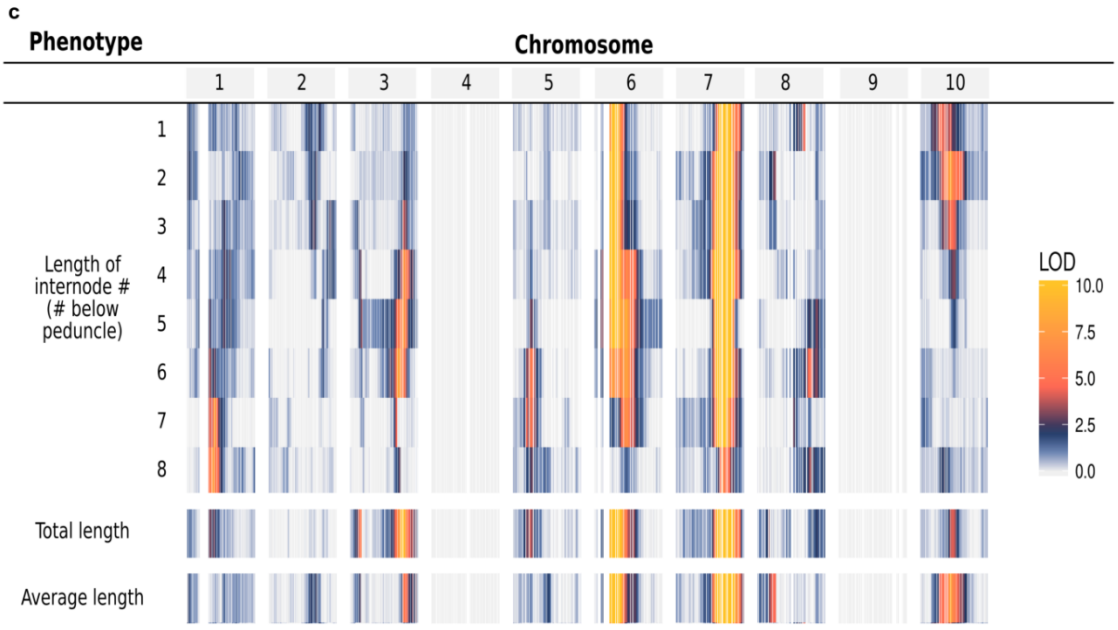

a
LOD
Internode Number
1
2
3
4
5
6
7
8
Length (mm)
7_61227548_SNP
BB
AA
1_56518269_indel
b
LOD
Internode Number
7_59847033_indel
5_5337348_SNP

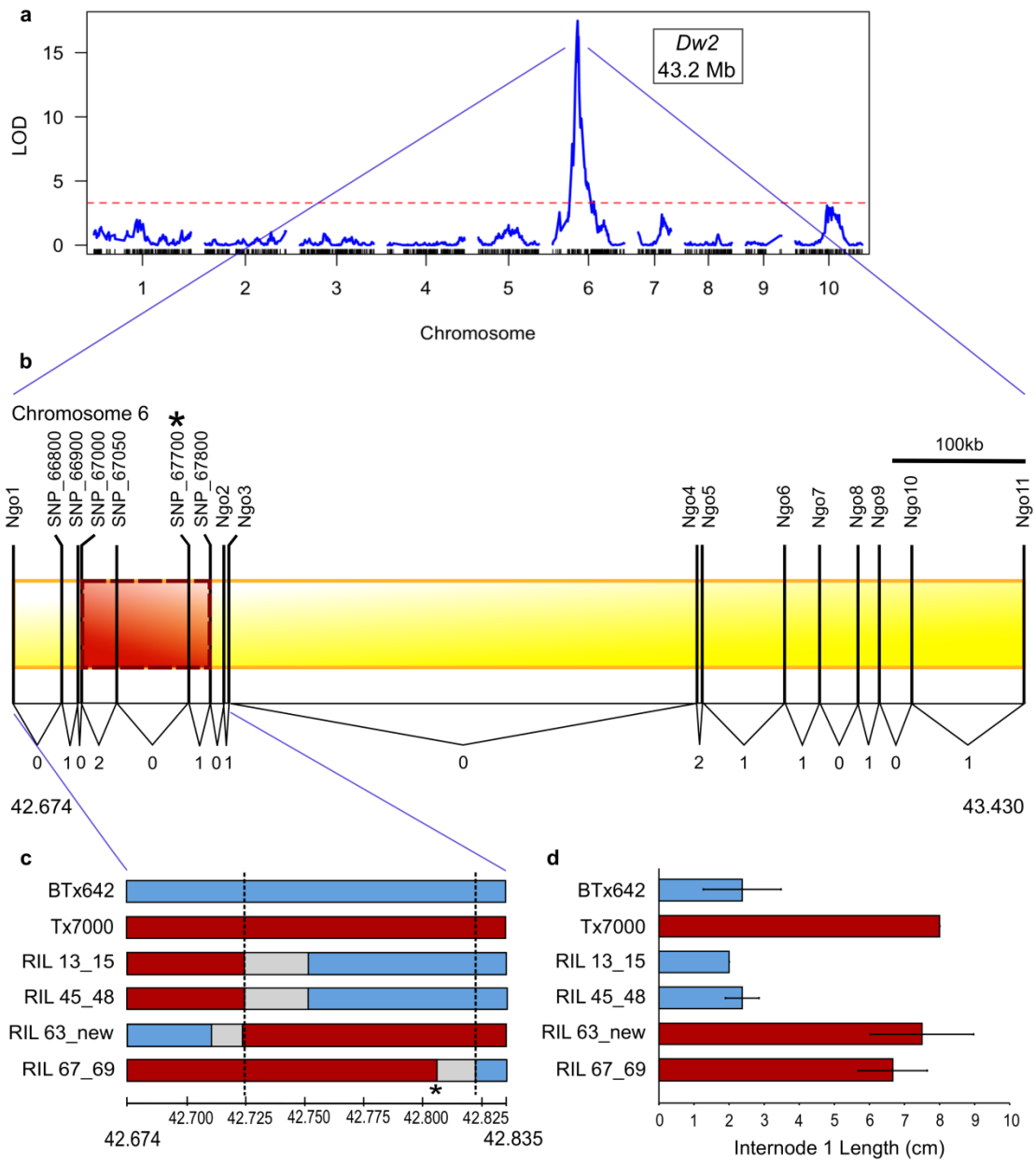
Fig 19. Continued.



Dw2 Fine Mapping and Gene Identification

Dw2 was fine mapped in a second RIL population derived from BTx642 x Tx7000 that was expected to segregate for alleles of *Dw2* in a background fixed for recessive *Dw1*, *Dw3*, and *Dw4*. QTL analysis of BTx642 x Tx7000 RILs for total plant height revealed a major QTL aligned with *Dw2* as expected (Fig 20a). The QTL corresponding to *Dw2* showed a peak located on chromosome 6 at ~43.2 Mb. The 2-LOD interval containing *Dw2* in the BTx642 x Tx7000 RIL population spanned a region of ~756 kb on chromosome 6. Eight RILs with recombination breakpoints in this region were identified and targeted for higher resolution analysis of breakpoint locations. Sequence polymorphisms within the target interval identified using high resolution DG analysis and by targeted gene sequencing were used to fine map the breakpoints in the eight fine mapping lines (Fig. 20b). Four RILs with breakpoints closest to *Dw2* were phenotyped in a greenhouse during the winter. Phenotyping in the winter under low light conditions revealed that *Dw2* had a large impact on the length of the internode below the peduncle. As a consequence, RILs containing *Dw2* could be readily distinguished from RILs encoding *dw2* by phenotyping eight plants from each genotype for the length of the internode below the peduncle (Fig 20c and d, Fig 21). The information from lines with breakpoints delimited the *Dw2* locus to a region spanning ~98.1 kb containing ten genes (Fig. 20, Table 13, Table 14). The genes within this region were annotated in Phytozome as encoding a PPR repeat protein, an rRNA N-glycosylase, an F-box protein, a glycogen branching enzyme, a phosphatase, a histone deacetylase, a kinase, and three genes of unknown function.

Fig 20. Fine mapping of *Dw2* in the BTx642 x Tx7000 RIL population. (**a**) QTL map of total plant height (2009) with *Dw2* labeled. Plant height was measured as the length of the plant from the base of the stem at ground level to the top of the panicle. Genetic map construction and QTL analysis were performed in R/qtl using IM. The x-axis is the markers along the chromosomes and the y-axis is the LOD value. (**b**) Diagram of fine mapping in BTx642 x Tx7000. The diagram shows the location of the recombination breakpoints in the 2-LOD region in the eight fine mapping lines (numbers at bottom), two of these lines had more than one recombination breakpoint in the region. The markers found through DG using NgoMIV are labeled as "Ngo_". The markers found with Sanger sequencing are labeled with "SNP_ _" with the last five digits of the gene name. The red, dashed-line box shows the refined region of *Dw2*. For both (**b**) and (**c**), asterisk indicates the approximate location of *Dw2*. (**c**) Diagram of the haplotypes of the four fine mapping lines with breakpoints closest to the refined region. The region between Ngo1 and Ngo3 is shown. Blue indicates that the RIL has the BTx642 allele, red is the Tx7000 allele, and grey is the region where the breakpoint is located. Dashed lines flank the refined region of *Dw2*. (**d**) The length of the first internode below the peduncle in the same lines shown in (**c**). Blue indicates that the line is *dw2* while red is *Dw2*. Average (n=4) and standard deviation is shown.


a
Dw2
43.2 Mb
LOD
Chromosome
b
Chromosome 6
100kb
Ngo1
SNP_66800
SNP_66900
SNP_67000
SNP_67050
SNP_67700
SNP_67800
Ngo2
Ngo3
Ngo4
Ngo5
Ngo6
Ngo7
Ngo8
Ngo9
Ngo10
Ngo11
42.674
43.430
c
BTx642
Tx7000
RIL 13_15
RIL 45_48
RIL 63_new
RIL 67_69
42.674
42.835
d
Internode 1 Length (cm)

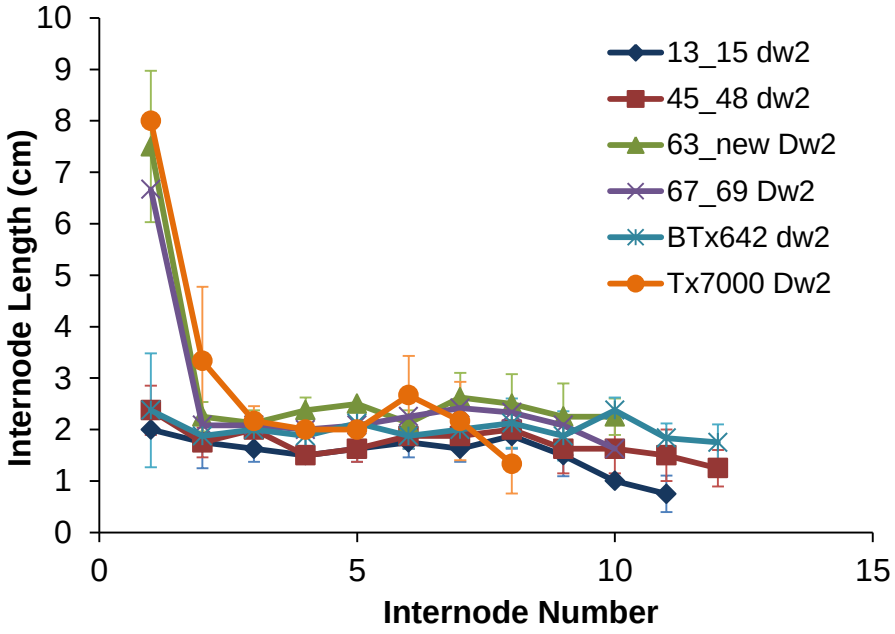


Fig 21. Internode length phenotypes for select BTx642 x Tx7000 RILs. The RILs (n=4 per line) are the lines that had a close breakpoint in the *Dw2* delimited region. At grain maturity, the length of each internode was measured with the average and standard deviation shown. These lines were grown in the winter under low light intensity in two different greenhouses. Only one greenhouse set is shown, though both are similar.

Table 13. Genes in the Delimited Region of *Dw2*.

Gene	Description	Location (v3.1)
Sobic.006G067000	PPR repeat	42,723,881-42,725,688
Sobic.006G067050	Unknown	42,751,421-42,752,998
Sobic.006G067100	rRNA N-glycosylase	42,753,303-42,756,717
Sobic.006G067150	Unknown	42,758,806-42,759,413
Sobic.006G067200	Unknown	42,760,512-42,761,535
Sobic.006G067300	F-box domain	42,769,007-42,770,832
Sobic.006G067400	1,4-alpha-glucan branching enzyme; Calcineurin-like phosphoesterase	42,774,078-42,778,987
Sobic.006G067500	Calcineurin-like phosphoesterase; Ser/Thr protein phosphatase family protein; Prespore protein DP87	42,781,244-42,785,442
Sobic.006G067600	Histone deacetylase	42,785,485-42,802,516
Sobic.006G067700	Ribosomal protein S6 kinase; Protein tyrosine kinase	42,803,037-42,807,134

Table 14. Additional Notes on the Genes in the Delimited Region of *Dw2*.

Gene	Description	Maize Homolog(s)	Notes
Sobic.006G067000	PPR repeat	GRMZM2G163043	
Sobic.006G067050	Unknown	None	Low levels of gene expression
Sobic.006G067100	rRNA N-glycosylase	GRMZM2G013331 & GRMZM2G022095	Maize homologs lack the first part of the gene
Sobic.006G067150	Unknown	GRMZM2G017933	BLAST match has limited percent identity
Sobic.006G067200	Unknown	None	
Sobic.006G067300	F-box domain	GRMZM2G015349 & GRMZM2G125954 & GRMZM2G435096	BLAST matches have limited percent identity; very low levels of gene expression
Sobic.006G067400	Calcineurin-like phosphoesterase	GRMZM2G128399	
Sobic.006G067500	Calcineurin-like phosphoesterase	GRMZM2G128399	First ~220 residues lack homology to maize homolog
Sobic.006G067600	Histone deacetylase	GRMZM2G119703	
Sobic.006G067700	Ribosomal protein S6 kinase	GRMZM2G412524 & GRMZM2G128319	

Sequence Analysis of Genes in the *Dw2* Locus

The gene corresponding to *dw2* is expected to contain a mutation(s) that decreases function; therefore, all of the genes in the delimited *Dw2* locus (Table 13) were sequenced from DYM and DDYM. Only one polymorphism was found in the delimited *Dw2* locus that distinguished DYM from DDYM, an INDEL in Sobic.006G067700 located in the first exon at 549 bp that causes a frameshift resulting in a stop codon at 573 bp. This mutation changed the amino acid sequence after E183 resulting in a truncated polypeptide containing 190 amino acids instead of the 809 amino acids present in the full-length protein. The INDEL mutation in Sobic.006G067700 that causes protein truncation was also present in BTx642 and IS3620c, genotypes that acquired *dw2* by introgression from DDYM, and not present in BTx623 (*Dw2*) and

Tx7000 (*Dw2*) (Table 15 and 16). None of the parental lines contain polymorphisms in the coding region of the histone deacetylase (Sobic.006G067600), a gene previously proposed as a candidate for *Dw2* [112]. A number of sequence variants in the *Dw2* delimited region were identified that distinguished the parental mapping lines (Table 17); however, none of these variants differentiated DYM (*Dw2*) from DDYM (*dw2*), the source of the recessive allele of *dw2*.

Table 15. Polymorphisms in Sobic.006G067700.

Number	Polymorphism	Location	Region	Result	SIFT
1	SNP; C > T	-138 bp	5'UTR		
2	11 bp INDEL	-132 -> -122bp	5'UTR		
3	INDEL; GA > -	549 bp	Exon 1	Stop codon at 573 bp	
4	SNP; G > A	650 bp	Exon 1	Glycine > Aspartic Acid	0.09=tolerated
5	SNP; A > C	1279 bp	Exon 1	Isoleucine > Leucine	0.17=tolerated
6	SNP; G > A	2561 bp	Exon 2	Cysteine > Tyrosine	0.11=tolerated

Table 16. Selected Genotypes Scored at the Polymorphisms Listed in Table 3.

Line	*Dw2*	1	2[a]	3	4	5	6
BTx623	*Dw2*	C	+	GA	G	A	G
IS3620c	*dw2*	T	-	-	A	C	A
Tx7000	*Dw2*	C	+	GA	G	A	G
BTx642	*dw2*	T	-	-	A	C	A
Standard Yellow Milo	*Dw2*	T	-	GA	A	C	A
Dwarf Yellow Milo	*Dw2*	T	-	GA	A	C	A
Double Dwarf Yellow Milo	*dw2*	T	-	-	A	C	A
80M	*dw2*	T	-	-	A	C	A
SC170	*dw2*	T	-	-	A	C	A
BTx406	*dw2*	T	-	-	A	C	A
Early White Milo	*Dw2*	C	+	GA	G	A	G
Texas Blackhull Kafir	*Dw2*	C	+	GA	G	A	G
Spur Feterita	*Dw2*	C	+	GA	G	C	A
Sumac	*Dw2*	C	+	GA	G	A	G

[a] The minus sign indicates that the genotype has the deletion.

Table 17. Polymorphisms between the Parental Genotypes in the Exons of the Genes in the *Dw2* Region. Parental genotypes are based on whole genome sequencing while the yellow milo genotypes are based on Sanger sequencing. The only polymorphism between DYM and DDYM is bolded.

Gene	#	Type	Polymorphism (bp; aa)	Location in Gene (bp)	Genotypes Same as Reference	Genotypes that Differ from Reference
Sobic.006G067000	none					
Sobic.006G067050	1	INDEL	- > TACCGA; T > IPT	273	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	2	SNP	C > A; F > L	308	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	3	SNP	C > T; T > I	1020	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	4	SNP	C > T; Q > stop	1043	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	5	SNP	G > C; S > T	1110	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	6	SNP	T > C; F > F	1138	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	7	SNP	A > G; K > E	1154	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	8	SNP	A > G; S > S	1174	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
Sobic.006G067100	1	First exon is missing or poorly aligned reads			Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	2	SNP	T > G; D > E	2656	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	3	INDEL	- > GATCTA; C > WIY	2787	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	4	SNP	A > C; V > V	3094	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	5	SNP	T > C; V > A	3105	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
Sobic.006G067150	1	SNP	G > T; E > stop	28	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	2	INDEL	- > 31bp sequence; premature stop (139aa > 82aa)	430	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	3	SNP	A > G; Y > C	582	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
Sobic.006G067200	none					

Table 17. Continued.

Gene	#	Type	Polymorphism (bp; aa)	Location in Gene (bp)	Genotypes Same as Reference	Genotypes that Differ from Reference
Sobic.006G067300	1	SNP	G > A; L > L	39	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	2	SNP	T > C; C > R	82	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	3	INDEL	- > T; premature stop (480aa > 234aa)	617	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	4	SNP	A > G; E > E	657	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	5	SNP	G > T; G > C	1553	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
Sobic.006G067400	1	INDEL	- > CTTCGCT; premature stop (527aa > 82aa)	12	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	2	SNP	G > C; G > R	73	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	3	SNP	C > A; R > R	82	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	4	SNP	C > G; L > V	106	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	5	SNP	T > C; F > F	111	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	6	SNP	C > G; Q > E	476	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	7	SNP	G > C; E > Q	497	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	8	SNP	C > T; C > C	559	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	9	INDEL	GTCCGA > -; VR > -	773	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	10	SNP	C > T; P > P	1162	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
Sobic.006G067500	1	SNP	T > C; F > S	17	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	2	SNP	C > A; L > I	184	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	3	SNP	C > G; R > R	405	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	4	SNP	A > G; K > K	444	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	5	SNP	G > C; L > L	450	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	6	SNP	A > C; S > S	516	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	7	SNP	C > A; R > R	2117	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	8	SNP	A > G; Q > Q	3511	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	9	SNP	T > C; I > I	3613	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c

Table 17. Continued.

Gene	#	Type	Polymorphism (bp; aa)	Location in Gene (bp)	Genotypes Same as Reference	Genotypes that Differ from Reference
Sobic.006G067600	none					
Sobic.006G067700	**1**	**INDEL**	**GA > -; aa sequence differs after 183, truncated polypeptide of 190 aa**	**549**	**DYM, Tx7000, BTx623**	**DDYM, BTx642, IS3620c**
	2	SNP	G > A; G > D	650	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	3	SNP	A > C; I > L	1279	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c
	4	SNP	G > A; C > Y	2561	Tx7000, BTx623	DYM, DDYM, BTx642, IS3620c

Dw2 Alleles in Sorghum Germplasm

The number and distribution of *Dw2* alleles in historically important sorghum genotypes was investigated by sequencing Sobic.006G067700 from the genotypes listed in Tables 15 and 16. All of the genotypes identified by Quinby and Karper [104] as *dw2* contain the INDEL in Sobic.006G067700 derived from DDYM. Several genotypes contained polymorphisms in exons that changed the protein sequence encoded by Sobic.006G067700. However, SIFT [144] analysis predicted that those polymorphisms would be tolerated and not disrupt function (Table 15). As expected, the haplotypes of the two *dw2* recessive RIL population parents, BTx642 and IS3620c, were the same as the progenitor lines DDYM and BTx406. The haplotypes of the two *Dw2* dominant RIL population parents, BTx623 and Tx7000, were the same as the progenitor line Texas Blackhull Kafir (Table 16).

Dw2 is Homologous to the AGCVIII Protein Kinase KIPK

Fine mapping, sequence analysis, and gene annotation indicates that Dw2 is a protein kinase encoded by Sobic.006G067700 (Phytozome). Genes in other plants with the greatest sequence similarity to Sobic.006G067700 include LOC_Os12g29580 (rice), GRMZM2G412524 (maize), GRMZM2G128319 (maize), and At3G52890 (Arabidopsis) (Phytozome). At3G52890 encodes an ACGVIII kinase called KIPK, a KCBP-interacting protein kinase [165]. In Arabidopsis there are 23 members of the AGCVIII kinase subfamily that has been further subdivided into four groups, AGC1–AGC4. KIPK and D6 PROTEIN KINASE/D6 PROTEIN KINASE LIKEs

(D6PK/D6PKLs) are members of the AGC1 group. A BLAST search of the Arabidopsis AGCVIII kinase gene family to the sorghum genome identified 21 sorghum homologs (Fig 22). Among these genes, Dw2 was the best BLAST hit for the Arabidopsis KIPK1, KIPK2 (AGC1-9) and AGC1-8. KIPK1 and KIPK2 also aligned well with a related gene in sorghum, Sobic.008G096200. Since the correspondence between AtKIPK1, AtKIPK2 and the two sorghum homologs could not be assigned, we designated Dw2 as SbKIPK and Sobic.008G096200 as SbKIPK-like. The relationship among the 21 members of the sorghum AGCVIII subfamily was analysed by constructing a phylogenetic tree (Fig 22). The sorghum genes clustered into four groups, as in Arabidopsis, though the closest sorghum homolog to AGC1-12 (Sobic.005G036500) groups with the AGC3s. If this gene is excluded from the sorghum AGC1 subfamily, then sorghum has three fewer members of the AGC1 group than Arabidopsis. Interestingly, while similar phylogenetic trees of the Arabidopsis AGC1 subfamily showed KIPK1 and KIPK2 grouping with AGC1-8 [166–168], the sorghum AGC1 family has only two genes on that branch, Sobic.006G067700 (Dw2, SbKIPK) and Sobic.008G096200 (SbKIPK-like) (Fig 22). The sorghum AGC1 group also includes a cluster of four sorghum genes that correspond to the four Arabidopsis genes that encode D6PK/D6PKLs. The sorghum AGC3 group has five members, including the AGC1-12 homolog, with two genes matching with PID and one gene corresponding with the WAGs. The remaining two groups of sorghum genes corresponding to AGC2 and AGC4 are similar to Arabidopsis (Fig 22).

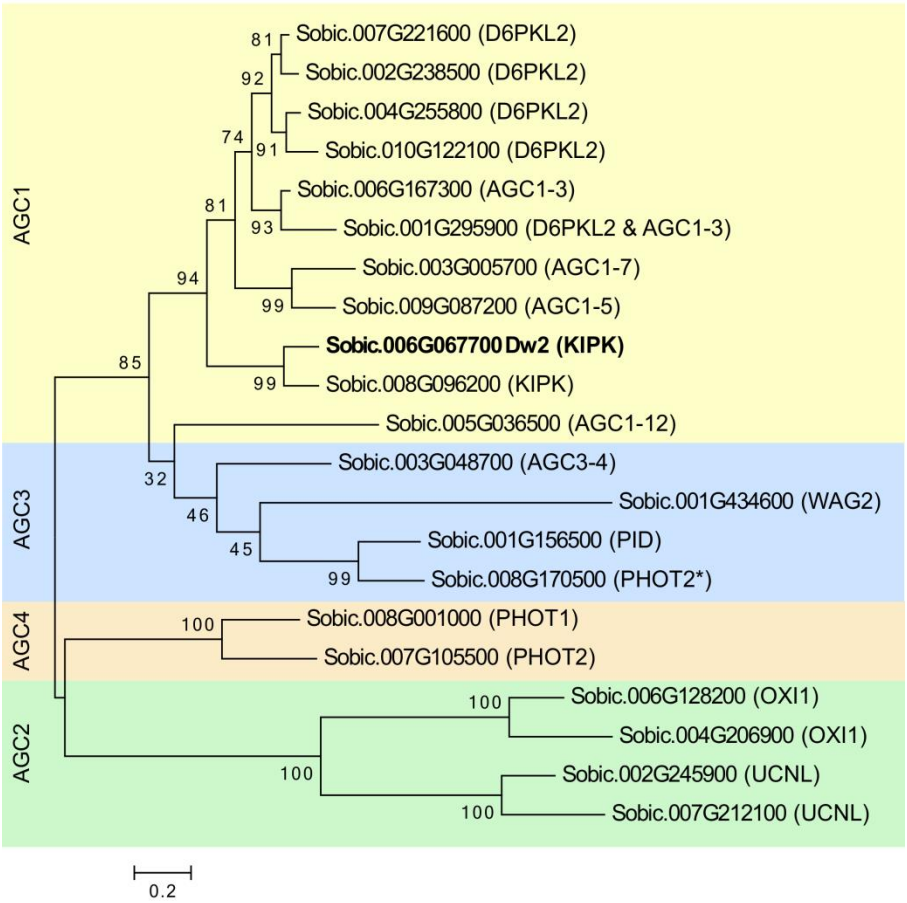


Fig 22. Phylogenetic tree of the AGCVIII subfamily in sorghum. The tree of the 21 sorghum AGCVIII genes was generated in MEGA6 using Maximum Likelihood. Dw2 is bolded. The four different groups, AGC1-4, are labeled and colored. The names in parenthesis are the best hit from a BLAST search of the Arabidopsis genome using that sorghum gene as a query. * The best hit for Sobic.008G170500 is PHOT2 but the score is much lower than Sobic.007G105500 to PHOT2 (203.4 vs. 1122.1 for the Dual Affine Smith Waterman alignment score). Further, Sobic.008G170500 is the best BLAST match of the maize PID homolog, BARREN INFLORESCENCE2, in sorghum.

Plant AGC kinases contain a catalytic core consisting of 12 conserved subdomains [36]. A comparison of Dw2 (Sobic.006G067700) with KIPK and other members of the AGC1-kinase group showed that Dw2 contains a conserved GxGxxG

sequence in the P-loop of sub-domain I of the N-lobe, an activation segment in the C-lobe that includes the Mg++ binding sequence DFDLS, an insertion domain typical of plant AGC-kinases, and a T-loop and activation domain [SxxSFVGTxYxAPE] that is a site of phosphorylation [36] (labeled in Fig 23). The protein has a C-terminal FxxF sequence found in many AGC-kinases that binds 3-phosphoinositide-dependent kinase 1 (PDK1), a highly conserved member of the AGC kinase family that phosphorylates several AGC kinases [167]. AGC kinases vary significantly in the length, sequence, and function of their N-terminal domains that often mediate interaction with other proteins. KIPK1 and 2 and AGC1-8 have N-terminal domains of 546-549 amino acids, significantly larger than other members of the AGC1 kinase subfamily [36]. When the N-terminal 423 amino acid domain of Dw2 was used to search for matches in the Arabidopsis genome (Phytozome), it aligned best with the N-terminal domain of KIPK and next best to KIPK2 (AGC1-9). Multiple sequence alignment of Dw2, rice and maize homologs of Dw2, and Arabidopsis KIPK showed regions of sequence similarity throughout the N-terminal domain and several deletions relative to Arabidopsis KIPK that explain the difference in overall length of the N-terminal protein sequences (423 versus 545 amino acids) (Fig 23).

Fig 23. Alignment of sorghum Dw2 with the maize, rice, and Arabidopsis homologs. A multiple sequence alignment was made in Jalview 2.0 with TCoffee between Dw2 (Sobic.006G067700), two closest maize homologs (GRMZM2G412524 & GRMZM2G128319), two closest rice homologs (LOC_Os12g29580 & LOC_Os04g33500), and the closest Arabidopsis homolog, KIPK (AT3G52890). The kinase domain is demarcated with solid brackets. Dashed brackets demarcate the insertion domain. A red box labels the highly conserved GxGxxG sequence that functions in ATP binding. A green box labels the DFDLS Mg^{2+} binding site and the orange box denotes the T-loop that is phosphorylated to activate the kinase. Finally, the purple box highlights the PDK-interacting fragment.


Sobic.006G067700.2/1-809
GRMZM2G412524_T01/1-799
GRMZM2G128319_T01/1-803
LOC_Os12g29580.1/1-787
LOC_Os04g33500.1/1-503
AT3G52890.2/1-934

Expression of *Dw2*

Dw2 RNA abundance was examined in tissues of BTx623 (*Dw2*) by analysis of RNAseq profiles that are part of the sorghum RNA Atlas (Phytozome). *Dw2* is annotated as having two transcripts that differ in the 5'UTR. The primary transcript (Sobic.006G067700.2) has a UTR with no introns that extends 537 bp before the start codon, while the secondary transcript (Sobic.006G067700.1) has one intron and extends 923 bp. The analysis of *Dw2* expression shown in Fig 24 utilized tissues collected from plants at ~10 days post-floral initiation, when upper leaves, leaf sheaths, internodes, nascent panicles and peduncles are growing. The expression of *Dw2* was relatively high in developing panicles, peduncles, growing internodes and leaf sheaths, with lower expression in fully expanded internodes, leaf blades and the lower portion of the root system that includes root tips and fully elongated roots (Fig 24). The expression of sorghum KIPK-like (Sobic.008G096200) was higher than *Dw2* in roots and lower in leaf tissues, the peduncle, and panicle (Fig 24).

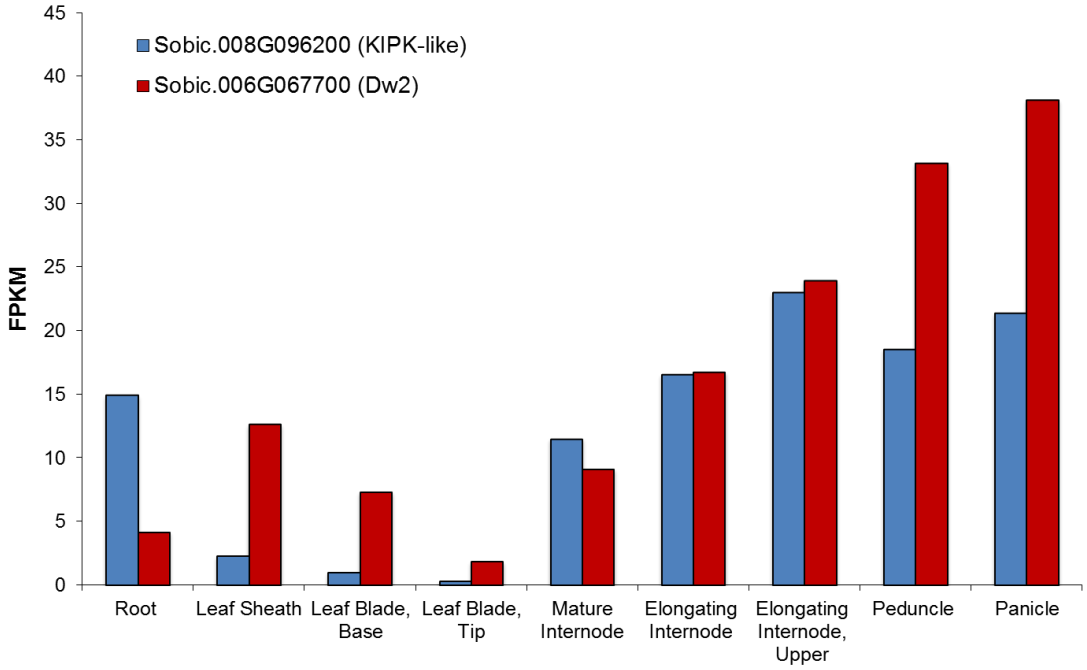


Fig 24. Expression of *Dw2* and Sobic.008G096200 in various tissues. Gene expression data is from the publicly available RNA-seq GeneAtlas on Phytozome v11. Tissues are from BTx623 (dominant *Dw2*) at 44 Days after Emergence (DAE). The leaf tissue was taken from the last ligulated leaf, so the base is still growing whereas the tip is maturing.

DISCUSSION

In this study, *Dw2*, an important dwarfing locus used in grain sorghum breeding, was mapped as a QTL in two populations. Using map-based cloning, the gene corresponding to *Dw2* was identified as a protein kinase whose closest homolog in Arabidopsis is KIPK, a member of the AGCVIII protein kinase family.

Dw2 QTL analysis and fine mapping were performed using two different RIL populations. In the first population derived from BTx623 x IS3620c, alleles of the dwarfing loci *Dw2* and *Dw3* were segregating. Analysis of average internode length

identified a QTL aligned with *Dw2* at ~42.7 Mb on chromosome 6 and a QTL corresponding to *Dw3* located on chromosome 7 at ~59.8 Mb. *Dw2* was the only dwarfing (*Dw*) locus segregating in the second population derived from BTx642 x Tx7000, genotypes recessive for *dw1dw3dw4*. Indeed, the only QTL segregating for total height in this population was a QTL corresponding to *Dw2* (~43.2 Mb). The location of the *Dw2* QTL mapped in this study corresponds to most previous reports of the location of *Dw2* [9,112]. Higgins et al [114] also identified QTL for plant height in this region of chromosome 6 with peaks at 44.3-44.5 Mb or 42.1 Mb depending on the population and QTL model. The authors suggested that variation in QTL location was due to the linkage between *Dw2* and *Ma1* since both influence plant height [114]. In the current study, the influence of *Ma1* alleles is minimal because the BTx642 x Tx7000 RIL population is segregating for a weak allele and null allele of *Ma1,* respectively [169], and BTx623 and IS3620c each contain null alleles of *Ma1* [9,151]. During the analysis of *Dw2* a nearby QTL located at 48.6 Mb on chromosome 6 was identified that modified the length of internode 6 according to single QTL mapping. MQM revealed that this QTL also affected the length of internodes 4-7; however, *Dw2* had a greater impact on the length of the fourth and fifth internode. This additional QTL could also have confounded the location of *Dw2* in the study of Higgins et al [114].

QTL analysis in the BTx623 x IS3620c population showed that *Dw2* and *Dw3* influence internode length differentially during development. *Dw2* had the greatest additive effect on the length of the internode immediately below the peduncle. The additive effects of *Dw2* and *Dw3* on the length of this internode were similar. The

influence of *Dw2* gradually decreased in the internodes below the top internode and there was no detectable impact of *Dw2* on the length of internodes 7-8 below the peduncle in this population. *Dw3* had a much greater effect than *Dw2* on internodes 2-5 below the peduncle with reduced but significant impact on the length of internodes 6-8 (Table 10, Fig 18). Similarly, in maize, *Br2*, the homolog of *Dw3* that encodes an ABCB1 auxin transporter, had a greater influence on elongation of the lower stem internodes compared to upper internodes that elongate post-floral initiation [109]. RILs from the BTx642 x Tx7000 population that are null for *Dw3* and differ in *Dw2* alleles showed a large difference in length of the internode below the peduncle when grown in low light in the greenhouse during the winter (Fig 21). A comparison of the yellow milos (DYM: *dw1**Dw2**Dw3dw4* and DDYM: *dw1**dw2**Dw3dw4*) showed that *Dw2* has an effect on the length of nearly all of the ~25 internodes produced by plants grown in the greenhouse during the summer under long day conditions (Fig 15). Delayed flowering due to increased photoperiod sensitivity in these genotypes caused more internodes to accumulate during the vegetative phase in DYM and DDYM. Taken together these results indicate that *Dw2* affects the length of internodes produced by plants during the vegetative phase and the last 6-7 internodes produced after floral initiation.

Fine mapping narrowed the region encoding *Dw2* to a ~98.1 kb region of chromosome 6 containing ten genes. One of the ten genes in the delimited *Dw2* locus encoded a histone deacetylase that was previously suggested to be a candidate for *Dw2* [112]. However, the deacetylase did not contain polymorphisms in the coding regions that distinguish the parental genotypes used for fine mapping, or DYM (*Dw2*) and

DDYM (*dw2*). DDYM was reported to have originated as a shorter mutant in a field of DYM [6]. Thus, these two yellow milos should be isogenic except at *Dw2*. All of the other genes in the delimited *Dw2* region were sequenced from DYM and DDYM. Only the kinase encoded by Sobic.006G067700 had a polymorphism that distinguished DDYM from DYM in the delimited *Dw2* locus. This polymorphism resulted in a frameshift mutation and a premature stop codon in the first exon. This results in a protein of only 190 amino acids instead of 809 amino acids found in DYM. The kinase domain is located between 424-763 amino acids; therefore, the mutant protein found in DDYM would lack kinase activity.

The closest homolog of sorghum Dw2 in Arabidopsis is KIPK, a member of the AGC family of kinases. The AGC family is named after the cAMP dependent protein kinases, cGMP dependent protein kinases, and protein kinase C and also includes PDK1 and the ribosomal protein S6 kinases. The plant-specific AGCVIII subfamily includes PID, PHOT1 and 2, and the D6PK/D6PKLs [35]. Each of these kinases has been shown to regulate auxin efflux transporters, including ABCB1 and PIN1, with PHOT1 and 2 doing so in a blue-light dependent manner [36,37]. In Arabidopsis, KIPK has a close homolog, KIPK2 (also known as AGC1-9 and At2g36350) and the closely related kinase, AGC1-8 [167,170]. In sorghum, Dw2 has one closely related homolog, Sobic.008G096200, and these two genes form their own branch on the phylogenetic tree (Fig 22). As some of the members of the AGCVIII subfamily have been shown to regulate auxin transport, Dw3, the sorghum homolog of Arabidopsis ABCB1, was initially considered a potential target of Dw2 action. However, while *Dw2* was

expressed in growing internodes, MQM analysis provides no genetic evidence for interaction between *Dw2* and *Dw3*. Furthermore, the *Dw2* allele positively affects the length of the upper most internode in a *dw3* background, indicating that Dw2 can act at least partially through pathways independent of Dw3.

In Arabidopsis KIPK was so named due to its interaction with KCBP, a plant-specific kinesin-like calmodulin binding protein that functions in cell division and trichome formation [165]. KCBP has a C-terminal motor and calmodulin-binding domain, and is unusual among kinesins in its ability to interact with microtubules and with actin, the latter interaction mediated by a MyTH4-FERM tandem that occurs in myosin [171]. Type-VI kinesin-14 dimers in *Physcomitrella patens*, homologs of KCBP, are highly processive, and transport vesicles/cargo long distances when clustered [172]. KCBP contains a calmodulin binding domain and is down-regulated by calcium via calmodulin as well as the KCBP interacting Ca^{2+}-binding protein (KIC) [173,174]. While KIPK did not phosphorylate the N-terminal end of KCBP under experimental conditions, it is possible that it phosphorylates KCBP under other conditions, and it is possible that KCBP transports KIPK within the cell [165].

Subsequent work has also shown that Arabidopsis KIPK1 and 2 directly interact with members of the proline-rich extensin-like receptor-like kinase (PERK) family, specifically PERK8, 9, 10, and 13 [170]. Other PERK-genes, such as PERK1, mediate growth inhibition, possibly in response to cell wall signals [175]. In Arabidopsis, KIPK1 and 2 double mutants did not produce shoot phenotypes although there were differences in root elongation when plants were grown on elevated sucrose [170].

Different parts of the N-terminal domain of KIPK1 and 2 mediate the direct interactions with KCBP and the various PERKs [170]. The 423 amino acid N-terminal sequence of Dw2 aligned well with the ~545 amino acid N-terminus of KIPK despite several deletions that account for the difference in overall length of this domain. The sequence similarity of the N-terminal domains of KIPK and Dw2 indicates that Dw2 has likely retained the ability to interact with one or more members of the PERK family. The best BLAST hits to Arabidopsis PERK8 and 10 (At5g38560 and At1g26150, respectively) in sorghum (Sobic.003G100700, Sobic.003G289800, and Sobic.009G000300) were expressed in stem internodes (Phytozome). Therefore, it will be of interest to determine if Dw2 interacts with sorghum PERK8 or 10 homologs.

If Dw2, like Arabidopsis KIPK, interacts with PERKs and KCBP, the interactions with these proteins may modulate growth regulation and serve other regulatory functions. For example, because KCBP transports vesicles/cargo long distances [172] potential Dw2 interactions with PERKs and KCBP in sorghum could regulate growth and the flow of materials to the cell wall during and after organ elongation. Alternatively, in trichomes KCBP has been found to organize cytoskeleton components [171], thus KIPK may be involved cytoskeletal regulation that is associated with cell elongation. This more general coordinating function may explain why Dw2 is expressed in growing zones of leaf blades, leaf sheaths, stems, and panicles. Lack of growth phenotypes in all organs where *Dw2* is expressed (i.e., peduncle) could be due to the presence of a second KIPK-like gene in sorghum (Sobic.008G096200). In fact, Sobic.008G096200 is more highly expressed than *Dw2* in the roots, and both genes are

highly expressed in the panicle, peduncle, and internodes (Fig 24). One other possibility could be that KIPK is involved in a PERK signalling pathway. Another member of the PERK family, PERK4, has been shown to regulate cell elongation in roots as part of an abscisic acid (ABA) signaling pathway [176].

While Dw2 is a homolog of Arabidopsis KIPK, Dw2 has an important role in regulating stem length in sorghum, a function not observed in Arabidopsis KIPK mutants [170]. This may be because grass stem growth occurs by sequentially elongating internodes adjacent to intercalary meristems located just above nodes, a mode of stem growth that is unique to grasses. The first sorghum dwarfing locus cloned, *Dw3*, also had a more severe stem phenotype than mutants affecting the Arabidopsis homolog, ABCB1. Multani et al [109] showed that mutation of *Dw3*, an auxin efflux carrier, results in short internodes in sorghum whereas the corresponding ABCB1 single mutant in Arabidopsis had little effect on stem length [177]. Knoller et al [110] showed that brachytic2, the maize homolog of sorghum Dw3, is expressed in stem nodes but not in stem internodes, whereas Arabidopsis lacks intercalary meristems. This difference in physiology between Arabidopsis and the grasses helps explain the differences in ABCB1 mutant phenotypes. It may also explain the differences in phenotypes between the Dw2 and KIPK mutants in sorghum and Arabidopsis, respectively. Alternatively, the difference in phenotype could be due to differences in functional redundancy and/or expression within the AGCVIII subfamily.

Dw2 has been used extensively in grain sorghum breeding in the U.S. to create lines and hybrids with reduced stem length. A recessive allele of *dw2* derived from

DDYM was used in the Sorghum Conversion Program to reduce the height of lines that were being converted for use in temperate grain sorghum breeding programs [9]. *Dw2* is linked to *Ma1*, another important gene in grain sorghum and energy sorghum development [151]. In addition to its historical significance, a better understanding of *Dw2* function may enable the design of improved sorghum crops.

CHAPTER IV

THE CURIOUS CASE OF *Dw4*

INTRODUCTION

Sorghum is an important cereal crop that has many different uses. Most sorghum grown in the U.S. is grown for grain that is generally used for animal feed. Sorghum is also grown for the sugar in the stem and the shoot for forage. More recently, the crop has also been used for bioenergy, which can be produced from the grain as with maize, the stem sugar, as with sugar cane, or the shoot for biomass. An additional interesting use of sorghum is in the production of brooms. A small group of sorghum lines called broomcorns have been bred for long panicle branches that can be made into brooms (Fig 25).

Fig 25. Representative panicles (heads) of Standard Broomcorn (top) and BTx623. Note the much longer panicle branches of SB compared to the shorter branches of the grain sorghum BTx623.

Sorghum is native to Africa where it is generally 3-4 meters tall and photoperiod sensitive. When it was introduced into the U.S., plants were selected for photoperiod insensitivity so they would flower in the temperate climate. Plants raised for grain were also bred to be shorter to reduce lodging and enable mechanical harvesting. This was accomplished through selection of naturally occurring mutants of short plants. However, plants grown for stem sugar or biomass are generally taller than those grown for grain.

In the 1950s, Quinby and Karper [104] used many different crosses to determine the number of genes that are responsible for the range of height seen in sorghum in the U.S. They determined that there are four genes which they termed dwarfing genes or *Dw1-Dw4*. At each gene, the recessive allele reduces height. Furthermore, the four genes segregated independently and so are not linked. The genes that had recessive alleles originating in the milos were labeled *Dw1* and *Dw2*. *Dw3* was the name for the gene whose recessive allele came from the kafirs. The dominant allele of *Dw4* was only found in the broomcorns, all other genotypes were recessive at *Dw4* [104]. The specific broomcorn genotypes used in this study were Japanese Dwarf Broomcorn (JDB) and Scarborough Dwarf Broomcorn. Another paper that looked at height in sorghum was an earlier study in broomcorns by Sieglinger [103]. This study used Standard, Acme, and Japanese Dwarf Broomcorns to determine that there were two genes segregating for height in broomcorns. Based on this and their own data, Quinby and Karper [104] determined that those genes were *Dw1* and *Dw2* and that the broomcorns were fixed recessive at *dw3* and dominant at *Dw4*.

The first *Dw* gene to be cloned was *Dw3*. Multani et al. [109] showed that it encodes an ATP-binding cassette type B (ABCB) auxin efflux transporter located on chromosome 7 at ~59.8 Mbp. More recently, *Dw1* and *Dw2* have been map-based cloned. *Dw1* is a highly conserved gene of unknown function on chromosome 9 [149,150] and *Dw2* is an AGCVIII protein kinase on chromosome 6. While the recessive alleles at *Dw1* and *Dw2* are each the result of a premature stop codon caused by a SNP and an INDEL respectively [149,150], the recessive allele at *Dw3* is an 882 bp tandem repeat in the last exon [109]. This mutation sometimes results in uneven recombination yielding a high rate of reversions from the recessive to dominant allele. However, *Dw4* has not been cloned. Morris et al. [112] found a fourth QTL segregating for height in an association mapping (GWAS) study using the Sorghum Association Panel (SAP) which does include three broomcorns [178]. They speculated that this additional QTL which is located at ~6.6 Mbp on chromosome 6, ~36 Mbps from *Dw2*, could be *Dw4*. Another group, also performing GWAS on the SAP, found a QTL on chromosome 4 at ~67 Mbp that they suggested could be *Dw4* [117].

As height is such an important trait, we wanted to further our understanding of the genetics and physiology of height in sorghum by cloning *Dw4*. To do this, several different crosses with two different broomcorns were made. The first three crosses were with Standard Broomcorn (SB). Surprisingly, no QTL was found that fit with Quinby and Karper's [104] description of *Dw4*. Furthermore, SB is dominant at *Dw3*, not recessive as described. An additional cross was made with Acme Broomcorn (AB) to confirm this result, which it did, indicating that AB is also dominant at *Dw3*.

METHODS

QTL Mapping in SC170 x SB

To determine the location of *Dw4*, a cross between SB and SC170 was made. Seed for each parent was obtained from Dr. W.L. Rooney (Dept of Soil and Crop Sciences, TAMU). Quinby and Karper [104] designated SB as *Dw1Dw2dw3Dw4*. Based on previous work in the Mullet laboratory, SC170 is *dw1dw2dw3dw4* [179], thus the population should be segregating for *Dw1*, *Dw2*, and *Dw4* (Table 18). The F_1 plants were checked to be sure they were F_1s with CAPS markers (Table A6). Plants that were heterozygous for the two parental alleles at the CAPS markers were selfed. The F_2 population was planted out in a higher light intensity greenhouse (HLG) in summer, 2012 (n=154). The F_2 plants were harvested at grain maturity and the days to anthesis, total length, and length of each internode were noted. DNA was extracted from leaf tissue using FastDNA Spin Kit (MP Biomedicals). The plants were genotyped using Digital Genotyping (DG) with the enzyme FseI used for digestion [124]. Reads were mapped to version 1 of the sorghum reference genome assembly (*Sorghum bicolor* v1.1 DOE-JGI, http://phytozome.jgi.doe.gov/) and were processed as described in Morishige et al (2013). The genetic map was made in MapMaker [125] using the Kosambi mapping function. QTL mapping was performed in QTL Cartographer [126] using Composite Interval Mapping (CIM) and a threshold of α=0.05 determined with 1000 permutations.

Table 18. Genotype at Each of the *Dw* Loci for Each of the Parents Used in This Study According to Quinby and Karper [104].

Cross	*Dw* genotypes	*Dw* loci segregating	n
SC170 x SB (204A)	*dw1dw2dw3dw4* x ***Dw1Dw2****dw3****Dw4***	*Dw1, Dw2, Dw4*	154
SC170 x SB (117)	*dw1dw2dw3dw4* x ***Dw1Dw2****dw3****Dw4***	*Dw1, Dw2, Dw4*	124
BTx623 x SB	*dw1****Dw2****dw3dw4* x ***Dw1Dw2****dw3****Dw4***	*Dw1, Dw4*	132
Hegari x SB	***Dw1****dw2****Dw3****dw4* x ***Dw1Dw2****dw3****Dw4***	*Dw2, Dw3, Dw4*	128
BTx623 x AB	*dw1****Dw2****dw3dw4* x ***Dw1****dw2dw3****Dw4***	*Dw1, Dw2, Dw4*	97
Hegari x JDB	***Dw1****dw2****Dw3****dw4* x *dw1****Dw2****dw3****Dw4***	*Dw1, Dw2, Dw3, Dw4*	100
SYM x JDB	***Dw1Dw2Dw3****dw4* x *dw1****Dw2****dw3****Dw4***	*Dw1, Dw3, Dw4*	100

QTL Mapping in Additional SB Populations

SB was also crossed with BTx623 and Hegari. Seed from each of these lines was obtained from Dr. W.L. Rooney (Dept of Soil and Crop Sciences, TAMU). BTx623 is *dw1Dw2dw3dw4* [154]; therefore, that population should be segregating for *Dw1* and *Dw4* (Table 18). Hegari was designated by Quinby and Karper [104] as *Dw1dw2Dw3dw4*; however, based on another population made with Hegari, Hegari x 80M, the Hegari used in this study is *Dw1Dw2Dw3dw4* [149]. Based on that, the cross with Hegari is expected to segregate for *Dw3* and *Dw4*. The F_1s were checked with CAPS markers and selfed (Table A6). The F_2s were planted in a lower light intensity greenhouse (LLG) in summer 2012 (n=132 for the BTx623 cross and n=128 for the Hegari cross). These plants were grown to grain maturity and the length of two fully expanded internodes was measured. The plants were genotyped using DG [124] with the enzyme FseI used for digestion. Reduced representation reads were mapped to the reference genome using BWA v0.7.12 [156] and indel realignment and joint variant calling were performed with the GATK using the naive pipeline of the RIG workflow

[157–160]. The genetic map was constructed in R/QTL [128] using the Kosambi function. QTL mapping was performed in QTL Cartographer [126] using CIM and a threshold of α=0.05 set with 1000 permutations.

An additional group of SC170 x SB F_2 plants (n=124) was grown at the same time as the BTx623 and Hegari populations. These plants were phenotyped, genotyped, and QTL mapped as with the BTx623 and Hegari populations.

QTL Mapping in Acme Broomcorn and Japanese Dwarf Broomcorn

Crosses in two additional broomcorns were made. Acme Broomcorn (AB) was designated as *Dw1dw2dw3Dw4* and Japanese Dwarf Broomcorn (JDB) was designated as *dw1Dw2dw3Dw4* (Table 18) [104]. Seed for both AB (PI 656014) and JDB (PI 30204) was obtained from USDA ARS-GRIN. AB was crossed with BTx623, while JDB was crossed with Standard Yellow Milo (SYM) and Hegari. Thus the BTx623 x AB population should be segregating for *Dw1*, *Dw2*, and *Dw4*. Both of the JDB crosses should be segregating for *Dw1*, *Dw3*, and *Dw4* [104]. The F_1 plants were checked with CAPS markers and selfed (Table A6). The F_2 plants were grown in the LLG in fall, 2014 (n=97 for BTx623 x AB; n=100 for SYM x JDB; n=100 for Hegari x JDB). The plants were harvested at grain maturity and phenotyped for total length and the length of each internode. Genotyping and QTL mapping was performed as with the SB x BTx623 or Hegari crosses. For all QTL found in this study, the physical locations of QTL were subsequently converted to version 3 sorghum reference genome locations (*Sorghum*

bicolor v3.1 DOE-JGI, http://phytozome.jgi.doe.gov/) using BLAST and Gramene (http://www.gramene.org**).**

Sequencing of *Dw3*

Whole genome sequencing of SB was performed to examine the sequence at *Dw3*. SB seeds were obtained from Dr. W.L. Rooney (Dept of Soil and Crop Sciences, TAMU). Seeds were soaked in 20% bleach for 20 minutes and washed extensively in distilled water for one hour. Seeds were germinated on water-saturated germination paper in a growth chamber (14 hr light; 30° C/10 hr dark; 24° C). Genomic DNA was isolated from 8-day old root tissue using a FastPrep DNA Extraction kit and FastPrep24 Instrument (MP Biomedicals LLC, Solon, OH, USA), according to the manufacturer's specifications. DNA template (350 bp average insert size) was prepared using a TruSeq® DNA PCR-Free LT Kit, according to the manufacturer's directions. Paired-end sequencing (125 x 125 bases) was performed on an Illumina HiSeq2500. Sequence reads were mapped to version 3 of the sorghum reference genome assembly (*Sorghum bicolor* v3.1 DOE-JGI, http://phytozome.jgi.doe.gov/) and processed as with QTL mapping.

Sanger capillary sequencing of *Dw3* in AB was performed, as well. The gene was amplified with Phusion DNA Polymerase (NEB) using the standard reaction conditions. The gene was sequenced using BigDye v3.1 (Invitrogen) and Sanger capillary sequencing. Primers for sequencing are listed in Table A7.

RESULTS

QTL Mapping in the SC170 x SB F_2

Table 19. SC170 x Standard Broomcorn F_2 Grown in HLG QTL. For the additive (Add) effect, a negative sign indicates that the SB allele increases length.

Trait	Chr	Peak (Mbp)	Peak LOD	Start (bp)	Stop(bp)	Add	Dom	R^2
Total Length	1	0.4	4.42	start	3358339	-17.30	2.618	0.05
	6	42.4	6.17	3470967	45508141	-15.86	13.79	0.09
	7	59.8	36.66	59749922	60105972	-49.59	30.71	0.48
	9	57.1	12.55	55101556	57957604	-27.85	9.55	0.19
Ave Internode Length	6	46.1	5.72	43828484	48365044	-13.83	16.62	0.07
	7	59.8	39.35	59749922	60105972	-51.32	35.08	0.43
	9	57.1	20.03	55101556	58889232	-38.78	7.46	0.27
Total Length w/o Peduncle	6	42.4	7.85	3470967	45247886	-16.63	13.44	0.11
	7	59.8	42.2	59749922	60105972	-50.26	31.08	0.64
	9	57.1	10.31	55101556	58889232	-23.01	7.59	0.14

The QTL maps for total height, total height without the peduncle, and average internode length are shown in Figure 26. There are four QTL segregating for total height: one on chromosome 1 at ~353 kb, one on chromosome 6 at 42.4 Mbp which corresponds to *Dw2*, one on chromosome 7 at 59.8 Mbp which corresponds to *Dw3*, and one on chromosome 9 at 57.1 Mbp which corresponds to *Dw1* (Table 19). Total height without the peduncle is segregating for three QTL: *Dw1*, *Dw2*, and *Dw3*. Average internode length is segregating for three QTL which correspond to *Dw1*, *Dw3*, and probably *Dw2*, though the peak is at 46.1 Mbp instead of ~42 Mbp (Table 19). *Dw4* is

expected to segregate for all of these traits; however, the three QTL that segregated for all of the traits corresponded to *Dw1*, *Dw2*, and *Dw3*. Thus, there is no QTL that matched Quinby and Karper's [104] description of *Dw4*.

QTL Mapping of Length of Fully Expanded Internodes in the SB Populations

The additional SB x SC170 F_2 plants that were only measured at two fully expanded internodes showed QTL at *Dw1* and *Dw3*, as well as a QTL on chromosome 2 at 65.7 Mbp for the average of the two internodes (Table 20). This table also shows the results of the other two SB populations. The cross with BTx623 has three QTL segregating for the average of the two internodes one at *Dw1*, one at *Dw3*, and one on chromosome 1 at 72.3 Mbp. The cross with Hegari also has three QTL segregating. One of these QTL is on chromosome 7 at 55.6 Mbp, another is on chromosome 1 at 63.3 Mbp, and the third is on chromosome 2 at 59.2 Mbp (Table 20). The QTL on chromosome 7 has been previously described [117,149]. None of the QTL was consistent across the populations, as would be expected for *Dw4*.

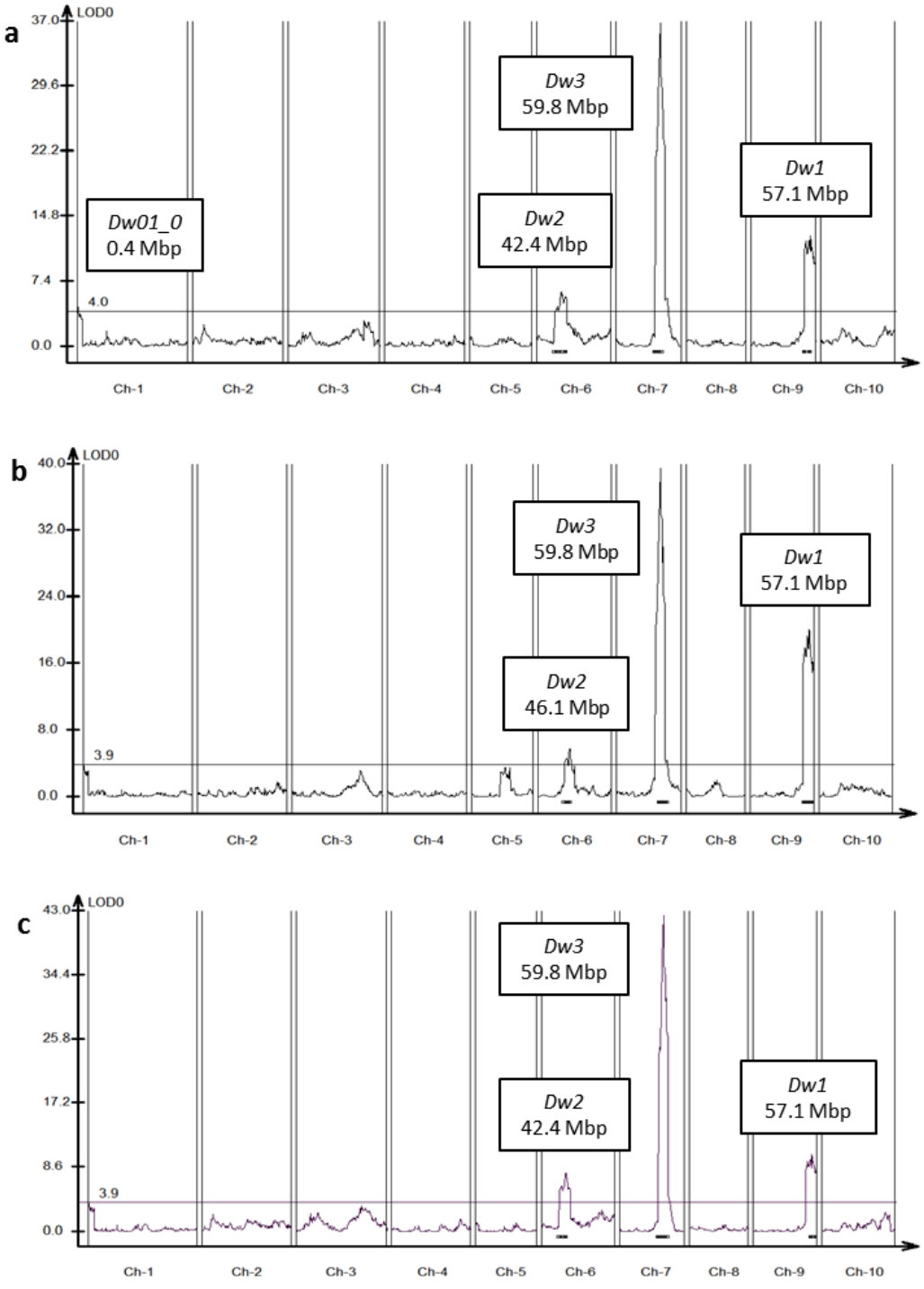


Fig 26. QTL maps of Standard Broomcorn x SC170 F_2 (HLG) for total height (a), average internode length (b), and length without the peduncle (c). *Dw1-Dw3* are noted on each graph.

Table 20. QTL Segregating for Average Internode Length for Each of the Remaining SB Populations. For the additive (Add) effect, a negative sign indicates that the SB allele increases length.

Cross	Chr	Peak (Mbp)	Peak LOD	Start (bp)	Stop (bp)	Add	Dom	R^2
SC x SB (LLG)	2	65.7	4	62,872,583	68,871,404	-1.68	38.02	0.07
	7	59.6	30.25	59,469,953	61,086,315	-57.91	73.93	0.77
	9	57.0	4.95	55,101,556	58,246,093	-34.88	-0.54	0.11
BTx623 x SB	1	72.2	4.16	70,308,570	76,022,320	-3.48	-33.44	0.01
	7	59.8	23.94	59,178,473	60,105,977	-57.74	43.20	0.69
	9	56.1	5.67	54,986,925	end	-26.84	11.45	0.12
Hegari x SB	1	63.3	4.23	60,931,109	64,358,053	-19.7	-9.59	0.13
	2	59.2	4.62	57,735,403	59,651,107	-20.25	3.25	0.08
	7	55.6	16.8	55,269,585	56,485,690	-29.83	35.78	0.03

QTL Mapping with the Additional Broomcorns

The QTL that are segregating in the populations derived from AB and JDB are shown in Table 21. The BTx623 x AB population is segregating for *Dw3*, an additional QTL on chromosome 7 at 63.7 Mbp, and a QTL on chromosome 4 at 67.5 Mbp for average length of all elongated internodes (Figure 27). The QTL on chromosome 4 has been documented before [117]. Fig 27b shows the additive effect of the QTL on chromosome 4 with the AB allele ("B" allele) decreasing height. This is contradictory to *Dw4*, for which AB should have the allele that increases height.

The SYM x JDB population segregated for one QTL on chromosome 6 at 40.2 Mbp for total height (Table 21). This is close in location to Ma1, which is at 40.3 Mbp [151]; indeed, this QTL is also segregating for days to flowering in this population. For average internode length, it was segregating for three QTL, one is on chromosome 6 at 183 kb, one on chromosome 8 at 52.1 Mbp, and one on chromosome 10 at 2.0 Mbp

(Table 21). The Hegari x JDB population is segregating for three QTL for total length and two for average internode length. For the former, the QTL are on chromosome 6 at 31.8 Mbp, chromosome 7 at 56.5 Mbp, and chromosome 10 at 57.4 Mbp. For the latter, the QTL are on chromosome 3 at 64.0 Mbp and chromosome 7 at 56.2 Mbp. The QTL on chromosome 7 correspond not to *Dw3* but to the nearby locus [117,149]. The QTL on chromosome 6 includes *Ma1*. A summary of all the QTL segregating for average internode length found in this study is shown in Figure 28.

Table 21. QTL Segregating for Total Length (TL) and Average Internode Length (AIL) in the Populations Derived from Other Broomcorns. For the additive (Add) effect, a negative sign indicates that the allele from the broomcorn parent increases length.

Cross	Trait	Chr	Peak (Mbp)	Peak LOD	Start (bp)	Stop (bp)	Add	Dom	R^2
BTx623 x AB	TL	4	66.7	13.77	66,357,351	67,930,048	39.1	32.93	0.06
		7	59.6	10.57	59,430,281	60,822,767	-38.94	32.78	0.42
	AIL	4	67.5	15.96	66,357,351	68,280,471	3.52	2.67	0.09
		7	59.6	19.58	59,430,281	60,032,260	-3.97	3.84	0.63
		7	63.7	4.64	62,633,856	end	1.92	1.46	0.01
Hegari x JDB	TL	6	31.8	7.15	3,471,066	42,399,256	31.92	51.10	0.00
		7	56.5	11.97	12,070,549	57,499,745	-48.99	57.00	0.52
		10	57.4	4.13	53,874,452	58,538,649	-9.51	-42.81	0.01
	AIL	3	64.0	4.73	61,487,086	66,580,756	0.96	1.67	0.00
		7	56.2	24.71	55,631,277	56,557,348	-3.46	3.12	0.39
SYM x JDB	TL	6	40.2	11.22	4,388,373	41,417,205	39.82	38.35	0.07
	AIL	6	0.2	4.14	start	1,783,595	-1.00	-0.13	0.09
		8	52.1	5.64	6,277,121	54,775,372	-1.13	0.18	0.17
		10	2.0	6.2	start	3,170,306	-1.49	0.06	0.18

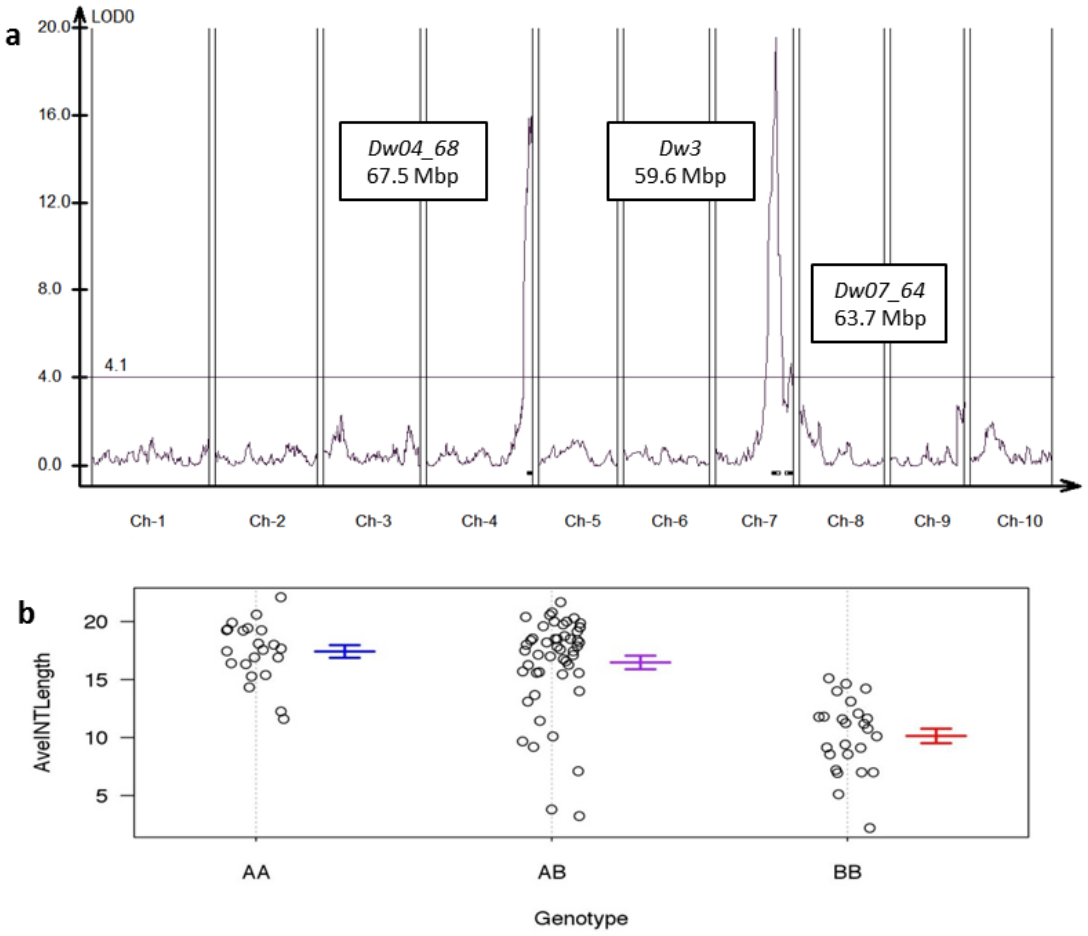


Fig 27. QTL for average internode length in BTx623 x Acme Broomcorn F_2. (**a**) The QTL map. (**b**) Dot plot of phenotype by genotype of the QTL on chromosome 4. The "A" allele is the allele from BTx623.

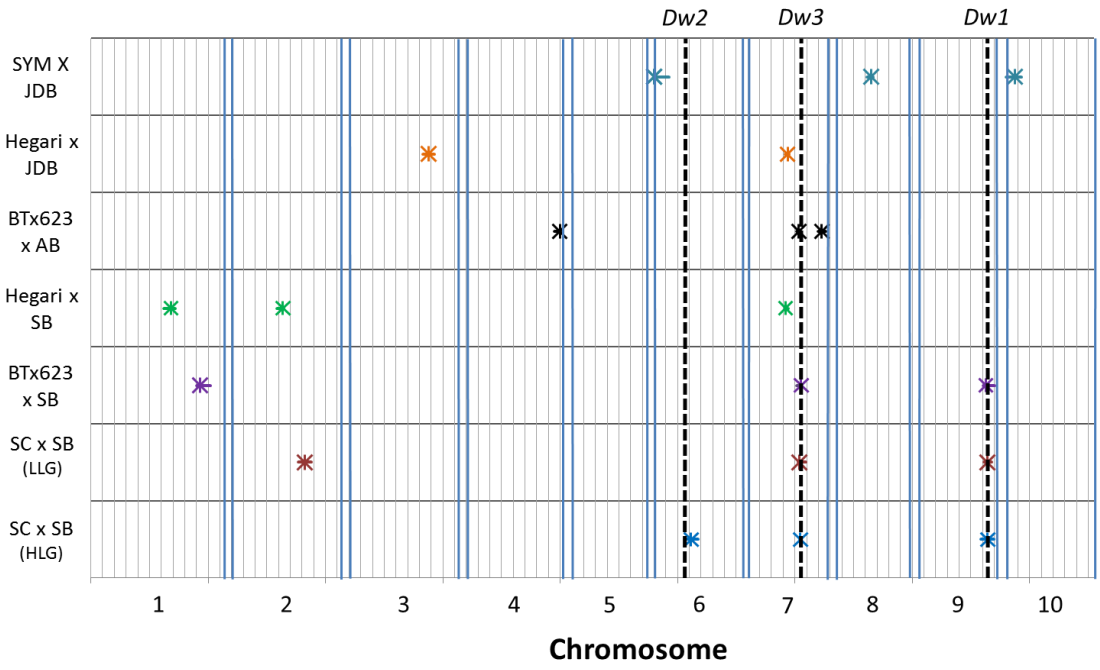


Fig 28. Diagram of all of the QTL for average internode length found in this study. The populations are listed along the y-axis. The x-axis is the genetic map of the SC170 x SB (HLG) along the genome with the double blue lines denoting the boundary between chromosomes. For each of the other populations, the genetic map coordinates for that population was converted to the physical location. The physical locations were then converted to the coordinates of the SC170 x SB (HLG) genetic map. For each QTL, the peak is marked with an asterisk and the LOD 2 interval is denoted with the lines extending from the asterisk. Location of *Dw1*, *Dw2*, and *Dw3* is shown.

Genotype at *Dw3*

Quinby and Karper [104] described the broomcorns as recessive at *dw3*. However, all of the populations used in this study surprisingly gave the opposite result of what that genotype would. In other words, *Dw3* was segregating in crosses between a broomcorn and a line that is recessive at *dw3* but not in crosses with lines that are dominant. Thus, it appears that all of these broomcorns are dominant at *Dw3*. To check

this, the tandem repeat that causes the recessive allele was amplified using previously published primers (Fig 29) [180]. All of the broomcorns have the PCR product size that results from lacking the repeat. The other parents in populations segregating for *Dw3* (SC170 and BTx623) do have the larger PCR product (Fig 29). Thus the broomcorns lack the classic recessive allele at *Dw3*.

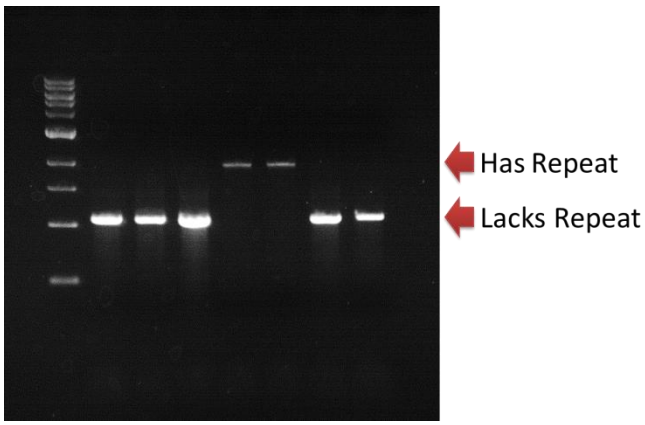


Fig 29. Identification of parental genotypes with repeat insertion in *Dw3*.
Photograph of gel of the PCR products from the PCR described in Farfan et al. [180]. From left to right: 1kb ladder (NEB), SB, AB, JDB, SC170, BTx623, Hegari, and SYM. Products that lack the repeat are at just over 1 kb and products that have the repeat insert are just short of 2 kb in length, both of which are annotated.

To check for any other polymorphisms in Sobic.007G163800 (*Dw3*) in the broomcorns, whole genome sequencing of SB and Sanger sequencing of AB was performed. There were seven polymorphisms in the exons found in Standard and Acme Broomcorn compared to BTx623 (Table 22). Each of these polymorphisms results in a

synonymous amino acid substitution. Thus, the broomcorns appear to have functional versions of *Dw3*.

Table 22. Polymorphisms from BTx623 Found in the Broomcorns at *Dw3* (Sobic.007G163800).

Number	Location (bp from start)	Polymorphism (DNA; AA)	Result
1	468	C > G; A > A	Synonymous
2	4102	G > A; G > G	Synonymous
3	4804	T > G; P > P	Synonymous
4	4924	G > C; T > T	Synonymous
5	5065	T > C; R > R	Synonymous
6	5119	C > G; G > G	Synonymous
7	5341	T > G; L > L	Synonymous

Check of JDB

JDB seed was obtained from USDA ARS GRIN for the creation of these populations. However, upon growing it out, the plants did not look how JDB was shown and described by Sieglinger [103]. Sieglinger [103] describes the line as shorter than AB and pictures show an obvious broomcorn head. However, our plants were taller than AB and did not have the head expected for a broomcorn (Fig 30).

Fig 30. Photograph of the three broomcorn parents. From left to right: SB, JDB, and AB. Note the large, broad broomcorn heads in SB and AB, though the head of AB is still emerging from the flag leaf and will get bigger and broader. The head of the other two genotypes has fully emerged.

Furthermore, the CAPS markers that worked in the other broomcorns did not work in this genotype. The genotypes of the parents were checked further with genotype data from DG using NgoMIV to perform the digest [124]. The lines that were used to make the populations discussed were compared with another seed source of the same line across the genome: PI 642997 for SB and PI 598119 for JDB [181] (USDA ARS GRIN for both). The two seed sources of SB/Evergreen are very similar (150 differences) and much of the difference is due to heterozygous calls in one line or the other. However, the JDB used to make crosses (PI 30204) and the alternative seed

source of JDB (PI 598119) were very different (5,551 differences). Furthermore, the polymorphisms were often fixed for the different alleles.

DISCUSSION

Dw4 is the remaining dwarfing gene that has not been cloned. Quinby and Karper [104] only found the dominant allele in the group of sorghums known as broomcorns. To locate and hopefully fine map *Dw4*, several crosses were made to two broomcorn parents: Standard and Acme. However, none of the QTL identified were segregating in all of the broomcorn crosses (Figure 28) as would be expected for *Dw4*. Furthermore, while Quinby and Karper [104] designated the broomcorns as recessive *dw3*, each of the broomcorn parents used in this study is dominant *Dw3*.

The first population examined was SC170 x SB which was predicted to segregate for *Dw1*, *Dw2*, and *Dw4*. However, for average internode length and length of the stem without the peduncle *Dw1*, *Dw2*, and *Dw3* were segregating instead. Total height had an additional minor QTL on chromosome 1 (*Dw01_0*). Plants from this population were also grown in a greenhouse with lower light intensity compared to the greenhouse of the original grow-out. This population showed QTL at *Dw1* and *Dw3* as well as a QTL on chromosome 2 (*Dw02_66*). It is also surprising that *Dw3* is segregating in this population. The SB allele is the allele at this QTL that increases height, which would mean that it is dominant at *Dw3*. Quinby and Karper [104] described SB as recessive at *dw3*.

The next two populations examined were BTx623 x SB and Hegari x SB. The cross with BTx623 was predicted to segregate for *Dw1* and *Dw4*; however, it was actually segregating for *Dw1* and *Dw3*. Also, it was not segregating for *Dw01_0* or *Dw02_66* found in the cross with SC170, though there was a novel QTL on chromosome 1 at 72 Mbp. Meanwhile, the cross with Hegari was predicted to segregate for *Dw2*, *Dw3*, and *Dw4*. However, the cross actually had a QTL near *Dw3* on chromosome 7 that has been previously described (*Dw07_56*) [117,149] and two additional QTL on chromosomes 1 and 2. None of the QTL on chromosome 1 overlaps. Thus none of them are candidates for *Dw4* since none of them are segregating in all of the populations as *Dw4* should be. The two QTL found on chromosome 2 also do not overlap.

The lack of a presumptive *Dw4* QTL that consistently segregated in the SB populations was very surprising; therefore, crosses with two other broomcorns were made. One of these crosses was BTx623 x AB which was found to be segregating for *Dw3*, a major QTL on chromosome 4 (*Dw04_68*), and a minor QTL downstream of *Dw3* on chromosome 7. For *Dw04_68*, the allele that increased length is the BTx623 allele, not the AB allele as would be expected if this QTL is *Dw4*. This QTL is in the same location as a QTL mentioned [117] as segregating in a GWAS study of the SAP. Indeed, the authors did suggest this could be *Dw4*. However, since the broomcorn allele decreases length and it is not found in any of the SB crosses, our results would argue against this being *Dw4*.

The two JDB populations were crosses with Hegari and Standard Yellow Milo (SYM). The Hegari cross was segregating for *Dw07_56* and a QTL on chromosome 3.

For the SYM cross, there are three QTL, none of which align with a known *Dw* locus or a QTL found in one of the other broomcorn crosses. Thus, there is no QTL segregating in these populations that fits with *Dw4* (Figure 28). Table 23 shows the genotype calls at each of the *Dw* loci for each of the parents used in this study based on the QTL mapping performed herein.

Table 23. *Dw* Genotype for Each of the Parents Used in This Study Based on Our QTL Mapping.

Cross	***Dw* genotypes**	***Dw* loci segregating**
SC170 x SB (HLG)	*dw1dw2dw3* x ***Dw1Dw2Dw3***	*Dw1, Dw2, Dw3*
SC170 x SB (LLG)	*dw1dw2dw3* x ***Dw1****dw2*‡***Dw3***	*Dw1, Dw3*
BTx623 x SB	*dw1****Dw2****dw3* x ***Dw1Dw2****dw3*	*Dw1, Dw3*
Hegari x SB	***Dw1Dw2Dw3*** x ***Dw1Dw2Dw3***	None
BTx623 x AB	*dw1****Dw2****dw3* x *dw1****Dw2Dw3***	*Dw3*
Hegari x JDB	***Dw1Dw2Dw3*** x ***Dw1Dw2Dw3***	None
SYM x JDB	***Dw1Dw2Dw3*** x ***Dw1Dw2Dw3***	None

‡There is no significant QTL in the area of *Dw2* in this lower light environment. However, since there is a QTL there in the higher light, that is presumably because of differences in the environment or the traits measured and SB is, in fact, *Dw2* dominant.

The phenotype of JDB used in this study did not look as described and pictured in Sieglinger [103] and Quinby and Karper [104]. The crosses were made anyway. However, when attempting to verify the F_1s with CAPS markers, the CAPS marker used for the other broomcorns could not be used for JDB. New CAPS markers were developed using SNPs found in the broomcorns in the SAP collection [178]. However, most of these did not work either, suggesting sequence differences between JDB and the other broomcorns used in this study. Additionally, the broomcorns used to generate the

populations were compared to other available accessions of the same genotype. While the SBs were essentially the same, JDB differed a lot from the alternative seed source of JDB, PI 598119. Thus, we conclude that the JDB seed used in this study is not JDB as previously described.

The inability to identify a QTL that matches the description of *Dw4* is difficult to explain. Furthermore, while Quinby and Karper [104] scored all of the broomcorns as recessive at *dw3*, the broomcorns used in this study are dominant at *Dw3*. It is possible that the SB and AB used herein are not the same as those used by earlier researchers. However, the phenotype of each of these is similar to that described and pictured in Sieglinger [103] and Quinby and Karper [104]. In any case, it would be useful to map QTL in populations using other broomcorns as parents, such as the alternative seed source of JDB, Scarborough Dwarf Broomcorn, or the other two broomcorns in the SAP collection. An alternative source of JDB and Scarborough Dwarf Broomcorn would be of the most interest as they are what were used by Quinby and Karper [104].

Another possibility is that there is an environmental influence. Quinby and Karper [104] performed their research in the field in Lubbock, TX while this study was conducted in a greenhouse. However, the phenotypes listed by Quinby and Karper [104] for AB are similar to the phenotype for that line we obtained in the greenhouse (112 vs. 135 cm for height). For SB, the total height obtained in the greenhouse is a bit taller than that listed by Quinby and Karper [104] (285 vs. 207 cm). Additionally, the Hegari grown in the greenhouse was a bit taller than that grown in the field (169 vs. 126 cm). It should be noted that Quinby and Karper [104] measured to the height to the flag leaf,

while this study used height to the base of the panicle; therefore, comparisons are approximate. Nonetheless, differences in environment cannot be ruled out as a cause for our inability to locate *Dw4*. The SC170 x SB populations could provide some precedence for this as there is a peak at *Dw2* in the higher light intensity greenhouse but not the lower, although this difference could also be due to differences in phenotyping. Our other surprising result is that SB and AB are dominant at *Dw3* instead of recessive. As *Dw3* has been shown to affect height in both environments, the difference in environments does not explain this result.

Based on these results, it appears that *Dw4* does not exist. Additionally, the broomcorns are dominant *Dw3* instead of recessive as previously described. Both of these are surprising results that contradict a seminal paper in sorghum genetics. Further QTL mapping studies in other broomcorns, and possibly in the field, would be useful in confirming this surprising result.

CHAPTER V

CONCLUSION

SUMMARY

Sorghum demonstrates a great deal of height variation, from less than one meter to over four. This variation is due to a combination of internode length, time to flowering, and rate of phytomer production. Researchers in the first part of the 20th century determined that variation in internode length was due to four major genes, *Dw1-Dw4* [104]. In the early 2000s, the first *Dw* gene to be cloned was *Dw3*, an efflux transporter of auxin [109]. While the function of the other genes has not been determined, the location of *Dw1* has been shown to be on chromosome 9 at ~57 Mbp and *Dw2* is on chromosome 6 at ~42 Mbp [9,111,112,114,152]. The goal of this work was to identify the genes that correspond to the remaining three *Dw* loci.

Dw1 was found to be Sobic.009G229800, a highly conserved gene of unknown function. The causative mutation was a SNP that resulted in a premature stop codon. As the reference sequence is BTx623, which is recessive *dw1*, the mutation has caused some problems in the gene annotation for the reference genome. The actual intron/exon structure was clarified here by sequencing RNA from both parents of the mapping population, Hegari and 80M. The QTL mapping of *Dw1* revealed some other interesting characteristics of *Dw1*. In addition to length, *Dw1* was found to be segregating for fresh and dry stem weight. Furthermore, *Dw1* interacts with a previously described QTL on chromosome 7, *Dw07_56* [117]. The statistical interaction found here resulted in *Dw1*

having limited effect on height when coupled with the 80M allele at *Dw07_56*. However, *Dw1* had an appreciably larger effect on height when coupled with the Hegari allele.

Dw2 was identified as Sobic.006G067700, whose closest homolog in Arabidopsis is KIPK. KIPK is an AGCVIII kinase that interacts with KCBP [165]. KCBP is a unique kinesin that is involved in the arrangement of microtubules and actin in trichomes [171]. Initial QTL mapping of *Dw2* revealed some interesting trends. *Dw2* and *Dw3* have similar additive effects for the internode immediately below the peduncle. *Dw2* has similar effects with a slight decreasing trend for the next four internodes. For the internode below that, a nearby, not previously described QTL has a greater influence on length. On the other hand, for the second through fifth internode, *Dw3* has almost twice as great an additive effect as *Dw2* on the same internodes. From there the effect of *Dw3* decreases though it influenced the length of each internode measured.

Dw4 could not be located. Furthermore, the genotype of broomcorns at *Dw3* is dominant, not recessive as previously described [104]. Three crosses with SB and one with AB were constructed in the hopes of locating *Dw4*. Previously, the broomcorns were described as the only group of sorghums in the U.S. that were dominant at *Dw4* [104]. Thus, all of the populations should be segregating for a QTL that does not align with any of the other *Dw* loci and at that QTL the broomcorn allele should increase height. No QTL was found that matched those criteria. Several additional QTL were identified including a QTL on chromosome 4 that was previously suggested to be *Dw4* [117]. However, the AB allele at that QTL decreased length thus it is not *Dw4*.

Table 24. QTL for Average Internode Length Described in This Work.

Locus	Chr.	Populations Found In	Location of Peak (bp)	Allele That Increases Length
Dw01_62	1	Hegari x 80M	61,856,846	80M
		Hegari x SB	63,346,273	SB
Dw01_72	1	BTx623 x SB	72,249,472	SB
Dw02_59	2	Hegari x SB	59,224,916	SB
Dw02_66	2	SC x SB (LLG)	65,713,407	SB
Dw03_64	3	Hegari x JDB	63,953,497	Hegari
Dw04_68	4	BTx623 x AB	67,516,202	BTx623
Dw06_0	6	SYM x JDB	183,471	JDB
Dw2	6	Hegari x 80M	42,691,024	Hegari
		BTx623 x IS3620c	42,691,080	BTx623
		SC x SB (HLG)	46,083,204	SB
Dw07_56	7	Hegari x 80M	56,464,933	80M
		Hegari x SB	55,631,355	SB
		Hegari x JDB	56,241,011	JDB
Dw3	7	BTx623 x IS3620c	59,830,285	IS3620c
		SC x SB (HLG)	59,828,318	SB
		SC x SB (LLG)	59,613,664	SB
		BTx623 x SB	59,828,212	SB
		BTx623 x AB	59,613,618	AB
Dw07_64	7	BTx623 x AB	63,672,014	BTx623
Dw08_52	8	SYM x JDB	52,051,135	JDB
Dw1	9	Hegari x 80M	56,636,487	Hegari
		SC x SB (HLG)	57,069,211	SB
		SC x SB (LLG)	56,996,129	SB
		BTx623 x SB	56,082,948	SB
Dw10_2	10	SYM x JDB	2,029,603	JDB

Height variation in sorghum is thought to be the result of four genes of large effect, the *Dw* genes. However, there is variation in height within each of the *Dw* classes. This has been suggested to be from modifiers or allelic series at the *Dw* genes. I could not locate *Dw4*, so there appear to be only three major *Dw* genes. I also found additional QTL, mostly of small effects, as well as an interesting interaction between *Dw1* and a QTL on chromosome 7. Table 24 summarizes all of the QTL for average

internode length found in this study. Thus the genetic control of height in sorghum is complex involving three classical *dwarfing* genes along with many additional QTL and interaction between the QTL.

FUTURE WORK

While I found the gene that underlies *Dw1* and *Dw2*, there is much work to be done on elucidating the functions of those genes and how they influence height. While homologs of *Dw1* are found in many different plants, both monocot and dicot, there is no annotated function for the gene. Therefore this gene could be part of a previously unknown pathway, or part of a pathway, to control height. After the study described here was published, another group published a map based cloning paper on *Dw1* [150]. This group also identified Sobic.009G229800 as *Dw1*. Additionally, they found that *Dw1* changed the number of cells but not the cell size [150]. Future work could investigate if *Dw1* is involved in height regulation through one of the phytohormones as well as looking into differences in global gene expression for the different *Dw1* alleles. It would also be interesting to investigate the location of *Dw1* within the cell, as it is annotated as localizing at the plasma membrane and in the nucleus in Arabidopsis. One more potentially fascinating line of future work would be to determine the gene or genes that underlie *Dw07_56*, the QTL that statistically interacts with *Dw1*.

On the other hand *Dw2* has a homolog in Arabidopsis that has been described to a degree, though much remains to be determined. The Arabidopsis homolog has been shown to interact with the kinesin KCBP [165] and some of the PERKs [170]. While

KIPK is a kinase it is not known if any of these proteins is phosphorylated by KIPK or if it is another protein entirely. In addition to determining the phosphorylation target of KIPK, it would be interesting to look into the downstream results of the KIPK mutation through global gene expression. Furthermore, little is known about the actions of the PERKs, though studies have implicated them in root growth, wound response, and stem branching and they appear to be at least somewhat redundant in function [170,175]. It would be beneficial to determine how well conserved these functions are in monocots like sorghum. Examining the function of both *Dw1* and *Dw2* may help to elucidate some new means of regulation of stem growth.

REFERENCES

1. Olson SN, Ritter K, Rooney W, Kemanian A, McCarl BA, Zhang Y, et al. High biomass yield energy sorghum: Developing a genetic model for C4 grass bioenergy crops. Biofuels, Bioprod Biorefining. 2012; 640–655. doi:10.1002/bbb.1357

2. Swigoňová Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, et al. Close split of sorghum and maize genome progenitors. Genome Res. 2004;14: 1916–1923. doi:10.1101/gr.2332504

3. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, et al. The *Sorghum bicolor* genome and the diversification of grasses. Nature. 2009;457: 551–556. doi:10.1038/nature07723

4. Rooney WL. Integrating traditional and new technology to produce improved genotypes. Adv Agron. 2004;83. doi:10.1016/S0065-2113(04)83002-5

5. Harlan JR, de Wet JMJ. A simplified classification of cultivated sorghum. Crop Sci. 1972;12: 172–176.

6. Quinby JR. The genetics of sorghum improvement. J Hered. 1975;66: 56–62.

7. Karper RE, Quinby JR. The history and evolution of milo in the United States. Agron J. 1946;38: 441–453. doi:10.2134/agronj1946.00021962003800050007x

8. Stephens JC, Miller FR, Rosenow DT. Conversion of alien sorghums to early combine genotypes. Crop Sci. 1967;7: 396. doi:10.2135/cropsci1967.0011183X000700040036x

9. Klein RR, Mullet JE, Jordan DR, Miller FR, Rooney WL, Menz MA, et al. The effect of tropical sorghum conversion and inbred development on genome diversity as revealed by high-resolution genotyping. Crop Sci. 2008;48: S12–S26. doi:10.2135/cropsci2007.06.0319tpg

10. Cosgrove DJ. Growth of the plant cell wall. Nat Rev Mol Cell Biol. 2005;6: 850–861. doi:10.1038/nrm1746

11. Vogel J. Unique aspects of the grass cell wall. Curr Opin Plant Biol. 2008;11: 301–307. doi:10.1016/j.pbi.2008.03.002

12. Cosgrove DJ. Re-constructing our models of cellulose and primary cell wall

assembly. Curr Opin Plant Biol. 2014;22: 122–131. doi:10.1016/j.pbi.2014.11.001

13. Cosgrove DJ. How do plant cell walls extend? Plant Physiol. 1993;102: 1–6. doi:10.1104/pp.102.1.1

14. Rayle DL, Cleland RE. The Acid Growth Theory of auxin-induced cell elongation is alive and well. Plant Physiol. 1992;99: 1271–1274. doi:10.1104/pp.99.4.1271

15. Hager A. Role of the plasma membrane H+-ATPase in auxin-induced elongation growth: Historical and new aspects. J Plant Res. 2003;116: 483–505. doi:10.1007/s10265-003-0110-x

16. Sánchez-Rodríguez C, Rubio-Somoza I, Sibout R, Persson S. Phytohormones and the cell wall in Arabidopsis during seedling growth. Trends Plant Sci. 2010;15: 291–301. doi:10.1016/j.tplants.2010.03.002

17. Kumar M, Campbell L, Turner S. Secondary cell walls: Biosynthesis and manipulation. J Exp Bot. 2016;67: 515–531. doi:10.1093/jxb/erv533

18. Vanneste S, Friml J. Auxin: A trigger for change in plant development. Cell. 2009;136: 1005–1016. doi:10.1016/j.cell.2009.03.001

19. Kim J, Harter K, Theologis A. Protein-protein interactions among the Aux/IAA proteins. Proc Natl Acad Sci. 1997;94: 11786–91. doi:10.1073/pnas.94.22.11786

20. Szemenyei H, Hannon M, Long JA. TOPLESS mediates auxin-dependent transcriptional repression during Arabidopsis embryogenesis. Science. 2008;319: 1384–1386. doi:1151461 [pii]\r10.1126/science.1151461

21. Gray WM, Kepinski S, Rouse D, Leyser O, Estelle M. Auxin regulates SCF(TIR1)-dependent degradation of AUX/IAA proteins. Nature. 2001;414: 271–276. doi:10.1038/35104500

22. Dharmasiri N, Dharmasiri S, Estelle M. The F-box protein TIR1 is an auxin receptor. Nature. 2005;435: 441–5. doi:10.1038/nature03543

23. Kepinski S, Leyser O. The Arabidopsis F-box protein TIR1 is an auxin receptor. Nature. 2005;435: 446–451. doi:10.1038/nature03542

24. Takahashi K, Hayashi K -i., Kinoshita T. Auxin activates the plasma membrane H+-ATPase by phosphorylation during hypocotyl elongation in Arabidopsis. Plant Physiol. 2012;159: 632–641. doi:10.1104/pp.112.196428

25. Leblanc N, David K, Grosclaude J, Pradier JM, Barbier-Brygoo H, Labiau S, et al. A novel immunological approach establishes that the auxin-binding protein,

Nt-abp1, is an element involved in auxin signaling at the plasma membrane. J Biol Chem. 1999;274: 28314–28320. doi:10.1074/jbc.274.40.28314

26. Chen X, Grandont L, Li H, Hauschild R, Paque S, Abuzeineh A, et al. Inhibition of cell expansion by rapid ABP1-mediated auxin effect on microtubules. Nature. 2014; doi:10.1038/nature13889

27. Gao Y, Zhang Y, Zhang D, Dai X, Estelle M, Zhao Y. Auxin binding protein 1 (ABP1) is not required for either auxin signaling or Arabidopsis development. Proc Natl Acad Sci USA. 2015;112: 2275–2280. doi:10.1073/pnas.1500365112

28. Yang Y, Hammes UZ, Taylor CG, Schachtman DP, Nielsen E. High-affinity auxin transport by the AUX1 influx carrier protein. Curr Biol. 2006;16: 1123–1127. doi:10.1016/j.cub.2006.04.029

29. Gälweiler L, Guan C, Müller A, Wisman E, Mendgen K, Yephremov A, et al. Regulation of polar auxin transport by AtPIN1 in Arabidopsis vascular tissue. Science. 1998;282: 2226–2230. doi:10.1126/science.282.5397.2226

30. Geisler M, Blakeslee JJ, Bouchard R, Lee OR, Vincenzetti V, Bandyopadhyay A, et al. Cellular efflux of auxin catalyzed by the Arabidopsis MDR/PGP transporter AtPGP1. Plant J. 2005;44: 179–94. doi:10.1111/j.1365-313X.2005.02519.x

31. Vieten A, Sauer M, Brewer PB, Friml J. Molecular and cellular aspects of auxin-transport-mediated development. Trends Plant Sci. 2007;12: 160–168. doi:10.1016/j.tplants.2007.03.006

32. Petrasek J, Friml J. Auxin transport routes in plant development. Development. 2009;136: 2675–2688. doi:10.1242/dev.030353

33. Titapiwatanakun B, Murphy AS. Post-transcriptional regulation of auxin transport proteins: Cellular trafficking, protein phosphorylation, protein maturation, ubiquitination, and membrane composition. J Exp Bot. 2009;60: 1093–107. doi:10.1093/jxb/ern240

34. Blakeslee JJ, Bandyopadhyay A, Lee OR, Mravec J, Titapiwatanakun B, Sauer M, et al. Interactions among PIN-FORMED and P-glycoprotein auxin transporters in Arabidopsis. Plant Cell. 2007;19: 131–47. doi:10.1105/tpc.106.040782

35. Zhang Y, McCormick S. AGCVIII kinases: At the crossroads of cellular signaling. Trends Plant Sci. 2009;14: 689–695. doi:10.1016/j.tplants.2009.09.006

36. Rademacher EH, Offringa R. Evolutionary adaptations of plant AGC kinases: From light signaling to cell polarity regulation. Front Plant Sci. 2012;3: 250. doi:10.3389/fpls.2012.00250

37. Barbosa ICR, Schwechheimer C. Dynamic control of auxin transport-dependent growth by AGCVIII protein kinases. Curr Opin Plant Biol. 2014;22: 8. doi:10.1016/j.pbi.2014.09.010

38. Friml J, Yang X, Michniewicz M, Weijers D, Quint A, Tietz O, et al. A PINOID-dependent binary switch in apical-basal PIN polar targeting directs auxin efflux. Science. 2004;306: 862–865. doi:10.1126/science.1100618

39. Dhonukshe P, Huang F, Galvan-Ampudia CS, Mahonen AP, Kleine-Vehn J, Xu J, et al. Plasma membrane-bound AGC3 kinases phosphorylate PIN auxin carriers at TPRXS(N/S) motifs to direct apical PIN recycling. Development. 2010;137: 3245–3255. doi:10.1242/dev.052456

40. Dai M, Zhang C, Kania U, Chen F, Xue Q, Mccray T, et al. A PP6-type phosphatase holoenzyme directly regulates PIN phosphorylation and auxin efflux in Arabidopsis. Plant Cell. 2012;24: 2497–2514. doi:10.1105/tpc.112.098905

41. Zourelidou M, Müller I, Willige BC, Nill C, Jikumaru Y, Li H, et al. The polarly localized D6 PROTEIN KINASE is required for efficient auxin transport in *Arabidopsis thaliana*. Development. 2009;636: 627–636. doi:10.1242/dev.028365

42. Zourelidou M, Absmanner B, Weller B, Barbosa ICR, Willige BC, Fastner A, et al. Auxin efflux by PIN-FORMED proteins is activated by two different protein kinases, D6 PROTEIN KINASE and PINOID. Elife. 2014;2014: 1–25. doi:10.7554/eLife.02860

43. Henrichs S, Wang B, Fukao Y, Zhu J, Charrier L, Bailly A, et al. Regulation of ABCB1/PGP1-catalysed auxin transport by linker phosphorylation. EMBO J. 2012;31: 2965–80. doi:10.1038/emboj.2012.120

44. Yamaguchi S. Gibberellin metabolism and its regulation. Annu Rev Plant Biol. 2008;59: 225–251. doi:10.1146/annurev.arplant.59.032607.092804

45. Ueguchi-Tanaka M, Ashikari M, Nakajima M, Itoh H, Katoh E, Kobayashi M, et al. GIBBERELLIN INSENSITIVE DWARF1 encodes a soluble receptor for gibberellin. Nature. 2005;437: 693–8. doi:10.1038/nature04028

46. Peng J, Carol P, Richards DE, King KE, Cowling RJ, Murphy GP, et al. The Arabidopsis GAI gene defines a signalling pathway that negatively regulates gibberellin responses. Genes Dev. 1997;11: 3194–3205. doi:10.1101/gad.11.23.3194

47. Silverstone AL, Jung HS, Dill A, Kawaide H, Kamiya Y, Sun TP. Repressing a repressor: Gibberellin-induced rapid reduction of the RGA protein in Arabidopsis. Plant Cell. 2001;13: 1555–66. doi:10.1105/TPC.010047

48. Ikeda A, Uefuchi-Tanaka M, Sonoda Y, Kitano H, Koshioka M, Futsuhara Y, et al. slender rice, a constitutive gibberellin response mutant, is caused by a null mutation of the SLR1 gene, an ortholog of the height-regulating gene GAI/RGA/RHT/D8. Plant Cell. 2001;13: 999–1010. doi:10.1105/tpc.13.5.999

49. Schwechheimer C. Understanding gibberellic acid signaling-Are we there yet? Curr Opin Plant Biol. 2008;11: 9–15. doi:10.1016/j.pbi.2007.10.011

50. Feng S, Martinez C, Gusmaroli G, Wang Y, Zhou J, Wang F, et al. Coordinated regulation of *Arabidopsis thaliana* development by light and gibberellins. Nature. 2008;451: 475–479. doi:10.1038/nature06448

51. de Lucas M, Davière J-M, Rodríguez-Falcón M, Pontin M, Iglesias-Pedraz JM, Lorrain S, et al. A molecular framework for light and gibberellin control of cell elongation. Nature. 2008;451: 480–484. doi:10.1038/nature06520

52. Kinoshita T, Cano-Delgado A, Seto H, Hiranuma S, Fujioka S, Yoshida S, et al. Binding of brassinosteroids to the extracellular domain of plant kinase BRI1. Nature. 2005;433: 167–171. doi:10.1038/nature03188

53. Tang W, Kim T-W, Oses-Prieto JA, Sun Y, Deng Z, Zhu S, et al. BSKs mediate signal transduction from the receptor kinase BRI1 in Arabidopsis. Science. 2008;321: 557–60. doi:10.1126/science.1156973

54. Kim TW, Guan S, Burlingame AL, Wang ZY. The CDG1 kinase mediates brassinosteroid signal transduction from BRI1 receptor kinase to BSU1 phosphatase and GSK3-like kinase BIN2. Mol Cell. 2011;43: 561–571. doi:10.1016/j.molcel.2011.05.037

55. Mora-Garcia S, Vert G, Yin YH, Cano-Delgado A, Cheong H, Chory J. Nuclear protein phosphatases with Kelch-repeat domains modulate the response to brassinosteroids in Arabidopsis. Genes Dev. 2004;18: 448–460. doi:10.1101/gad.1174204

56. Kim T-W, Guan S, Sun Y, Deng Z, Tang W, Shang J-X, et al. Brassinosteroid signal transduction from cell-surface receptor kinases to nuclear transcription factors. Nat Cell Biol. 2009;11: 1254–60. doi:10.1038/ncb1970

57. He J-X, Gendron JM, Yang Y, Li J, Wang Z-Y. The GSK3-like kinase BIN2 phosphorylates and destabilizes BZR1, a positive regulator of the brassinosteroid signaling pathway in Arabidopsis. Proc Natl Acad Sci. 2002;99: 10185–90. doi:10.1073/pnas.152342599

58. He J-X, Gendron JM, Sun Y, Gampala SSL, Gendron N, Sun CQ, et al. BZR1 is a transcriptional repressor with dual roles in brassinosteroid homeostasis and

growth responses. Science. 2005;307: 1634–1638. doi:10.1126/science.1107580

59. Vert G, Chory J. Downstream nuclear events in brassinosteroid signalling. Nature. 2006;441: 96–100. doi:10.1038/nature04681

60. Gampala SS, Kim TW, He JX, Tang W, Deng Z, Bai MY, et al. An essential role for 14-3-3 proteins in brassinosteroid signal transduction in Arabidopsis. Dev Cell. 2007;13: 177–189. doi:10.1016/j.devcel.2007.06.009

61. Tang W, Yuan M, Wang R, Yang Y, Wang C, Oses-Prieto JA, et al. PP2A activates brassinosteroid-responsive gene expression and plant growth by dephosphorylating BZR1. Nat Cell Biol. 2011;13: 124–131. doi:10.1038/ncb2151

62. Zhu JY, Sae-Seaw J, Wang ZY. Brassinosteroid signalling. Development. 2013;140: 1615–20. doi:10.1242/dev.060590

63. Wang W, Bai MY, Wang ZY. The brassinosteroid signaling network-A paradigm of signal integration. Curr Opin Plant Biol. 2014;21: 147–153. doi:10.1016/j.pbi.2014.07.012

64. Wen X, Zhang C, Ji Y, Zhao Q, He W, An F, et al. Activation of ethylene signaling is mediated by nuclear translocation of the cleaved EIN2 carboxyl terminus. Cell Res. 2012;22: 1613–6. doi:10.1038/cr.2012.145

65. Merchante C, Alonso JM, Stepanova AN. Ethylene signaling: Simple ligand, complex regulation. Curr Opin Plant Biol. 2013;16: 554–560. doi:10.1016/j.pbi.2013.08.001

66. Hwang I, Sheen J, Muller B. Cytokinin signaling networks. Annu Rev Plant Biol. 2012;63: 353–80. doi:10.1146/annurev-arplant-042811-105503

67. Lorrain S, Allen T, Duek PD, Whitelam GC, Fankhauser C. Phytochrome-mediated inhibition of shade avoidance involves degradation of growth-promoting bHLH transcription factors. Plant J. 2008;53: 312–323. doi:10.1111/j.1365-313X.2007.03341.x

68. Jung JH, Domijan M, Klose C, Biswas S, Ezer D, Gao M, et al. Phytochromes function as thermosensors in Arabidopsis. Science. 2016;354: 886–889. doi:10.1126/science.aaf6005

69. Legris M, Klose C, Burgie ES, Rojas CCR, Neme M, Hiltbrunner A, et al. Phytochrome B integrates light and temperature signals in Arabidopsis. Science. 2016;354: 897–900. doi:10.1126/science.aaf5656

70. Pedmale U V., Huang SSC, Zander M, Cole BJ, Hetzel J, Ljung K, et al.

Cryptochromes interact directly with PIFs to control plant growth in limiting blue light. Cell. Elsevier Inc.; 2016;164: 233–245. doi:10.1016/j.cell.2015.12.018

71. Ma D, Li X, Guo Y, Chu J, Fang S, Yan C, et al. Cryptochrome 1 interacts with PIF4 to regulate high temperature-mediated hypocotyl elongation in response to blue light. Proc Natl Acad Sci. 2015;113: 224–229. doi:10.1073/pnas.1511437113

72. Keller MM, Jaillais Y, Pedmale U V., Moreno JE, Chory J, Ballaré CL. Cryptochrome 1 and phytochrome B control shade-avoidance responses in Arabidopsis via partially independent hormonal cascades. Plant J. 2011;67: 195–207. doi:10.1111/j.1365-313X.2011.04598.x

73. Keuskamp DH, Sasidharan R, Vos I, Peeters AJM, Voesenek L a CJ, Pierik R. Blue-light-mediated shade avoidance requires combined auxin and brassinosteroid action in Arabidopsis seedlings. Plant J. 2011;67: 208–217. doi:10.1111/j.1365-313X.2011.04597.x

74. Pedmale U V., Liscum E. Regulation of phototropic signaling in Arabidopsis via phosphorylation state changes in the phototropin 1-interacting protein NPH3. J Biol Chem. 2007;282: 19992–20001. doi:10.1074/jbc.M702551200

75. Roberts D, Pedmale U V, Morrow J, Sachdev S, Lechner E, Tang X, et al. Modulation of phototropic responsiveness in Arabidopsis through ubiquitination of phototropin 1 by the CUL3-Ring E3 ubiquitin ligase CRL3(NPH3). Plant Cell. 2011;23: 3627–3640. doi:10.1105/tpc.111.087999

76. Christie JM, Yang H, Richter GL, Sullivan S, Thomson CE, Lin J, et al. Phot1 inhibition of ABCB19 primes lateral auxin fluxes in the shoot apex required for phototropism. PLoS Biol. 2011;9. doi:10.1371/journal.pbio.1001076

77. Oh E, Zhu JY, Bai MY, Arenhart RA, Sun Y, Wang ZY. Cell elongation is regulated through a central circuit of interacting transcription factors in the Arabidopsis hypocotyl. Elife. 2014;2014: 1–19. doi:10.7554/eLife.03031

78. Bai M-Y, Shang J-X, Oh E, Fan M, Bai Y, Zentella R, et al. Brassinosteroid, gibberellin and phytochrome impinge on a common transcription module in Arabidopsis. Nat Cell Biol. 2012;14: 810–817. doi:10.1038/ncb2546

79. Locascio A, Blázquez MA, Alabadí D. Genomic analysis of DELLA protein activity. Plant Cell Physiol. 2013;54: 1229–1237. doi:10.1093/pcp/pct082

80. Ikeda M, Fujiwara S, Mitsuda N, Ohme-takagi M. A triantagonistic basic helix-loop-helix system regulates cell elongation in Arabidopsis. Plant Cell. 2013;24: 4483–4497. doi:10.1105/tpc.112.105023

81. Bai MY, Fan M, Oh E, Wang ZY. A triple helix-loop-helix/basic helix-loop-helix cascade controls cell elongation downstream of multiple hormonal and environmental signaling pathways in Arabidopsis. Plant Cell. 2012;24: 4917–4929. doi:10.1105/tpc.112.105163

82. Chaiwanon J, Wang W, Zhu JY, Oh E, Wang ZY. Information integration and communication in plant growth regulation. Cell. Elsevier Inc.; 2016;164: 1257–1268. doi:10.1016/j.cell.2016.01.044

83. Haruta M, Sabat G, Stecker K, Minkoff BB, Sussman MR. A peptide hormone and its receptor. Science. 2014;343: 408–411. doi:10.1126/science.1244454

84. Polyn S, Willems A, De Veylder L. Cell cycle entry, maintenance, and exit during plant development. Curr Opin Plant Biol. 2015;23: 1–7. doi:10.1016/j.pbi.2014.09.012

85. Haga N, Kobayashi K, Suzuki T, Maeo K, Kubo M, Ohtani M, et al. Mutations in MYB3R1 and MYB3R4 cause pleiotropic developmental defects and preferential down-regulation of multiple G2/M-specific genes in Arabidopsis. Plant Physiol. 2011;157: 706–717. doi:10.1104/pp.111.180836

86. Schoof H, Lenhard M, Haecker A, Mayer KF, Jürgens G, Laux T. The stem cell population of Arabidopsis shoot meristems in maintained by a regulatory loop between the CLAVATA and WUSCHEL genes. Cell. 2000;100: 635–644. doi:10.1016/S0092-8674(00)80700-X

87. Aichinger E, Kornet N, Friedrich T, Laux T. Plant stem cell niches. Annu Rev Plant Biol. 2012;63: 615–636. doi:10.1146/annurev-arplant-042811-105555

88. Gaillochet C, Daum G, Lohmann JU. O cell, where art thou? The mechanisms of shoot meristem patterning. Curr Opin Plant Biol. 2015;23: 91–97. doi:10.1016/j.pbi.2014.11.002

89. Zhao Z, Andersen SU, Ljung K, Dolezal K, Miotk A, Schultheiss SJ, et al. Hormonal control of the shoot stem-cell niche. Nature. 2010;465: 1089–92. doi:10.1038/nature09126

90. Hay A, Kaur H, Phillips A, Hedden P, Hake S, Tsiantis M. The gibberellin pathway mediates KNOTTED1-type homeobox function in plants with different body plans. Curr Biol. 2002;12: 1557–1565. doi:10.1016/S0960-9822(02)01125-9

91. Jasinski S, Piazza P, Craft J, Hay A, Woolley L, Rieu I, et al. KNOX action in Arabidopsis is mediated by coordinate regulation of cytokinin and gibberellin activities. Curr Biol. 2005;15: 1560–1565. doi:10.1016/j.cub.2005.07.023

92. Besnard F, Refahi Y, Morin V, Marteaux B, Brunoud G, Chambrier P, et al. Cytokinin signalling inhibitory fields provide robustness to phyllotaxis. Nature. 2014;505: 417–21. doi:10.1038/nature12791

93. Gendron JM, Liu J-S, Fan M, Bai M-Y, Wenkel S, Springer PS, et al. Brassinosteroids regulate organ boundary formation in the shoot apical meristem of Arabidopsis. Proc Natl Acad Sci. 2012;109: 21152–7. doi:10.1073/pnas.1210799110

94. Bell EM, Lin W, Husbands AY, Yu L, Jaganatha V, Jablonska B, et al. Arabidopsis lateral organ boundaries negatively regulates brassinosteroid accumulation to limit growth in organ boundaries. Proc Natl Acad Sci. 2012;109: 21146–51. doi:10.1073/pnas.1210789109

95. Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, et al. "Green Revolution" genes encode mutant gibberellin response modulators. Nature. 1999;400: 256–261. doi:10.1038/22307

96. Monna L, Kitazawa N, Yoshino R, Suzuki J, Masuda H, Maehara Y, et al. Positional cloning of rice semidwarfing gene, sd-1: Rice "Green Revolution gene" encodes a mutant enzyme envolved in gibberellin synthesis. DNA Res. 2002;9: 11–17. doi:10.1093/dnares/9.1.11

97. Sasaki A, Ashikari M, Ueguchi-Tanaka M, Itoh H, Nishimura A, Swapan D, et al. A mutant gibberellin-synthesis gene in rice. Nature. 2002;291: 1–2. doi:10.1038/416701a

98. Spielmeyer W, Ellis MH, Chandler PM. Semidwarf (sd-1), "Green Revolution" rice, contains a defective gibberellin 20-oxidase gene. Proc Natl Acad Sci. 2002;99: 9043–9048. doi:10.1073/pnas.132266399

99. Chono M, Honda I, Zeniya H, Yoneyama K, Saisho D, Takeda K, et al. A semidwarf phenotype of barley uzu results from a nucleotide substitution in the gene encoding a putative brassinosteroid receptor. Plant Physiol. 2003;133: 1209–1219. doi:10.1104/pp.103.026195

100. Hattori Y, Nagai K, Furukawa S, Song X-J, Kawano R, Sakakibara H, et al. The ethylene response factors SNORKEL1 and SNORKEL2 allow rice to adapt to deep water. Nature. 2009;460: 1026–1030. doi:10.1038/nature08258

101. Xu K, Xu X, Fukao T, Canlas P, Maghirang-Rodriguez R, Heuer S, et al. Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. Nature. 2006;442: 705–708. doi:10.1038/nature04920

102. Fukao T, Bailey-Serres J. Submergence tolerance conferred by Sub1A is mediated

by SLR1 and SLRL1 restriction of gibberellin responses in rice. Proc Natl Acad Sci U S A. 2008;105: 16814–16819. doi:10.1073/pnas.0807821105

103. Sieglinger JB. Inheritance of height in broomcorn. J Agric Res. 1932;44: 13–20.

104. Quinby JR, Karper RE. Inheritance of height in sorghum. Agron J. 1954;46: 211–216.

105. Graham D, Lessman KJ. Effect of height on yield and yield components of two isogenic lines of *Sorghum vulgare* Pers. Crop Sci. 1966;6: 372–374. doi:10.2135/cropsci1966.0011183X000600040024x

106. Pereira MG, Lee M. Identification of genomic regions affecting plant height in sorghum and maize. Theor Appl Genet. 1995;90: 380–388. doi:10.1007/BF00221980

107. Casady AJ. Effect of a single height gene (*Dw3*) of *Sorghum vulgare* Pers. on certain culm and leaf blade characteristics. Crop Sci. 1967;7: 595–598.

108. Truong SK, McCormick RF, Rooney WL, Mullet JE. Harnessing genetic variation in leaf angle to increase productivity of *Sorghum bicolor*. Genetics. 2015;201: 1229–1238. doi:10.1534/genetics.115.178608

109. Multani DS, Briggs SP, Chamberlin MA, Blakeslee JJ, Murphy AS, Johal GS. Loss of an MDR transporter in compact stalks of maize *br2* and sorghum *dw3* mutants. Science. 2003;302: 81–84. doi:10.1126/science.1086072

110. Knöller AS, Blakeslee JJ, Richards EL, Peer WA, Murphy AS. Brachytic2/ZmABCB1 functions in IAA export from intercalary meristems. J Exp Bot. 2010;61: 3689–96. doi:10.1093/jxb/erq180

111. Brown PJ, Rooney WL, Franks C, Kresovich S. Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes. Genetics. 2008;180: 629–37. doi:10.1534/genetics.108.092239

112. Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, et al. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proc Natl Acad Sci. 2013;110: 453–8. doi:10.1073/pnas.1215985110

113. Thurber CS, Ma JM, Higgins RH, Brown PJ. Retrospective genomic analysis of sorghum adaptation to temperate-zone grain production. Genome Biol. 2013;14: R68. doi:10.1186/gb-2013-14-6-r68

114. Higgins RH, Thurber CS, Assaranurak I, Brown PJ. Multiparental mapping of plant height and flowering time QTL in partially isogenic sorghum families. G3

Genes, Genomes, Genet. 2014;4: 1593–602. doi:10.1534/g3.114.013318

115. Ordonio RL, Ito Y, Hatakeyama A, Ohmae-Shinohara K, Kasuga S, Tokunaga T, et al. Gibberellin deficiency pleiotropically induces culm bending in sorghum: An insight into sorghum semi-dwarf breeding. Sci Rep. 2014;4: 5287. doi:10.1038/srep05287

116. Reid JB, Ross JJ, Swain SM. Internode length in Pisum-A new, slender mutant with elevated levels of C19 gibberellins. Planta. 1992;188: 462–467. doi:10.1007/BF00197036

117. Li X, Li X, Fridman E, Tesso TT, Yu J. Dissecting repulsion linkage in the dwarfing gene *Dw3* region for sorghum plant height provides insights into heterosis. Proc Natl Acad Sci. 2015;112: 11823–11828. doi:10.1073/pnas.1509229112

118. Mullet J, Morishige D, McCormick R, Truong S, Hilley J, McKinley B, et al. Energy sorghum-A genetic model for the design of C4 grass bioenergy crops. J Exp Bot. 2014;65: 3479–3489. doi:10.1093/jxb/eru229

119. Rooney WL, Blumenthal J, Bean B, Mullet JE. Designing sorghum as a dedicated bioenergy feedstock. Biofuels, Bioprod Biorefining. 2007;1: 147–157. doi:10.1002/bbb.15

120. Gill JR, Burks PS, Staggenborg SA, Odvody GN, Heiniger RW, Macoon B, et al. Yield results and stability analysis from the sorghum regional biomass feedstock trial. BioEnergy Res. 2014;7: 1026–1034. doi:10.1007/s12155-014-9445-5

121. Casady AJ. Effect of a single height (*Dw3*) gene of sorghum on grain yield, grain yield components, and test weight. Crop Sci. 1965;5: 385–388. doi:10.2135/cropsci1965.0011183X000500050002x

122. Murray SC, Sharma A, Rooney WL, Klein PE, Mullet JE, Mitchell SE, et al. Genetic improvement of sorghum as a biofuel feedstock: I. QTL for stem sugar and grain nonstructural carbohydrates. Crop Sci. 2008;48: 2165–2179. doi:10.2135/cropsci2008.01.0016

123. Quinby J. Sorghum improvement and the genetics of growth. College Station, TX: Texas A&M University Press; 1974.

124. Morishige DT, Klein PE, Hilley JL, Sahraeian SME, Sharma A, Mullet JE. Digital genotyping of sorghum-A diverse plant species with a large repeat-rich genome. BMC Genomics. 2013;14: 448. doi:10.1186/1471-2164-14-448

125. Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, et al.

MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics. 1987;1: 174–181. doi:10.1016/0888-7543(87)90010-3

126. Wang S, Basten C, Zeng Z. Windows QTL Cartographer 2.5 [Internet]. Raleigh, NC: Department of Statistics, North Carolina State University; 2012. Available: http://statgen.ncsu.edu/qtlcart/WQTLCart.htm

127. Manichaikul A, Moon JY, Sen Ś, Yandell BS, Broman KW. A model selection approach for the identification of quantitative trait loci in experimental crosses, allowing epistasis. Genetics. 2009;181: 1077–1086. doi:10.1534/genetics.108.094565

128. Broman KW, Wu H, Sen Ś, Churchill GA. R/qtl: QTL mapping in experimental crosses. Bioinformatics. 2003;19: 889–890. doi:10.1093/bioinformatics/btg112

129. Tuinstra MR, Ejeta G, Goldsbrough PB. Heterogeneous inbred family (HIF) analysis: A method for developing near-isogenic lines that differ at quantitative trait loci. Theor Appl Genet. 1997;95: 1005–1011. doi:10.1007/s001220050654

130. Bookout AL, Mangelsdorf DJ. Quantitative real-time PCR protocol for analysis of nuclear receptor signaling pathways. Nucl Recept Signal. 2003;1: e012. doi:10.1621/nrs.01012

131. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25: 1189–1191. doi:10.1093/bioinformatics/btp033

132. Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol. 2000;302: 205–17. doi:10.1006/jmbi.2000.4042

133. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30: 2725–2729. doi:10.1093/molbev/mst197

134. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acid Res. 2004;32: 1792–1797. doi:10.1093/nar/gkh340

135. Edgar RC. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5: 113. doi:10.1186/1471-2105-5-113

136. Buchan DWA, Minneci F, Nugent TCO, Bryson K, Jones DT. Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res.

2013;41: 349–357. doi:10.1093/nar/gkt381

137. Nugent T, Jones DT. Transmembrane protein topology prediction using support vector machines. BMC Bioinformatics. 2009;10: 159. doi:10.1186/1471-2105-10-159

138. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol. 2004;337: 635–645. doi:10.1016/j.jmb.2004.02.002

139. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins Struct Funct Genet. 2001;42: 38–48. doi:10.1002/1097-0134(20010101)42:1<38::AID-PROT50>3.0.CO;2-3

140. Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, et al. FoldIndex: A simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics. 2005;21: 3435–3438. doi:10.1093/bioinformatics/bti537

141. Ritter KB, Jordan DR, Chapman SC, Godwin ID, Mace ES, Lynne McIntyre C. Identification of QTL for sugar-related traits in a sweet x grain sorghum (*Sorghum bicolor* (L.) Moench) recombinant inbred population. Mol Breed. 2008;22: 367–384. doi:10.1007/s11032-008-9182-6

142. Mocoeur A, Zhang YM, Liu ZQ, Shen X, Zhang LM, Rasmussen SK, et al. Stability and genetic control of morphological, biomass and biofuel traits under temperate maritime and continental conditions in sweet sorghum (*Sorghum bicolour*). Theor Appl Genet. 2015;128: 1685–1701. doi:10.1007/s00122-015-2538-5

143. Karper RE. A dominant mutation of frequent recurrence in sorghum. Am Nat. 1932;66: 511–529. doi:10.2307/2678832

144. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res. 2001;11: 863–74. doi:10.1101/gr.176601

145. Benschop JJ, Mohammed S, O'Flaherty M, Heck AJR, Slijper M, Menke FLH. Quantitative phosphoproteomics of early elicitor signaling in Arabidopsis. Mol Cell Proteomics. 2007;6: 1198–1214. doi:10.1074/mcp.M600429-MCP200

146. Quinby JR, Karper RE. Inheritance of duration of growth in the milo group of sorghum. Crop Sci. 1961;1: 8–10. doi:10.2135/cropsci1961.0011183X000100010004x

147. Slewinski TL. Non-structural carbohydrate partitioning in grass stems: A target to

increase yield stability, stress tolerance, and biofuel production. J Exp Bot. 2012;63: 4647–4670. doi:10.1093/jxb/ers124

148. McKinley B, Rooney W, Wilkerson C, Mullet J. Dynamics of biomass partitioning, stem gene expression, cell wall biosynthesis, and sucrose accumulation during development of *Sorghum bicolor*. Plant J. 2016;88: 662–680. doi:10.1111/tpj.13269

149. Hilley J, Truong S, Olson S, Morishige D, Mullet J. Identification of *Dw1* , a regulator of sorghum stem internode length. PLoS One. 2016;11: e0151271. doi:10.1371/journal.pone.0151271

150. Yamaguchi M, Fujimoto H, Hirano K, Araki-Nakamura S, Ohmae-Shinohara K, Fujii A, et al. Sorghum *Dw1*, an agronomically important gene for lodging resistance, encodes a novel protein involved in cell proliferation. Sci Rep. 2016;6: 28366. doi:10.1038/srep28366

151. Murphy RL, Klein RR, Morishige DT, Brady JA, Rooney WL, Miller FR, et al. Coincident light and clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum. Proc Natl Acad Sci. 2011;108: 16469–16474. doi:10.1073/pnas.1106212108

152. Thurber CS, Ma JM, Higgins RH, Brown PJ. Retrospective genomic analysis of sorghum adaptation to temperate-zone grain production. Genome Biol. 2013;14: R68. doi:10.1186/gb-2013-14-6-r68

153. Burow GB, Klein RR, Franks CD, Klein PE, Schertz KF, Pederson GA, et al. Registration of the BTx623/IS3620c recombinant inbred mapping population of sorghum. J Plant Regist. 2011;5: 141. doi:10.3198/jpr2010.04.0219crmp

154. Hart GE, Schertz KF, Peng Y, Syed NH. Genetic mapping of *Sorghum bicolor* (L.) Moench QTLs that control variation in tillering and other morphological characters. Theor Appl Genet. 2001;103: 1232–1242. doi:10.1007/s001220100582

155. Truong SK, McCormick RF, Morishige DT, Mullet JE. Resolution of genetic map expansion caused by excess heterozygosity in plant recombinant inbred populations. G3 Genes, Genomes, Genet. 2014;4: 1963–1969. doi:10.1534/g3.114.012468

156. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25: 1754–1760. doi:10.1093/bioinformatics/btp324

157. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al.

The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20: 1297–1303. doi:10.1101/gr.107524.110

158. DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43: 491–8. doi:10.1038/ng.806

159. van der Auwera G, Carneiro M, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From FastQ data to high-confidence variant calls: The Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinforma. 2013;43: 11.10.1-11.10.33.

160. McCormick RF, Truong SK, Mullet JE. RIG: Recalibration and interrelation of genomic sequence data with the GATK. G3 Genes, Genomes, Genet. 2015;5: 655–665. doi:10.1534/g3.115.017012

161. Peng B, Yu RK, Dehoff KL, Amos CI. Normalizing a large number of quantitative traits using empirical normal quantile transformation. BMC Proc. 2007;1: S156. doi:10.1186/1753-6561-1-s1-s156

162. Kwak IY, Moore CR, Spalding EP, Broman KW. A simple regression-based method to map quantitative trait loci underlying function-valued phenotypes. Genetics. 2014;197: 1409–1416. doi:10.1534/genetics.114.166306

163. Weers BD. Integrated analysis of phenology, traits, and QTL in the drought tolerant sorghum genotypes BTx642 and RTx7000. Dissertation, Texas A&M University. 2011.

164. Le SQ, Gascuel O. An improved general amino acid replacement matrix. Mol Biol Evol. 2008;25: 1307–1320. doi:10.1093/molbev/msn067

165. Day IS, Miller C, Golovkin M, Reddy ASN. Interaction of a kinesin-like calmodulin-binding protein with a protein kinase. J Biol Chem. 2000;275: 13737–13745. doi:10.1074/jbc.275.18.13737

166. Bögre L, Okrész L, Henriques R, Anthony RG. Growth signalling pathways in Arabidopsis and the AGC protein kinases. Trends Plant Sci. 2003;8: 424–31. doi:10.1016/S1360-1385(03)00188-2

167. Zegzouti H, Li W, Lorenz TC, Xie M, Payne CT, Smith K, et al. Structural and functional insights into the regulation of Arabidopsis AGC VIIIa kinases. J Biol Chem. 2006;281: 35520–35530. doi:10.1074/jbc.M605167200

168. Galván-Ampudia CS, Offringa R. Plant evolution: AGC kinases tell the auxin

tale. Trends Plant Sci. 2007;12: 541–547. doi:10.1016/j.tplants.2007.10.004

169. Yang S, Weers BD, Morishige DT, Mullet JE. CONSTANS is a photoperiod regulated activator of flowering in sorghum. BMC Plant Biol. 2014;14: 148. doi:10.1186/1471-2229-14-148

170. Humphrey T V., Haasen KE, Aldea-Brydges MG, Sun H, Zayed Y, Indriolo E, et al. PERK-KIPK-KCBP signalling negatively regulates root growth in *Arabidopsis thaliana*. J Exp Bot. 2015;66: 71–83. doi:10.1093/jxb/eru390

171. Tian J, Han L, Feng Z, Wang G, Liu W, Ma Y, et al. Orchestration of microtubules and the actin cytoskeleton in trichome cell shape determination by a plant-unique kinesin. Elife. 2015;4: e09351. doi:10.7554/eLife.09351

172. Jonsson E, Yamada M, Vale RD, Goshima G. Clustering of a kinesin-14 motor enables processive retrograde microtubule-based transport in plants. Nat Plants. 2015;1: 1–7. doi:10.1038/nplants.2015.87

173. Vinogradova M V, Malanina GG, Reddy ASN, Fletterick RJ. Structure of the complex of a mitotic kinesin with its calcium binding regulator. Proc Natl Acad Sci. 2009;106: 8175–8179.

174. Vinogradova M V, Malanina GG, Waitzman JS, Rice SE, Fletterick RJ. Plant kinesin-like calmodulin binding protein employs its regulatory domain for dimerization. PLoS One. 2013;8: e66669. doi:10.1371/journal.pone.0066669

175. Borassi C, Sede AR, Mecchia MA, Salgado Salter JD, Marzol E, Muschietti JP, et al. An update on cell surface proteins containing extensin-motifs. J Exp Bot. 2016;67: 477–487. doi:10.1093/jxb/erv455

176. Bai L, Zhang G, Zhou Y, Zhang Z, Wang W, Du Y, et al. Plasma membrane-associated proline-rich extensin-like receptor kinase 4, a novel regulator of Ca2+ signalling, is required for abscisic acid responses in *Arabidopsis thaliana*. Plant J. 2009;60: 314–327. doi:10.1111/j.1365-313X.2009.03956.x

177. Noh B, Murphy AS, Spalding EP. Multidrug resistance-like genes of Arabidpsis required for auxin transport and auxin-mediated development. Plant Cell. 2001;13: 2441–2454. doi:10.1105/tpc.010350

178. Casa AM, Pressoir G, Brown PJ, Mitchell SE, Rooney WL, Tuinstra MR, et al. Community resources and strategies for association mapping in sorghum. Crop Sci. 2008;48: 30–40. doi:10.2135/cropsci2007.02.0080

179. Olson S. Designing an ideal energy crop: The case for *Sorghum bicolor*. Dissertation, Texas A&M University. 2012.

180. Farfan IDB, Bergsma BR, Johal G, Tuinstra MR. A stable *dw3* allele in sorghum and a molecular marker to facilitate selection. Crop Sci. 2012;52: 2063–2069. doi:10.2135/cropsci2011.12.0631

181. Miller FR, Dahlberg JA, Morgan PW. Twenty A3/B3 sorghum maturity and height parental lines. Crop Sci. 1999;39: 306–307.

APPENDIX

PRIMERS USED IN THIS STUDY

PRIMERS USED IN CHAPTER II

Table A1. Primers for Fine Mapping of *Dw1*.

Marker Name	Forward Primer Sequence	Reverse Primer Sequence	Marker Type	RE (CAPS only)	Location (bp)
CAPS_2	GGCAAGCTTAGTTGAAGTTGTT	GTCCCAATGACTTGGCTATCT	CAPS	CviQI	56,763,365
SNP_110	CAAGGTTTCTCTGCCACTAGAC	TGCTTGGGTAACGTGGTAATC	SNP		56,925,217
SNP_180	GGTGTTTCATCGTCCTCCTATC	CCGTACCTGATGATGGGATTAG	SNP		56,975,859
SNP_210	CGGGTGGCAATTAGAAGTAAGG	TCCATCCATGCTGACCTTAAC	SNP		56,998,902
SNP_220	TGCTCCTGAAACTGCTAACC	GAGGAGGTTCCAGGTTGAAAT	SNP		57,020,309
SNP_230	TGTTTAGGAAGGCTCCATGTC	TCCAACGCCACAACTGTAA	SNP		57,025,222
SNP_250	GATCGCCTAACAGCATGTAATTC	TACCCTACGCATGAGGATAAGA	SNP		57,033,287
SNP_270	GGACTAACACACGCTTCTCTAC	CATCTTGCTTCTCCCTGGATAC	SNP		57,043,969
SNP_300	TCTGATGCGACCGATCTTTG	CCTGAAGCAGGTCTCTGAATG	SNP		57,053,808
SNP_310	CCATGCACATGGTCGTTATG	GGACGTACTCGTAGCTGAAG	SNP		57,058,513
SNP_320	GGTGCTATTCCCGTTACCTTAC	GATGCGTTCCAGGTCTTTCA	SNP		57,063,444
CAPS_9	TTCGGTGAAGCTGGAGAAAC	GACGTGACCCAAACCAATCT	CAPS	HaeIII	57,073,201

Table A2. Primers for Sequencing of *Dw1* Candidate Genes. Bolded primers were used to amplify the gene.

Gene	Primer	Primer Sequence
Sobic.009G229500	**240_AmpFor3**	CTTGGCGCTAGTTCCTACTT
	240_AmpRev4	GCAGTTGGAGGAGCTAAGAAA
	240_AmpFor2	CTGCTGAGCTGAGTATGGATATG
	SNP_240B Rev	GGGAGAAGGCCGTGATATAAA
	240_AmpRev3	TGGAACTGTGGAAGGCAATAG
Sobic.009G229600	**250_AmpFor5**	AGCTGACCTGGCAATACTTAC
	250_SeqRev6	GGCAGACTCTCTAAGCTGATTT
	SNP_250D For	GAGCTGAAGAGCTTCCCTTG
	SNP_250F Rev	CTGTCGAAGTTCTTCTCGATGT
	SNP_250D Rev	CTCTTAGCCAGCACTAGCAATC
	250_SeqFor8	AATCCTCTGTCCTGCCATTC
	SNP_250 For	GCGGACATCCAACTCTGATAG
	250_SeqRev1	CACATTGCATCACCAACATCAA
	SNP_250 Rev	AGTCTTCAGGTTGCTCCATTAC
	SNP_250G For	CCATAACTGCAGTGCATGATTC
	SNP_250G Rev	TATGTGCCTCACCTTCCTTTC
	250_SeqFor4	TCAGTAGCCCACAGGAGAATAG
	250_SeqFor5	GATCGCCTAACAGCATGTAATTC
	250_SeqRev7	CAAATGGCACCAGGACCTATTA
	250_SeqRev8	TACCCTACGCATGAGGATAAGA
	250_SeqFor1	GCGGCTAGTGTTGAGGATTTA
	SNP_250C Rev	CTGTTTAGCCCGTCCTTCTT
	250_SeqFor6	GCTTTACTCCAGATGCACAAATAG
	250_SeqRev9	GAAGTTCACTGGCCTGAAGTATTA
	250_SeqFor9	CGCCCTGAACAAGATGTTATG
	250_SeqRev11	CCAATGCCTCAGCCTCTTTA
	SNP_250C For	GGTTGTTTCAGGCTGCTTTC
	250_SeqRev3	TGGCAGAGTTCACCCAAATAA
	SNP_250E For	CTAAAGTTCCACTTCCCGATCA
	SNP_250E Rev	CACAGGTTGGCAGCAGATA
	250_SeqFor7	TGCAGGAAGCAGAAGCTAAA
	250_SeqRev10	GGAGTCATGGTCCTCAGATAATAC
	250_AmpFor3	CCTATACCTCCCACGTTCAAATAC
	250_AmpFor4	GCTGTCTAGTTCTGGCAGTATAA
	250_SeqFor3	GTCTTCCTGGCTAACTTCTACTG
	250_AmpRev4	CAAGAATGGAACTGGCAACATAC
	SNP_250B For	GCTGAATTGGAAGCTCTGAAAC
	250_SeqFor10	GCAGTTGGAGGAGCTAAGAAA
Sobic.009G229700	**SNP_260C For**	CACGATACGATTCCACCGAATTA
	SNP_260C Rev	CAACCAAGCAGTTAGGCTCATA

Table A2. Continued.

Gene	Primer	Primer Sequence
Sobic.009G229800	**270_AmpFor3**	GCACGTACGTACAATCAAGTTATG
	270_AmpRev1	CACAGCCTACATCATCAGTAAGA
	270_AmpFor2	GAGCAACCGTGTGTGTTTAC
	270_SeqRev1	GTAAAGATGCCCAGTTTCAAGTC
	SNP_270 For	GATATGTGGACGACAGGATCAG
	SNP_270B For	GGACTAACACACGCTTCTCTAC
	SNP_270 Rev	ATTGAGCAGTCGAAGGAAGG
	270_SeqFor1	CAGGCATCCTACCCACTTTAC
	SNP_270B Rev	CATCTTGCTTCTCCCTGGATAC
	270_SeqFor2	ACCAACTCTCCATTGATTCTCC
	270_SeqRev3	CCAGCTGCAAATAGCCAAATAG
	270_SeqRev4	GCCCATCTACTTTGCTGTTTAG
	270_SeqRev5	GGAACCTCTTGCTCAGGTATAG
	270_SeqFor2	CGATACACTCCCACCCATTT
	270_AmpFor4	CTCTCACTCAGCTCTCTCTTTC
	270_SeqRev2	CCTGCCATTTGAGAACAGAAAC
Sobic.009G229900	**280_AmpFor1**	CGTGCTCAGTGCTCTTTATATTTG
	SNP_280B Rev	CAGGAACCTCCATTTCCATGA
	280_SeqRev4	CGCCTGAACGAGAACCTTT
	SNP_280B For	CTGTCCAACGCCATCACTAA
	280_SeqRev3	CAGTGGTGTTTAACGCTGTATTG
	SNP_280 For	CCTTGACAGTTTCGAGGGTAAG
	280_AmpRev2	CAGCAAGGGTAGCATTAGAAGAG
	280_SeqFor3	CCCAGTCGTCCCTAGACATAA
	SNP_280 Rev	CTGCGTTCTTGGGATCTTGT
	280_SeqFor4	GGCCTGTGGGATTTGTACTT
Sobic.009G230000	**290_AmpFor4**	CCAGCATCGTCAACGTAACT
	290_AmpRev1	CTCTTAATCGGTGGATGAGTACAA
	290_AmpFor1	CTCGTGAACCGACGATTTCT
	290_SeqRev2	GTGGGCGGTGGGATTTATAG
	SNP_290 For	TGGTCCACCTGCTCTACA
	290_SeqRev1	CGGGCTCCAGTATCTCCA
	290_SeqFor1	TTGGCTCGTCCCATGATTT

Table A2. Continued.

Gene	Primer	Primer Sequence
Sobic.009G230100	**300_AmpFor2**	CTGGACTAGTTTCTGGTTCGTTAC
	SNP_300B Rev	CCTGAAGCAGGTCTCTGAATG
	300_SeqRev3	GCAGGGTAGATTGAGAGCTTAC
	300_SeqRev 4	CCATGTAGAGCCACCTCATAGA
	SNP_300B For	TCTGATGCGACCGATCTTTG
	300_AmpRev1	CCAATGGGTTTACCGTCTACTG
	300_SeqFor3	GACACCCTGTCGCGAATAAA
	SNP_300 For	CCTGAGTTGTTCCTGCAGATAG
	SNP_300 Rev	GGCGCGTGTCATTAGTAGAA
	300_SeqFor4	CGAGATCTATGAGGTGGCTCTA
	300_SeqRev5	GCAGTATGCTAGTCCCATGATAA
	300_SeqFor5	CCAGCATCGTCAACGTAACT
	300_SeqRev6	GATCCAGCAAGGAGGCTATAC

Table A3. Primers for Amplifying cDNA of Sobic.009G229800.

Primer name	Primer sequence
270_AmpFor4	CTCTCACTCAGCTCTCTCTTTC
270_AmpRev2	ACGATTGGAGTGTCTACAAAGAG
SNP_270 Rev	ATTGAGCAGTCGAAGGAAGG
270_SeqFor1	CAGGCATCCTACCCACTTTAC
SNP_270B Rev	CATCTTGCTTCTCCCTGGATAC
270_SeqFor2	ACCAACTCTCCATTGATTCTCC
qRT_4R	CAAGAATGGCCAGGAAGAGAT
qRT_9R	CCCAACTGAAGACATCTCTGAC
qRT_20F	GCGGTCCAACGTCTAATATGT
qRT_10F	GCAGGACAGGCAAAGTAGAT
qRT_21R	CATCTTGCTTCTCCCTGGATAC
qRT_29F	GCGGTCCAACGTCTAATATGT
qRT_28R	TGGTCTCTTGCTCAGGAATTG

PRIMERS USED IN CHAPTER III

Table A4. Primers for *Dw2* Fine Mapping SNPs.

Name	Forward Primer	Reverse Primer	SNP Location
SNP_66800	CACTCATAGCTGAGGAGAAACC	TAACCAGGATGCCCAAACTC	42,710,479
SNP_66900	CTTCTTTCGAGACCTCCTTCATT	TCTGGTTATTGGCAGGAGATTAC	42,723,163
SNP_67000	CGCCGAATGCTGTTACCTATAA	GCCATAGCTTAGTTCCTCCTAAC	42,724,389
SNP_67050	CAACACTAAACACCAGCACAAC	GGCCAGGCTTCTAAATAGTAGAG	42,751,429
SNP_67700	TCGGTGGAGGATGATCTTGA	TTCCGAAACATTGGCCTCACCA	42,806,049
SNP_67800	ATGGTGACATGTGAGGTCTATTT	GTTACTGGACTGAAGAACCAGAG	42,822,513

Table A5. Primers for Sanger Sequencing of Genes in the Delimited *Dw2* Region.
Bolded primers were used to amplify the genes.

Gene	Name	Sequence
Sobic.006G067000	**7000_ForAmp2**	TTCAAGCGCCACAATACAAATC
	7000_RevAmp5	CGGTTGCCCATGCCTATAA
	7000_ForSeq1	GTCTGATCCTCCTTGAGCTATTC
	7000_ForSeq2	CCCACAAGGCCACAACTATT
	7000_ForSeq3	CGCCGAATGCTGTTACCTATAA
	7000_RevSeq2	GCCATAGCTTAGTTCCTCCTAAC
	7000_RevSeq3	GAATCACGGCACAAGCAATC
	7000_RevSeq4	GAATCGCAGAGCATCCAAATG
	7000_RevAmp4	GGAGCAGGGTCAGTACATATTTC
	7000_ForSeq4	CACTCTTCACCACGCTTATCT
	7000_ForSeq5	CTGAAGCTGTTGGTTGATCTTG

Table A5. Continued.

Gene	Name	Sequence
Sobic.006G067050	**SNP_7100B For**	GCAGTGGTCTTACCCATTCA
	7050_RevSeq7	AATCTATTTAGTAAGAGGCACCTG
	7050_ForSeq6	GAAGCAAGTTCCGTGAGTTTC
	7050_RevAmp2	GGCCAGGCTTCTAAATAGTAGAG
	SNP_7100A Rev	CGCACCAGTGACCTTACTATTT
	7050_ForAmp2	CATTCGCCTGCCCATCTATTA
	7050_RevSeq3	GGTGACCCTTCTATCCATTTGT
	7050_RevSeq4	CACATGAGGATGAGACCAATGA
	7050_ForSeq4	CAACACTAAACACCAGCACAAC
	7050_ForSeq5	AGGCATGGCAAGTAGTATCAAG
	7050_RevSeq6	ATCTGGCCAAGCAGGAAAC
	7050_ForSeq8	CGACGCCACATTTCACAAATAC
	7050_RevSeq8	GTGCTGGTGTTTAGTGTTGTG
Sobic.006G067100	**7100_ForAmp4**	CCACACAAGCATCGATCATTTAC
	SNP_7100B Rev	CTTCTTATGATGCGCTCCATTTAC
	7100_ForAmp7	AGCAGCCGTGCTTATTAGTC
	SNP_7100A For	CATTCTCAAGCACACTACCCTAC
	7100_RevSeq1	GGTTACCTTGTCTTCCTTCTCTT
	SNP_7100B For	GCAGTGGTCTTACCCATTCA
	7100_RevSeq2	CAGGGTCGTAGGTTGCTAATTC
	SNP_7100A Rev	CGCACCAGTGACCTTACTATTT
Sobic.006G067150	**7150_ForAmp2**	GACTCCACCATAATCCAGCTTAG
	7150_RevAmp1	GGAGTTTCTCGAGGTCGTTTAC
	7150_RevAmp3	CGTGCTAACGCTACGGATTTA
	7150_RevSeq1	AACCATCAGCCAGCAGAAA
	7150_ForSeq1	GGCAGCAACATCCACAATTC
	7150_ForSeq2	GTAACGGTTGTCTGGGACATTA
Sobic.006G067200	**7200_ForAmp3**	CCCTCTCTGAGAACACACATTC
	7200_RevAmp1	GCCAACCTATATCAGAGGCTAAA
	7200_ForAmp2	GCATCAAACTCTCTACCCTCTC
	7200_ForSeq1	GACTCTGTATCTCTGCCGTCTA
	7200_RevSeq1	GAACAGCTCTAGGGTTCCATAAT
	7200_RevSeq4	TGCTGCTACCAGAGATAATAACC
	7200_RevSeq3	GCAACAGACAGGGACTCAAA
	7200_ForSeq2	TGAGTCCCAATCTGACCTCTAA

Table A5. Continued.

Gene	Name	Sequence
Sobic.006G067300	**390AmpFor3**	CACCACCACCGTATAATCCATC
	390AmpRev2	AGTAATAAGCAAGTGCCGAGGGTC
	390AmpFor1	TCGACAGGCTGACGTATTTCTTCG
	390AmpFor2	GACAGGCTGACGTATTTCTTCGCA
	390AmpRev1	GTAATAAGCAAGTGCCGAGGGTCA
	390IntFor1	ATTATCACAGCGGCTTTGCTGC
	390IntFor2	CTCTAACCAGCTTCGTCGTTTC
	390IntRev1	CGATTGTCCAGCAGAACATGGA
	390IntRev2	CAGCATTGTCAAAGTCCCATTC
Sobic.006G067400	**400AmpFor1**	GTACACGTTGCTCCCACCATTATC
	400IntRev2	GAGCAACATCTCTTGCTTGAATAC
	400IntFor2	CCGCTATGTTACACGGATACTC
	400AmpRev2	CTCTCGTCCAAGTTGACAGTATC
	400AmpFor2	GTGATTGAAGGCACTGATGAAAC
	400IntFor1	TGCTCCAAAGGTATTCAGGTTATG
	400IntFor4	ATAATGGCTTCGCTCCTCTG
	400IntFor6	GACTCTTGATACCACCCACTC
	400IntFor7	AAACCTCCGCTTCTGTCATAG
	400IntFor8	TGGGACGAGTACATCCACTA
	400IntRev4	GTTCACCCAAGGGATGATGAG
	400IntRev5	ACGCGGATACACACTTTCTC
	400IntRev6	GTTCCACCTCTCCTTGATGAC
	400IntRev7	GCTCTGGCTGCCTTACATTA
	400IntRev8	ATGAGGTCACAGAATGCGATAA
Sobic.006G067500	**410ForAmp4**	CGAAGTCCGAAGTGGAGTAATAAG
	410AmpRev2	CCAAGTCGACAGAGATTCAAATAG
	410IntFor1	CAGAGGAGCACAAGAGGTTTC
	410RevAmp3	CCAAGATGGTCTCTTCCCTAAAT
	410ForAmp3	GGTGTGTATTGTCCCTCAGTAAT
	410AmpRev1	GGTCTGGTTCACCAACTTATTTC
	410IntFor2	ACGAGACTGCATGAACCATAAG
	410IntFor3	GACTCTTGATTCCACCCACTC
	410IntFor4	CAGGGAAGCATGAGTGCTATAC
	410IntFor5	GCAGCACACCAAAGGATAGA
	410IntRev1	CGCCCACTAAGCATTTGAAC
	410IntRev2	GGCCTGTCCTGCAGTTAATATG
	410IntRev3	AACCTTCCACCAGCAGATTC
	410IntRev4	GTTCCACCTCTCCTTGATGAC
	410IntRev5	GTTCCTTGCATGCTCGATTTC

Table A5. Continued.

Gene	Name	Sequence
Sobic.006G067600	**420E1AmpFor2**	CAATACACACCGTTGGATCTTATG
	420E1AmpRev2	CAGGAACTCCATCATCCTCTTAC
	420E1AmpRev1	AATCTAAGCCTGTTCGGCTATTC
	420E1IntFor1	GGACTTGCTCGGTTTGAATTG
	420E1IntFor2	AAGCCTCATCGTCTCTGTATG
	420E1IntRev1	CCAGGAGCTCTGGAACTATAC
	420E1IntRev2	GGTGAATGTACCGTCGTAGAAG
	420E2AmpFor1	CCATAATCTGCCTATCTGACACC
	420E2AmpRev1	GTAGCCCAAAGGGCCATAATAG
	420E2IntFor1	AGTGTCTTCAAGGTGCTTCAG
	420E2IntRev1	AGGCTTCTTCAACTCCATCTC
	420E3AmpFor1	CATCATATGACACTGCTCCTACAG
	420E3AmpRev1	CTACTGTCGACTGACAAACTACTC
	420E3IntFor1	ACATATCTGCCTGGTGCTATTG
	420E3IntFor2	TCCCAGTCACATCGAATCTTG
	420E3IntFor3	CCCTGTTGGCCTTTATTAACC
	420E3IntFor4	GTCTCTAGCAGCCATTCACATAG
	420E3IntFor5	GGCTATTCACCTAGCTTCCTTAG
	420E3IntRev2	TGTATGGACCGCAAACTCTC
	420E3IntRev3	CTAGGCTACTAGCTGCTTCAC
	420E3IntRev5	CATGAGTCTTGCCGATTCTCTC
	420E3IntRev6	GAACAATAGCACCAGGCAGATA
	420E3IntRev7	GCAACAAGACTGGAAAGCTAAT
Sobic.006G067700	**430AmpFor2**	ACGTTGGACACCAAGATCTACAGG
	430AmpRev1	TACCCAGGAATTTCCCAACCGT
	430AmpRev2	ATCCTTGGGACTACAGCAGTGA
	430IntFor1	TTCAGTGGGTAAGCCAAGTGGA
	430IntFor2	TGGCAGCTTCAGTGCTAATGGA
	430IntFor3	AGTGAGCCCAATGCTTGTGAGA
	430IntFor4	CACAAGCATTGTGAGACGAAAG
	430IntFor5	CCGCTATTGTAATCCTCCTGTG
	430IntFor6	TCGGTGGAGGATGATCTTGA
	430IntFor7	CTTCCACGCACGAGATCTTATC
	430IntRev1	TTCTCACAAGCATTGGGCTCAC
	430IntRev2	TTCCGAAACATTGGCCTCACCA
	430IntRev3	TCACAGACTCCACAATCTCCGA
	430IntRev4	CCAAGCTCCCTTGTTGGATAG
	430IntRev5	GGGATCAATGCAGCTTTGTG
	430IntRev6	CCAGGGCAGAACTCCATTAC

PRIMERS USED IN CHAPTER IV

Table A6. CAPS Markers Used to Determine the Parentage of the F_1 Plants of Broomcorn Crosses.

Broomcorn Parent	For Primer	Rev Primer	RE
Standard & Acme Broomcorn	CTTGAGCCCTTGACT GGACAAAGA	TCACAAGATGCCA AGCTCTGATCG	BstUI
Acme Broomcorn*	GGAGCATCCAAGAA GACAGAAC	CGGTCGTGCGAGTT TATGATAC	HhaI
Japanese Dwarf Broomcorn	GGCAACAGGAACAG AACAAAG	GATTTCTGATTGCG CGTTCTT	HaeIII

* The marker that was used for SB became inconsistent with the AB F_1s, so a different marker was used for the rest of the AB F_1s.

Table A7. Primers Used to Sequence Sobic.007G163800.

Primer	Sequence
Dw3 AmpFor1	**TTCGTGACGACACTGATAGAAC**
Dw3 AmpRev2	**TGCTTCTATCTGTTCCACATCTC**
Dw3 AmpFor2	CAAGTACTGCTACCTGCTCATC
Dw3 IntFor1	GGCACTACTCATCCATCACATAG
Dw3 IntFor2	GCGCCAATGACAACAAGAAG
Dw3 IntFor3	GCACCTACTTCACCGTCTTC
Dw3 IntFor4	AGCTAGTCAACCAAGCATCC
Dw3 IntFor5	AGAGCAGGGCCTTGTTTAG
Dw3 IntFor6	CCATCTTCGCCTACATCCTC
Dw3 IntFor7	AGCATCCACGACAACATCG
Dw3 IntRev1	TGCTCACCATCCATTCATCTC
Dw3 IntRev2	GTCGGGATGGTGCTTGAG
Dw3 IntRev3	ACACCATGAGCACCATGAAC
Dw3 IntRev4	AAGGTTGGCCTCGAAGAG
Dw3 IntRev5	GGATGGCAGGGTTCTTGAG
Dw3 IntRev6	AACAGCGTCGGCTCCTG
Dw3 IntRev7	TGGTCGATGATGCGGAAG
Dw3 IntRev8	ACGTCGGTGTCGAAGAAG