GENETIC DISSECTION OF PLANT ARCHITECTURE AND LIFE HISTORY TRAITS SALIENT TO CLIMATE-RESILIENT SUSTAINABLE INTENSIFICATION OF AGRICULTURE

by

WENQIAN KONG

(Under the Direction of Andrew Paterson)

ABSTRACT

Multiple-harvest ratooning or perennial crops have recently regained recognition with increased projected demand for food and fuel accompanied by anticipated world population growth, and decline of arable land due to soil erosion, nutrient leaching and decreasing organic matter. The genus Sorghum is especially attractive to study perenniality and related traits, being an important source of food, feed and fuel; a weakly ratooning plant at tropical and sub-tropical regions; and with the possibility to cross with two perennial relatives. We developed a series of sorghum populations to study perenniality and life history related traits using forward genetics. Notably, two novel tetraploid backcross populations (BC_1F_1) developed by interspecific crosses derived from *Sorghum bicolor* × *S. halepense* displayed rich transgressive variation for some life history traits. Patterns of gene transmission have been characterized by constructing genetic maps and resolving segregation patterns. A single nucleotide polymorphism (SNP) profile among *S. halepense* and its progenitors, *S. bicolor* and *S. propinquum*, has elucidated the evolutionary history of *S. halepense* with deduction of its genomic composition from its progenitors and from mutation. Quantitative trait studies have discovered important

chromosomal regions responsible for agronomically important and perenniality related traits, enriching knowledge of quantitative trait loci (QTL) in sorghum. Comparison of *S. halepense* derived QTLs to those in two other sorghum populations and to those on paleo-duplicated chromosomal regions provide early insight into the extent of genetic novelty that may have been associated with the evolution of polyploid *S. halepense* following 96 million years of abstinence from polyploidy in the sorghum lineage.

INDEX WORDS: Perenniality, ratoon, transmission genetics, quantitative study, plant height, flowering time, tillering, vegetative branching

GENETIC DISSECTION OF PLANT ARCHITECTURE AND LIFE HISTORY TRAITS SALIENT TO CLIMATE-RESILIENT SUSTAINABLE INTENSIFICATION OF AGRICULTURE

by

WENQIAN KONG

B.S., Yangzhou University, China, 2010

M.S., The University of Georgia, Athens, GA, 2013

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2017

GENETIC DISSECTION OF PLANT ARCHITECTURE AND LIFE HISTORY TRAITS SALIENT TO CLIMATE-RESILIENT SUSTAINABLE INTENSIFICATION OF AGRICULTURE

by

WENQIAN KONG

Major Professor:	Andrew Paterson
Committee:	Jaxk Reeves
	John Burke
	Kelly Dawe
	Rodney Mauricio

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2017

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor Dr. Andrew Paterson for his continuous support and encouragement to finish this work, no matter how depressed I was. I also thank Dr. Stan Cox and Dr. Pheonah Nabukalu from The Land Institute for population construction and phenotyping; their dedication to breeding for perenniality in sorghum enlightened this dissertation.

Thanks to all members of the Plant Genome Mapping Lab for their help toward this dissertation and also thanks to my family for their never-ending support on my finishing longer than expected school days.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Toward breeding for perenniality

Rationale

The world's population is projected to reach 8.5 billion by 2030 and 9.7 billion by 2050, with demand for food predicted to double by 2030. At the same time, the area of agricultural land and arable land per capita has been decreasing (www.fao.org). Forty years of modern agriculture practices relying heavily on tilling and synthetic fertilizer has been suggested to have resulted in erosion of one third of the world's arable land, continuing at a rate of 10 million hectares per year (Pimentel et al. 1995). It is long been recognized that perennial plants play an excellent role in preventing soil erosion and water run-off especially on hillsides (Kosmas et al. 1997). While many fruit, forage, and fiber crops are perennial, breeding for perennial grain crops is relatively new but has gained recognition for its potential especially in marginal land to improve agroecological conditions while ensuring food security (Jackson and Jackson 1999; Glover et al. 2010; Perennial crops for food security 2013) .

Perennial plants have four advantages over annuals: a longer growing season; better access to water and nutrients; more conservative use of nutrients; and better adaptation to marginal lands (Cox et al. 2002; Glover et al. 2010) . In developing countries where most farms are small and family-owned, growing perennial plants may be particularly attractive, being less time consuming and less costly. Also, production of biofuel or chemical feed stocks makes a

strong case for cultivation of perennial plants on agriculturally degraded land (Tilman et al. 2006).

All major grain crops are cultivated as annuals, though many wild progenitors from which they were domesticated are perennial and others have perennial relatives. The general association of perennial plants with wild species and annual plants with cultivated crops suggests that perennial habit is ancestral to annual habit. Perennial progenitors or relatives of our grain crops generally share some key characteristics, particularly abundant roots and complex underground stems (rhizomes) that store nutrition and energy, and enable vegetative reproduction in the subsequent growing season. Rhizomes, subterranean stems that grow diageotropically (perpendicular to the force of gravity), develop from axillary buds on the basal portion of seedling shoots (Gizmawy et al. 1985). The depth and extensiveness of rhizomes are key components in winter hardiness and survival, while regrowth from the crown of a plant may also be utilized to breed for 'ratooning' in tropical or subtropical areas where two growth cycles are possible.

Regrowth of the vegetative parts of a perennial plant is closely related to both tillering and rhizomatousness, as regrowth might be initiated from rhizomes and/or the crown of the plant (Paterson et al. 1995b). Both tillers and rhizomes develop from axillary meristems at the lowest nodes of a plant, but the mechanism of their differentiation remains unknown. Interestingly, each of three studies of rhizomatousness (Paterson et al. 1995b; Hu et al. 2003; Westerbergh and Doebley 2004) have suggested QTL correspondence between tillering and rhizomatousness, and a detailed comparison of tillering and above-ground rhizomatousness has been conducted in a recombinant inbred line (RIL) population of sorghum (Kong 2013). These studies all indicate that QTL regions affecting plant branching and rhizomatousness may show pleotropic effects,

suggesting that these two traits are likely to involve some of the same or related causal genes. Details of current research on genes affecting tillering and high-order vegetative branching will be reviewed in the following section.

Current achievement

Development of perennial crops ranges from its infancy, such as maize, to intermediate, such as wheat, to almost mature such as rice and sorghum. The difficulties of breeding for perenniality of a species has partly depended on the level of genetic complexity and barrier between the domesticated and related wild species and on the duration of effort.

Progress in breeding for perenniality in *Zea mays* seems still at early stages and requires a long –term effort, possibly due to complexity of the genome and more recessive quantitative traits needed to pyramid (Murray and Jessup 2013). However, many traits of interest could be useful for farmers found in the germplasm used for breeding for perenniality, such as increased 'staygreen', prolificacy— the ability to produce many ears on the same plant, and indeterminant regrowth— the ability to produce biomass after initial flowering and seed set (Murray and Jessup 2013).

Perennial wheat has been developed from interspecific crosses between cultivars and their relatives such as *Thinopyrum elongatum* (tall wheatgrass), *Th. ponticum, and Th. intermedium* (Murphy et al. 2009; Larkin and Newell 2013). Another approach developed by The Land Institute is to directly domesticate may perennial species, and success has been made to domesticate a wheat relative, *Th. Intermedium*, using recurrent selection. Future perennial wheat breeding may aim at working on simultaneously and eventually combining both methods.

In rice, effort has been made to evaluate performances of crosses between *Oryza sativa* and *Oryza longistaminata* or *Oryza rufipogon* (Sacks et al. 2003a, 2003b). A successful lowland

rice cultivar, PR23, has been developed by backcrossing an F1 of *Oryza sativa* and *Oryza longistaminata* to cultivated rice parents. Extensive work in genetics has characterized rhizome specific QTLs and genes responsible for perenniality (Hu et al. 2003; Hu et al. 2011). Future work would focus on adaptation of perennial rice to certain environments and increasing yield potential for the ratoon crop.

Progress of research in sorghum

Despite occasional use as forage, *Sorghum halepense* is considered to be among the world's most noxious weeds, having spread from its west Asian center of diversity across much of Asia, Africa, Europe, North and South America, and Australia (Holm et al. 1977). Moreover, spontaneous hybridization between *S. bicolor* × *S. halepense*, is potentially risky with the possibility of 'gene escape' from genetically modified crops (Morrell et al. 2005) which may improve fitness of the weed and exacerbate weed problems (Arriola and Ellstrand 1996; Snow 2002). However, many attributes of *S. halepense*, such as early flowering, insect and disease resistant, winterhardiness and extensive rhizomes may be beneficial to sorghum cultivars, and its perenniality offers potential for alternative cropping systems (Paterson et al. 2013). Indeed, breeding for perennial sorghum launched in the early 1980s by making hybrids between induced tetraploids of *Sorghum bicolor* and various types of winterhardy, rhizomatous *S. halepense* at The Land Institute (Piper and Kulakow 1994; Cox et al. 2006). A naturally-formed polypoid derived from interspecific hybridization of *S. bicolor* (2n=20) and *S. propinquum* (2n=20), *S. halepense*, has more extensive rhizomes than its rhizomatous progenitor, perhaps indicating some 'recruitment' of genes from *S. bicolor* to expression in rhizomes (Jang et al. 2009).

Sorghum has become an excellent model in which to dissect the genetic basis of 'ratooning', rhizomatousness and overwintering, largely due to crosses made with *S.*

propinquum, a diploid tropical plant that has abundant tillers and rhizomes (Paterson et al. 2013). Paterson et al. (1995b) identified rhizomes and perenniality associated QTLs in a diploid population derived from *S. bicolor* × *S. propinquum*. A follow-up study in recombinant inbred line (RIL) population from the same cross, advanced in a temperate environment that resulted in selection against short-day alleles, elucidated additional QTLs responsible for rhizomatousness in Bogart, GA (Kong et al. 2015). Both studies also suggested some correspondence between tillering and rhizomatousness (Paterson et al. 1995b; Kong et al. 2014), possibly because they both develop from axillary buds at the lowermost nodes of the erect leafy shoot of the plant. Populations developed from crosses between *S. bicolor* and *S. halepense* provide new opportunities to benefit from many traits that differentiate between the parents, both perenniality-related and otherwise.

Comparative studies reveal that QTLs conferring rhizomatousness fall in corresponding genomic regions of the maize and sorghum genomes in a few cases (Westerbergh and Doebley 2004), while QTLs responsible for rhizomatousness in rice and sorghum fall largely in corresponding genomic regions (Hu et al. 2003), supporting the notion that independent but convergent mutations at corresponding loci might exert large effects during grass domestication (Paterson et al. 1995a). Correspondence between divergent grasses such as the panicoid sorghum and oryzoid rice also indicates that the locations of genes controlling rhizomatousness may be extrapolated to many other species. However, it still remains unclear which specific genes are responsible for development and metabolism of rhizomes.

Genes associated with rhizome expression are somewhat enriched in the genomic intervals containing rhizome-related QTLs and cover a wide range of functional categories (Jang et al. 2006). Interestingly, the expression of most rhizome-specific genes in the rhizomatous *S.*

propinquum does not differ dramatically from the counterparts of non-rhizomatous *S. bicolor*. The aggressiveness of *S. halepense* might have come from the novel alleles formed after its occurrence, or some 'recruitment' of S. bicolor genes (Jang et al. 2009). A recent RNA-sequencing experiment in *S. halepense* recently identified 262 genes with different expression patterns between buds induced to develop rhizomes and leafy shoots, with 168 enriched in rhizome buds and 94 enriched in shoot buds. Rhizome-bud enriched genes are associated with functions such as rapid cell division and maturation, while shoot-bud enriched genes are associated with cell recognition (Paterson et al. 2017).

Future perspective

Breeding for perennial grain crops opens new doors to bringing the world's marginal land into food or biomass production, while arresting and preventing loss of ecological capital. Genetics and breeding of perennial types has been initiated in various crops such as rice, maize, wheat, and sorghum. Study of perenniality relies on help from many other fields such as agronomy, ecology, and policy and economics. As we are facing unprecedented challenges with rapid population growth, degradation of world's food-producing soil and a possible water crisis, continuous effort toward perennial agriculture is urgently needed.

Genetic basis of shoot branching

Introduction

Plant architecture is the three dimensional organization of a plant body, determined by the sizes and shapes of plant organs and patterns of above-ground vegetative branching. One of the most important components that determine plant architecture, vegetative branching, is shaped by the interactions of genetics, hormonal and environmental factors. Vegetative branching frequently contributes to classification of different genotypes into taxa and genera. A remarkable

feature of plant development is that the branching patterns of a plant also respond greatly to changes of environmental factors, such as density, humidity and nutrition, for better adaptation to the ecological niche in which it grows.

Modifying tillering and vegetative branching patterns has been historically important for plant breeders to achieve high yields during the 'Green Revolution' (Conway 1998), and will continue to be important. Understanding the genetics of vegetative branching has taken on new importance with invigorated efforts to develop plant genotypes optimized for production of biomass for use in fuels or chemical feedstocks. Increases in yield of one of the best-studied biomass crops, sugarcane, have been achieved primarily by increasing source and sink capacity (Moore and Maretzki 1996; Moore et al. 1997). While tillering is an important element of sink capacity, additional factors including higher-order branching (i.e. 'secondary' branches from tillers) as well as stalk (tiller) dimensions, must also be considered.

Knowledge of vegetative branching may take on new importance in crop production and resilience to climate change. A degree of early-season branching may confer some resilience to weather variations such as transient temperature extremes, for example by providing for some compensatory seed set if pollen viability on the primary inflorescence(s) is damaged. Late-season branching or post-harvest regrowth in the tropics may contribute toward a 'ratoon' crop, but in temperate climates where cold temperatures prohibit maturation is likely to be a futile waste of resources.

There have been voluminous studies to discover genes for vegetative branch initiation and outgrowth, genetic pathways, hormonal regulations and gene-gene interactions (Leyser 2003; Ward and Leyser 2004; Wang and Li 2006; Doust 2007; McSteen 2009; Shimizu-Sato et al. 2009; Waldie et al. 2014). Recent studies in monocots have brought new insights into

discovering more tillering and branching genes and regulatory mechanisms (Wang and Li 2011; Kebrom et al. 2013), which may facilitate wide and direct applications, such as improving crop production. Comparisons of those genes, hormonal responses and related biochemical pathways in monocots and dicots will further deepen our understanding of how genes and hormonal factors interact and fit into the broad picture of vegetative branching regulation, and will facilitate manipulation of genes and their related pathways to benefit a variety of applications.

Tillering/Branching patterns in eudicots and monocots

Plant architecture results from combined developmental control of the shoot apical meristem (SAM) and axillary meristems (AM). SAM, a group of cells at the tip of the primary axis initiated during embryogenesis, controls the development and elaboration of the primary axis; AM is a group of cells formed in the axils of leaves (McSteen and Leyser 2005). Therefore, SAM controls the development of the primary axis, whereas the majority of diversity and variation of branching are influenced by AM.

Vegetative branching patterns and growth habits vary dramatically among different species. Both belonging to the clade of eudicots, Arabidopsis and tomato show striking differences with respect to their branching patterns. Arabidopsis exhibits a monopodial development: the SAM of Arabidopsis continues to provide rosette leaves until center?? photoperiod cues trigger flowering. After transitioning to reproductive growth, new stems elongate vertically at which cauline leaves form and new axillary meristems form buds and spread basipetally. In contrast with Arabidopsis, tomato exhibits sympodial development: after producing a few leaves by the SAM, the growth of the primary shoots is terminated by the first inflorescence. The sympodial meristem continues shoot growth, providing three more leaves until it produces a second inflorescence (Pnueli et al. 1998), and this pattern continues (Figure 1).

In domesticated crops, at least two distinct branching patterns are found. Crops like rice and wheat, exemplifying the ehrhartoid and pooid grasses, are selected for synchronized proliferation of tillers. Vegetative branching patterns are more complicated in the panicoid family (Doust 2007). Crops such as sorghum and pearl millet are selected for apical dominance and severe reduction of both tillers and axillary branches. Like sorghum and pearl millet, maize tillers and axillary branches have been selected against compared with its ancestor, teosinte, while maize ears, usually only one or two, are in fact axillary branches derived from the AMs a few nodes higher than the basal nodes (Doebley et al. 1997).

To date, even for Arabidopsis, a model plant with a small genome, a complete picture that depicts gene functions, gene-gene interactions, hormonal regulation and signaling pathways for vegetative branching is still lacking, due to complex interactions of genetic, hormonal, and environmental effects. The beauty of this complexity lies in its non-deterministic development along with its rapid responses to the changing environment. Nevertheless, progress has been made for the past two decades in identifying some genes that control axillary meristem initiation and outgrowth, and deciphering the biochemical pathways in which those genes are involved. In this review, I will summarize genes that have been discovered in both monocots and eudicots, comparing their functions and explaining how they fit in the big picture of the vegetative branching pathways.

Genes for axillary meristem initiation

One classical genetic pathway that controls axillary meristem initiation shared by monocots and dicots involves a gene coding the GRAS transcription factors, that is, *LATERAL SUPPRESSOR* (*LAS*) in Arabidopsis (Greb et al. 2003), *LATERAL SUPPRESSOR* (*LS*) in tomato (Groot et al. 1994) and *MONOCULM1* (*MOC1*) in rice (Li et al. 2003). Mutants of this gene fail

to initiate axillary meristems during vegetative development, so that plants with mutant phenotypes are usually less branched. In Arabidopsis, *LAS* is upstream to many other genes for axillary meristem development that are locally expressed, such as *REVOLUTA* and *AXR1* (Greb et al. 2003). *MOC1* in rice has pleotropic functions in reducing plant height and maintaining axillary meristems of both inflorescence and vegetative branches.

The NAC family proteins represent one of the largest families of transcription factors with more than 100 members in Arabidopsis (Olsen et al. 2005) and rice. Three *CUP-SHAPED COTYLEDON* genes in Arabidopsis, *CUC1*, *CUC2* and *CUC3*, regulate embryonic shoot and axillary meristem formation and boundary specification with partially overlapping functions with *LAS* (Takada et al. 2001; Vroemen et al. 2003{Hibara, 2006 #746; Hibara et al. 2006). A rice ortholog, OsTIL1/OsNAC2, however, inhibits shoot branching through regulating axillary meristem outgrowth, not initiation (Mao et al. 2007).

Another gene family shared by monocots and dicots encodes CLASS III homeodomain (HD) leucine zipper proteins, represented by *REVOLUTA* (*REV*) in Arabidopsis (Otsuga et al. 2001) and *OSHB3* in rice (Itoh et al. 2008). *REV* is expressed early in embryogenesis at both shoot apical meristem and axillary meristem and has a pleiotropic effect involved in patterning the leaf and secondary branches (Otsuga et al. 2001). Phenotypes of *rev* mutant plants include reduced secondary branches and many other defects at patterning. CLASS III HD-ZIP includes five classes of genes. Two other genes in this family with redundant functions with *REV*, *PHB* and *PHV*, have also been reported (McConnell et al. 2001; Byrne 2006). A REV homolog in Populus, *popREVOLUTA*, has been proven to be involved in the initiation of cambium in regulating the patterning of secondary vascular tissues (Robischon et al. 2011). All evidence has

suggested that this CLASS III HD ZIP gene family might be conserved among many angiosperm species.

One gene that has not been found and proven in the monocots is *BLIND* (*BL*) in tomato. This gene, encoding a Myb gene product of the R2R3 class of transcription factors, is a general regulator of shoot branching affecting axillary meristems for both lateral and inflorescence shoots in tomato (Schmitz et al. 2002). Discovery of the *BL* homolog in Arabidopsis, *REGULATORS OF AXILLARY MERISTEMS 1-3* (*RAX1-3*) (Muller et al. 2006), has suggested that *RAX* and *LAS* are involved in two different pathways affecting axillary meristem initiation along the whole axis of Arabidopsis (Figure 2).

A basic helix-loop-helix transcription factor, *BARREN STALK 1* (*BA1*) is found to affect axillary meristem initiation for both vegetative and inflorescence branching in maize (Ritter et al. 2002; Gallavotti et al. 2004). A homolog of *BA1* in rice is *LAX PANICLE 1* (*LAX1*) (Komatsu et al. 2003). Although *lax* mutants only affect inflorescence branching in rice, plants with the double mutant of *lax* and *spa* (*small panicle*) showed severe reduction in both vegetative and inflorescence branching. *BA1/LAX1* was originally considered to be conserved in monocots before the recent discovery of its homologous gene, *RECULATOR OF AXILLARY MERISTEM FORMATION* (*ROX*) in Arabidopsis (Yang et al. 2012). Different from *LAX*, and *BA1*, *ROX* only functions during early vegetative development and does not interfere with inflorescence development in Arabidopsis. A recently identified gene, *LAX PANICLE 2* (*LAX2*) in rice, is also proven to maintain axillary meristem by interacting with *LAX1*, but in a different pathway from *LAX1* and *MOC1*, because double mutants of either *lax2 lax1* or *lax2 moc1* showed more severe reduction of branching (Tabuchi et al. 2011).

Recent studies in Arabidopsis have demonstrated that genes for flowering time usually have pleiotropic effects and contribute to shoot branching (Huang et al. 2013). For example, *AGAMOUS-LIKE 6* (*AGL6*), encoding a MADS box transcription factor, is known for controlling flowering time by regulating *FT* and *FLC* (Yoo et al. 2011). A novel function of this gene has recently been unraveled using forward genetics, controlling shoot branching (Huang et al. 2012). *AGL6* is a positive regulator of axillary meristem formation and promotes stem branching in the axils of cauline leaves (Huang et al. 2012). Similar regulators of shoot branching are also found in genes such as *CONSTANS-LIKE 7* (Wang et al. 2013) and *FT* (Navarro et al. 2015). These recent discoveries have suggested that genes controlling flowering time in Arabidopsis might have pleiotropic effects in influencing vegetative branching, which might also be extrapolated to many other species.

Hormones play an important role in controlling apical dominance and axillary branch initiation (McSteen 2009), exemplified by the role of auxin and its interaction with cytokinin (Leyser 2003, 2006; Shimizu-Sato et al. 2009). Auxin has been identified for 70 years (Thimann and Skoog 1934), and proved to affect all organ primordia, including both vegetative and floral meristems. Auxin biosynthesis, transportation, storage, degradation and signaling are all essential for shaping plant architecture. Mutants at each step of the auxin signaling pathway will result in defects in organ formation, apical dominance and vascular development (Gallavotti 2013). Biosynthesis of auxin is mediated by the *YUCCA* gene in Arabidopsis, with mutant plants showing multiple defects due to lack of auxin (Cheng et al. 2006). Recent findings have suggested that two major gene families, *TAA* and *YUC*, are essential in auxin biosynthesis in Arabidopsis (Mashiguchi et al. 2011). In addition, *AUXIN RESISTANT 1* (*AXR1*) is responsible for auxin response in Arabidopsis. Mutants of *axr1* cannot respond to auxin and are associated

with many phenotypic defects, including excess lateral shoots, decreased plant height and hypocotyl elongation (Lincoln et al. 1990).

The *PIN-FORMED* family of auxin carriers, prevalent in both monocots and eudicots (Hoshino et al. 2004; Xu et al. 2005; Gallavotti et al. 2008), regulate the efflux of auxin and function in organ initiation in the shoot (Galweiler et al. 1998). A recent study has suggested that depletion of auxin locally is required to form axillary meristems in leaf axils, mainly accomplished by upregulation of *PIN1* in Arabidopsis and tomato (Wang et al. 2014). Mutants of the *PIN1* gene in Arabidopsis are unable make flowers, and result in a pin-shaped inflorescence. Mutants of another gene family, *PINOID*, display similar phenotypes with those of the *PIN* family. *PINOID* encodes a Ser/Thr kinase that phosphorylates the *PIN* gene. In maize, *BARREN INFLORESCENCE 2* (*BIF2*) is analogous to the *PID* gene in Arabidopsis (McSteen et al. 2007). Not only does *BIF2* interact with *PIN*, it also phosphorylates *BA1* (Skirpan et al. 2008), with double mutants of *bif2* and *ba1* showing significant reduction of vegetative and inflorescence branching. Only recently did researchers discover the *BA1* ortholog, *ROX*, in Arabidopsis. However, the interaction between *PID* and *ROX* is unknown (Yang et al. 2012).

Recent studies have unraveled the functions of miRNA and related pathways affecting SAM and axillary meristem initiation (Raman et al. 2008; Zhou et al. 2015). miRNA is known to be recruited by a AGO protein to form a RNA induced silencing complex (RISC) to repress or disrupt gene products and their functions. miRNA164 is reported to negatively regulate two NAC domain proteins, *CUC1* and *CUC2*, which regulate downstream *LAS* to promote axillary meristem (Raman et al. 2008). Mutants of miRNA164 develop accessory buds in leaf axils. Another miRNA pathway, miRNA 165/166 maintains the SAM by regulating the downstream HD-ZIP III transcription factors, such as *REVOLUTA* (Zhou et al. 2015). A conserved miRNA

pathway shared by monocots and dicots is mediated by miRNA 156 check format, I think miRNA identifiers generally do not use a space before the number. Mutants of miRNA 156 in rice will result in reduction of tillering, while overexpression of miRNA 156 will lead to increased tillering by regulating the rice SQUAMOSA Promoter-binding Protein-Like gene (OsSPL14) (Xie et al. 2006). Similar results have also been found in maize (Chuck et al. 2007) and Arabidopsis (Schwarz et al. 2008).

Although voluminous studies have identified genes and clarified their functions in axillary meristem initiation, a broad picture of genetic pathways affecting this aspect of plant development is not yet clear. Nonetheless, double or triple mutant phenotypes has elucidated some parts of genetic pathways. These results have also suggested that genes controlling axillary meristem initiation generally exerted redundant functions, which might be a fail-safe mechanism because of its importance for plant development. With the available information about vegetative branch initiation, a schematic and tentative genetic pathway for controlling axillary meristem initiation is shown in Figure 2 1.2? using Arabidopsis as a model.

Axillary meristem outgrowth

Usually, a plant produces more axillary meristems than it actually uses. Once an axillary meristem has formed, both genetic and environmental factors will determine whether and when it will remain dormant or continue to outgrow. It has been known that auxin can inhibit the activation of axillary buds, and mutants of genes controlling auxin and related products have been reported to cause a bushy architecture (Lincoln et al. 1990; Xu et al. 2005; Cheng et al. 2006). Direct application of cytokinin has been shown to promote axillary bud outgrowth, but whether it works independently to auxin (Chatfield et al. 2000), or interacts with auxin (Nordstrom et al. 2004) is still under debate (Mueller and Leyser 2011). The recently discovered

hormone strigolactone (Gomez-Roldan et al. 2008; Umehara et al. 2008) acts in regulating above-ground shoot branching, and also interacts actively with auxin (Figure 2).

The strigolactone class of hormones (SLs) were originally identified to be involved in producing root exudates in the rhizosphere. However, recently, its role in regulating shoot branching outgrowth has been unraveled and the backbone of its genetic pathway has been gradually improved (Umehara et al. 2008; Beveridge and Kyozuka 2010; Rameau 2010; Waldie et al. 2010; Waldie et al. 2014). A number of studies have been focused on genes controlling axillary meristem outgrowth in Arabidopsis and rice with response to SLs, such as genes involved in the *more axillary growth* (*max*) pathway in Arabidopsis (Stirnberg et al. 2002; Sorefan et al. 2003; Booker et al. 2004), *dwarf* (*d*) and *high tillering dwarf* (*htd*) in *Oryza sativa* (Ishikawa et al. 2005; Zou et al. 2005; Arite et al. 2007; Arite et al. 2009; Lin et al. 2009; Wang and Li 2011; Waters et al. 2012) and *ramous* (*rms*) in *Pisum sativum* (Sorefan et al. 2003), and in many other species (Vogel et al. 2010). This indicates that the SL pathway is conserved among monocots and dicots. Mutants of these genes show excess branches due to inability of branching outgrowth inhibition and interference of the SL signaling pathway.

The SL pathway starts with the biosynthesis of SL, mediated by genes *CLEAVAGE DIOXYGENASES* [*MAX3(CCD7)/D10/RMS5*, *MAX4(CCD8)/D17/RMS1*] and *D27* that encodes an ion-containing protein (Lin et al. 2009) in the chloroplast, producing a mobile SL precursor or intermediate (Sorefan et al. 2003; Booker et al. 2004; Arite et al. 2007; Vogel et al. 2010). Mutants of these genes can be rescued by applying exogenous analogs of SL, such as GR24. *MAX1*, which encodes a cytochrome P450 protein, is downstream of *CCD7/CCD8* and is required for conversion and synthesis into active SLs (Stirnberg et al. 2002; Booker et al. 2005), while its homolog in other species has not been found yet. A recent study has shown that *D14* in

rice is involved in perception of strigolactone and functions as a SL receptor in inhibiting shoot branching (Arite et al. 2009). Its orthologs in Petunia, *DAD2* (Hamiaux et al. 2012), and in Arabidopsis, *AtD14*, both have the same SL-deficient phenotype, resulting in excess axillary branches. A downstream gene is represented by *MAX2/D3/RMS4* recruited by D14, which mediates the SL signaling transduction. Different from CCD7/CCD8, mutants of *D14/DAD2/AtD14* and *MAX2/RMS4/D3* genes cannot be rescued by exogenous SL analogs. This difference in exogenous SL response demonstrates the functional difference of CCD7/CCD8 with D14 and MAX2, for the former functions in the biosynthesis of SLs, and the latter in the signaling perception and transduction of SLs.

A recent study has identified one target protein of the SL pathway, *DWARF 53* in rice, which belongs to the small family of eight SMXL (*SUPPRESSOR OF MAX2*) proteins. Together with *D14* and *D3*, they form a SCF^{D3} complex in regulating SL signaling with *D53* acting as a repressor. Perception of SLs will lead to degradation of *D53* to activate downstream targets, while mutants of *d53* block the SL pathway and result in highly branched phenotypes (Jiang et al. 2013). A homolog of *D53* in Arabidopsis has been recently identified (Umehara et al. 2015).This study has provided new insight into *MAX*/*DWARF* gene interactions in regulating the SL pathway.

Strigolactone, auxin and cytokinin act in a dynamic feedback loop, however this interaction is still under debate (Domagalska and Leyser 2011). Specifically, it is unclear whether axillary bud outgrowth resulted from direct auxin transportation (auxin canalization) or is controlled by a second messenger such as cytokinin or strigolactone. Both theories have been supported by experimental evidence. Auxin acts upstream to control the SL level and cytokinin signals entering into the axillary buds (Brewer et al. 2009). Meanwhile, to activate an axillary

bud, auxin needs to be exported locally from the bud, which is regulated downstream by SLs (Crawford et al. 2010). Although a detailed explanation of these interactions remains elusive, empirical evidence has pieced together a dynamic picture of auxin, cytokinin and strigolactone interactions.

The downstream targets of the SL pathways remain an open question. One possibility could be the TCP transcription factors represented by *TEOSINTE BRANCHED 1* (*TB1*) in maize, which is a major regulator controlling axillary meristem outgrowth (Guan et al. 2012). *TB1* is a well-characterized gene for vegetative branching in maize and was involved in domestication from its wild relative, teosinte (Doebley et al. 1997; Hubbard et al. 2002). Maize has experienced selection for severe apical dominance while teosinte is highly branched. *TB1* encodes a TCP transcription factor family member of which an increasing level will suppress bud outgrowth. The orthologs of maize *TB1*, *OsTB1 (FC1)* in rice and *BRC1* in Arabidopsis function similarly to *TB1*, promoting growth arrest of axillary buds (Takeda et al. 2003; Aguilar-Martinez et al. 2007). Downstream interactions of SLs with *TB1* orthologs remain diverse among different species. In rice, GR24 treatment did not affect the expression of *FC1*, while in pea and Arabidopsis, expression of *TB1* orthologs *BRC1* and *psBRC1* were upregulated by addition of GR24 (Minakuchi et al. 2010; Braun et al. 2012). In maize, *TB1* has been shown to work in a SL-independent manner, but a clear pathway of how TB1 and SLs interact remains to be tested (Guan et al. 2012). Clearly, *TB1*/*OsTB1*/*BRC1*/*psBRC1* are involved in a conserved pathway in monocots and dicots, though the growth habit of maize, rice and Arabidopsis vary. Therefore, this similar set of genes may reflect the common evolutionary origin of branching pathways in plants. Further comparison of orthologous genes in other species may verify this hypothesis and identify elements for their different growth habits.

Environmental response

It is well known that vegetative branching is greatly affected by environmental factors, such as density and nutrition (Doust and Kellogg 2006; Kebrom et al. 2006; Whipple et al. 2011). The fact that SL mutants are still responding to environmental factors such as photoperiod and density, indicates the existence of pathways other than the SL pathway (Finlayson et al. 2010). An example can be *SORGHUM PHYTOCHROME B* (*PHYB*), which perceives shade by the red to infrared ratio (R:FR) and regulates the expression of downstream target, *TB1* (Kebrom et al. 2006). When density is high and plants are in shade, *PHYB* increases the expression of *SbTB1*, with axillary buds remaining dormant. *phyB-1* mutants are associated with reduced expression of *SbTB1*, suggesting that *phyB-1* regulates axillary branch outgrowth by mediating the expression of *SbTB1* (Kebrom et al. 2006; Kebrom et al. 2010). Similar results have been elucidated in Arabidopsis (Su et al. 2011). Moreover, it is suggested that *PHYB* interacts with auxin by showing that *phyB* mutant phenotypes can be rescued by compromising the auxin signaling pathway in Arabidopsis, which also regulates the expression of TCP transcription factors (Krishna Reddy and Finlayson 2014). However, the signaling molecules for light and underlying pathways that connect *PHYB* and other downstream genes is still unknown.

Another breakthrough is the discovery of maize *GRASSY TILLER* (*GT1*), which responds to shade by enhancing the expression of *GT1* in developing buds and flowers, leading to decreased branching and enhanced apical dominance. It encodes a class I HD-ZIP transcription factor, and its Arabidopsis ortholog is not clear to date. Further, *GT1* acts downstream of *TB1*, and both genes suppress axillary bud outgrowth in response to light capture (Whipple et al. 2011).

Concluding remarks

Vegetative branching is a remarkable feature during plant development as it adapts the body plan of a plant to respond to changing environments. The mechanisms controlling vegetative branching are complex. A voluminous amount of literature has been published to identify mutants for axillary meristem initiation and outgrowth, to elucidate the role of hormones, to reveal responses to environmental factors, and to depict signaling pathways combining all those factors for vegetative branching. The recent discoveries of strigolactone and its involvement in controlling vegetative branching, together with its interactions with auxin and cytokinin, have been a major advance. However, our understanding of vegetative branching is still not complete. Continuous efforts to identify new genes and characterize gene functions, with further understanding of their roles in signaling pathways and their responses to environments may be crucial to unravel this highly dynamic system.

References

Aguilar-Martinez, J.A., C. Poza-Carrion, and P. Cubas, 2007 Arabidopsis BRANCHED1 acts as an integrator of branching signals within axillary buds. *Plant Cell* 19 (2):458-472.

Arite, T., H. Iwata, K. Ohshima, M. Maekawa, M. Nakajima *et al.*, 2007 DWARF10, an RMS1/MAX4/DAD1 ortholog, controls lateral bud outgrowth in rice. *Plant Journal* 51 (6):1019-1029.

Arite, T., M. Umehara, S. Ishikawa, A. Hanada, M. Maekawa *et al.*, 2009 d14, a Strigolactone-Insensitive Mutant of Rice, Shows an Accelerated Outgrowth of Tillers. *Plant and Cell Physiology* 50 (8):1416-1424.

Arriola, P.E., and N.C. Ellstrand, 1996 Crop-to-weed gene flow in the genus Sorghum (Poaceae): Spontaneous interspecific hybridization between johnsongrass, Sorghum halepense, and crop sorghum, S-bicolor. *American Journal of Botany* 83 (9):1153-1159.

Beveridge, C.A., and J. Kyozuka, 2010 New genes in the strigolactone-related shoot branching pathway. *Current Opinion in Plant Biology* 13 (1):34-39.

Booker, J., M. Auldridge, S. Wills, D. McCarty, H. Klee *et al.*, 2004 MAX3/CCD7 is a carotenoid cleavage dioxygenase required for the synthesis of a novel plant signaling molecule. *Current Biology* 14 (14):1232-1238.

Booker, J., T. Sieberer, W. Wright, L. Williamson, B. Willett *et al.*, 2005 MAX1 encodes a cytochrome P450 family member that acts downstream of MAX3/4 to produce a carotenoid-derived branch-inhibiting hormone. *Developmental Cell* 8 (3):443-449.

Braun, N., A. de Saint Germain, J.P. Pillot, S. Boutet-Mercey, M. Dalmais *et al.*, 2012 The pea TCP transcription factor PsBRC1 acts downstream of Strigolactones to control shoot branching. *Plant Physiol* 158 (1):225-238.

Brewer, P.B., E.A. Dun, B.J. Ferguson, C. Rameau, and C.A. Beveridge, 2009 Strigolactone acts downstream of auxin to regulate bud outgrowth in pea and Arabidopsis. *Plant Physiol* 150 (1):482-493.

Byrne, M.E., 2006 Shoot meristem function and leaf polarity: the role of class III HD-ZIP genes. *Plos Genetics* 2 (6):e89.

Chatfield, S.P., P. Stirnberg, B.G. Forde, and O. Leyser, 2000 The hormonal regulation of axillary bud growth in Arabidopsis. *Plant Journal* 24 (2):159-169.

Cheng, Y.F., X.H. Dai, and Y.D. Zhao, 2006 Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in Arabidopsis. *Genes & Development* 20 (13):1790-1799.

Chuck, G., A.M. Cigan, K. Saeteurn, and S. Hake, 2007 The heterochronic maize mutant Corngrass1 results from overexpression of a tandem microRNA. *Nat Genet* 39 (4):544-549.

Conway, G., 1998 *The doubly green revolution : food for all in the twenty-first century*. Ithaca, N.Y.: Comstock Pub. Associates.

Cox, T.S., M. Bender, C. Picone, D.L. Van Tassel, J.B. Holland *et al.*, 2002 Breeding perennial grain crops. *Critical Reviews in Plant Sciences* 21 (2):59-91.

Cox, T.S., J.D. Glover, D.L. Van Tassel, C.M. Cox, and L.R. DeHaan, 2006 Prospects for developing perennial grain crops. American Institute of Biological Sciences.

Crawford, S., N. Shinohara, T. Sieberer, L. Williamson, G. George *et al.*, 2010 Strigolactones enhance competition between shoot branches by dampening auxin transport. *Development* 137 (17):2905-2913.

Doebley, J., A. Stec, and L. Hubbard, 1997 The evolution of apical dominance in maize. *Nature* 386 (6624):485-488.

Domagalska, M.A., and O. Leyser, 2011 Signal integration in the control of shoot branching. *Nat Rev Mol Cell Biol* 12 (4):211-221.

Doust, A.N., 2007 Grass architecture: genetic and environmental control of branching. *Current Opinion in Plant Biology* 10 (1):21-25.

Doust, A.N., and E.A. Kellogg, 2006 Effect of genotype and environment on branching in weedy green millet (Setaria viridis) and domesticated foxtail millet (Setaria italica) (Poaceae). *Molecular Ecology* 15 (5):1335-1349.

Finlayson, S.A., S.R. Krishnareddy, T.H. Kebrom, and J.J. Casal, 2010 Phytochrome Regulation of Branching in Arabidopsis. *Plant Physiology* 152:1914-1927.

Gallavotti, A., 2013 The role of auxin in shaping shoot architecture. *Journal of Experimental Botany* 64 (9):2593-2608.

Gallavotti, A., Y. Yang, R.J. Schmidt, and D. Jackson, 2008 The relationship between auxin transport and maize branching. *Plant Physiology* 147 (4):1913-1923.

Gallavotti, A., Q. Zhao, J. Kyozuka, R.B. Meeley, M. Ritter *et al.*, 2004 The role of barren stalk1 in the architecture of maize. *Nature* 432 (7017):630-635.

Galweiler, L., C. Guan, A. Muller, E. Wisman, K. Mendgen *et al.*, 1998 Regulation of polar auxin transport by AtPIN1 in Arabidopsis vascular tissue. *Science* 282 (5397):2226-2230.

Gizmawy, I., J. Kigel, D. Koller, and M. Ofir, 1985 Initiation, Orientation and Early Development of Primary Rhizomes in Sorghum-Halepense (L) Pers. *Annals of Botany* 55 (3):343-350.

Glover, J.D., J.P. Reganold, L.W. Bell, J. Borevitz, E.C. Brummer *et al.*, 2010 Increased Food and Ecosystem Security via Perennial Grains. *Science* 328 (5986):1638-1639.

Gomez-Roldan, V., S. Fermas, P.B. Brewer, V. Puech-Pages, E.A. Dun *et al.*, 2008 Strigolactone inhibition of shoot branching. *Nature* 455 (7210):189-U122.

Greb, T., O. Clarenz, E. Schafer, D. Muller, R. Herrero *et al.*, 2003 Molecular analysis of the LATERAL SUPPRESSOR gene in Arabidopsis reveals a conserved control mechanism for axillary meristem formation. *Genes & Development* 17 (9):1175-1187.

Groot, S.P.C., L.C.P. Keizer, W. Deruiter, and J.J.M. Dons, 1994 Seed and Fruit-Set of the Lateral Suppressor Mutant of Tomato. *Scientia Horticulturae* 59 (2):157-162.

Guan, J.C., K.E. Koch, M. Suzuki, S. Wu, S. Latshaw *et al.*, 2012 Diverse roles of strigolactone signaling in maize architecture and the uncoupling of a branching-specific subnetwork. *Plant Physiol* 160 (3):1303-1317.

Hamiaux, C., R.S.M. Drummond, B.J. Janssen, S.E. Ledger, J.M. Cooney *et al.*, 2012 DAD2 Is an alpha/beta Hydrolase Likely to Be Involved in the Perception of the Plant Branching Hormone, Strigolactone. *Current Biology* 22 (21):2032-2036.

Hibara, K., M.R. Karim, S. Takada, K. Taoka, M. Furutani *et al.*, 2006 Arabidopsis CUP-SHAPED COTYLEDON3 regulates postembryonic shoot meristem and organ boundary formation. *Plant Cell* 18 (11):2946-2957.

Holm, L.G., D.L. Plucknett, J.V. Pancho, and J.P. Herberger, 1977 *The world's worst weeds*: University Press.

Hoshino, T., K. Miyamoto, and J. Ueda, 2004 Automorphosis and auxin polar transport of etiolated pea seedlings under microgravity conditions. *Biol Sci Space* 18 (3):94-95.

Hu, F.Y., D.Y. Tao, E. Sacks, B.Y. Fu, P. Xu *et al.*, 2003 Convergent evolution of perenniality in rice and sorghum. *Proceedings of the National Academy of Sciences of the United States of America* 100 (7):4050-4054.

Hu, F.Y., D. Wang, X.Q. Zhao, T. Zhang, H.X. Sun *et al.*, 2011 Identification of rhizome-specific genes by genome-wide differential expression Analysis in Oryza longistaminata. *BMC Plant Biology* 11.

Huang, X., S. Effgen, R.C. Meyer, K. Theres, and M. Koornneef, 2012 Epistatic natural allelic variation reveals a function of AGAMOUS-LIKE6 in axillary bud formation in Arabidopsis. *Plant Cell* 24 (6):2364-2379.

Huang, X.Q., J. Ding, S. Effgen, F. Turck, and M. Koornneef, 2013 Multiple loci and genetic interactions involving flowering time genes regulate stem branching among natural variants of Arabidopsis. *New Phytologist* 199 (3):843-857.

Hubbard, L., P. McSteen, J. Doebley, and S. Hake, 2002 Expression patterns and mutant phenotype of teosinte branched1 correlate with growth suppression in maize and teosinte. *Genetics* 162 (4):1927-1935.

Ishikawa, S., M. Maekawa, T. Arite, K. Onishi, I. Takamure *et al.*, 2005 Suppression of tiller bud activity in tillering dwarf mutants of rice. *Plant and Cell Physiology* 46 (1):79-86.

Itoh, J.I., K.I. Hibara, Y. Sato, and Y. Nagato, 2008 Developmental role and auxin responsiveness of class III homeodomain leucine zipper gene family members in rice. *Plant Physiology* 147 (4):1960-1975.

Jackson, W., and L. Jackson, 1999 Developing high seed yielding perennial polycultures as a mimic of mid-grass prairie. *CURRENT PLANT SCIENCE AND BIOTECHNOLOGY IN AGRICULTURE* 37:xvii-xlviii.

Jang, C.S., T.L. Kamps, D.N. Skinner, S.R. Schulze, W.K. Vencill *et al.*, 2006 Functional classification, genomic organization, putatively cis-acting regulatory elements, and relationship to quantitative trait loci, of sorghum genes with rhizome-enriched expression. *Plant Physiology* 142 (3):1148-1159.

Jang, C.S., T.L. Kamps, H. Tang, J.E. Bowers, C. Lemke *et al.*, 2009 Evolutionary fate of rhizome-specific genes in a non-rhizomatous Sorghum genotype. *Heredity* 102 (3):266-273.

Jiang, L., X. Liu, G.S. Xiong, H.H. Liu, F.L. Chen *et al.*, 2013 DWARF 53 acts as a repressor of strigolactone signalling in rice. *Nature* 504 (7480):401-+.

Kebrom, T.H., T.P. Brutnell, and S.A. Finlayson, 2010 Suppression of sorghum axillary bud outgrowth by shade, phyB and defoliation signalling pathways. *Plant Cell and Environment* 33 (1):48-58.

Kebrom, T.H., B.L. Burson, and S.A. Finlayson, 2006 Phytochrome B represses Teosinte Branched1 expression and induces sorghum axillary bud outgrowth in response to light signals. *Plant Physiology* 140 (3):1109-1117.

Kebrom, T.H., W. Spielmeyer, and E.J. Finnegan, 2013 Grasses provide new insights into regulation of shoot branching. *Trends in Plant Science* 18 (1):41-48.

Komatsu, K., M. Maekawa, S. Ujiie, Y. Satake, I. Furutani *et al.*, 2003 LAX and SPA: Major regulators of shoot branching in rice. *Proceedings of the National Academy of Sciences of the United States of America* 100 (20):11765-11770.

Kong, W., 2013 Genetic Analysis of Plant Architecture in Sorghum in *Crop and Soil Sciences*. The University of Georgia, Athens, GA.

Kong, W., H. Guo, V. Goff, T.-H. Lee, C. Kim *et al.*, 2014 Genetic analysis of vegetative branching in sorghum. *Theoretical and Applied Genetics* 127 (11):2387-2403.

Kong, W.Q., C. Kim, V.H. Goff, D. Zhang, and A.H. Paterson, 2015 Genetic Analysis of Rhizomatousness and Its Relationship with Vegetative Branching of Recombinant Inbred Lines of Sorghum Bicolor X S-Propinquum. *American Journal of Botany* 102 (5):718-724.

Kosmas, C., N. Danalatos, L.H. Cammeraat, M. Chabart, J. Diamantopoulos *et al.*, 1997 The effect of land use on runoff and soil erosion rates under Mediterranean conditions. *Catena* 29 (1):45-59.

Krishna Reddy, S., and S.A. Finlayson, 2014 Phytochrome B promotes branching in Arabidopsis by suppressing auxin signaling. *Plant Physiol* 164 (3):1542-1550.

Larkin, P.J., and M.T. Newell, 2013 Perennial wheat breeding:current germplasm anda way forward for breeding and global cooperation in *Perennial Crops for Food Security*, edited by C. Batello, L. Wade, S. Cox, N. Pogna, A. Bozzini *et al.*

Leyser, O., 2003 Regulation of shoot branching by auxin. *Trends in Plant Science* 8 (11):541-545.

Leyser, O., 2006 Dynamic integration of auxin transport and signalling. *Current Biology* 16 (11):R424-R433.

Li, X.Y., Q. Qian, Z.M. Fu, Y.H. Wang, G.S. Xiong *et al.*, 2003 Control of tillering in rice. *Nature* 422 (6932):618-621.

Lin, H., R.X. Wang, Q. Qian, M.X. Yan, X.B. Meng *et al.*, 2009 DWARF27, an Iron-Containing Protein Required for the Biosynthesis of Strigolactones, Regulates Rice Tiller Bud Outgrowth. *Plant Cell* 21 (5):1512-1525.

Lincoln, C., J.H. Britton, and M. Estelle, 1990 Growth and Development of the Axr1 Mutants of Arabidopsis. *Plant Cell* 2 (11):1071-1080.

Mao, C.Z., W.N. Ding, Y.R. Wu, J. Yu, X.W. He *et al.*, 2007 Overexpression of a NAC-domain protein promotes shoot branching in rice. *New Phytologist* 176 (2):288-298.

Mashiguchi, K., K. Tanaka, T. Sakai, S. Sugawara, H. Kawaide *et al.*, 2011 The main auxin biosynthesis pathway in Arabidopsis. *Proc Natl Acad Sci U S A* 108 (45):18512-18517.

McConnell, J.R., J. Emery, Y. Eshed, N. Bao, J. Bowman *et al.*, 2001 Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. *Nature* 411 (6838):709-713.

McSteen, P., 2009 Hormonal Regulation of Branching in Grasses. *Plant Physiology* 149 (1):46-55.

McSteen, P., and O. Leyser, 2005 Shoot branching. *Annual Review of Plant Biology* 56:353-374.

McSteen, P., S. Malcomber, A. Skirpan, C. Lunde, X.T. Wu *et al.*, 2007 barren inflorescence2 encodes a co-ortholog of the PINOID serine/threonine kinase and is required for organogenesis during inflorescence and vegetative development in maize. *Plant Physiology* 144 (2):1000-1011.

Minakuchi, K., H. Kameoka, N. Yasuno, M. Umehara, L. Luo *et al.*, 2010 FINE CULM1 (FC1) works downstream of strigolactones to inhibit the outgrowth of axillary buds in rice. *Plant and Cell Physiology* 51 (7):1127-1135.

Moore, P.H., F.C. Botha, R.T. Furbank, and C.R.L. Grof, 1997 Potential for overcoming physio-biochemical limits to sucrose accumulation, pp. 141-156 in *Intensive sugarcane production: meeting the challenge beyond 2000*

Sugar 2000 symposium, Intensive sugarcane production: meeting the challenge beyond 2000; 141-156. CAB International, Wallingford, UK.

Moore, P.H., and A. Maretzki, 1996 *Photoassimilate Distribution in Plants and Crops -- Source-Sink relationships.* : Marcel Dekker, .

Morrell, P.L., D. Williams-Coplin, J.E. Bowers, J.M. Chandler, and A.H. Paterson, 2005 Crop-to-weed introgression has impacted allelic composition of johnsongrass populations with and without recent exposure to cultivated sorghum. . *Molecular Ecology* 14:2143-2154.

Mueller, D., and O. Leyser, 2011 Auxin, cytokinin and the control of shoot branching. *Annals of Botany* 107 (7):1203-1212.

Muller, D., G. Schmitz, and K. Theres, 2006 Blind homologous R2R3 Myb genes control the pattern of lateral meristem initiation in Arabidopsis. *Plant Cell* 18 (3):586-597.

Murphy, K.M., L.A. Hoagland, P.G. Reeves, B.K. Baik, and S.S. Jones, 2009 Nutritional and quality characteristics expressed in 31 perennial wheat breeding lines. *Renewable Agriculture and Food Systems* 24 (4):285-292.

Murray, S.C., and R. Jessup, 2013 Breeding and genetics of perennial maize: progress, opportunities and challenges in *Perennial Crops for Food Security*, edited by C. Batello, L. Wade, S. Cox, N. Pogna, A. Bozzini *et al.*

Navarro, C., E. Cruz-Oro, and S. Prat, 2015 Conserved function of FLOWERING LOCUS T (FT) homologues as signals for storage organ differentiation. *Current Opinion in Plant Biology* 23:45-53.

Nordstrom, A., P. Tarkowski, D. Tarkowska, R. Norbaek, C. Astot *et al.*, 2004 Auxin regulation of cytokinin biosynthesis in Arabidopsis thaliana: A factor of potential importance for

auxin-cytokinin-regulated development. *Proceedings of the National Academy of Sciences of the United States of America* 101 (21):8039-8044.

Olsen, A.N., H.A. Ernst, L.L. Leggio, and K. Skriver, 2005 NAC transcription factors: structurally distinct, functionally diverse. *Trends Plant Sci* 10 (2):79-87.

Otsuga, D., B. DeGuzman, M.J. Prigge, G.N. Drews, and S.E. Clark, 2001 REVOLUTA regulates meristem initiation at lateral positions. *Plant Journal* 25 (2):223-236.

Paterson, A., H., T.S. Cox, W. Kong, and M. Navarro, 2013 Viewpoint: multiple-harvest sorghums toward improved food securit in *Perennial Crops for Food Security*, edited by C. Batello, L. Wade, S. Cox, N. Pogna, A. Bozzini *et al.*

Paterson, A., H., W. Kong, R. Johnston, P. Nabukalu, G. Wu *et al.*, 2017 The evolution of an invasive plant, Sorghum halepense L. ('Johnsongrass').

Paterson, A.H., Y.R. Lin, Z.K. Li, K.F. Schertz, J.F. Doebley *et al.*, 1995a Convergent Domestication of Cereal Crops by Independent Mutations at Corresponding Genetic-Loci. *Science* 269 (5231):1714-1718.

Paterson, A.H., K.F. Schertz, Y.R. Lin, S.C. Liu, and Y.L. Chang, 1995b The Weediness of Wild Plants - Molecular Analysis of Genes Influencing Dispersal and Persistence of Johnsongrass, Sorghum-Halepense (L) Pers. *Proceedings of the National Academy of Sciences of the United States of America* 92 (13):6127-6131.

Perennial crops for food security, 2013, edited by C. Batello, L. Wade, S. Cox, N. Pogna, A. Bozzini *et al.*

Pimentel, D., C. Harvey, P. Resosudarmo, K. Sinclair, D. Kurz *et al.*, 1995 Environmental and economic costs of soil erosion and conservation benefits. *Science* 267 (5201):1117-1123.

Piper, J.K., and P.A. Kulakow, 1994 Seed yield and biomass allocation in Sorghum bicolor and F1 and backcross generations of S. bicolor× S. halepense hybrids. *Canadian Journal of Botany* 72 (4):468-474.

Pnueli, L., L. Carmel-Goren, D. Hareven, T. Gutfinger, J. Alvarez *et al.*, 1998 The SELF-PRUNING gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of CEN and TFL1. *Development* 125 (11):1979-1989.

Raman, S., T. Greb, A. Peaucelle, T. Blein, P. Laufs *et al.*, 2008 Interplay of miR164, CUP-SHAPED COTYLEDON genes and LATERAL SUPPRESSOR controls axillary meristem formation in Arabidopsis thaliana. *Plant Journal* 55 (1):65-76.

Rameau, C., 2010 Strigolactones, a novel class of plant hormone controlling shoot branching. *Comptes Rendus Biologies* 333 (4):344-349.

Ritter, M.K., C.M. Padilla, and R.J. Schmidt, 2002 The maize mutant barren stalk1 is defective in axillary meristem development. *American Journal of Botany* 89 (2):203-210.

Robischon, M., J. Du, E. Miura, and A. Groover, 2011 The Populus class III HD ZIP, popREVOLUTA, influences cambium initiation and patterning of woody stems. *Plant Physiol* 155 (3):1214-1225.

Sacks, E.J., J.P. Roxas, and M.T.S. Cruz, 2003a Developing perennial upland rice I: Field performance of Oryza sativa/O. rufipogon F-1, F-4, and BC1F4 progeny. *Crop Science* 43 (1):120-128.

Sacks, E.J., J.P. Roxas, and M.T.S. Cruz, 2003b Developing perennial upland rice II: Field performance of S-1 families from an intermated Oryza sativa/O. longistaminata population. *Crop Science* 43 (1):129-134.

Schmitz, G., E. Tillmann, F. Carriero, C. Fiore, F. Cellini *et al.*, 2002 The tomato Blind gene encodes a MYB transcription factor that controls the formation of lateral meristems. *Proceedings of the National Academy of Sciences of the United States of America* 99 (2):1064-1069.

Schwarz, S., A.V. Grande, N. Bujdoso, H. Saedler, and P. Huijser, 2008 The microRNA regulated SBP-box genes SPL9 and SPL15 control shoot maturation in Arabidopsis. *Plant Mol Biol* 67 (1-2):183-195.

Shimizu-Sato, S., M. Tanaka, and H. Mori, 2009 Auxin-cytokinin interactions in the control of shoot branching. *Plant Molecular Biology* 69 (4):429-435.

Skirpan, A., X.T. Wu, and P. McSteen, 2008 Genetic and physical interaction suggest that BARREN STALK1 is a target of BARREN INFLORESCENCE2 in maize inflorescence development. *Plant Journal* 55 (5):787-797.

Snow, A.A., 2002 Transgenic crops[mdash]why gene flow matters. *Nat Biotech* 20 (6):542-542.

Sorefan, K., J. Booker, K. Haurogne, M. Goussot, K. Bainbridge *et al.*, 2003 MAX4 and RMS1 are orthologous dioxygenase-like genes that regulate shoot branching in Arabidopsis and pea. *Genes & Development* 17 (12):1469-1474.

Stirnberg, P., K. van de Sande, and H.M.O. Leyser, 2002 MAX1 and MAX2 control shoot lateral branching in Arabidopsis. *Development* 129 (5):1131-1141.

Su, H., S.D. Abernathy, R.H. White, and S.A. Finlayson, 2011 Photosynthetic photon flux density and phytochrome B interact to regulate branching in Arabidopsis. *Plant Cell and Environment* 34 (11):1986-1998.

Tabuchi, H., Y. Zhang, S. Hattori, M. Omae, S. Shimizu-Sato *et al.*, 2011 LAX PANICLE2 of rice encodes a novel nuclear protein and regulates the formation of axillary meristems. *Plant Cell* 23 (9):3276-3287.

Takada, S., K. Hibara, T. Ishida, and M. Tasaka, 2001 The CUP-SHAPED COTYLEDON1 gene of Arabidopsis regulates shoot apical meristem formation. *Development* 128 (7):1127-1135.

Takeda, T., Y. Suwa, M. Suzuki, H. Kitano, M. Ueguchi-Tanaka *et al.*, 2003 The OsTB1 gene negatively regulates lateral branching in rice. *Plant Journal* 33 (3):513-520.

Thimann, K.V., and F. Skoog, 1934 On the Inhibition of Bud Development and other Functions of Growth Substance in Vicia Faba. *Proceedings of the Royal Society of London B: Biological Sciences* 114:317-339.

Umehara, M., A. Hanada, S. Yoshida, K. Akiyama, T. Arite *et al.*, 2008 Inhibition of shoot branching by new terpenoid plant hormones. *Nature* 455 (7210):195-U129.

Umehara, M., C. Mengmeng, K. Akiyama, T. Akatsu, Y. Seto *et al.*, 2015 Structural Requirements of Strigolactones for Shoot Branching Inhibition in Rice and Arabidopsis. *Plant and Cell Physiology*:pcv028.

Vogel, J.T., M.H. Walter, P. Giavalisco, A. Lytovchenko, W. Kohlen *et al.*, 2010 SlCCD7 controls strigolactone biosynthesis, shoot branching and mycorrhiza-induced apocarotenoid formation in tomato. *Plant Journal* 61 (2):300-311.

Vroemen, C.W., A.P. Mordhorst, C. Albrecht, M.A.C.J. Kwaaitaal, and S.C. de Vries, 2003 The CUP-SHAPED COTYLEDON3 gene is required for boundary and shoot meristem formation in Arabidopsis. *Plant Cell* 15 (7):1563-1577.

Waldie, T., A. Hayward, and C.A. Beveridge, 2010 Axillary bud outgrowth in herbaceous shoots: how do strigolactones fit into the picture? *Plant Molecular Biology* 73 (1-2):27-36.

Waldie, T., H. McCulloch, and O. Leyser, 2014 Strigolactones and the control of plant development: lessons from shoot branching. *Plant Journal* 79 (4):607-622.

Wang, H., Z. Zhang, H. Li, X. Zhao, X. Liu *et al.*, 2013 CONSTANS-LIKE 7 regulates branching and shade avoidance response in Arabidopsis. *J Exp Bot* 64 (4):1017-1024.

Wang, Q., W. Kohlen, S. Rossmann, T. Vernoux, and K. Theres, 2014 Auxin Depletion from the Leaf Axil Conditions Competence for Axillary Meristem Formation in Arabidopsis and Tomato. *Plant Cell* 26 (5):2068-2079.

Wang, Y.H., and J.Y. Li, 2006 Genes controlling plant architecture. *Current Opinion in Biotechnology* 17 (2):123-129.

Wang, Y.H., and J.Y. Li, 2011 Branching in rice. *Current Opinion in Plant Biology* 14 (1):94-99.

Ward, S.P., and O. Leyser, 2004 Shoot branching. *Current Opinion in Plant Biology* 7 (1):73-78.

Waters, M.T., P.B. Brewer, J.D. Bussell, S.M. Smith, and C.A. Beveridge, 2012 The Arabidopsis Ortholog of Rice DWARF27 Acts Upstream of MAX1 in the Control of Plant Development by Strigolactones. *Plant Physiology* 159 (3):1073-1085.

Westerbergh, A., and J. Doebley, 2004 Quantitative trait loci controlling phenotypes related to the perennial versus annual habit in wild relatives of maize. *Theoretical and Applied Genetics* 109 (7):1544-1553.

Whipple, C.J., T.H. Kebrom, A.L. Weber, F. Yang, D. Hall *et al.*, 2011 grassy tillers1 promotes apical dominance in maize and responds to shade signals in the grasses. *Proceedings of the National Academy of Sciences of the United States of America* 108 (33):E506-E512.

Xie, K.B., C.Q. Wu, and L.Z. Xiong, 2006 Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice. *Plant Physiology* 142 (1):280-293.

Xu, M., L. Zhu, H.X. Shou, and P. Wu, 2005 A PIN1 family gene, OsPIN1, involved in auxin-dependent adventitious root emergence and tillering in rice. *Plant and Cell Physiology* 46 (10):1674-1681.

Yang, F., Q. Wang, G. Schmitz, D. Muller, and K. Theres, 2012 The bHLH protein ROX acts in concert with RAX1 and LAS to modulate axillary meristem formation in Arabidopsis. *Plant Journal* 71 (1):61-70.

Yoo, S.K., X.L. Wu, J.S. Lee, and J.H. Ahn, 2011 AGAMOUS-LIKE 6 is a floral promoter that negatively regulates the FLC/MAF clade genes and positively regulates FT in Arabidopsis. *Plant Journal* 65 (1):62-76.

Zhou, Y., M. Honda, H. Zhu, Z. Zhang, X. Guo *et al.*, 2015 Spatiotemporal Sequestration of miR165/166 by Arabidopsis Argonaute10 Promotes Shoot Apical Meristem Maintenance. *Cell Rep*.

Zou, J.H., Z.X. Chen, S.Y. Zhang, W.P. Zhang, G.H. Jiang *et al.*, 2005 Characterizations and fine mapping of a mutant gene for high tillering and dwarf in rice (Oryza sativa L.). *Planta* 222 (4):604-612.

CHAPTER 2

GENOTYPING BY SEQUENCING OF 393 SORGHUM BICOLOR BTX623 × IS3620C RECOMBINANT INBRED LINES IMPROVES SENSITIVITY AND RESOLUTION OF QTL DETECTION[1]

[1] Kong, WQ, Goff, V and Paterson, AH, 2017. To be submitted to G3: *Genes* | *Genomics* | *Genetics*

Abstract

We describe a genetic map with a total 381 bins of 616 GBS-based SNP markers in a F_6-F_8 recombinant inbred line (RIL) population of 393 individuals derived from crossing *S. bicolor* BTx623 to *S. bicolor* IS3620C, a guinea line largely diverged from BTx623. Five segregation distorted regions were found with four showing enrichment for *S. bicolor* alleles, suggesting possible selection force during formation of this RIL population. A quantitative trait locus (QTL) study with tripled number of individuals provided resources, validate previous findings, and demonstrated improved power of this genetic map to detect plant height and flowering time related QTLs by comparing to genes and QTLs found in other published studies. An unexpected low correlation between flowering time and plant height permit us to separate QTLs for each trait and provide evidence against pleiotropy. Ten non- random syntenic regions conferring QTLs of the same trait suggest that those QTL may harbor genes functioning the same manner after the duplication event happening 96 million years ago, while syntenic regions of QTLs for different trait may suggest sub-functionalization after the duplication. Together, this study provides resources for marker-assisted breeding, framework for fine mapping, and subsequent cloning of major genes for plant height and flowering time in sorghum.

Introduction

The most drought resistant of the world's top five cereal crops, sorghum, contributes 26-29 percent of calories in the human diet in semi-arid areas of Africa (FAO), on some of the world's most degraded soils and often with limited water inputs. A multi-purpose crop, sorghum has been traditionally used as grain and straw, and is also a promising crop for bioenergy production from starch, sugar, or cellulose on marginal lands with limited water and other resources (Rooney et al. 2007). Botanically, *Sorghum bicolor* is a model for plants that use C4 photosynthesis, improving carbon assimilation especially at high temperature, complementary to C3 model plants such as *Oryza sativa*. The sequenced ~730 megabase sorghum genome (Paterson et al. 2009) has not experienced genome duplication in an estimated ~96 million years (Wang et al. 2015) , making it particularly useful to study other C4 plants with large polyploid genomes such as maize and many other grasses in the *Saccharinae* clade including Miscanthus and Saccarum (sugarcane).

Evolution, natural selection and human improvement of sorghum have contributed to great morphological diversity of *Sorghum bicolor* spp. Cultivated forms of this species can be classified into five botanical races, bicolor, guinea, caudatum, kafir and durra; with ten intermediate races recognized based on inflorescence architecture and seed morphology (Dewet and Huckabay 1967; Harlan and Dewet 1972). *Sorghum bicolor* also has many wild relatives such as *S. propinquum* (2n=2x=20), diverged from *S. bicolor* 1-2 million years ago (Feltus et al. 2004); *Sorghum halepense* (2n=4x=40), an invasive and weedy species formed by unintentional crossing of *S. bicolor* and *S. propinquum* (Paterson et al. 1995a); and many hybrids between these species. These many possible intra and inter specific crosses have made sorghum a particular interesting model for dissecting the genetic control of complex traits such as plant

height and maturity (Lin et al. 1995; Feltus et al. 2006; Takai et al. 2012; Hart et al. 2001), tillering and vegetative branching (Alam et al. 2014; Kong et al. 2014; Kong et al. 2015), perenniality related traits (Washburn et al. 2013; Paterson et al. 1995b), sugar composition (Ritter et al. 2008; Murray et al. 2008a; Murray et al. 2008b; Shiringani et al. 2010; Vandenbrink et al. 2013), stay-green (Haussmann et al. 2002; Kassahun et al. 2010), drought resistance (Sanchez et al. 2002; Kebede et al. 2001; Tuinstra et al. 1998), disease and insect resistance (Totad et al. 2005; Katsar et al. 2002).

S. bicolor IS3620C is a representative of botanical race 'guinea', and is substantially diverged from *S. bicolor* BTx623 among the sorghum races. Prior workers described a genetic map with 323 RFLP and 147 SSR markers (Bhattramakki et al. 2000; Kong et al. 2000), and a quantitative trait locus (QTL) study with 137 F6-8 RILs demonstrated its usefulness for QTL detection by discovering as many as 27 QTLs traits of agronomical importance, such as plant height, maturity, number of basal tillers and panicle length (Hart et al. 2001; Feltus et al. 2006).

Next generation sequencing (NGS) has brought new power to revealing allelic differences among individuals by detecting large numbers of single nucleotide polymorphisms (SNP). While it is neither necessary nor practical to generate whole genome sequences for every individual in a designed mapping population, reduced representation library (RRL) sequencing has been widely utilized and proved to be both a cost and labor efficient tool for genotyping (Kim et al. ; Andolfatto et al. 2011b). Although current genotyping by sequencing (GBS) platforms and pipelines still face various issues regarding accuracy (especially in polyploids with high levels of heterozygosity), GBS is still a very powerful tool to generate genetic maps with high quality and resolution in nearly homozygous populations with reference genome sequences available (Kim et al. 2015).

In this paper, we describe a genetic map with a total of 381 bins of 616 GBS-based SNP markers in a recombinant inbred line (RIL) population of 393 individuals derived from two divergent *S. bicolor* genotypes, BTx623 and IS3620C. While recent work has utilized digital genotyping to saturate the genetic map of 137 individuals (Morishige et al. 2013) and has validated QTLs for height more precisely, this 6-fold improvement of the previous genetic map showed little increase in power to detect QTLs relative to prior analysis using a lower density of markers (Hart et al. 2001). Our study has tripled the number of individuals described in previous studies (Bhattramakki et al. 2000; Kong et al. 2000; Hart et al. 2001; Morishige et al. 2013), increasing both the number of discernable recombination events and the number of individuals carrying each parental allele, increasing the power to detect QTLs. A quantitative trait analysis of days to flowering and components of plant height demonstrate the improved power to detect quantitative trait loci (QTL) in this expanded population. Syntenic relationship of QTLs within the sorghum may reflect the homologous regions retaining the same or similar function after the duplication events 96 million years ago (Wang et al. 2015). Results of this new QTL mapping build on those of many other studies of these traits (Zhang et al. 2015; Morris et al. 2013; Hart et al. 2001; Shiringani et al. 2010; Lin et al. 1995; Kebede et al. 2001; Srinivas et al. 2009; Brown et al. 2006), enriching current resources and providing a better understanding of genetic control of plant height and days to flowering of *S. bicolor*.

Materials and Methods

Genetic stocks:

The mapping population is comprised of 399 F7-8 RILs derived by selfing a single F_2 plant from *S. bicolor* BTx623 and IS3620C as described (Hart et al. 2001; Kong et al. 2000). This RIL population was planted at the University of Georgia Plant Science Farm, Watkinsville,

GA, USA on 10 May 2011 and 18 May 2012. Single 3-m plots of each RIL were machine planted in a completely randomized design.

Genotyping:

Leaf samples of the RIL population was frozen at -80 C and lyophilized for 48 hours. Genomic DNA was extracted from the lyophilized leaf sample based on Aljanabi et al. (1999).

Our GBS platform is a slightly modified version of Multiplex Shotgun Genotyping (MSG) (Andolfatto et al. 2011a) combined with the Tassel GBS analysis pipeline. The basic sequencing platform is an in-house Illumina MiSeq that generates up to 25 million reads of 150 base pairs (bps) fragments per run with single-end sequencing. With one sequencing run of this platform, we obtained 7103 raw SNPs and 691 polymorphic SNPs that were sufficient for genetic mapping for this RIL population (Kim et al. 2015).

SNP 'calling' (inference) was based on the reference genome sequence of *Sorghum bicolor* (Paterson et al. 2009). In TASSEL-GBS, the first 64 bps of each reads were mapped onto a reference genome to decide the position of the reads. SNPs were called based on the alignment of reads to the reference genome. Heterozygosity at a locus is called if two alleles are each inferred to be present at a probability greater than that of sequencing error. Raw SNP data from the Tassel GBS pipeline were further filtered based on several criteria: (a) SNPs were removed if the minor allele frequency is less than 5% or the proportion of missing genotypes greater than 40%; (b) In order to reduce the number of redundant SNPs in studies where strong linkage disequilibrium necessitates only 5–10 cM resolution, we merged SNPs for which pairwise linkage disequilibrium (r^2) is greater than 0.9, deriving consensus genotypes in a manner minimizing missing genotypes. SNPs are further merged if the Pearson's correlation between them is larger than 0.95; (c) For bi-parental populations, the missing genotype of one parental

line can be imputed by offspring genotypes if the genotype of the other parent at the locus is known. After these filtering steps, SNP data are used for genetic mapping.

Map construction:

A genetic map using 616 SNP markers was firstly created using R/qtl (Broman et al. 2003). We further assigned bins for each chromosome to merge markers within 1 cM in genetic distance. Bin genotypes were defined as follows: If there was only one marker in the bin, the bin genotype would be the same as the marker genotype; if there were more than one marker in the bin, bin genotypes would be determined by merging marker genotypes to minimize missing data points. For example, for a particular individual if there were three SNP markers in a bin, and if the marker genotypes for all three SNPs agree, the bin genotype will be the same with the marker genotypes; if the marker genotypes showed discrepancy but not due to missing data, the bin genotype would be missing data, if more than one genotype were missing, the bin genotype would be the same with the non-missing genotype. Following this method, we obtained a total of 381 bins for map construction. Marker ordering used both *de novo* and reference based methods, i.e., the physical positions of SNPs. The 'Ripple' function was used to assist and validate ordering of the genetic map.

We used a chi-squared test to calculate the deviation from expected ratio (1:1) for each marker with both raw and imputed data as an indicator for segregation distortion. To account for multiple comparisons across the genome, the significance level was adjusted using Bonferroni correction. The imputed data was generated using R/qtl (Broman et al. 2003).

QTL mapping

QTLs were detected for five traits of interest: plant height (**PH**), the overall length of a plant; base to flag length (**BTF**), the length of from base of the plant to the flag leaf; flag to

rachis length (**FTR**), the length from the flag leaf to rachis (a positive sign was assigned if the position of the rachis is taller than the flag leaf; a negative sign was assigned if the rachis was 'buried' in the flag leaf); number of nodes (**ND**); and days to flowering (**FL**), the average days to flowering for the first five plants for each genotype.

We combined the phenotypic data using Best Linear Unbiased Prediction (BLUP) by treating individuals, years, replications nested within years and the interactions between individuals and years as random, since heritability for the traits of interest were relatively high. In 2011, we observed and recorded a soil type change within the experimental fields, which was treated as a covariate to calculate BLUP values for each genotype. A genome scan with the interval mapping method was first conducted with 1000 permutation tests; the putative QTLs were then selected and fit into a multiple QTL model. We added additional QTLs to the model if they exceeded the threshold of 3.0 after fixing the effect of QTLs included in the first genome scan. A multiple QTL model (MQM) was used to determine the final model for each trait. All statistical analyses and QTL mapping used R (R Core Team 2016) and the R/QTL package (Broman et al. 2003).

QTL nomenclature used a system that was previously described in rice (McCouch et al. 1997), starting with a 'q', followed by an abbreviation for each trait (**PH**, **BTF**, **FTR**, **ND** and **FL**), then the chromosome number and a decimal number to differentiate multiple QTLs on the same chromosome.

Results

Genetic map

A total of 399 RILs were genotyped with 690 SNP markers. Six individuals with three times more than the average number of recombination events were removed from the analysis.

Marker ordering first follows the published sorghum genome sequence (Paterson et al. 2009). A *de novo* marker ordering method is also used to compare the order of the genetic map with the reference-based method, but no obvious differences were observed for these two methods in terms of the LOD scores. We excluded 74 unlinked SNPs and obtained an initial genetic map with a total of 616 markers on the ten sorghum chromosomes. As detailed in the methods, we combined SNPs that are within 1cM in genetic distance to construct a genetic map with 381 bins (bin map) with varying SNP numbers in each bin (Supplementary document 1). The bin map collectively spans a genetic distance of 1404.8 cM, with average spacing 3.8 cM between loci and the largest gap being 27.6 cM on chromosome 5 (Table 2.1). The percentage of missing genotypes is 24% and 18.4% in total for the initial and bin maps, respectively. About 54.2% of the alleles of the RIL population come from *S. bicolor* BTx623, and 45.8% from *S. bicolor* IS3620C.

There are several advantages of the bin mapping strategy. First, it reduces the percentage of missing genotypes by combining the genotypes of adjacent markers. For this experiment, the percentage of missing genotypes is reduced by 5.6% with the bin map. Moreover, QTL intervals are usually 5-10 cM for a typical bi-parental QTL experiment; high marker density does not significantly increase the power of detecting QTLs (Lander and Botstein 1989) while combining markers may increase computational processing speed.

The order of the genetic map agrees closely with the physical positions of loci on the genome sequence (Figure 2.1), suggesting that the GBS method used yields a high quality genetic map in this nearly-homozygous diploid population. Markers are generally more concentrated in the distal regions of each chromosome than central regions (consistent with the general distribution of low-copy DNA sequences in sorghum: Paterson et al 2008), although

distributions of markers on each chromosome vary. For example, we have observed a much larger pericentrimeric region on chromosome 7 than on chromosome 1.

Segregation distortion

Segregation distortion occurs when the segregation ratio of offspring at a locus deviates from the Mendelian expectation. In a RIL population, we expect to see half of the alleles come from one parent and the other half from the other parent, i.e., the expected segregation ratio is 1:1 for each marker locus. A deviation from this ratio may be a result of gametic or zygotic selection. The BTx623 x IS3620C genetic map reveals several clusters of markers experiencing segregation distortion on chromosomes 1, 4, 5, 8, and 9 (Figure 2.2), peaking at 48.0, 153.6, 64.4, 47.5, and 122.4 cM, corresponding to 14.8, 65.1, 49.9,27.8-38.5 and 59.5 in physical distance, respectively, with the imputed data. All regions but the one on chromosome 8 show enrichment of *S. bicolor* alleles. The most extreme segregation distorted region is on chromosome 1, spanning 0-105 cM (Figure 2.2) with a ratio of 358:35 (p = 1.10E-59) at the peak marker, S1_14765342. Peng et al. (1999) also observed this long-spanning segregation distorted region on chromosome 1, despite using a population with a much smaller sample size and marker numbers.

Interestingly, the same general region of extreme segregation distortion on chromosome 1 is also found in two *S. bicolor* × *S. propinquum* derived populations (Kong 2013; Bowers et al. 2003), and remarkably, all three populations mapped this distortion peak at ~14 Mb in physical distance. This correspondence between populations suggests that alleles from *S. bicolor* might be selected for at this genomic position, however, the exact mechanism and the genes involved are unclear.

QTL Mapping:

The BTx623 x IS3620C RIL map provides high power for detecting QTLs. As examples, we have investigated QTLs for five phenotypic traits, plant height (**PH**), base to flag length (**BTF**), flag to rachis length (**FTR**), number of nodes (**ND**) and days to flowering (**FL**). Means, standard deviation and other summary statistics are shown in Table 2.2. Broad-sense heritability estimates for all five traits are relatively high (Table 2.2). It is interesting that the average **PH** of the progenies is 98.62 cm, greater than the average of either parents, 94.40 for BTx623 and 84.49 for IS3620C . The same was observed for **BTF**. Both height components show substantial genetic variations, indicating that each parent contributes different alleles for **PH** to their progenies (Table 2.2). Moreover, the relatively high amount of genetic variations in **PH** in this population fosters discovery of QTLs, despite that the difference between the two parents is relatively small.

PH and **BTF** are highly correlated in this population (Table 2.3), with a correlation coefficient of 0.8983 ($p<0.001$). We detect a total of seven and five QTLs for **PH** and **BTF**, accounting for 40.13% and 41.58% of the total phenotypic variances, respectively (Table 2.4). Three QTLs on chromosomes 3, 6 and 7 overlap for these two traits. The QTLs on chromosomes 6 and 7 account for the majority of the phenotypic variance explained for these traits, ~28% and ~32% for **PH** and **BTF**, respectively. These two large effect QTLs might be related to previously defined **PH** genes, presumably *dw2* on chromosome 6 and *dw3* (Sb07g023730) on chromosome 7 (Multani et al. 2003; Quinby and Karper 1945).

Compared to prior analysis using a subset of the individuals from this population Hart et al. (2001), our study largely increases the power of QTL detection. Not only do we detect more QTLs than the previous study, seven QTLs for **PH** and nine QTLs for **FL** in our study, compared

to five and three from the previous study, , some of the QTL intervals also significantly decrease (Table 2.4). For example, the 1-lod interval of QTL on chromosome 7 is narrowed from ~26 cM previously to only about 3 cM in this study (~57.7 Mb to ~59.5 Mb in physical distance) , , which harbors the gene Sb07g023730 (*DW3*)at ~58.6 Mb. This example indicates that nearly tripling numbers of individuals and increasing marker density greatly increased the power of QTL detection in this study.

An interesting phenotype that we observed is distance from the flag leaf to the rachis. We distinguished whether the rachis is buried in the flag leaf in our phenotyping system (see Materials and Methods). The correlation coefficient (Table 2.3) between **FTR** and **PH** ($r=0.0778$), though significant at $p<0.01$, is not nearly high as the correlation between the **BTF** and **PH** ($r=0.8983$), suggesting that the genetic control of these two traits might be different. Indeed, QTL mapping suggests that the genetic control of **FTR** is quite different from that of **PH** and **BTF**, with only one QTL (qFTR7.1) on chromosome 7 overlapping with QTL for **PH** and **BTF** (qPH7.1 and qBTF7.1). While the one-LOD QTL interval for **PH** qPH10.1 overlaps with qFTR10.1 for **FTR** to some extent, there is no solid evidence to conclude that they are controlled by the same genetic factors, given that the likelihood peaks of the QTLs for these two traits are ~10 cM apart. We detected a total of five QTLs for **FTR**, explaining 28.21 % of the total phenotypic variance (Table 2.4 and Figure 2.3). The additive effect of **FTR** needs to be carefully interpreted, especially when comparing to the additive effect of **PH**. Since the average value of **FTR** is negative, a negative number for the additive effect for this trait indicates increased **FTR** for a particular allele. For example, the additive effect for qFTR7.1 is -1.19, which indicates that **FTR** of plants carrying the IS3620 alleles are actually longer than those of plants carrying the

BTx623 alleles. Both alleles from IS3620C for qPH7.1 and qFTR7.1 have the same effect of increasing length, although the sign of their additive effects is different.

We have detected a total of six QTLs for number of nodes (**ND**) in this population (Figure 2.3), collectively explaining 32.07% of the phenotypic variance. The largest effect QTL is qND8.1, with a LOD score of 12.96 and explaining 11.15% of the variance. QTLs for the **ND** rarely overlap with other **PH** related traits—the only QTL that shows some correspondence with other height related traits is qND10.1, marginally overlapping with qFTR10.1.

In this study, **PH** and **FL** were not significantly correlated (Table 2.4 QTL mapping plant height, base to flag length, flag to rachis length, number of nodes and days to flower in the *S. bicolor* BTx623× IS3620C RILs.), an unexpected finding compared to many other sorghum studies (Ritter et al. 2008; Lin et al. 1995; Murray et al. 2008b). A total of nine QTLs are detected for **FL**, substantially more than the 4-6 conventionally thought to influence this trait in a wide range of sorghum genotypes (Quinby and Karper 1945), although only collectively explaining 46.33% of the total phenotypic variance. QTL intervals for **FL** rarely overlap with those for **PH**, **BTF** and **FTR**. However, five out of six QTLs for **ND** overlap with QTLs for **FL**, and the sign of the allelic effect suggests that plants with more nodes usually flower late. This result indicates that some genes might have pleiotropic effects on these traits or genes for these two traits are linked and have been selected simultaneously.

QTL correspondence with other studies

Traits related to sorghum **PH** and **FL** have been extensively studied in many QTL experiments (Lin et al. 1995; Brown et al. 2006; Hart et al. 2001; Ritter et al. 2008; Shiringani et al. 2010; Srinivas et al. 2009; Kebede et al. 2001) and two genome wide association studies (Morris et al. 2013; Kong et al. 2015). The Comparative Saccharinae Genome Resource QTL

database (Zhang et al. 2013) aids comparisons of QTL intervals across different studies in sorghum and facilitates validation of QTLs for traits of interest. All our QTLs detected for **PH** have been found in other studies. Three **PH** QTLs, qPH6.1, qPH7.1 and qPH9.1, also found in two GWAS studies (Morris et al. 2013; Zhang et al. 2015), are likely to correspond to three dwarf genes in sorghum, *dw2*, *dw3* and *dw1* (Quinby and Karper 1945). The accuracy of our QTL study can be demonstrated by the gene known to cause the *dw3* phenotype, Sb07g023730, on chromosome 7. Our QTL study narrowed the 1-lod interval for this QTL to 2Mb (57.7-59.5Mb), with a peak at 58.4Mb, close to the 58.6 Mb location of the causal gene (Multani et al. 2003).

A total of 6 QTLs controlling **FL** in the BTx623 x IS3620C RILs, qFL1.2, qFL3.1, qFL6.1, qFL8.2, qFL9.1 and qFL10.1, showed correspondence with QTLs found in other studies, (Yang et al. 2014; Brown et al. 2006; Lin et al. 1995; Hart et al. 2001; Shiringani et al. 2010) and two QTLs, qFL1.1 and qFL6.1, are novel. The QTLs with the largest effects on **FL**, qFL8.2 with a LOD score of 16.1 and explaining 11.16% of the phenotypic variance; and qFL9.1 with a LOD score of 11.4 and explaining 7.69% of phenotypic variance, have been consistently found in many independent QTL and GWAS studies (Lin et al. 1995; Brown et al. 2008; Zhang et al. 2015; Morris et al. 2013). Identification of the genes underlying these QTL regions might be especially important. The peak of qFL9.1 is located at ~59.3Mb in our study, close to significant peaks at ~58.7Mb and ~59.0Mb for **FL** found in a GWAS study (Zhang et al. 2015).

Syntenic study

A total of 30 out of 202 genomic regions contain QTLs found in this study were located in the colinear locations within sorghum (Paterson et al. 2009), possibly being experienced genome duplication events (Table 2.5 and Figure 2.4), thanks to the Plant Genome Duplication

Database (Lee et al. 2013). Among these, a total of five regions on chromosomes 1 (1), 3 (3), 9 (1) are duplicated within the same chromosome. Among the 25 duplicated genomic regions located on different chromosomes, ten syntenic regions contain the same trait (2 for ND, 3 for PH, 4 for FL and 1 for BTF), which is significant than observed by chance (p=0.0002), and 15 regions contain different traits (Figure 2.4).

Discussion

The present study providing a relatively dense GBS-based map for 393 individuals of the singularly-important *S. bicolor* BTx623× IS3620C identifies new QTLs and increases precision of mapping previously-known QTLs, providing both an important resource and new information about the genetic control of important sorghum traits. Benefiting both from an increased sample size and GBS, our study has demonstrated increased power and accuracy of detecting QTLs, relative to previous studies of a total of 137 individuals (Hart et al. 2001; Feltus et al. 2006). We also discovered a total of five regions with segregation distortion, possibly due to gameto or zygotic selection during the formation of this RIL population.

Within this single population, we now find more QTLs for **PH** and **FL** than have been classically thought to segregate in all forms of *Sorghum bicolor* (Quinby and Karper 1945), reiterating a conclusion from meta-analysis of multiple populations (Zhang et al. 2015) that these traits are more complex than previously thought. Most of the QTLs we mapped here correspond to QTLs found in other studies, improving confidence in our result. An example is re-identification of the *dw3* locus on chromosome 7, a P-glycoprotein auxin transporter (Multani et al. 2003), which proved the power and accuracy of our QTL study by narrowing the QTL intervals from ~26 cM to 3 cM and harboring the *dw3* locus. As a complement, intersections with published data may demonstrate on the resolution of QTL mapping, toward identification of

causal genes. For example, the candidate gene on chromosome 6, possibly *dw2*, has been refined to a 5 cM interval in genetic distance in this study, peaking at ~42.4 Mb in physical distance. The GWAS study (Morris et al. 2013) proposed that the location of *dw2* lies between 39.7 Mb- 42.6 Mb, and a recently study proposed it to be Sobic.006G067700 (Sb06g15430, (Hilley et al. 2017)). However, an independent study suggests that the strongest allele variation with **PH** may be at Sb06g007330 (Cuevas et al. 2016).

In addition to **PH** and **FL**, we have also identified QTLs for three other traits which are not extensively studied, **BTF**, **FTR** and **ND**. QTL intervals for **FTR** mostly differ from those associated with **PH**, suggesting that the genetic control of these traits might be different. We also found that the genetic control of the **ND** is correlated with **FL**, demonstrated by the fact that five of six QTLs for **ND** correspond to QTLs for **FL**.

An unexpected low correlation between **FL** and **PH** in this population and the high power of this genetic map, together permit us to differentiate **FL** and **PH** QTLs. In fact, no corresponding QTLs for these two traits were identified in the present study, an extremely unusual finding. This lack of overlap strongly supports a hypothesis (Cuevas et al. 2016(Lin et al 1995) that the *dw2* trait affecting **PH** and the *ma1* trait affecting flowering, each mapping very close together on chromosome 6), are determined by different genes. This is proven by the fact that *ma1* alleles are not included in this study (IS3620C was 'converted' to day-neutral flowering) (Stephens et al. 1967) while a strong signal has been detected for **PH** on chromosome 6 at ~42.4cM, in the vicinity to the recently published *dw2* gene (Hilley et al. 2017), albeit further functional analysis is needed. A similar example is within the general area of *dw1* (Yamaguchi et al. 2016) on chromosome 9 (qPH9.1) where we also find a QTL controlling **FL**,

qFL9.1, that is ~30cM from qPH9.1. This result again suggests that two separate QTLs control these **PH** and **FL**, a conclusion that is also supported by another study (Thurber et al. 2013).

We discovered a total of 30 the syntenic regions containing QTLs within the sorghum genome sequence (Lee et al. 2013; Paterson et al. 2009), with 10 regions containing QTLs responsible for the same trait (Table 2.5 and Figure 2.4). This non-random correspondence between regions of genome conferring the same traits indicates that the ten syntenic regions contain corresponding genes may still function in the same ways at the homologous locations, despite duplicated 96 million years ago (Wang et al. 2015), while the syntenic regions with different trait may suggest potential sub-functionalization of genes after the duplication.

Components of **PH** and **FL** have been and will continue to be important for sorghum breeding programs. The past century has witnessed breeding for modern varieties with a particular plant type, for example a semi-dwarf type, to realize striking increases in production such as those which led to the "Green Revolution" (Evenson and Gollin 2003). The concept of ideotype breeding (Donald (1968), is still an ongoing priority for many breeding programs to increase food and feed production, adapt to climate change and minimize inputs. Genetic components discovered for plant height related traits and flowering time in this study, together with closely-linked diagnostic DNA markers that permit their selection at seedling stages or in non-target environments, may benefit breeding for plant types idealized for the different purposes that sorghum is used. Specifically, little correspondence between **PH** and **FL**, together with narrowed QTL intervals, facilitate accurate selection for each trait. The QTLs found in this study and their correspondence with those from many other studies also provides a framework for fine mapping or subsequent cloning of major genes for **PH** and **FL** in sorghum.

Acknowledgments

The late K. Schertz, USDA-ARS and Texas A&M University, was instrumental in producing the BTx623× IS3620C RILs. We thank the members of the Paterson lab for valuable help and precious suggestions in this work. We appreciate financial support from the AFRI Plant Growth and Development Program (2009-03477), DOE-USDA Plant Feedstock Genomics Program (2012-03304), and USDA Biotechnology Risk Assessment Program (2012-01658).

References

Alam, M.M., E.S. Mace, E.J. van Oosterom, A. Cruickshank, C.H. Hunt *et al.*, 2014 QTL analysis in multiple sorghum populations facilitates the dissection of the genetic and physiological control of tillering. *Theoretical and Applied Genetics* 127 (10):2253-2266.

Aljanabi, S.M., L. Forget, and A. Dookun, 1999 An improved and rapid protocol for the isolation of polysaccharide- and polyphenol-free sugarcane DNA. *Plant Molecular Biology Reporter* 17 (3):281-281.

Andolfatto, P., D. Davison, D. Erezyilmaz, T.T. Hu, J. Mast *et al.*, 2011a Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Research* 21 (4):610-617.

Andolfatto, P., D. Davison, D. Erezyilmaz, T.T. Hu, J. Mast *et al.*, 2011b Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Research* 21 (4):610-617.

Bhattramakki, D., J.M. Dong, A.K. Chhabra, and G.E. Hart, 2000 An integrated SSR and RFLP linkage map of Sorghum bicolor (L.) Moench. *Genome* 43 (6):988-1002.

Bowers, J.E., C. Abbey, S. Anderson, C. Chang, X. Draye *et al.*, 2003 A high-density genetic recombination map of sequence-tagged sites for Sorghum, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses. *Genetics* 165 (1):367-386.

Broman, K.W., H. Wu, S. Sen, and G.A. Churchill, 2003 R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19 (7):889-890.

Brown, P.J., P.E. Klein, E. Bortiri, C.B. Acharya, W.L. Rooney *et al.*, 2006 Inheritance of inflorescence architecture in sorghum. *Theoretical and Applied Genetics* 113 (5):931-942.

Brown, P.J., W.L. Rooney, C. Franks, and S. Kresovich, 2008 Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes. *Genetics* 180 (1):629-637.

Cuevas, H.E., C. Zhou, H. Tang, P.P. Khadke, S. Das *et al.*, 2016 The Evolution of Photoperiod-Insensitive Flowering in Sorghum, A Genomic Model for Panicoid Grasses. *Molecular Biology and Evolution* 33 (9):2417-2428.

Dewet, J.M.J., and J.P. Huckabay, 1967 Origin of Sorghum Bicolor .2. Distribution and Domestication. *Evolution* 21 (4):787-&.

Donald, C.M., 1968 The breeding of crop ideotypes. *Euphytica* 17 (3):385-403.

Evenson, R.E., and D. Gollin, 2003 Assessing the impact of the green revolution, 1960 to 2000. *Science* 300 (5620):758-762.

Feltus, F.A., G.E. Hart, K.F. Schertz, A.M. Casa, P. Brown *et al.*, 2006 Genetic map alignment and QTL correspondence between inter- and intra-specific sorghum populations. *Theoretical and Applied Genetics* 112:1295-1305.

Feltus, F.A., J. Wan, S.R. Schulze, J.C. Estill, N. Jiang *et al.*, 2004 An SNP resource for rice genetics and breeding based on subspecies indica and japonica genome alignments. *Genome Research* 14 (9):1812-1819.

Harlan, J.R., and J.M.J. Dewet, 1972 Simplified Classification of Cultivated Sorghum. *Crop Science* 12 (2):172-&.

Hart, G.E., K.F. Schertz, Y. Peng, and N.H. Syed, 2001 Genetic mapping of Sorghum bicolor (L.) Moench QTLs that control variation in tillering and other morphological characters. *Theoretical and Applied Genetics* 103 (8):1232-1242.

Haussmann, B.I.G., V. Mahalakshmi, B.V.S. Reddy, N. Seetharama, C.T. Hash *et al.*, 2002 QTL mapping of stay-green in two sorghum recombinant inbred populations. *Theoretical and Applied Genetics* 106 (1):133-142.

Hilley, J.L., B.D. Weers, S.K. Truong, R.F. McCormick, A.J. Mattison *et al.*, 2017 Sorghum Dw2 Encodes a Protein Kinase Regulator of Stem Internode Length. *Sci Rep* 7 (1):4616.

Kassahun, B., F.R. Bidinger, C.T. Hash, and M.S. Kuruvinashetti, 2010 Stay-green expression in early generation sorghum [Sorghum bicolor (L.) Moench] QTL introgression lines. *Euphytica* 172 (3):351-362.

Katsar, C.S., A.H. Paterson, G.L. Teetes, and G.C. Peterson, 2002 Molecular analysis of Sorghum resistance to the greenbug (Homoptera : Aphididae). *Journal of Economic Entomology* 95 (2):448-457.

Kebede, H., P.K. Subudhi, D.T. Rosenow, and H.T. Nguyen, 2001 Quantitative trait loci influencing drought tolerance in grain sorghum (Sorghum bicolor L. Moench). *Theoretical and Applied Genetics* 103 (2-3):266-276.

Kim, C., H. Guo, W. Kong, R. Chandnani, L.-S. Shuang *et al.*, 2015 Application of genotyping by sequencing technology to a variety of crop breeding programs. *Plant Science* (0).

Kong, L., J. Dong, and G.E. Hart, 2000 Characteristics, linkage-map positions, and allelic differentiation of Sorghum bicolor (L.) Moench DNA simple-sequence repeats (SSRs). *Theoretical and Applied Genetics* 101 (3):438-448.

Kong, W., 2013 Genetic Analysis of Plant Architecture in Sorghum in *Crop and Soil Sciences*. The University of Georgia, Athens, GA.

Kong, W., H. Guo, V. Goff, T.-H. Lee, C. Kim *et al.*, 2014 Genetic analysis of vegetative branching in sorghum. *Theoretical and Applied Genetics* 127 (11):2387-2403.

Kong, W.Q., C. Kim, V.H. Goff, D. Zhang, and A.H. Paterson, 2015 Genetic Analysis of Rhizomatousness and Its Relationship with Vegetative Branching of Recombinant Inbred Lines of Sorghum Bicolor X S-Propinquum. *American Journal of Botany* 102 (5):718-724.

Lander, E.S., and D. Botstein, 1989 Mapping Mendelian Factors Underlying Quantitative Traits Using Rflp Linkage Maps. *Genetics* 121 (1):185-199.

Lee, T.H., H.B. Tang, X.Y. Wang, and A.H. Paterson, 2013 PGDD: a database of gene and genome duplication in plants. *Nucleic Acids Research* 41 (D1):D1152-D1158.

Lin, Y.R., K.F. Schertz, and A.H. Paterson, 1995 Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population. *Genetics* 141 (1):391-411.

McCouch, S.R., Y.G. Cho, M. Yano, E. Paul, M. Blinstrub *et al.*, 1997 Report on QTL nomenclature. *Rice Genetics Newsletter* 14:11-13.

Morishige, D.T., P.E. Klein, J.L. Hilley, S.M. Sahraeian, A. Sharma *et al.*, 2013 Digital genotyping of sorghum - a diverse plant species with a large repeat-rich genome. *Bmc Genomics* 14:448.

Morris, G.P., P. Ramu, S.P. Deshpande, C.T. Hash, T. Shah *et al.*, 2013 Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences of the United States of America* 110 (2):453-458.

Multani, D.S., S.P. Briggs, M.A. Chamberlin, J.J. Blakeslee, A.S. Murphy *et al.*, 2003 Loss of an MDR transporter in compact stalks of maize br2 and sorghum dw3 mutants. *Science* 302 (5642):81-84.

Murray, S.C., W.L. Rooney, S.E. Mitchell, A. Sharma, P.E. Klein *et al.*, 2008a Genetic Improvement of Sorghum as a Biofuel Feedstock: II. QTL for Stem and Leaf Structural Carbohydrates. *Crop Science* 48 (6):2180-2193.

Murray, S.C., A. Sharma, W.L. Rooney, P.E. Klein, J.E. Mullet *et al.*, 2008b Genetic Improvement of Sorghum as a Biofuel Feedstock: I. QTL for Stem Sugar and Grain Nonstructural Carbohydrates. *Crop Science* 48 (6):2165-2179.

Paterson, A., H., J.E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood *et al.*, 2009 The Sorghum bicolor genome and the diversification of grasses *Nature* 457:551-556.

Paterson, A.H., K.F. Schertz, Y.R. Lin, S.C. Liu, and Y.L. Chang, 1995a The Weediness of Wild Plants - Molecular Analysis of Genes Influencing Dispersal and Persistence of Johnsongrass, Sorghum-Halepense (L) Pers. *Proceedings of the National Academy of Sciences of the United States of America* 92 (13):6127-6131.

Paterson, A.H., K.F. Schertz, Y.R. Lin, S.C. Liu, and Y.L. Chang, 1995b The Weedniss of Wild Plants- Molecular Analysis of Genes Influencing Dispersal and Persistence of Johngrass, Sorghum-Halepense (L.) Pers. *Proceedings of the National Academy of Sciences of the United States of America* 92 (13):6127-6131.

Peng, Y., K.F. Schertz, S. Cartinhour, and G.E. Hart, 1999 Comparative genome mapping of Sorghum bicolor (L.) Moench using an RFLP map constructed in a population of recombinant inbred lines. *Plant Breeding* 118 (3):225-235.

Quinby, J., and R. Karper, 1945 The inheritance of three genes that influence time of floral initiation and maturity date in milo. *Jour. Amer. Soc. Agron.* 37 (916).

R Core Team, 2016 R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ritter, K.B., D.R. Jordan, S.C. Chapman, I.D. Godwin, E.S. Mace *et al.*, 2008 Identification of QTL for sugar-related traits in a sweet x grain sorghum (Sorghum bicolor L. Moench) recombinant inbred population. *Molecular Breeding* 22 (3):367-384.

Rooney, W.L., J. Blumenthal, B. Bean, and J.E. Mullet, 2007 Designing sorghum as a dedicated bioenergy feedstock. *Biofuels Bioproducts & Biorefining-Biofpr* 1 (2):147-157.

Sanchez, A.C., P.K. Subudhi, D.T. Rosenow, and H.T. Nguyen, 2002 Mapping QTLs associated with drought resistance in sorghum (Sorghum bicolor L. Moench). *Plant Molecular Biology* 48 (5):713-726.

Shiringani, A.L., M. Frisch, and W. Friedt, 2010 Genetic mapping of QTLs for sugar-related traits in a RIL population of Sorghum bicolor L. Moench. *Theoretical and Applied Genetics* 121 (2):323-336.

Srinivas, G., K. Satish, R. Madhusudhana, R.N. Reddy, S.M. Mohan *et al.*, 2009 Identification of quantitative trait loci for agronomically important traits and their association with genic-microsatellite markers in sorghum. *Theoretical and Applied Genetics* 118 (8):1439-1454.

Stephens, J.C., F.R. Miller, and D.T. Rosenow, 1967 Conversion of Alien Sorghums to Early Combine Genotypes. *Crop Science* 7 (4):396-&.

Takai, T., J. Yonemaru, H. Kaidai, and S. Kasuga, 2012 Quantitative trait locus analysis for days-to-heading and morphological traits in an RIL population derived from an extremely late flowering F-1 hybrid of sorghum. *Euphytica* 187 (3):411-420.

Thurber, C.S., J.M. Ma, R.H. Higgins, and P.J. Brown, 2013 Retrospective genomic analysis of sorghum adaptation to temperate-zone grain production. *Genome Biol* 14 (6):R68.

Totad, A.S., B. Fakrudin, and M.S. Kuruvinashetti, 2005 Isolation and characterization of resistance gene analogs (RGAs) from sorghum (Sorghum bicolor L. Moench). *Euphytica* 143 (1-2):179-188.

Tuinstra, M.R., G. Ejeta, and P. Goldsbrough, 1998 Evaluation of near-isogenic sorghum lines contrasting for QTL markers associated with drought tolerance. *Crop Science* 38 (3):835-842.

Vandenbrink, J.P., V. Goff, H. Jin, W. Kong, A.H. Paterson *et al.*, 2013 Identification of bioconversion quantitative trait loci in the interspecific cross Sorghum bicolor x Sorghum propinquum. *Theor Appl Genet* 126 (9):2367-2380.

Wang, X., J. Wang, D. Jin, H. Guo, T.H. Lee *et al.*, 2015 Genome Alignment Spanning Major Poaceae Lineages Reveals Heterogeneous Evolutionary Rates and Alters Inferred Dates for Key Evolutionary Events. *Mol Plant* 8 (6):885-898.

Washburn, J.D., S.C. Murray, B.L. Burson, R.R. Klein, and R.W. Jessup, 2013 Targeted mapping of quantitative trait locus regions for rhizomatousness in chromosome SBI-01 and analysis of overwintering in a Sorghum bicolor x S. propinquum population. *Molecular Breeding* 31 (1):153-162.

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

Yang, S.S., B.D. Weers, D.T. Morishige, and J.E. Mullet, 2014 CONSTANS is a photoperiod regulated activator of flowering in sorghum. *BMC Plant Biology* 14.

Zhang, D., H. Guo, C. Kim, T.H. Lee, J.P. Li *et al.*, 2013 CSGRqtl, a Comparative Quantitative Trait Locus Database for Saccharinae Grasses. *Plant Physiology* 161 (2):594-599.

Zhang, D., W. Kong, J. Robertson, V.H. Goff, E. Epps *et al.*, 2015 Genetic analysis of inflorescence and plant height components in sorghum (Panicoidae) and comparative genetics with rice (Oryzoidae). *BMC Plant Biol* 15:107.

Chapter 2 Tables and figures

Table 2.1 Summary Statistics for the *S. bicolor* BTx623× IS3620C Initial Genetic Map and Bin Map.

Original Map					Bin Map				
Chr	Marker No.	Length (cM)	Avg Spacing (cM)	Max Spacing	Chr	Marker No.	Length	Avg Spacing	Max Spacing
1	140	186.6	1.3	8.5	1	66	189.7	2.9	8.6
2	83	173.2	2.1	8.9	2	50	180.2	3.7	9.7
3	80	158.6	2.0	11.0	3	43	160.7	3.8	11.8
4	57	158.9	2.8	21.1	4	42	164.9	4.0	21.3
5	23	119.2	5.4	27.6	5	21	123.7	6.2	27.6
6	53	106.9	2.1	14.6	6	39	111.8	2.9	14.9
7	44	125	2.9	17.5	7	30	127.8	4.4	17.7
8	35	100.8	3.0	17.5	8	21	101.6	5.1	17.8
9	36	120.5	3.4	21.4	9	23	122.4	5.6	21.6
10	65	120.6	1.9	5.3	10	46	126.1	2.8	5.9
Overall	616	1370.3	2.3	27.6	Overall	381	1408.8	3.8	27.6

Table 2.2 Summary Statistics for Components of Plant Height and Flowering for the *S. bicolor* BTx623× IS3620C RIL population and parents.

	RIL						BTx 623					IS3620C				
Trait	N	Mean	Std	Min	Max	Broad Sense Heritability (%)	N	Mean	Std	Min	Max	N	Mean	Std	Min	Max
Plant height (PH)	1523	98.62	24.0256	43.0	193.0	78.63	28	94.40	8.2106	74.8	111.4	29	84.49	9.4353	67.6	103.5
Base to flag length (BTF)	1509	72.12	23.5039	13.7	54.7	79.71	28	69.41	7.9703	49.2	86.3	28	64.86	9.4520	40.5	80.7
Flag to rachis length (FTR)	1509	-2.60	9.2664	-55.1	31.7	68.24	28	-3.68	5.4898	-14.5	9.7	28	-13.68	6.1065	-20.5	5.9
Number of nodes (ND)	1515	9.16	1.8573	4.0	16.0	73.61	28	11.75	1.3229	8.0	14.0	29	7.72	1.3065	6.0	10.0
Days to flower (FL)	776	59.01	9.2530	40.8	99.2	83.65	26	72.11	5.3668	63.4	80.0	15	58.35	6.1307	49.4	66.6

Table 2.3 Correlation Coefficients of five phenotypic traits in the *S. bicolor* BTx623× IS3620C RILs.

	Plant height	Base to flag length	Flag to rachis length	Number of nodes	Days to flower
Plant height	-	0.8983 ***(1508)	0.0778 ** (1508)	0.2255 ***(1514)	-0.0360 (770)
Base to flag length		-	-0.2459 *** (1508)	0.3331 *** (1500)	0.0038 (770)
Flag to rachis length			-	-0.3228 *** (1500)	-0.0795 *(770)
Number of nodes				-	0.5806 ***(770)
Days to flower					-

Numbers in the parenthesis refer to the sample sizes.

Table 2.4 QTL mapping plant height, base to flag length, flag to rachis length, number of nodes and days to flower in the *S. bicolor* BTx623× IS3620C RILs.

	Peak (cM)	Peak (Mb)	LOD	% Var Explained	Additive [1] Effect	One-LOD interval (cM)	Flanking SNP Physical positions (Mb)		QTLs found in other studies	GWAS studies
Plant height (**PH**)										
qPH2.1	121.8	65.5	3.1	2.20	2.37	114-133	63.7	68.8	Shiringani,2008;	-
qPH3.1	19.0	3.8-4.0	4.4	3.19	-2.86	15-30	3.2	5.6	Hart,2001; Brown, 2006	-
qPH6.1	21.0	42.4	22.1	17.68	-6.83	19-24	39.6	44.9	Kebede,2001; Brown, 2006; Srinivas, 2009	Zhang,2015; Morris,2013
qPH7.1	89.0	58.4	14.5	11.12	5.22	88-91	57.7	59.5	Hart,2001; Srinivas,2009	Zhang,2015; Morris, 2013
qPH8.1	69.0	50.2	3.1	2.21	2.53	62-75	47.5	51.8	Shiringani,2008;	
qPH9.1	93.0	52.6	3.3	2.35	2.36	91-101	52.0	55.8	Lin, 1995;	Zhang,2015; Morris, 2013
qPH10.1	55.0	12.3	4.3	3.06	-2.87	46-70	8.1	52.2	Hart,2001;	-
Full model				**40.13**						
Base to flag length (**BTF**)										
qBTF3.1	19.6	3.8-4.0	4.6	3.07	-2.75	17-25	3.2	4.9	NA	-
qBTF4.1	124.4	60.5	3.1	2.03	-2.09	117-133	58.9	61.9	NA	-
qBTF5.1	30.6	4.8	3.6	2.36	2.43	25-40	1.5	7.4	NA	-
qBTF6.1	23.0	44.7	22.8	16.92	-6.68	19-26	39.6	45.5	NA	Zhang, 2015
qBTF7.1	88.9	58.4	32.5	25.65	8.17	87-90	57.7	59.5	NA	Zhang, 2015
Full model				**41.58**						
Flag to rachis length (**FTR**)										
qFTR1.1	18.5	5.8	5.2	4.50	1.17	12-27	1.9	8.1	NA	
qFTR3.1	106.0	59.3	11.3	10.14	-1.51	104-115	58.9	63.0	NA	
qFTR7.1	88.0	58.4	7.1	6.23	-1.19	83-93	57.7	59.5	NA	

qFTR8.1	15.0	2.7	4.6	3.96	0.95	10-20	2.4	4.3	NA	
qFTR10.1	46.8	8.1	8.8	7.79	-1.32	44-50	6.7	9.9	NA	
Full model				28.21						
Number of nodes (**ND**)										
Full model				**32.07**						
qND1.1	69.0	23.1	5.6	4.65	0.30	66-75	21.3	45.3	NA	-
qND1.2	126.0	57.6	5.2	4.31	-0.22	123-131	57.6	59.4	NA	Zhang,2015
qND3.1	115.0	62.5	4.9	3.98	-0.22	108-121	59.3	64.9	NA	-
qND6.1	83.0	56.6	4.6	3.76	-0.21	48-93	48.9	59.0	NA	Zhang,2015
qND8.1	80.0	51.8	13.0	11.15	-0.36	78-84	50.2	52.4	NA	-
qND10.1	14.0	2.8	4.1	3.35	0.20	7-39	2.1	7.3	NA	Zhang,2015
Days to flowering (**FL**)										
qFL1.1	70.0	23.1	9.6	6.37	1.56	68-73	22.0	26.2	-	-
qFL1.2	128.0	58.2	8.1	5.35	-1.13	120-130	55.7	59.4	Ritter,2008; Yang,2014	-
qFL3.1	116.6	62.5	8.9	5.88	-1.15	108-118	59.3	63.4	Shiringani,2010	Zhang, 2015
qFL4.1	151.0	64.5	5.4	3.51	0.92	148-156	63.7	65.8	-	-
qFL6.1	81.0	56.0	4.2	2.71	-0.78	77-85	55.5	57.9	Shiringani,2010	-
qFL8.1	31.0	4.3	4.4	2.86	-0.79	27-36	2.7	6.2	-	-
qFL8.2	84.6	52.4	16.1	11.16	-1.70	83.3-86	52.3	52.8	Brown,2006, Kong 2013	-
qFL9.1	122.4	59.3	11.4	7.69	1.37	120-122.4	56.1	59.5	Hart, 2001; Lin, 1995; Kong, 2013	Zhang, 2015
qFL10.1	91.0	55.9	5.5	3.60	0.89	87-95.2	55.2	56.5	Hart,2001;	-
Full model				**46.33**						

[1] A positive additive effect indicates that alleles from *S. bicolor* IS3620C increase the trait

Table 2.5 Syntenic block containing QTLs in the *S. bicolor* BTx623× IS3620C RILs

Block				Syntenic block			
Chr	Start	End	QTL	Chr	Start	End	QTL
1	4,138,373	4,814,717	qFTR1.1	1	6,747,755	7,335,602	qFTR1.1
1	3,320,514	4,555,286	qFTR1.1	3	3,674,155	4,726,885	qPH3.1,qBTF3.1
1	18,939,586	22,223,201	qND1.1,qFL1.1	6	51,031,041	52,399,134	qND6.1
1	5,863,915	6,401,188	qFTR1.1	8	49,781,451	51,087,643	qPH8.1,qND6.1
1	55,589,953	56,460,161	qFL1.2	10	5,327,229	6,414,985	qND10.1
2	62,172,404	63,725,249	qPH2.1,qFTR3.1	4	58,971,695	60,034,016	qBTF4.1
2	65,268,178	65,446,128	qPH2.1	6	54,707,353	54,916,820	qND6.1
3	61,582,073	61,947,654	qFTR3.1,qND3.1,qFL3.1	3	62,453,853	62,923,059	qFTR3.1,qND3.1,qFL3.1
3	58,614,522	59,299,720	qFTR3.1	3	62,654,858	63,400,537	qFTR3.1,qND3.1,qFL3.1
3	2,976,412	4,063,507	qPH3.1,qBTF3.1	3	61,025,693	61,509,214	qFTR3.1,qND3.1,qFL3.1
3	64,491,930	65,019,090	qND3.1	6	57,729,502	58,436,727	qND6.1,qFL6.1
3	53,773,334	69,736,676	qFTR3.1,qND3.1,qFL3.1	9	49,537,193	58,851,665	qPH9.1,qFL9.1
3	1,573,379	3,383,712	qPH3.1	10	45,831,374	50,640,112	qPH10.1
3	3,967,608	4,602,545	qPH3.1	10	51,456,066	52,370,000	qPH10.1
4	60,670,676	63,663,336	qBTF4.1	6	52,577,923	56,419,085	qND6.1,qFL6.1
4	63,724,547	65,721,547	qFL4.1	7	55,662,744	58,222,839	qPH7.1,qBTF7.1,qFTR7.1
4	56,891,502	60,846,142	qBTF4.1	10	7,362,992	34,318,417	qND10.1
4	63,433,491	66,326,642	qFL4.1	10	3,698,395	7,399,451	qFTR10.1,qND10.1
5	3,313,848	4,261,872	qBTF5.1	6	42,609,247	44,723,752	qPH6.1,qBTF6.1
5	3,002,284	9,567,784	qBTF5.1	8	2,928,748	6,852,099	qFTR8.1,qFL8.1
5	927,332	2,974,384	qBTF5.1	8	892,171	2,887,041	qFTR8.1,qFL8.1
5	56,030	2,561,726	qBTF5.1	9	52,374,244	53,685,564	qPH9.1
6	42,609,247	44,850,780	qPH6.1,qBTF6.1	8	3,243,188	4,337,088	qFTR8.1,qFL8.1
6	57,454,242	57,849,435	qND6.1	8	52,354,729	52,809,711	qND8.1,qFL8.1
6	57,354,553	58,493,055	qND6.1,qFL6.1	10	56,445,974	57,858,406	qFL10.1
6	56,444,847	57,021,984	qND6.1,qFL6.1	10	55,667,130	56,035,826	qFL10.1
7	55,569,988	58,222,839	qPH7.1,qBTF7.1,qFTR7.1	10	5,013,735	7,346,097	qFTR10.1,qND10.1
8	1,173,574	2,859,840	qFTR7.1,qFL8.1	9	52,819,933	53,685,564	qPH9.1
9	57,346,672	57,494,840	qFL9.1	9	57,814,499	58,269,454	qFL9.1
9	54,677,505	55,546,184	qPH9.1	10	47,541,173	49,960,942	qPH10.1

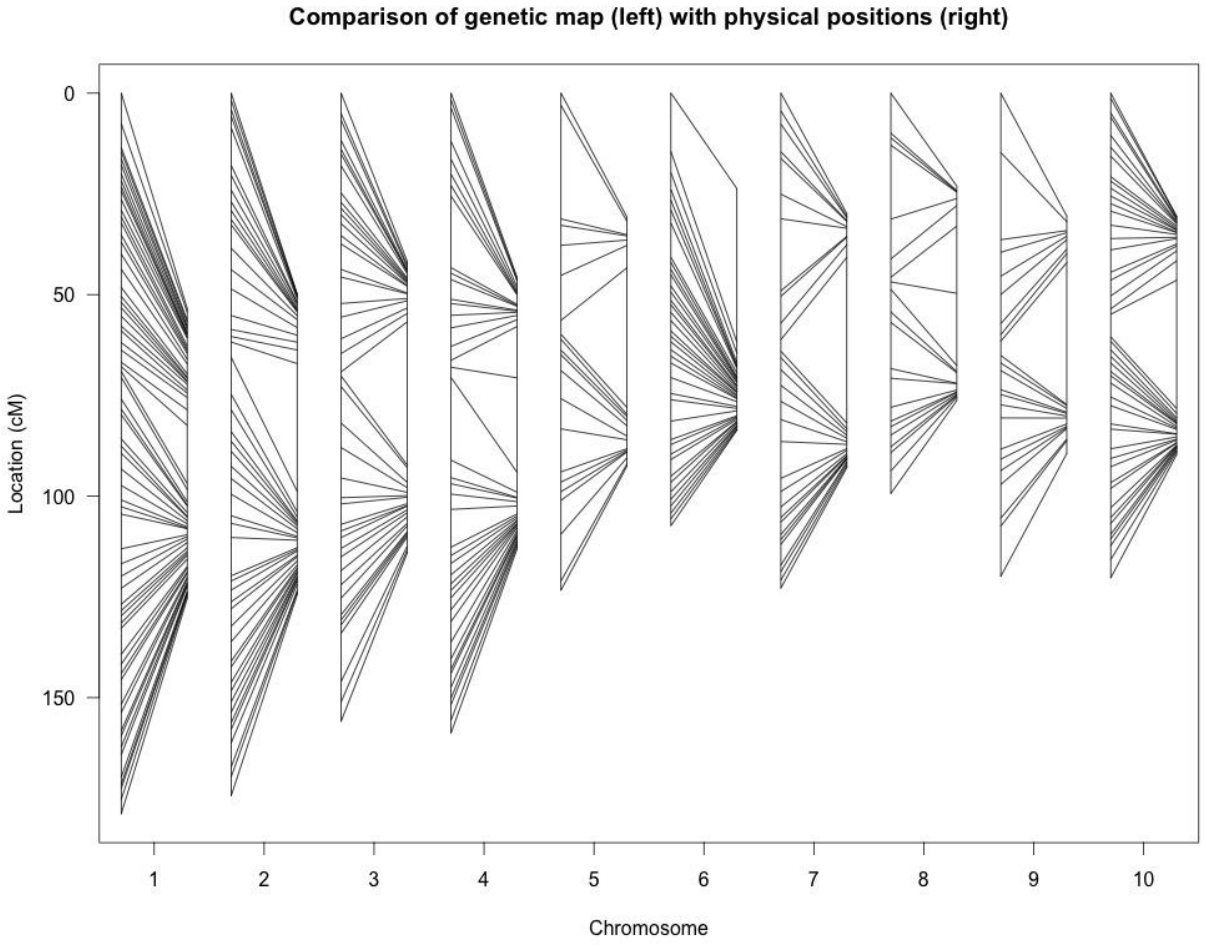


Figure 2.1 Comparisons between genetic map (left) with physical map (right)

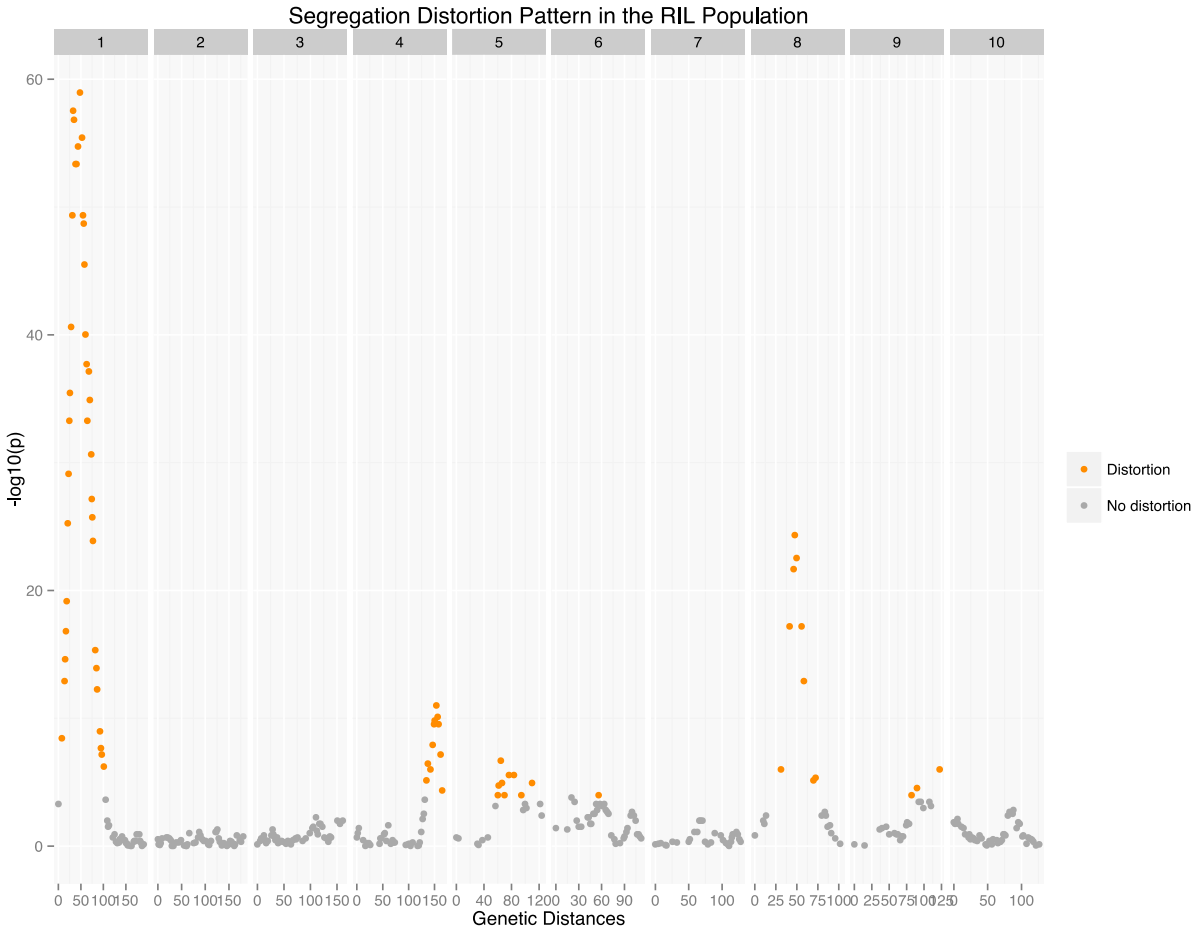


Figure 2.2 Segregation distortion pattern in the *S. bicolor* BTx623 × IS3620C RIL population. y-axis is negative logarithm p value with a base of 10, x-axis is the genetic distance. Distorted regions are plotted in orange.

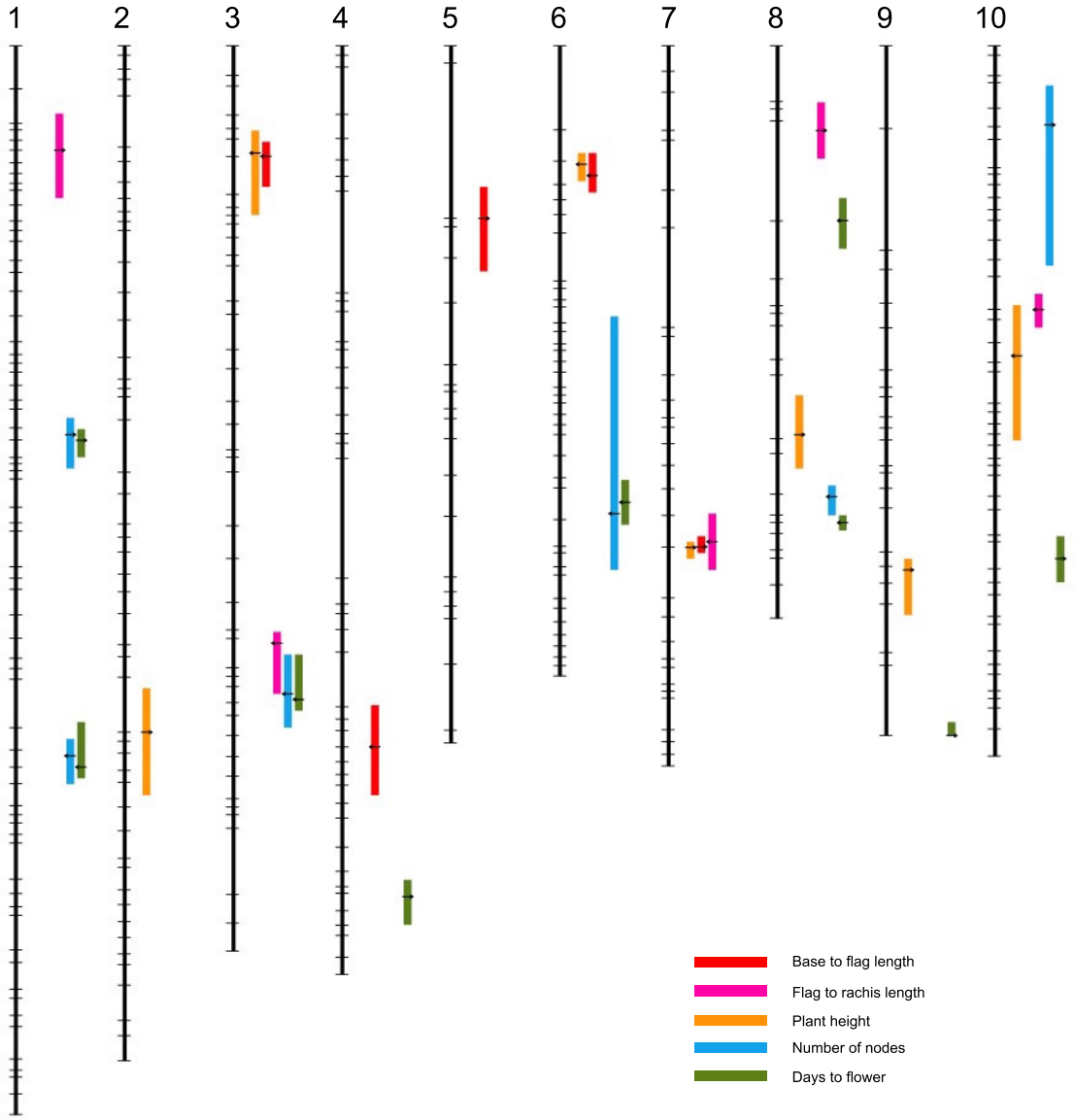


Figure 2.3 QTLs of five traits of interest, base to flag length (red), flag to rachis length (magenta), plant height (orange), number of nodes (blue) and days to flower (green) on ten sorghum chromosomes.

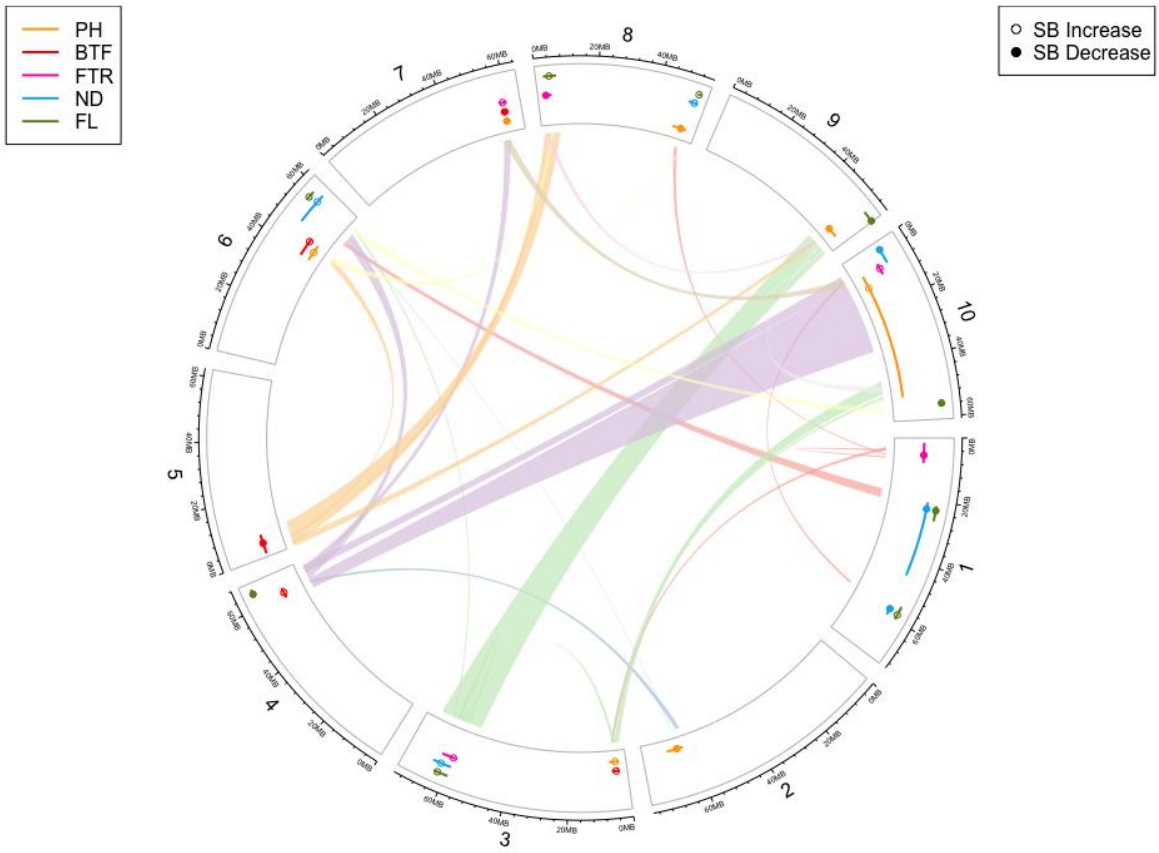


Figure 2.4 Syntenic relationship of QTLs in the *S. bicolor* BTx623 × IS3620C population

CHAPTER 3

TRANSMISSION GENETICS OF A *SORGHUM BICOLOR* × *S. HALEPENSE* BACKCROSS POPULATION[2]

[2] Kong, WQ., Nabukalu, P., Robertson, J., Goff, V., Pierce, G., Lemke, C., Compton, R., Cox, S. Paterson, A.H. To be submitted to Theoretical and Applied Genetics.

Abstract

The spread across six continents of *S. halepense* ('Johnsongrass'), a plant with the unusual distinction of being both a noxious weed and an invasive species, has been accompanied and perhaps aided by ongoing gene exchange with closely related sorghum (*S. bicolor*), also conferring risks of 'escape' of sorghum genes that could make *S. halepense* more difficult to control. We describe the transmission genetics of two BC_1F_1 populations derived from crosses of *Sorghum bicolor* and *S. halepense*, resembling those expected to occur naturally near sorghum fields. Genetic maps of these populations with totals of 726 and 799 single nucleotide polymorphism (SNP) markers span 38 and 36 linkage groups corresponding to the ten basic sorghum chromosomes. Both *S. halepense* and *S. bicolor* enriched linkage groups were consistently found for each sorghum chromosome, though the latter is unexpected. The number of linkage groups per chromosome segregating with the average ratio of one expected for singleton polymorphisms varies from 2-6, possibly due to fragmented pieces covering different portions of the chromosome or independent segregation from different *S. halepense* homologs. Segregation distortion favored *S. halepense* alleles on chromosomes 2 (1.06-4.68 Mb), 7 (1.20-6.16 Mb), 8 (1.81-5.33 Mb), and 9 (47.5-50.1 Mb); and *S. bicolor* alleles on chromosome 6 (0-40Mb). The segregation-distorted region on chromosome 2 is near a fertility restoration gene (*RF2*); on chromosome 8 might be associated with gene conversion; and on chromosome 6 is heterochromatic and may reject *S. halepense* alleles. Its SNP profile suggests that *S. halepense* has experienced extensive homogenization of chromosome segments from its respective progenitors, with polymorphic SNP alleles coming 36.64 % from *S. bicolor*, 15.35% from *S. propinquum*, 9.22% remaining heterozygous and 38.61% comprising new mutations. This study clarifies the evolutionary genetics of *S. halepense* and its genome-wide transmission genetics

with cultivated sorghum, suggesting regions of the genome with minimal ‘crop-to-weed’ gene-flow, and providing genetic maps and DNA markers suitable for further marker-trait association analysis.

Introduction

Sorghum halepense (Johnsongrass), a plant with the unusual distinction of being both a noxious weed and an invasive species (Quinn et al. 2013), has adapted to a variety of habitats and is widespread on six continents. The invasiveness of *S. halepense* is mainly owing to effective propagation by rapid flowering and disarticulation of mature inflorescences, while simultaneously developing underground rhizomes that can account for up to 70% of an individual plant's dry weight (Oyer et al. 1959), store nutrients and quickly produce new vegetative growth after quiescent periods (cold or drought). Its ability to cross with cultivated sorghum (*S. bicolor)* makes *S. halepense* a paradigm for the dangers of crop 'gene escape' (Morrell et al. 2005; Ellstrand 2001; Dale 1992). To date, no herbicide has been found to eradicate *S. halepense* without damaging sorghum – moreover, at least 24 herbicide-resistant *S. halepense* biotypes (Heap 2012) are known.

A tetraploid (2n=4x=40), *S. halepense* is thought to have formed by naturally occurring hybridization between annual *S. bicolor* (2n=2x=20) and perennial *Sorghum propinquum* (2n=2x=20) (Paterson et al. 1995), diploids thought to have last shared common ancestry 1-2 million years ago (Feltus et al. 2004). Although *S. bicolor* and *S. halepense* differ in ploidy, *S. bicolor* can serve as the pollen parent of triploid or tetraploid hybrids (Hoang-Tang and Liang 1988; Warwick and Black 1983). Gene 'escape' from *S. bicolor* to *S. halepense* raises concerns about the potential to increase persistence and/or spread of this weedy and invasive plant (Arriola and Ellstrand 1996; Tesso et al. 2008). Indeed, the 'Johnsongrass' of North America has been extensively affected by introgression from *S. bicolor* (Morrell et al. 2005) like *S. almum*, commonly known as Columbus Grass (Warwick et al. 1984). Introgression from *S. bicolor* to *S. halepense* has persisted in non-random regions of the genome, associated with seed size,

rhizomatousness, and levels of lutein, an antioxidant implicated in cold tolerance (Paterson et al. 2017).

Increased knowledge of *S. halepense* offers numerous opportunities to increase the productivity and sustainability of agriculture, while also improving management of burgeoning formerly agricultural areas of the US that are ripe for colonization by weedy/invasive species due to diminishing production of cotton and tobacco and the shift of pulp and paper production to South America. Understanding growth and development of *S. halepense*, especially its rhizomes, might lead to strategies for effective control of rhizomatous weeds even in closely related crops, such as Johnsongrass in sorghum fields. From a different perspective, however, *S. halepense* harbors many characteristics that may increase agricultural productivity if transferred to sorghum (Sangduen and Hanna 1984). It flowers and produces seeds rapidly, is resistant to many diseases and insects, and adapts to a wider range of environments than both of its progenitors. *S. halepense* might also contribute to breeding of genotypes suitable for multiple harvests from single plantings, either by ratooning or perenniality, a topic that has recently regained momentum (Glover et al. 2010; Cox et al. 2002; Paterson et al. 2013).

Genetic analysis is a powerful means to investigate differences in growth and development among genotypes, however the polyploidy level of *S. halepense* imposes complications. Genotyping by sequencing (GBS) employs reduced representation DNA libraries to detect large numbers of single nucleotide polymorphisms (SNP) that comprise allelic differences between individuals. GBS is a powerful method for genetic mapping (Elshire et al. 2011), study of population diversity (Lu et al. 2013), genome selection (Poland et al. 2012a) and ecology and conservation genomics (Narum et al. 2013). However, GBS in polyploids is complicated by its accuracy, missing data and inference of allele dosages (Kim et al. 2015),

although it has been applied in allopolyploids such as switchgrass (Lu et al. 2013) and cotton (Logan-Young et al. 2015) and autotetraploids such as alfalfa (Li et al. 2014) and potato (Hackett et al. 2013). A recent study in autotetraploid potato suggested that to accurately infer allele dosage, i.e. to differentiate among simplex, duplex and triplex markers, a minimum depth of 60~80 mapped reads is needed (Uitdewilligen et al. 2013).

Here, we report genetic maps of two BC_1F_1 populations derived from different F_1 tetraploid progenies from a cross of *S. bicolor* BTx623 (recurrent parent) × *S. halepense* (Gypsum 9E), and reveal chromosomal characteristics and segregation patterns using GBS. In comparison to its progenitors *S. bicolor* and *S. propinquum*, the chromosomal composition of *S. halepense* sheds light on its evolution. Patterns of transmission of alleles from *S. bicolor* and *S. halepense* to interspecific progenies provide evidence of genomic regions that may respectively be favorable or recalcitrant to gene flow between these species. This information identifies potential locations for transgenes or other genetic modifications ('edited' alleles) that may minimize crop-to-weed gene flow. These two populations are also of potential agronomic importance: identifying and incorporating novel alleles conferring yield potential, nitrogen fixation, insect or disease resistance and rhizomatousness may benefit current or future sorghum breeding programs.

Materials and Methods

Genetic Stocks

Two tetraploid F_1 hybrids (named H4 and H6) derived from crossing *S. bicolor* BTx623 (with unreduced gametes) × *S. halepense* (G9E) were backcrossed to the tetraploid recurrent parent, *S. bicolor* BTx623: two BC_1F_1 mapping populations, of 146 H4-derived and 108 H6-derived individuals respectively, were developed. BF_1F_2 rows derived from selfed seed of a

single BC_1F_1 plant were planted at the University of Georgia Plant Science Farm, Watkinsville, GA, USA on May 28th 2013 and May 9th 2014, and at The Land Institute, Salina, KS, USA on Jun 3rd 2013, and Jun 17th 2014.

Genotyping by Sequencing

Leaf samples of the BC_1F_1 individuals were frozen at -80 C and lyophilized for 48 hours. Genomic DNA was extracted from the lyophilized leaf sample based on Aljanabi et al. (1999).

The GBS platform used a slightly modified version of Multiplex Shotgun Genotyping (MSG) (Andolfatto et al. 2011) combined with the Tassel GBS5 v2 analysis pipeline (Glaubitz et al. 2014a). Sequencing used an Illumina HiSeq 2500, Rapid V2 kit that generated about 150 million reads of 100 base pair fragments per run with single-end sequencing.

Genotype calling and filtering

Genotypes were determined by single nucleotide polymorphism (SNP) 'calling' based on the reference genome of *S. bicolor* BTx623 *v1.4* (Paterson et al. 2009). Using Tassel-GBS 5 (Glaubitz et al. 2014b), the first 90 base pairs (bp) of each read were mapped onto the reference genome. SNPs were 'called' based on alignment of the reads to the reference genome. An in-house pipeline was used to determine the genotypes for these two populations, as follows:

1. Raw SNPs was firstly thinned out within 100 base pairs, since SNP sites close to each other or on the same read provide little non-redundant information in early generations following crossing.
2. Biallelic SNP markers with an average depth of 10 were selected.
3. The PL (phred-scaled genotype likelihoods) field from the raw VCF file consisted of three floating point log10-scaled likelihoods for AA, AB and BB genotypes where A is the reference allele and B is the alternative allele (Danecek et al. 2011). Genotype calling

used the field with the minimum PL value from the VCF file with some minor adjustments. The PL field was firstly transformed into probability scales by $10^{(-PL/10)}$. A missing genotype was assigned if the second largest probability of a genotype is greater than 0.05 for each individual at each locus.

4. Homozygous genotypes with lower than 6x coverage were considered missing data.
5. Heterozygous genotypes were called if the depth of the alternative alleles from the Allele Depth (AD) fields of the VCF file was greater than 2.

Map construction

For each sorghum chromosome, we clustered markers based on a minimum LOD score of 10. Genetic distances were first estimated based on the physical orders of markers in the published sorghum genome (Paterson et al. 2009), then markers within 1cM bins were combined. Bin genotypes were defined as follows: If there was only one marker in the bin, the bin genotype would be the same as the marker genotype; if there were more than one marker in the bin, bin genotypes would be determined by merging marker genotypes to minimize missing data points. Using the combined genotype file, a *de novo* marker ordering was implemented for each corresponding sorghum chromosome and the final genetic map was constructed using R/qtl with the Kosambi mapping function. The map distance was calculated with an error probability of 0.01 (Broman et al. 2003).

Analysis of segregation

A total of 2240 polymorphic markers were obtained after the *genotyping and filtering* steps described above for both H4 and H6-derived populations, and used to analyze patterns of segregation distortion. Using the R program (R Core Team 2013), a chi-squared test was applied to each marker to test the hypothesis that it deviated significantly from a ratio of 5:1.

Whole genome polymorphism analysis

A total of four genotypes, *S. bicolor* IS3620C (SRX2158431), *S. propinquum* from Univ. GA (SRX030701 and SRX030703), *S. propinquum* from Australia (SRX208587 and SRX208588), and *S. halepense* (SRX142088), were included in whole-genome SNP analysis against the *S. bicolor* BTx623 v1.4 reference genome. The Burrows-Wheeler Aligner (BWA) MEM algorithm was used for read alignment (Li and Durbin 2009). Variant calling used samtools/Bcftools (Li 2011). Data were filtered with a minimum phred score of Q20, and a minimum depth of 10 with a maximum missing data of 30% for each SNP locus.

Results

Genotyping by sequencing

A total of 717.5 million reads were obtained from the Illumina Hiseq 2500 platform. The first 96 individuals were sequenced twice with 179.9 million and 180.8 million reads obtained, respectively. A total of 181.2 million and 175.6 million reads were obtained for the second and third plates, respectively, each including 96 individuals. A total of 689,684 raw SNP markers were initially obtained from the Tassel GBS 5 pipeline. After thinning SNP markers within 100bp, a total of 215, 341 SNP markers remained. Of the 254 genotyped individuals, 8 were deleted due to very low sequence coverage. A total of 141 of the 246 sequenced individuals were derived from the H4 population, and 105 from the H6. After filtering steps (see Materials and Methods), we selected the same 2,240 polymorphic markers for H4 and H6-derived populations with a minimum average depth of 10 at each locus for genetic mapping.

Linkage maps and patterns of segregation

An autotetraploid species can segregate in a variety of manners, including random chromosome segregation, random chromatid segregation, and maximum equational segregation,

and can be further complicated by degrees of double reduction (Gupta 2007). Random chromosome segregation assumes that a gene is so close to the centromere that crossing over between the gene and centromere never occurs. The other extreme is maximum equational segregation where crossing over always occurs between the gene and the centromere. The intermediate state between random chromosome segregation and maximum equational segregation is often observed more frequently than the two extremes (Gupta 2007). With random chromosome segregation (Muller 1914), the expected segregation ratios for these two populations is 1:1 (heterozygotes:homozygotes) for simplex markers and 5:1 for duplex markers (Figure 3.1). With random chromatid segregation, where a chromatid can end up with any chromatid in a gamete with equal frequency, the segregation ratio can be 13:15 for simplex markers, and 11:3 for duplex markers (Haldane 1930). With maximum equational segregation (Mather 1935), the segregation ratio can be 11:13 for simplex markers and 7:2 for duplex markers.

A histogram of ratios of heterozygotes to homozygotes with all mapped markers after square root transformation (Figure 3.2) suggests that the distribution of segregation ratios in the two BC_1F_1 mapping populations is continuous. In other words, there are many markers with intermediate segregation ratios, leading to a mixture of disomic and polysomic inheritance, which is expected and observed in other studies of tetraploids (Jannoo et al. 2004; Stift et al. 2008).

We grouped all 2240 selected SNP markers based on pair-wise recombination fractions using relatively stringent thresholds in R/qtl to reveal appropriate numbers of linkage groups for each sorghum chromosome. We mapped a total of 726 and 799 SNP markers covering a total of 38 and 36 linkage groups, spanning 3902.4 and 6075.1 cM for the H4 and H6-derived

populations, respectively (Table 3.1Table 3.2). For individual sorghum chromosomes, we obtained 2-6 homologous groups versus a maximum of five expected, possibly due to complexity of segregation patterns, lack of statistical power due to relatively low numbers of individuals in each population, and fragmented linkage groups. Markers are more concentrated on the distal than the central regions of each chromosome, possibly due to use of a methylation sensitive restriction enzyme and higher sequence complexity in terminal regions.

Transmission genetics of each chromosome

We define linkage groups as either *S. halepense* - or *S. bicolor* enriched based on statistically significant deviation from the expected segregation ratio of 1:1 for the average of all markers in the group. Linkage groups were defined as *S. halepense* enriched if the segregation ratio is greater than 1.82 or *S. bicolor* enriched if it is smaller than 0.55 (calculated based on 105 individuals from the H6–derived population with a Chi-squared test with 1 degree of freedom of 1 and an alpha value of 0.001).

For chromosome 1, the total of six H4-derived linkage groups had average segregation ratios of 3.84 (1A), 1.72 (1B), 1.73 (1C), 1.64 (1D), 0.92 (1E) and 0.28 (1F); and four H6-derived groups had 3.82 (1A), 2.72 (1B), 1.48 (1C) and 0.21 (1D), Linkage groups 1B, 1C and 1D segregated with similar ratios and covered different parts of chromosome 1. The two *S. halepense* enriched groups (1A for both populations) and the two *S. bicolor* enriched groups (1F for H4, 1D for H6) both covered nearly all of chromosome 1. No groups with average segregation ratios not significantly different from 1:1 were found covering the short arm of chromosome 1 for the H6-derived population (Table 3.1Table 3.2, Figure 3.4).

For chromosome 2, the total of four H4-derived linkage groups had average segregation ratios of 3.80 (2A), 2.54 (2B), 1.23 (2C) and 0.33 (2D); and three H6-derived groups had 5.67

(2A), 1.63 (2B) and 0.22 (2C). *S. halepense* enriched linkage groups 2A and 2B in the H4-derived population largely covered euchromatic regions of chromosome 2, but absence of markers from possible pericentromeric regions (40-60Mb) makes it difficult to coalesce these two linkage groups (Figure 3.4 Physical coverage of each *S. bicolor* BTx623 × *S. halepense* G9E linkage group to the sorghum genome. The H4 population is on the left of the black line and the H6 population is on the right. The x-axis is the segregation ratio after the square root transformation and that of the H4 population (left) is assigned a negative sign.). The two *S. halepense* enriched groups (2A and 2B) in H4 and one (2A) in the H6- derived population covered most of chromosome 2, and the two *S. bicolor* enriched groups (2D from H4, 2C from H6) also covered most of chromosome 2. Linkage groups segregating with ratios of 1 (2B and 2C in H4, 2B in H6) are only concentrated on the long arms of chromosome 2, indicating segregation distortion and possible enrichment of *S. halepense* alleles on the short arms.

For chromosome 3, the total of four H4-derived linkage groups had average segregation ratios of 3.51 (3A), 1.52 (3B), 1.02 (3C) and 0.45 (3D); and five H6-derived groups had 3.93 (3A), 3.39 (3B), 1.04 (3C), 0.95 (3D) and 0.26 (3E). The one *S. halepense* enriched group (3A) from the H4-derived population and two (3A and 3B) from the H6-derived population, and the two *S. bicolor* enriched groups (3D from H4 and 3E from H6) both covered most of chromosome 3 with markers concentrated in euchromatic regions. Linkage groups 3C and 3D from the H6-derived population are more concentrated in the central part of chromosome 3 compared to *S. bicolor* and *S. halepense* enriched groups.

For chromosome 4, the total of four H4-derived linkage groups had average segregation ratios of 3.51 (4A), 1.52 (4B), 1.02 (4C) and 0.45 (4D); and four H6-derived groups had 4.33 (4A), 2.64 (4B), 0.58 (4C) and 0.27 (4D). Linkage groups 4B and 4C, each covering different

portions of chromosome 4 with similar segregation ratios, might come from the same homologous group. Two *S. halepense* enriched groups (4A for both H4 and H6) and two *S. bicolor* enriched groups (4D for both H4 and H6) largely covered chromosome 4 with markers concentrated in euchromatic regions. No chromosome 4 linkage groups segregating with an average ratio not significantly different from 1:1 were found in the H6-derived population.

For chromosome 5, the total of three H4-derived linkage groups had average segregation ratios of 2.17 (5A), 0.89 (5B) and 0.39 (5C); and three H6-derived groups had 6.56 (5A), 0.70 (5B) and 0.19 (5C). The one *S. halepense* enriched group (5A) in the H4-derived population only covers a small portion of the short arm of chromosome 5, while the one in the H6-derived population covers the long arm (Figure 3.4). The two *S. bicolor* enriched groups (5C for both H4 and H6) cover most of chromosome 5 with markers mostly in the euchromatic regions. Linkage groups segregating with average ratio of 1 were more concentrated in the middle of the chromosome compared to *S. bicolor* and *S. halepense* enriched groups.

For chromosome 6, the total of three H4-derived linkage groups had average segregation ratios of 2.91 (6A), 1.70 (6B), and 0.42 (5C); and five H6-derived groups had 1.91 (6A), 1.24 (6B), 1.24 (6C), 0.91 (6D) and 0.37 (6E). Linkage groups 6B, 6C and 6D from the H6-derived population, each covering a separate portion of chromosome 6 with similar segregation ratios, might come from the same homologous chromosome. The *S. halepense* enriched groups (6A) only cover 50-60Mb of sorghum chromosome 6 and were only segregating with ratios of 2.91 and 1.9 for the H4 and H6-derived populations respectively, lower than the other linkage groups. An *S. bicolor* enriched group in the H4 population (6C) covered 40-60Mb, while a *S. bicolor* enriched group in the H6 population (6E) covered all of chromosome 6. Linkage groups

segregating with ratios not significantly different from 1 covered the entire physical chromosome.

For chromosome 7, the total of four H4-derived linkage groups had average segregation ratios of 3.34 (7A), 1.15 (7B), 0.97 (7C) and 0.39 (7D), and four H6-derived groups had 5.63 (7A), 1.66 (7B), 1.48 (7C) and 0.47 (7D). Linkage groups 7B and 7C of the H6 population had similar segregation ratios and covered different chromosomal regions, indicating that they might come from the same homologous chromosome. The two *S. halepense* enriched groups (7A for both population) and the two *S. bicolor* enriched groups (7D for both populations) were largely concentrated in the euchromatic regions of sorghum chromosome 7. Linkage groups segregating with average ratios not significantly different from 1 cover most of chromosome 7.

For chromosome 8, the total of four H4-derived linkage groups had average segregation ratios of 3.78 (8A), 1.48 (8B), 1.18 (8C) and 0.27 (8D), and two H6-derived groups had 5.32 (8A) and 0.61 (8B). Both *S. halepense* enriched groups (8A in both populations) and *S. bicolor* enriched groups (8D in H4, 8C in H6) covered most of chromosome 8. Interestingly, linkage groups segregating with average ratios not significantly different from 1 were absent from the H6-derived population and that from the H4 population only covered a small distal portion of chromosome 8.

For chromosome 9, the total of two H4-derived linkage groups had average segregation ratios of 4.40 (9A) and 0.36 (9B); and three H6-derived groups had 4.82 (9A), 1.32 (9B) and 0.27 (9C). The two *S. halepense* enriched groups (9A in both populations) covered chromosome 9 with more markers on its long arm, while two *S. bicolor* enriched groups largely covered chromosome 9 with most markers on the euchromatic regions. No linkage group segregated

with average ratio not significantly different from 1 in the H4-derived population while in H6, the only one (9B) was concentrated on the long arm of chromosome 9.

For chromosome 10, the total of four H4-derived linkage groups had average segregation ratios of 3.83 (10A), 1.38 (10B), 0.94 (10C) and 0.32 (10D); and three H6-derived groups had 3.14 (10A), 1.45 (10B) and 0.33 (10C). Both *S. halepense* enriched groups (10A in both populations) and *S. bicolor* enriched groups (10D and 10C in H4 and H6-derived populations, respectively) covered chromosome 10 with markers mostly in the euchromatic regions. The marker distribution patterns of linkage groups segregating with an average ratio not significantly different from 1 is similar to that of *S. halepense* and *S. bicolor* enriched groups.

Segregation distortion

We detected a total of 53 and 79 SNP markers enriched in *S. halepense* alleles in the H4 and H6-derived populations from the total 2240 filtered markers, respectively, with heterozygote versus homozygote ratios significantly higher than 5:1 ($P<0.05$, $df=1$). A more compelling case can be made for segregation distortion based on 22 markers that are found significant at an alpha level of 0.05 for both populations. Finally, 57 markers are found significant from pooling the result of two populations (Table 3.3 and Figure 3.5). Regions on chromosomes 2 (1.06-4.68 Mb), 7 (1.20-6.16 Mb), 8 (1.81-5.33 Mb), and 9 (47.5-50.1 Mb) harbored at least three markers showing significant segregation distortion in each population. Interestingly, those regions also lack markers segregating with ratios of 1:1, indicating aberrant transmission affected by selection or illegitimate recombination, hypotheses that warrant further investigation.

Genomic composition of S. halepense

Noting that *S. halepense* is thought to be a naturally occurring polyploidy derived from crossing between *S. bicolor* and *S. propinquum*, to investigate its genomic composition, we

performed SNP 'calling' with four genotypes: *S. bicolor* IS3620C, a race 'guinea' accession that is highly diverged from BTx623 as a control; two *S. propinquum* accessions; and *S. halepense*. After filtering (see Materials and Methods), we obtained a total of 8,703,936 SNP markers genome-wide with a total of 1,777,782 (36.64%) marker genotypes identical to BTx623, 744,924 (15.35%) identical to *S. propinquum*, 447,479 (9.22%) heterozygous loci with one allele each from *S. bicolor* and *S. propinquum* (Table 3.4), 478,804 (38.61%) non-progenitor alleles, presumably arising from new mutation, and 3,852,257 unknown alleles due to missing data or polymorphism between *S. bicolor*. A much smaller sample with SNP markers only from the genetic maps were also categorized into groups matching *S. bicolor, S. propinquum* alleles and new mutations. For all non-polymorphic loci between BTx623 and IS3620C with mapped SNP markers, we found a total of 36.72 % and 42.41% of the *S. halepense* loci retaining the *S. propinquum* alleles, while 52.48% and 46.56% of the alleles are novel in the H4 and H6 -derived populations, respectively (Table 3.4).

The overall genomic composition reflects that the distribution of *S. halepense* alleles derived from *S. bicolor* and *S. propinquum* shows a largely random pattern— indicating that chromosomes from *S. propinquum* have gone through many recombination events with *S. bicolor*-derived chromosomes. Specifically, there are only 26 (16 in H4, 19 in H6, and 9 in both) non-random 'runs' of consecutive mapped loci with *S. propinquum* derived alleles, covering roughly 18.7% (H4) and 11.3% (H6) of the genome (Table 3.5).

Discussion

Genetic maps of two novel BC_1F_1 populations derived from crossing of *S. bicolor* BTx623 and *S. halepense* G9E provide important new information about the genome-wide transmission genetics of crosses resembling those expected to occur naturally near sorghum

fields, which may have aided and abetted the spread across six continents of *S. halepense* ('Johnsongrass'), and confer risks to 'escape' of sorghum genes that could make *S. halepense* more difficult to control. Identification of DNA markers and construction of genetic maps will facilitate marker-trait association analysis and comparative studies with other sorghum populations. Revealing chromosomal characteristics, especially identifying non-random patterns of DNA marker distribution, provides information about underlying features of sorghum genome organization. A SNP profile sampling the breadth of the Sorghum genus illustrates the evolutionary path and fate of alleles from progenitors of *S. halepense*.

GBS and Genetic mapping in polyploids

We applied genotyping by sequencing (GBS) in the two BC_1F_1 populations, which is a cost- and time- efficient method of finding SNP markers (Elshire et al. 2011; Poland et al. 2012b). In this study, our coverage of each locus is not high enough to differentiate heterozygous genotypes with different dosages, which might lower the statistical power of linkage mapping and subsequent QTL analysis. Despite our inability to differentiate heterozygous genotypes with different dosages, we obtained adequate numbers of SNP markers to construct linkage maps in these two populations, and the remnant unmapped markers might nonetheless be useful in analysis of marker-trait association.

Genetic maps of two BC_1F_1 populations derived from *S. bicolor* and *S. halepense*, with totals of 726 and 799 SNP markers span 38 and 36 linkage groups formed with assistance from the sorghum genome sequence (Paterson et al. 2009). Marker distribution patterns of the H4 and H6-derived populations are generally similar (Figure 3.4), meaning that chromosomal characteristics and re-arrangements may not be biased by constructing two different populations. For each of the basic sorghum chromosomes, we expect to find one linkage group segregating

with a ratio of 5:1 (heterozygotes: homozygotes) derived from homozygous *S. halepense* loci with genotypes different from *S. bicolor*, and varying numbers of linkage groups segregating with ratios of 1:1, possibly from different *S. halepense* homologous chromosomes. We consistently found at least one linkage group for each sorghum chromosome enriched with *S. halepense* alleles, segregating with ratios of approximately 3-5 (heterozygotes: homozygotes), as expected (Figure 3.1). We found 1-3 linkage groups segregating with average ratios near 1 for most chromosomes, which may be due to either fragmented pieces covering different portion of the chromosome or independent segregation from different homologous *S. halepense* chromosomes.

An unexpected discovery is linkage groups segregating with average ratios of 0.2-0.33 (heterozygotes:homozygotes) for every sorghum chromosome. In principle, markers segregating with ratios of approximately 5:1 and 1:5 might be in the repulsion phase. To test this hypothesis, we reversed the genotyping of groups segregating with patterns of 1:5 and tried merging and ordering them together with groups segregating near 5:1. Such pairs of linkage groups either failed to coalesce or were loosely connected to each other with relatively large genetic distances (data not shown). Therefore, linkage groups segregating with average ratios of approximately 5:1 and 1:5 appear not to be in the repulsion state, although it remains possible that the sample sizes of the two populations are not large enough to detect the linkage (Wu et al. 1992).

Chromosomes 5, 6, 8 and 9 are of particular interest, in that we fail to find linkage groups for certain ratios or markers only cover parts of the corresponding sorghum chromosome, suggesting aberrant chromosomal behavior caused by factors such as selection or preferential pairing. For chromosomes 5 and 6, markers enriched with *S. halepense* alleles only cover a small portion of the chromosome. A previous study (Bowers et al. 2003) of an F2 population

from *S. bicolor* BTx623 × *S. propinquum* discovered a ribosomal DNA enriched region with *S. propinquum*-dominated loci spanning 32.3-40cM on chromosome 5, corresponding to 5-20Mb in physical distance (Zhang et al. 2013). We find few *S. halepense* alleles in this region (Figure 3.4), possibly also due to the enrichment of rDNA. Similarly, a large heterochromatin block on sorghum chromosome 6 is enriched for *S. bicolor* alleles. For chromosomes 8 and 9, we found fewer markers segregating with a ratio of 1 than for the other chromosomes. Further investigation is needed to understand these biases of marker distribution across the genome.

Segregation distortion

The overall distributions of the segregation (Figure 3.2) in H4 and H6-derived population suggest that more intermediate segregation ratios have been observed than extreme ones (chromosome segregation and maximum equational segregation), consistent with findings from other studies (Stift et al. 2008; Jannoo et al. 2004). Segregation distorted regions in these two populations are particularly interesting, possibly due to illegitimate recombination, unusual chromosomal events such as translocation and gene conversion (Wang et al. 2009), and gametic or zygotic selection. For example, chromosomal regions enriched for *S. halepense* alleles in this study might be related to fitness of the progenies, a hypothesis which might be further investigated by QTL mapping. In addition, regions enriched with *S. halepense* alleles may provide guidance as to relatively 'safe landing sites' for which a transgene— to reduce crop-to-weed gene flow from cultivated sorghum to *S. halepense* (Arriola and Ellstrand 1996). With many different segregation patterns occurring in these populations, testing for segregation distorted regions is unusually challenging. Nonetheless, we still find four general regions, on chromosomes 2 (1.06-4.68 Mb), 7 (1.20-6.16 Mb), 8 (1.81-5.33 Mb), and 9 (47.5-50.1 Mb), with

more than expected *S. halepense* alleles, and one region on chromosome 6 (0-40Mb) with fewer than expected *S. halepense* alleles (Table 3.3).

Markers displaying segregation distortion might be linked to genes affecting fitness, for example controlling fertility. To date, three sorghum genes controlling fertility have been located, *Rf1* on the long arm of chromosome 8, *Rf2* on the short arm of chromosome 2 and and *Rf5* on the short arm of chromosome 5 (Klein et al. 2005; Jordan et al. 2010a; Jordan et al. 2011). All three *RF* genes were proposed to encode pentatricopeptide repeat proteins (PPR) that are essential in the post-transcriptional process (Schmitz-Linneweber and Small 2008). The interval on chromosome 2 (1.06-4.68 Mb) enriched for *S. halepense* alleles might be associated with the *Rf2* gene, which is located within the region from 5.4-5.7 Mb (Jordan et al. 2010b). Similarly, the interval (43.98-55.35 Mb in H6) of chromosome 8 enriched for *S. halepense* alleles in the H6 population harbors the *Rf1* gene, based on flanking SSR markers Xtxp18-Xtxp250 located at an interval from 50.5 to 51.0 Mb. Although the chromosome 8 interval seems longer than expected, there have been studies in many species (Xu et al. 2009; Kato et al. 2007; Lurin et al. 2004) indicating that the *Rf1* genes encoding pentatricopeptide repeat (PPR) protein families tend to duplicate and cluster (Kato et al. 2007; Jordan et al. 2011). Segregation distortion on the short arm of chromosome 8 (1.81-5.33 Mb) also overlaps with a region that has experienced frequent gene conversion (0.94-2.8Mb), a mechanism that may cause segregation distortion (Wang et al. 2009; Wang et al. 2011).

Evolution of S. halepense

The distribution in *S. halepense* of alleles from its progenitors, *S. bicolor* and *S. propinquum*, suggests that it is an autotetraploid. The *S. halepense* chromosomes consist of largely-random distributions of *S. bicolor*-derived, *S. propinquum*-derived, and novel alleles,

which indicates extensive recombination between *S. bicolor* and *S. propinquum*-derived 'subgenomes'. It has been controversial whether *S. halepense* is an allo- or auto- tetraploid (Fernandez et al. 2013; Hoangtang and Liang 1988; Endrizzi 1957; de Wet 1978). Since progenies of *S. bicolor* × *S. propinquum* are fertile and show near-normal recombination, our previous studies (Paterson et al. 1995; Kong et al. 2013) have favored that *S. halepense* could be an auto-tetraploid. Comparing segregation patterns among two mapping populations and SNP distributions across the entire genome each support the hypothesis that *S. halepense* is an autotetraploid, with its chromosomes a mosaic of alleles from *S. bicolor, S. propinquum* and novel mutations (Table 3.4). Nevertheless, we found a total of 26 regions of the genome in either H4 or H6-derived population with nonrandom distribution of consecutive *S. propinquum* alleles in both populations (Table 3.5), including a total of eight regions on chromosomes 1 (3 regions), 3 (2), 4 (1), and 6 (2).

Acknowledgments

We appreciate the support of the USDA Biotechnology Risk Assessment Program (2012-01658 to AHP and TSC), USAID Feed The Future (AID-OAA-A-13-00044 to AHP, TSC) program, and NIFA Global Food Security CAP (2015-68004-23492 to AHP, JNB).

References

Aljanabi, S.M., L. Forget, and A. Dookun, 1999 An improved and rapid protocol for the isolation of polysaccharide- and polyphenol-free sugarcane DNA. *Plant Molecular Biology Reporter* 17 (3):281-281.

Andolfatto, P., D. Davison, D. Erezyilmaz, T.T. Hu, J. Mast *et al.*, 2011 Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Research* 21 (4):610-617.

Arriola, P.E., and N.C. Ellstrand, 1996 Crop-to-weed gene flow in the genus Sorghum (Poaceae): Spontaneous interspecific hybridization between johnsongrass, Sorghum halepense, and crop sorghum, S-bicolor. *American Journal of Botany* 83 (9):1153-1159.

Bowers, J.E., C. Abbey, S. Anderson, C. Chang, X. Draye *et al.*, 2003 A high-density genetic recombination map of sequence-tagged sites for Sorghum, as a framework for comparative structural and evolutionary genomics of tropical grains and grasses. *Genetics* 165 (1):367-386.

Broman, K.W., H. Wu, S. Sen, and G.A. Churchill, 2003 R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19 (7):889-890.

Cox, T.S., M. Bender, C. Picone, D.L. Van Tassel, J.B. Holland *et al.*, 2002 Breeding perennial grain crops. *Critical Reviews in Plant Sciences* 21 (2):59-91.

Dale, P.J., 1992 Spread of Engineered Genes to Wild Relatives. *Plant Physiology* 100 (1):13-15.

Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks *et al.*, 2011 The variant call format and VCFtools. *Bioinformatics* 27 (15):2156-2158.

de Wet, J.M.J., 1978 Systematics and Evolution of Sorghum Sect. Sorghum (Gramineae). *American Journal of Botany* 65 (4):477-484.

Ellstrand, N.C., 2001 When transgenes wander, should we worry? *Plant Physiology* 125 (4):1543-1545.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto *et al.*, 2011 A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *Plos One* 6 (5).

Endrizzi, J.E., 1957 Cytological Studies of some Species and Hybrids in the Eu-Sorghums. *Botanical Gazette* 119 (1):1-10.

Feltus, F.A., J. Wan, S.R. Schulze, J.C. Estill, N. Jiang *et al.*, 2004 An SNP resource for rice genetics and breeding based on subspecies indica and japonica genome alignments. *Genome Research* 14 (9):1812-1819.

Fernandez, L., L.A. de Haro, A.J. Distefano, M.C. Martinez, V. Lia *et al.*, 2013 Population genetics structure of glyphosate-resistant Johnsongrass (Sorghum halepense L. Pers) does not support a single origin of the resistance. *Ecology and Evolution* 3 (10):3388-3400.

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire *et al.*, 2014a TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *Plos One* 9 (2):e90346.

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire *et al.*, 2014b TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *Plos One* 9 (2).

Glover, J.D., J.P. Reganold, L.W. Bell, J. Borevitz, E.C. Brummer *et al.*, 2010 Increased Food and Ecosystem Security via Perennial Grains. *Science* 328 (5986):1638-1639.

Gupta, P.K., 2007 *Cytogenetics (2)*. Meerut, IN: Rastogi Publications.

Hackett, C.A., K. McLean, and G.J. Bryan, 2013 Linkage analysis and QTL mapping using SNP dosage data in a tetraploid potato mapping population. *Plos One* 8 (5):e63939.

Haldane, J.B.S., 1930 Theoretical genetics of autopolyploids. *Journal of Genetics* 22 (3):359-372.

Heap, I., 2012 The International Survey of Herbicide Resistant Weeds.

Hoang-Tang, and G.H. Liang, 1988 The genomic relationship between cultivated sorghum [Sorghum bicolor (L.) Moench] and Johnsongrass [S. halepense (L.) Pers.]: a re-evaluation. *Theoretical and Applied Genetics* 76 (2):277-284.

Hoangtang, and G.H. Liang, 1988 The Genomic Relationship between Cultivated Sorghum [Sorghum-Bicolor (L) Moench] and Johnsongrass [Sorghum-Halepense (L) Pers] - a Re-Evaluation. *Theoretical and Applied Genetics* 76 (2):277-284.

Jannoo, N., L. Grivet, J. David, A. D'Hont, and J.C. Glaszmann, 2004 Differential chromosome pairing affinities at meiosis in polyploid sugarcane revealed by molecular markers. *Heredity* 93 (5):460-467.

Jordan, D.R., R.R. Klein, K.G. Sakrewski, R.G. Henzell, P.E. Klein *et al.*, 2011 Mapping and characterization of Rf 5 : a new gene conditioning pollen fertility restoration in A1 and A2 cytoplasm in sorghum (Sorghum bicolor (L.) Moench). *Theoretical and Applied Genetics* 123 (3):383-396.

Jordan, D.R., E.S. Mace, R.G. Henzell, P.E. Klein, and R.R. Klein, 2010a Molecular mapping and candidate gene identification of the Rf2 gene for pollen fertility restoration in sorghum [Sorghum bicolor (L.) Moench]. *Theor Appl Genet* 120 (7):1279-1287.

Jordan, D.R., E.S. Mace, R.G. Henzell, P.E. Klein, and R.R. Klein, 2010b Molecular mapping and candidate gene identification of the Rf2 gene for pollen fertility restoration in sorghum [Sorghum bicolor (L.) Moench]. *Theoretical and Applied Genetics* 120 (7):1279-1287.

Kato, H., K. Tezuka, Y.Y. Feng, T. Kawamoto, H. Takahashi *et al.*, 2007 Structural diversity and evolution of the Rf-1 locus in the genus Oryza. *Heredity (Edinb)* 99 (5):516-524.

Kim, C., H. Guo, W. Kong, R. Chandnani, L.-S. Shuang *et al.*, 2015 Application of genotyping by sequencing technology to a variety of crop breeding programs. *Plant Science* (0).

Klein, R.R., P.E. Klein, J.E. Mullet, P. Minx, W.L. Rooney *et al.*, 2005 Fertility restorer locus Rf1 [corrected] of sorghum (Sorghum bicolor L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12. *Theor Appl Genet* 111 (6):994-1012.

Kong, W., H. Jin, C.D. Franks, C. Kim, R. Bandopadhyay *et al.*, 2013 Genetic Analysis of Recombinant Inbred Lines for *Sorghum bicolor* × *Sorghum propinquum*. *G3: Genes*| *Genomes*| *Genetics* 3:101-108.

Li, H., 2011 A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27 (21):2987-2993.

Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25 (14):1754-1760.

Li, X., Y. Wei, A. Acharya, Q. Jiang, J. Kang *et al.*, 2014 A saturated genetic linkage map of autotetraploid alfalfa (Medicago sativa L.) developed using genotyping-by-sequencing is highly syntenous with the Medicago truncatula genome. *G3 (Bethesda)* 4 (10):1971-1979.

Logan-Young, C.J., J.Z. Yu, S.K. Verma, R.G. Percy, and A.E. Pepper, 2015 Snp Discovery in Complex Allotetraploid Genomes (Gossypium Spp., Malvaceae) Using Genotyping by Sequencing. *Applications in Plant Sciences* 3 (3).

Lu, F., A.E. Lipka, J. Glaubitz, R. Elshire, J.H. Cherney *et al.*, 2013 Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol. *Plos Genetics* 9 (1).

Lurin, C., C. Andres, S. Aubourg, M. Bellaoui, F. Bitton *et al.*, 2004 Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* 16 (8):2089-2103.

Mather, K., 1935 Reductional and equational separation of the chromosomes in bivalents and multivalents. *Journal of Genetics* 30 (1):53-78.

Morrell, P.L., D. Williams-Coplin, J.E. Bowers, J.M. Chandler, and A.H. Paterson, 2005 Crop-to-weed introgression has impacted allelic composition of johnsongrass populations with and without recent exposure to cultivated sorghum. . *Molecular Ecology* 14:2143-2154.

Muller, H.J., 1914 A New Mode of Segregation in Gregory's Tetraploid Primulas. *The American Naturalist* 48 (572):508-512.

Narum, S.R., C.A. Buerkle, J.W. Davey, M.R. Miller, and P.A. Hohenlohe, 2013 Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology* 22 (11):2841-2847.

Oyer, E.B., G. Gries, and B. Rogers, 1959 The seasonal development of Johnson grass plants. *Weeds* 7 (1):13-19.

Paterson, A., H., J.E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood *et al.*, 2009 The Sorghum bicolor genome and the diversification of grasses *Nature* 457:551-556.

Paterson, A., H., T.S. Cox, W. Kong, and M. Navarro, 2013 Viewpoint: multiple-harvest sorghums toward improved food securit in *Perennial Crops for Food Security*, edited by C. Batello, L. Wade, S. Cox, N. Pogna, A. Bozzini *et al.*

Paterson, A., H., W. Kong, R. Johnston, P. Nabukalu, G. Wu *et al.*, 2017 The evolution of an invasive plant, Sorghum halepense L. ('Johnsongrass').

Paterson, A.H., K.F. Schertz, Y.R. Lin, S.C. Liu, and Y.L. Chang, 1995 The Weediness of Wild Plants - Molecular Analysis of Genes Influencing Dispersal and Persistence of Johnsongrass, Sorghum-Halepense (L) Pers. *Proceedings of the National Academy of Sciences of the United States of America* 92 (13):6127-6131.

Poland, J., J. Endelman, J. Dawson, J. Rutkoski, S.Y. Wu *et al.*, 2012a Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. *Plant Genome* 5 (3):103-113.

Poland, J.A., P.J. Brown, M.E. Sorrells, and J.L. Jannink, 2012b Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. *Plos One* 7 (2).

Quinn, L., J.N. Barney, J. McCubbins, and A. Endres, 2013 Navigating the "Noxious" and "Invasive" Regulatory Landscape: Suggestions for Improved Regulation. *Bioscience* 63 (2):124-131.

R Core Team, 2013 R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Sangduen, N., and W.W. Hanna, 1984 Chromosome and fertility studies on reciprocal crosses between two species of autotetraploid sorghum: Sorghum bicolor (L.) Moench and S. halepense (L.) Pers. *Journal of Heredity* 75 (4):293-296.

Schmitz-Linneweber, C., and I. Small, 2008 Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci* 13 (12):663-670.

Stift, M., C. Berenos, P. Kuperus, and P.H. van Tienderen, 2008 Segregation Models for Disomic, Tetrasomic and Intermediate Inheritance in Tetraploids: A General Procedure Applied to Rorippa (Yellow Cress) Microsatellite Data. *Genetics* 179 (4):2113-2123.

Tesso, T., I. Kapran, C. Grenier, A. Snow, P. Sweeney *et al.*, 2008 The potential for crop-to-wild gene flow in sorghum in Ethiopia and Niger: A geographic survey. *Crop Science* 48 (4):1425-1431.

Uitdewilligen, J.G.A.M.L., A.M.A. Wolters, B.B. D'hoop, T.J.A. Borm, R.G.F. Visser *et al.*, 2013 A Next-Generation Sequencing Method for Genotyping-by-Sequencing of Highly Heterozygous Autotetraploid Potato. *Plos One* 8 (5).

Wang, X., H. Tang, and A.H. Paterson, 2011 Seventy million years of concerted evolution of a homoeologous chromosome pair, in parallel, in major Poaceae lineages. *Plant Cell* 23 (1):27-37.

Wang, X.Y., H.B. Tang, J.E. Bowers, and A.H. Paterson, 2009 Comparative inference of illegitimate recombination between rice and sorghum duplicated genes produced by polyploidization. *Genome Research* 19 (6):1026-1032.

Warwick, S., and L. Black, 1983 THE BIOLOGY OF CANADIAN WEEDS.: 61. Sorghum halepense (L.) PERS. *Canadian Journal of Plant Science* 63 (4):997-1014.

Warwick, S.I., B. Thompson, and L. Black, 1984 Population variation in Sorghum halepense, Johnson grass, at the northern limits of its range. *Canadian Journal of Botany* 62 (9):1781-1790.

Wu, K.K., W. Burnquist, M.E. Sorrells, T.L. Tew, P.H. Moore *et al.*, 1992 The Detection and Estimation of Linkage in Polyploids Using Single-Dose Restriction Fragments. *Theoretical and Applied Genetics* 83 (3):294-300.

Xu, X.B., Z.X. Liu, D.F. Zhang, Y. Liu, W.B. Song *et al.*, 2009 Isolation and Analysis of Rice Rf1-Orthologus PPR Genes Co-segregating with Rf3 in Maize. *Plant Molecular Biology Reporter* 27 (4):511-517.

Zhang, D., H. Guo, C. Kim, T.H. Lee, J.P. Li *et al.*, 2013 CSGRqtl, a Comparative Quantitative Trait Locus Database for Saccharinae Grasses. *Plant Physiology* 161 (2):594-599.

Chapter 3 Tables and Figures

Table 3.1 Characteristics of the Genetic Map of the H4-derived BC_1F_1 population of *S. bicolor* BTx623 × *S. halepense* G9E

LG	Marker No.	Length (cM)	Avg Spacing (cM)	Max spacing (cM)	Avg No. AA	Avg No. AB	AB/AA Ratio
1A	49	271.1	5.65	34.5	24.63	94.61	3.84
1B	4	5.9	1.97	3.5	43.25	74.50	1.72
1C	16	74.8	4.98	18.9	42.94	74.31	1.73
1D	9	35.9	4.49	7.1	47.11	77.11	1.64
1E	5	10.8	2.71	3.1	56.40	52.00	0.92
1F	20	110.6	5.82	27.4	95.15	27.10	0.28
2A	11	63.4	6.34	32.5	23.55	89.55	3.80
2B	19	39.3	2.18	8.0	33.74	85.74	2.54
2C	22	91.7	4.37	14.2	52.32	64.41	1.23
2D	47	164.2	3.57	18.1	89.04	29.28	0.33
3A	31	146.8	4.89	27.7	28.94	88.65	3.06
3B	35	166.7	4.90	22.3	45.43	72.17	1.59
3C	22	182.3	8.68	43.0	60.59	60.18	0.99
3D	33	112.0	3.50	14.4	82.73	34.27	0.41
4A	29	208.8	7.46	24.9	25.62	89.93	3.51
4B	9	67.7	8.47	24.4	46.00	69.89	1.52
4C	18	198.8	11.69	56.3	56.56	57.50	1.02
4D	29	149.6	5.34	19.5	81.14	36.66	0.45
5A	4	17.1	5.71	9.6	39.75	86.25	2.17
5B	7	37.8	6.30	16.4	62.71	55.57	0.89
5C	21	121.0	6.05	21.1	87.19	33.67	0.39
6A	26	99.3	3.97	11.2	30.35	88.35	2.91
6B	23	135.4	6.16	21.1	44.26	75.13	1.70
6C	11	61.6	6.16	15.7	80.36	33.91	0.42
7A	24	236.0	10.26	38.8	27.04	90.29	3.34
7B	5	9.6	2.40	5.7	58.20	67.20	1.15
7C	25	101.6	4.23	9.9	58.60	56.68	0.97
7D	11	41.2	4.12	12.0	84.82	33.36	0.39
8A	12	110.5	10.05	26.0	24.50	92.50	3.78
8B	7	46.7	7.78	12.2	44.43	65.86	1.48
8C	4	22.0	7.33	8.4	53.50	63.00	1.18
8D	8	23.6	3.37	6.9	94.13	25.25	0.27
9A	13	121.7	10.14	32.6	20.92	92.15	4.40
9B	26	98.2	3.93	18.9	86.50	31.27	0.36

10A	15	79.9	5.71	24.9	24.20	92.80	3.83
10B	7	40.2	6.70	14.0	49.14	67.57	1.38
10C	35	248.9	7.32	18.8	62.91	59.29	0.94
10D	34	149.6	4.53	23.2	88.74	28.82	0.32

LG: linkage group.

AA: homozygous genotypes matching S. bicolor BTx623

AB: heterozygous genotypes

Table 3.2 Summary of the Genetic Map of the H6-derived BC_1F_1 population of *S. bicolor* BTx623 × *S. halepense* G9E

LG	Marker No.	Length (Genetic)	Avg Spacing	Max spacing	Avg No. AA	Avg No.AB	Ratio
1A	38	257.5	6.96	40.72	17.97	68.63	3.82
1B	22	124.6	5.94	16.66	26.73	60.68	2.27
1C	31	172.9	5.76	21.01	34.23	50.77	1.48
1D	18	166.6	9.80	28.76	70.06	14.83	0.21
2A	48	266.3	5.67	18.33	12.90	73.10	5.67
2B	11	93.6	9.36	17.29	31.91	52.00	1.63
2C	22	173.5	8.26	32.11	67.45	14.82	0.22
3A	52	344.7	6.76	17.23	17.38	68.38	3.93
3B	30	147.8	5.10	15.71	19.37	65.63	3.39
3C	7	34.3	5.71	11.11	42.43	44.14	1.04
3D	5	26.6	6.65	9.12	45.00	42.80	0.95
3E	33	225.0	7.03	17.26	67.21	17.24	0.26
4A	15	148.7	10.62	19.71	15.33	66.47	4.33
4B	42	289.9	7.07	17.61	23.71	62.60	2.64
4C	13	138.7	11.56	33.11	51.62	29.92	0.58
4D	38	327.7	8.86	24.14	63.89	17.08	0.27
5A	5	59.6	14.89	18.80	11.00	72.20	6.56
5B	16	78.1	5.21	14.50	49.13	34.31	0.70
5C	19	135.6	7.53	22.71	76.16	15.21	0.20
6A	27	215.4	8.28	23.06	30.00	57.19	1.91
6B	17	142.9	8.93	19.86	38.24	47.59	1.24
6C	7	59.6	9.93	13.65	39.00	48.29	1.24
6D	4	26.6	8.88	11.25	45.00	40.75	0.91
6E	27	264.9	10.19	27.41	60.56	22.22	0.37
7A	24	182.5	7.93	16.07	12.96	72.92	5.63
7B	4	30.5	10.18	13.25	34.25	57.00	1.66
7C	3	19.5	9.75	16.09	30.67	45.33	1.48
7D	26	223.5	8.94	20.07	54.58	25.50	0.47
8A	31	269.2	8.97	26.05	12.84	68.29	5.32
8B	22	299.0	14.24	29.05	48.23	29.50	0.61
9A	34	231.7	7.02	22.51	14.00	67.50	4.82
9B	13	95.5	7.96	16.56	35.77	47.15	1.32
9C	32	260.0	8.39	30.14	63.78	17.38	0.27
10A	20	126.7	6.67	19.38	21.35	67.10	3.14
10B	19	250.6	13.92	48.90	34.89	50.74	1.45
10C	24	165.3	7.19	36.34	65.00	21.29	0.33

LG: linkage group.

AA: homozygous genotypes matching S. bicolor BTx623

AB: heterozygous genotypes

Table 3.3 Loci with non-Mendelian segregation combined for two BC_1F_1 populations of *S. bicolor* BTx623 × *S. halepense* G9E

			H4		H6			Pooled	
SNP	Chr	Physical_dist	AB/AA	Sig	AB/AA	Sig	Both	AB/AA	Sig
S1_10311755	1	10.31	10.67	*	7.88	NS	N	9.35	*
S1_16645888	1	16.65	9.11	NS	19.5	**	N	12.3	**
S1_40908384	1	40.91	36	***	15.6	**	Y	23.3	***
S1_46405932	1	46.41	9.33	*	2.54	**	N	4.81	NS
S1_49998086	1	50	9.5	*	2.8	*	N	4.97	NS
S1_55269769	1	55.27	24	***	12.17	*	Y	16.9	***
S1_61787869	1	61.79	1.66	***	11.14	*	N	3.04	NS
S1_70587720	1	70.59	8	NS	11.5	*	N	9.31	*
S2_1057029	2	1.06	8.67	NS	10.63	*	N	9.45	**
S2_1810453	2	1.81	10.38	*	6.33	NS	N	8.24	*
S2_1826192	2	1.83	10.91	*	10.38	*	Y	10.7	**
S2_2112837	2	2.11	12.25	*	8	NS	N	10	**
S2_2592696	2	2.59	9.55	*	5.77	NS	N	7.5	NS
S2_3982139	2	3.98	17.6	**	6.44	NS	N	10.4	**
S2_4685750	2	4.69	18.67	***	6.08	NS	N	10.3	**
S2_26594145	2	26.59	24.25	***	26	**	Y	25	***
S2_59755853	2	59.76	4.8	NS	15.75	*	N	6.63	NS
S2_63532297	2	63.53	2.22	***	36.5	**	N	4.24	NS
S2_66082501	2	66.08	3.06	*	10.56	*	N	4.63	NS
S2_67318879	2	67.32	2.76	**	10.75	*	N	4.32	NS
S2_70855137	2	70.86	3.56	NS	21	**	N	6.05	NS
S2_74686312	2	74.69	2.74	**	10.75	*	N	4.26	NS
S2_74899873	2	74.9	8.36	NS	10	*	N	9	**
S2_76168371	2	76.17	12	*	6.75	NS	N	9.2	*
S2_76789433	2	76.79	2.92	*	15.5	*	N	4.66	NS
S3_700571	3	0.7	10.18	*	2.71	**	N	5.06	NS
S3_1062926	3	1.06	10.08	*	2.43	**	N	4.73	NS
S3_1135026	3	1.14	23.2	***	27	**	Y	24.6	***
S3_6722016	3	6.72	5.8	NS	33.5	**	N	9.06	*
S3_20644695	3	20.64	13.38	**	19.5	**	Y	15.4	***
S3_34250295	3	34.25	31	***	21.25	**	Y	26.1	***
S3_34250708	3	34.25	26.5	***	-	***	N	44.8	***
S3_57669990	3	57.67	1.54	***	14.6	*	N	3.18	NS
S3_66256314	3	66.26	16.86	***	6	NS	N	9.62	**

S3_66505027	3	66.51	13.57	**	3.82	NS	N	6.67	NS
S3_66508407	3	66.51	10.8	*	7.55	NS	N	9.1	**
S4_805881	4	0.81	17.4	**	7	NS	N	11	**
S4_806010	4	0.81	11	*	4.53	NS	N	7.12	NS
S4_6311035	4	6.31	14.29	**	17.75	**	Y	15.5	***
S4_15434708	4	15.43	14.63	**	17	**	Y	15.5	***
S4_32653887	4	32.65	23.5	***	11.83	*	Y	16.5	***
S4_45556066	4	45.56	10.8	*	6.5	NS	N	8.65	*
S4_50435891	4	50.44	10.5	*	3.33	NS	N	5.83	NS
S4_57722969	4	57.72	13.63	**	28	***	Y	17.5	***
S5_17017428	5	17.02	22	***	3.48	NS	N	7.31	NS
S5_17837434	5	17.84	12.1	**	2.17	***	N	4.65	NS
S5_17839323	5	17.84	11.13	*	2.29	**	N	4.72	NS
S5_20824045	5	20.82	25.75	***	17.25	*	Y	21.5	***
S5_35996091	5	36	5.67	NS	11.43	*	N	7.28	NS
S5_61639044	5	61.64	1.93	***	14.83	**	N	3.51	NS
S6_5884544	6	5.88	8.3	NS	11.83	*	N	9.63	*
S6_16623205	6	16.62	23.25	***	11.67	*	Y	16.3	***
S6_25520097	6	25.52	24.75	***	24	**	Y	24.4	***
S6_43674309	6	43.67	3.26	NS	11.43	*	N	4.94	NS
S7_915336	7	0.92	10	*	1.41	***	N	3.51	NS
S7_1198601	7	1.2	10.88	*	24.67	**	Y	14.6	***
S7_1624694	7	1.62	10.09	*	8.1	NS	N	9.14	**
S7_2805740	7	2.81	4.53	NS	13	*	N	6.56	NS
S7_3048171	7	3.05	7.44	NS	23.25	***	N	10.6	**
S7_3612606	7	3.61	4.24	NS	17.6	**	N	6.47	NS
S7_6161515	7	6.16	8.23	NS	20.75	**	N	11.2	**
S7_6604014	7	6.6	3.88	NS	13.83	*	N	5.87	NS
S7_6735921	7	6.74	1.73	***	15	*	N	3.2	NS
S7_6858502	7	6.86	1.4	***	11.43	*	N	2.67	NS
S7_7548864	7	7.55	14.38	**	2.36	**	N	5.27	NS
S7_14222884	7	14.22	4	NS	14.6	*	N	6.04	NS
S7_16072821	7	16.07	3.46	NS	12	*	N	5.27	NS
S7_17895235	7	17.9	8.2	NS	12.8	*	N	9.73	*
S7_18949389	7	18.95	4.13	NS	11.86	*	N	5.87	NS
S7_37697004	7	37.7	4.48	NS	18.2	**	N	6.93	NS
S7_37954160	7	37.95	3.48	NS	10.71	*	N	5.17	NS
S7_56916581	7	56.92	16.71	***	8.33	NS	N	12	***
S7_57890358	7	57.89	1.51	***	19.4	**	N	3	NS
S8_1813552	8	1.81	7.57	NS	11.14	*	N	8.76	*
S8_3566394	8	3.57	4.21	NS	33.5	**	N	7	NS

S8_4847529	8	4.85	24.4	***	10.25	*	Y	15.7	***
S8_5330663	8	5.33	5.93	NS	20	*	N	8.41	*
S8_11551047	8	11.55	9	NS	4.67	NS	N	6.59	NS
S8_14556229	8	14.56	7.83	NS	13.8	*	N	9.59	**
S8_41273697	8	41.27	10.11	*	8.71	NS	N	9.5	*
S8_43980562	8	43.98	7.87	NS	13.14	*	N	9.55	**
S8_52724145	8	52.72	1.67	***	19	**	N	3.28	NS
S8_53308674	8	53.31	0.45	***	10.88	*	N	1.28	NS
S8_54355956	8	54.36	1.51	***	11.13	*	N	2.86	NS
S8_54678607	8	54.68	1.68	***	12.71	*	N	3.04	NS
S8_55352351	8	55.35	1.54	***	12.83	*	N	3.12	NS
S9_10225531	9	10.23	9.1	NS	30.5	**	N	12.7	**
S9_11731715	9	11.73	10.18	*	12.43	*	Y	11.1	***
S9_11977317	9	11.98	4.15	NS	24.5	***	N	6.87	NS
S9_14882885	9	14.88	4.65	NS	16	*	N	6.81	NS
S9_15158378	9	15.16	2.48	**	11.33	*	N	4	NS
S9_31580899	9	31.58	20.33	***	8.6	NS	N	13	***
S9_40134888	9	40.13	3.16	*	30	***	N	5.53	NS
S9_43649628	9	43.65	4.4	NS	15.33	**	N	6.52	NS
S9_47539817	9	47.54	15.25	**	32	***	Y	19.8	***
S9_49554277	9	49.55	14.13	**	17	**	Y	15.2	***
S9_50198667	9	50.2	7.8	NS	13.5	*	N	9.43	**
S9_50719802	9	50.72	6.25	NS	11.67	*	N	7.73	NS
S9_51718314	9	51.72	4.43	NS	28.33	***	N	7.19	NS
S9_55264760	9	55.26	2.33	***	11.13	*	N	3.93	NS
S9_55329149	9	55.33	0.81	***	15.8	**	N	1.98	NS
S9_56692240	9	56.69	0.98	***	29	***	N	2.57	NS
S9_57335220	9	57.34	6	NS	16	*	N	8.35	*
S9_57582717	9	57.58	5.89	NS	15.17	**	N	8.12	*
S9_58062014	9	58.06	3.52	NS	12.17	*	N	5.19	NS
S10_4382269	10	4.38	11.86	*	21.33	**	Y	14.7	***
S10_8603485	10	8.6	15.38	**	6.23	NS	N	9.71	**
S10_8803504	10	8.8	10.63	*	20.33	*	Y	13.3	**
S10_23708355	10	23.71	30	***	3.63	NS	N	8.22	*
S10_28674405	10	28.67	17.17	**	18	**	Y	17.5	***
S10_36575703	10	36.58	52	***	9	NS	N	17.6	***

Table 3.4 Inferred SNP origins in the H4 and H6 populations of *S. bicolor* BTx623 × *S. halepense* G9E

SNP Types	B	H-BP	H-PM	N-M	P	Unknown
Counts H4	7	79	1	423	296	300
(Proportion H4)	(0.87%)	(9.80%)	(0.12%)	(52.48%)	(36.72%)	-
Counts H6	5	71	1	325	296	296
(Proportion H6)	(0.72%)	(10.17%	(0.14%)	(46.56%)	(42.41%)	-
Whole Genome	1,777,782	447,479	8,379	478,804	744,924	3,852,257
(Proportion Whole Genome)	36.64%	9.22%	0.17%	38.61%	15.35%	-

B: Alleles matching *S. bicolor* but not *S. propinquum*

H-BP: Heterozygotes, matching both *S. bicolor* and *S. propinquum*

H-PM: Heterozygotes, matching *S. propinquum* and a new allele

N-M: Alleles not matching *S. bicolor* of *S. propinquum*, inferred to be new mutations

P: Alleles matching *S. propinquum*, but not *S. bicolor*

Unknown: missing data from either *S. propinquum* or *S. halepense*, or polymorphism between *S. bicolor*BTx623 and IS3620C. Not included in calculate the percentage

Table 3.5 *S. halepense* G9E genomic regions with non-random 'runs' of more than three consecutive *S. propinquum* alleles

Chr	LGH4	LGH6	H4 Start (Mb)	H4 End (Mb)	H4 Range	H6 Start (Mb)	H6 End (Mb)	H6 Range
1	1A	1A	1,002,293	2,075,903	1,073,610	1,002,293	2,075,903	1,073,610
1	-	1A	-	-	-	4,118,965	5,421,700	1,302,735
1	1A	1A	9,156,903	9,747,197	590,294	9,156,903	9,747,197	590,294
1	1A	1A	10,172,236	11,113,096	940,860	10,172,236	11,113,096	940,860
1	1A	-	58,398,011	60,856,958	2,458,947	-	-	-
1	-	1B	-	-	-	60,795,489	61,898,805	1,103,316
1	-	1C	-	-	-	68,343,440	69,007,210	663,770
2	-	2A	-	-	-	37,375,688	49,746,108	12,370,420
2	2B	-	70,855,137	71,665,944	810,807	-	-	-
2	-	2A	-	-	-	72,907,102	73,093,086	185,984
3	3A	3A	7,974,628	10,505,659	2,531,031	7,974,628	10,505,659	2,531,031
3	-	3A	-	-	-	57,155,624	57,669,990	514,366
3	-	3A	-	-	-	66,256,314	66,512,680	256,366
3	3B	-	57,050,572	57,258,601	208,029	-	-	-
3	3B	3A	71,350,072	73,236,656	1,886,584	72,084,184	73,236,656	1,152,472
3		3E	-	-	-	5,529,567	7,162,059	1,632,492
3	3C	-	8,639,654	60,878,509	52,238,855	-	-	-
4	4A	4B	60,169,823	66,355,590	6,185,767	58,910,354	64,899,883	5,989,529
4	4B	-	64,423,184	64,876,657	453,473	-	-	-
4	4D		61,407,366	64,036,600	2,629,234	-	-	-
6	6A	6A	55,862,094	64,036,600	8,174,506	55,862,094	57,182,238	1,320,144
6	6A	6A	55,862,094	64,036,600	8,174,506	58,155,004	58,738,149	583,145
6	6B	6B	3,622,183	37,245,941	33,623,758	3,327,280	37,245,941	33,918,661
9	9A		57,335,220	58,856,483	1,521,263	-	-	-
9	-	9A	-	-	-	55,778,912	57,267,091	1,488,179
10		10B	-	-	-	52,116,222	59,317,021	7,200,799

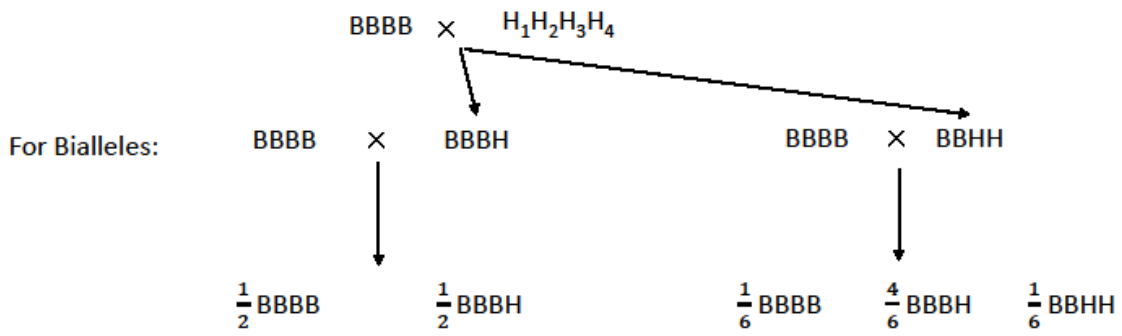


Figure 3.1 Theoretical Chromosomal Segregation Patterns in the BC_1F_1 Population

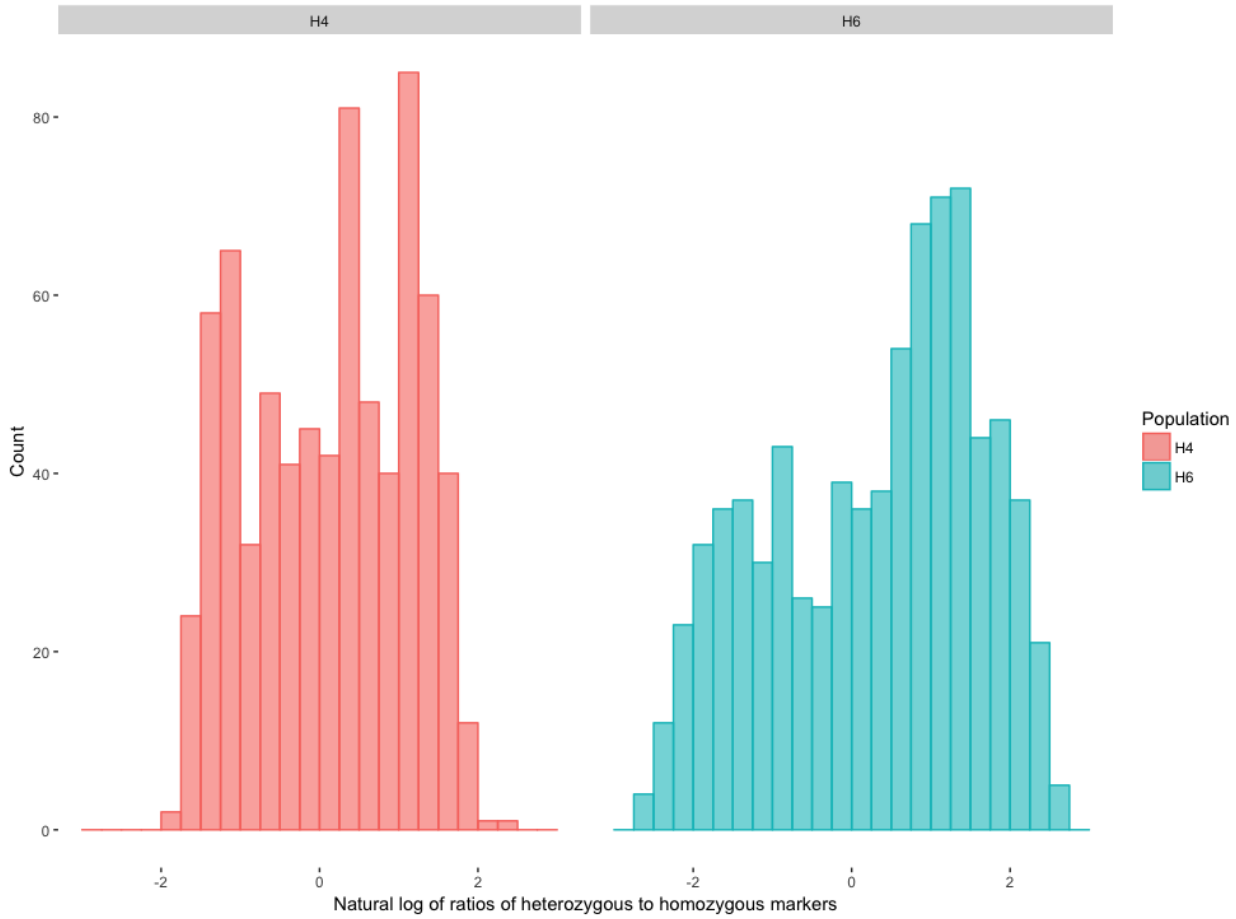


Figure 3.2 Histogram of Square Root of AB/AA for H4 and H6 populations of *S. bicolor* BTx623 × *S. halepense* G9E with mapped markers. AA is the homozygous genotype while AB is the heterozygous genotype

Figure 3.3 Genetic Maps of the BC_1F_1 Population of *S. bicolor* BTx623 × *S. halepense* G9E. The maps of the H4-derived population are on the left of the black line and H6 maps are on the right for each sorghum chromosome. The x-axis is the average segregation ratio of each linkage group after square root transformation. Negative values are assigned for the H4 population

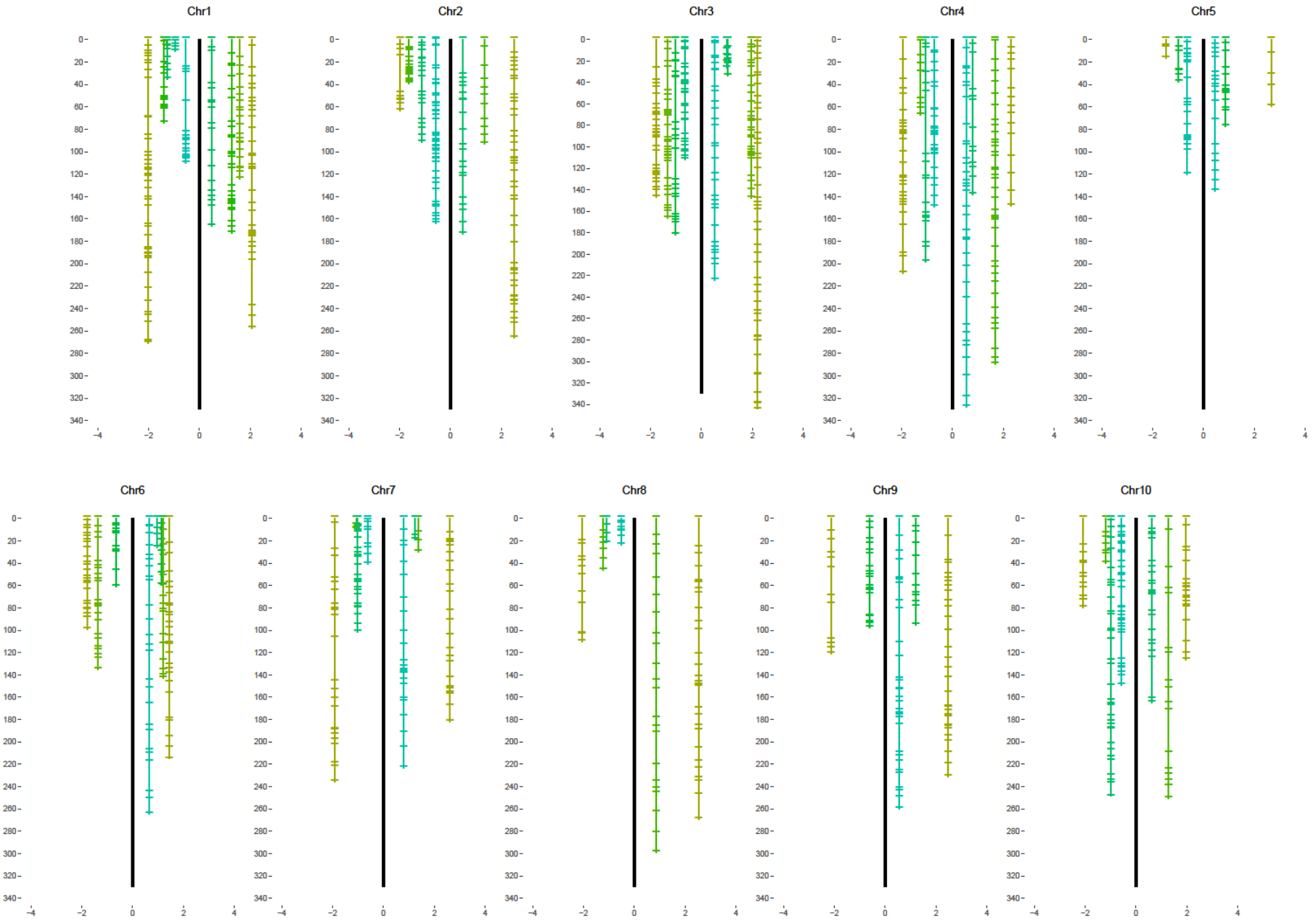
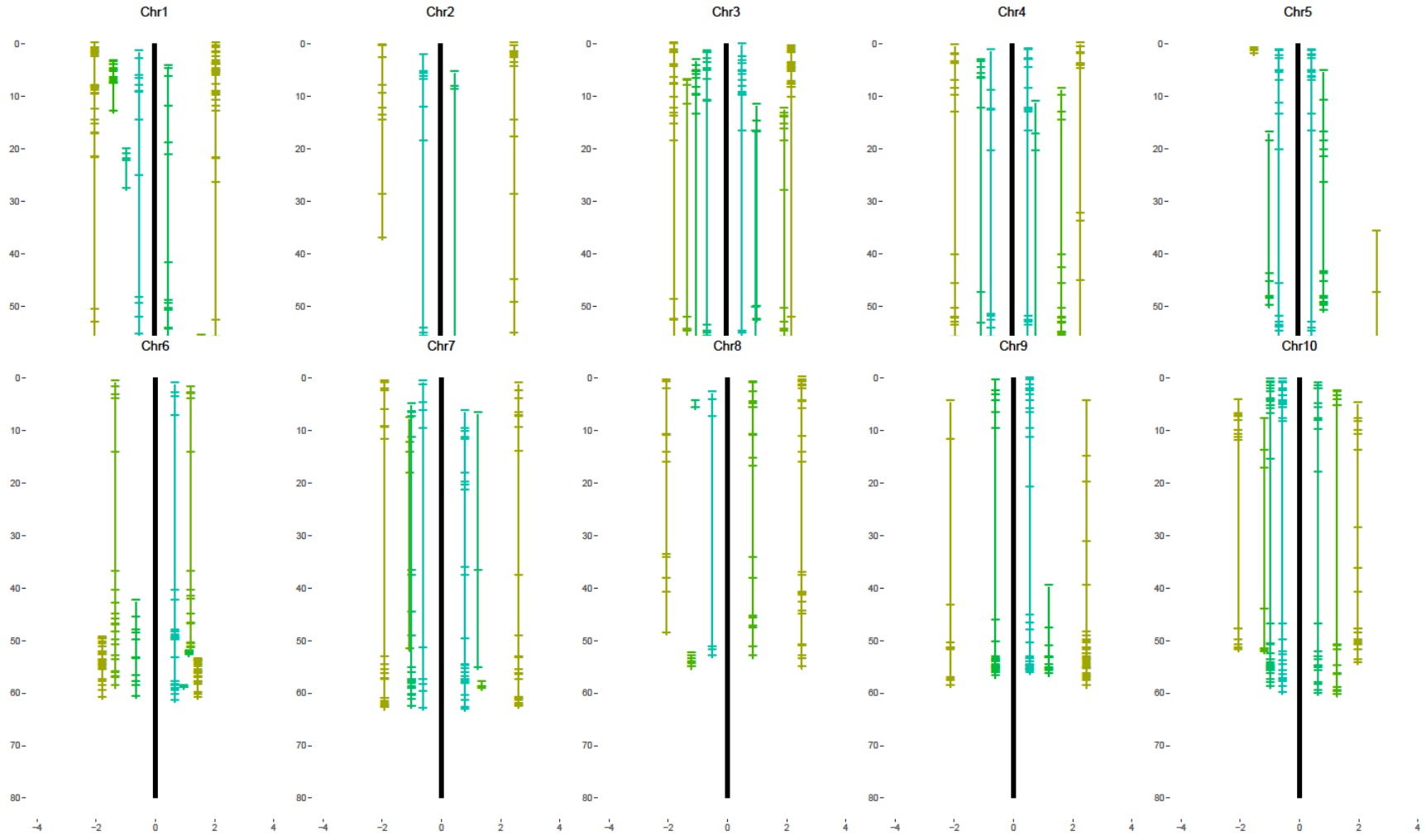
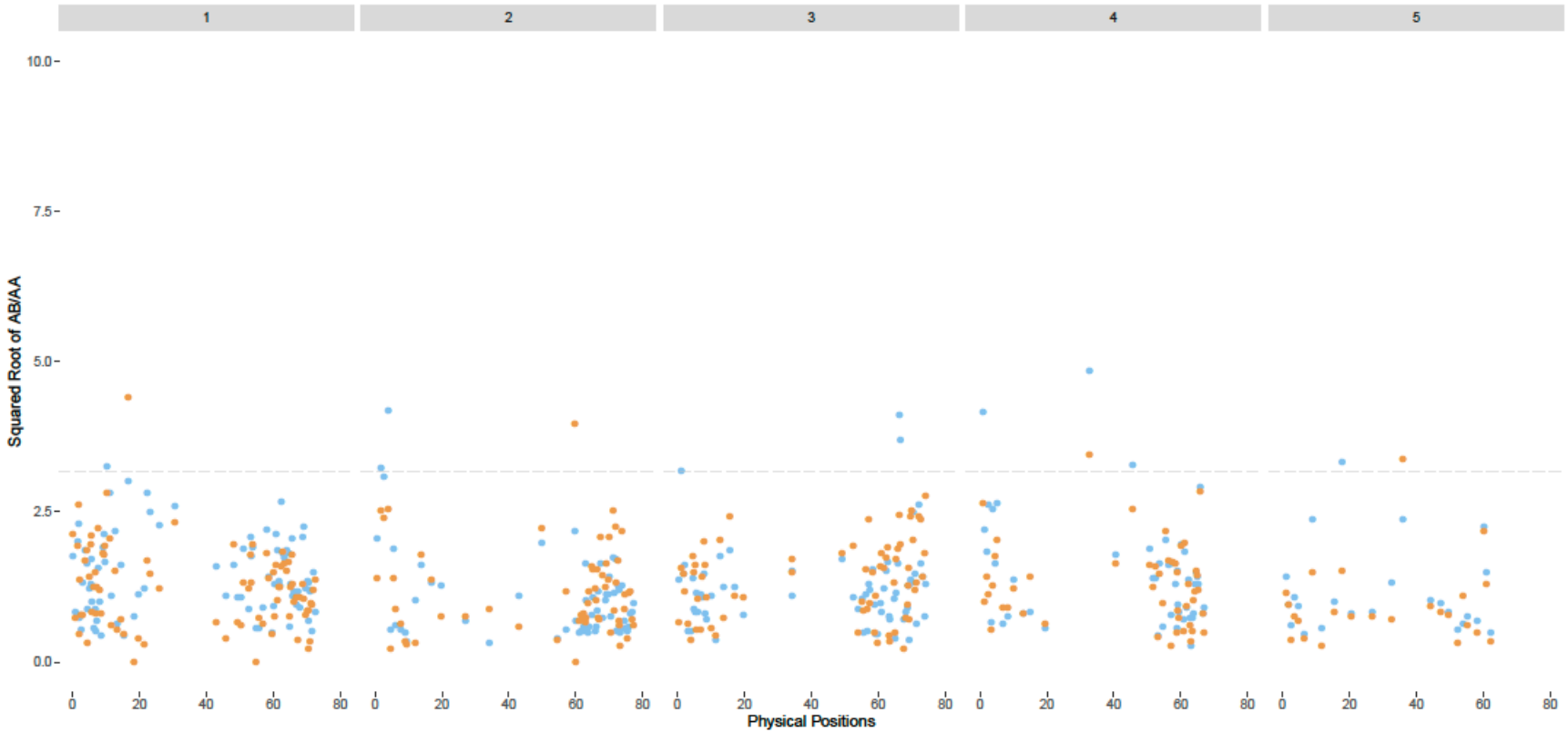

Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10

Figure 3.4 Physical coverage of each *S. bicolor* BTx623 × *S. halepense* G9E linkage group to the sorghum genome. The H4 population is on the left of the black line and the H6 population is on the right. The x-axis is the segregation ratio after the square root transformation and that of the H4 population (left) is assigned a negative sign.


Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10


1
2
3
4
5
10.0
7.5
5.0
2.5
0.0
Squared Root of AB/AA
0
20
40
60
80
Physical Positions

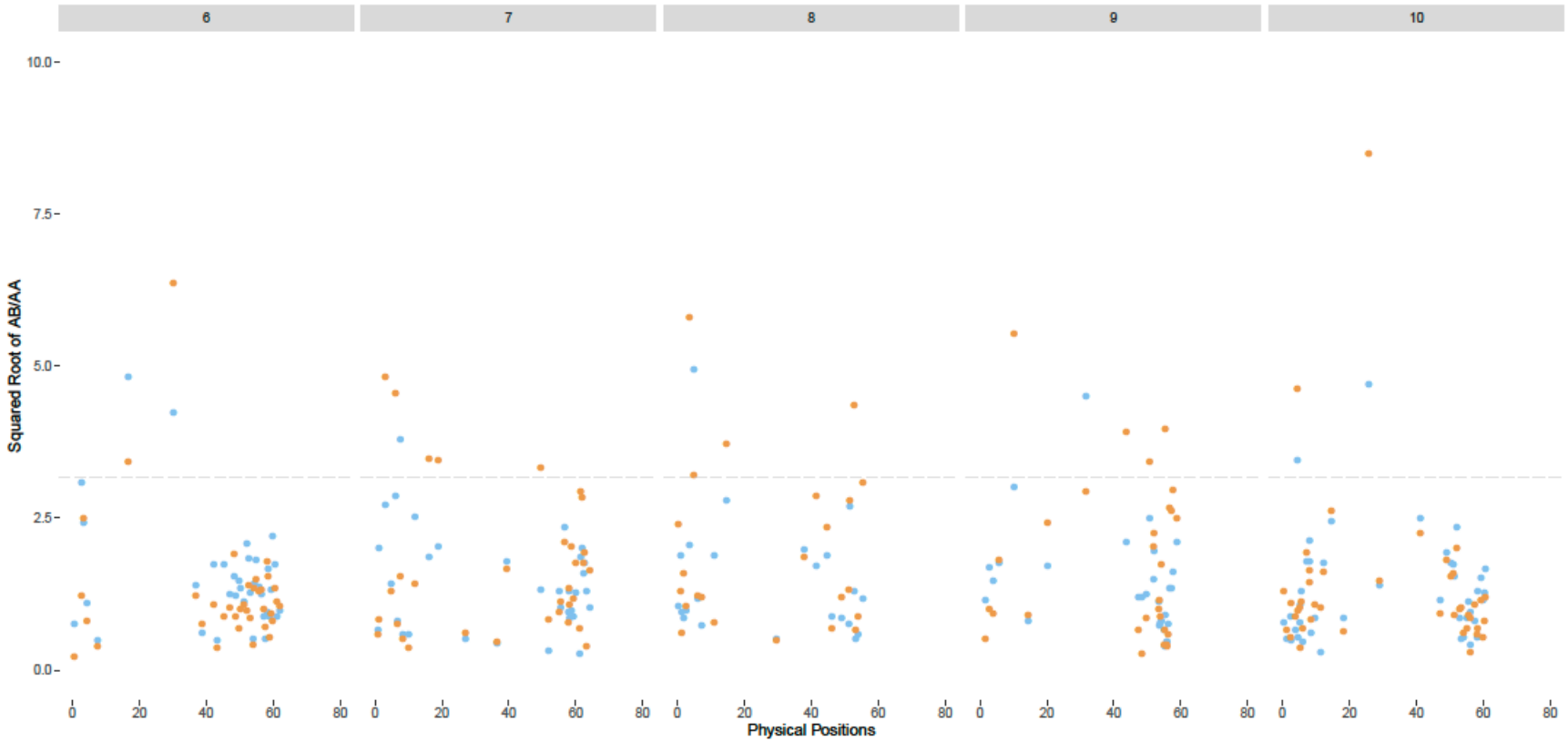


Figure 3.5 Patterns of segregation of the BC_1F_1 populations of *S. bicolor* BTx623 × *S. halepense* G9E based on the sorghum chromosomes. Square root transformation of the ratio of AB/AA for each marker are plotted in blue (H4-derived population) or orange (H6).

CHAPTER 4

QUANTITATIVE TRAIT MAPPING OF PLANT ARCHITECTURE IN TWO BC_1F_2 POPULATIONS OF *SORGHUM BICOLOR* × *S. HALEPENSE* AND COMPARISONS TO TWO OTHER SORGHUM POPULATIONS[3]

[3] Kong, WQ., Nabukalu, P., Robertson, J., Goff, V., Pierce, G., Lemke, C., Compton, R., Cox, S. Paterson, AH.. To be submitted to Theoretical and Applied Genetics.

Abstract

We describe quantitative trait analysis of plant height and flowering time in two novel BC_1F_2 populations totaling 246 genotypes derived from backcrossing two *Sorghum bicolor* x *S. halepense* F_1 plants to a tetraploidized *S. bicolor*. Phenotyping for two years each in Bogart, GA and Salina, Kansas allowed us to dissect variance into narrow-sense genetic (QTLs) and environmental components. Overlapping plant height and flowering time QTLs suggest either pleiotropy or co-selection of multiple nearby genes. In crosses with a common *S. bicolor* BTx623 parent, correspondence of QTLs in *S. halepense*, its rhizomatous progenitor *S. propinquum*, and *S. bicolor* race guinea which is highly divergent from BTx623 suggests QTLs at which new alleles have been associated with improvement of elite sorghums, while additional QTLs unique to the *S. halepense* populations may be recent mutations associated with its polyploidy. While resolving linkage maps for QTL mapping for mostly auto-tetraploid populations such as those studied here is time consuming and difficult, we developed a fast genome-wide association study (GWAS)-based method for QTL detection and visualization.

Introduction

Vegetative growth and the timing of the vegetative-to-reproductive transition are critical to a plant's fitness, directly and indirectly determining when and how a plant lives, grows and reproduces. Modification of two important traits, plant height and flowering time, has been central to plant domestication and crop improvement. For example, the 'Green Revolution' was based largely upon short and lodging resistant plants suitable for increased fertilization and mechanical harvesting. Increasing demand for lignocellulosic biomass has motivated increasing plant height or developing dual purpose plants with both grain and biomass potential (Fernandez et al. 2009). Adjusting flowering time is important for hybrid development, for utilization of germplasm in non-native latitudes, and for adaptation to changing climatic conditions (Jung and Müller 2009).

One of the best-studied components of plant architecture is plant height, because of its importance, reliable phenotyping and relatively high heritability (Salas Fernandez et al. 2009; Wang and Li 2006). However, it took nearly 40 years after the 'Green Revolution' for the underlying genes to be identified, semi dwarf (*SD1*) in rice and Reduced height (*Rht*) in wheat, both involved in gibberellin (GA) pathways (Peng et al. 1999; Sasaki et al. 2002; Wang et al. 2017). In addition to GA pathways, the brassinosteroid (BR) pathway is responsible for cell elongation by cell wall loosening (Kutschera and Wang 2012). Mutants of the BR pathway show dwarf phenotypes, while increased BR levels can increase plant size (Bishop 2003). Quantitative studies suggested that control of plant height is polygenic in maize (Peiffer et al. 2014; Ku et al. 2015). In sorghum, a canonical model suggested four general loci for plant height, *DW1- DW4* (Quinby and Karper 1945b), but multiple quantitative studies (Lin et al. 1995; Hart et al. 2001; Brown et al. 2008; Upadhyaya et al. 2012; Murray et al. 2008; Zhang et al. 2015; Ritter et al.

2008; Zhang et al. 2013; Morris et al. 2013) support that at least six non-overlapping loci contribute to plant height (Zhang et al. 2015). To date, *Dw1* and *Dw3* have been cloned (Multani et al. 2003; Yamaguchi et al. 2016; Hilley et al. 2016a); *Dw2* is located in or near a large heterochromatin region on sorghum chromosome 6 (Brown et al. 2008; Murray et al. 2009; Higgins et al. 2014; Zhang et al. 2015; Cuevas et al. 2016; Murphy et al. 2011; Hilley et al. 2017); and *Dw4*, may be near ~6.6Mb on chromosome 6 (Morris et al. 2013) but strong signals on chromosome 4 have also been considered as indicators of *Dw4* (Li et al. 2015; Zhang et al. 2015). Additional loci with small effects responsible for plant height might explain residual variance.

Genetic manipulation of flowering time is important in crop domestication, and directly or indirectly influences plant architecture (Hill and Li 2016). In most Poaceae taxa, plants flowering late are taller, as flowering terminates apical growth. Indeed, many studies reported positive correlations between plant height and flowering time in sorghum and often discovered QTLs for these two traits located in the same general genomic regions (Lin et al. 1995; Morris et al. 2013; Brown et al. 2008; Zhang et al. 2015; Higgins et al. 2014). A total of six maturity genes controlling days to flowering have been denoted *Ma1 – Ma6* (Quinby and Karper 1945a; Quinby 1966; Brady 2006). An intriguing early candidate for *Ma1* on sorghum chromosome 6 (Cuevas et al. 2016; Murphy et al. 2011) has recently been found unlikely, and replaced by a more probable candidate supported by numerous lines of evidence (Cuevas et al. 2016). *Ma3* is proposed to be the phytochrome B on sorghum chromosome 1 (Yang et al. 2014a; Childs et al. 1997), while *Ma6* is hypothesized to be an ortholog of a rice grain number, plant height and heading date gene (GHD7) (Murphy et al. 2014), albeit needing further functional validation.

Sorghum provides many avenues to study traits related to plant architecture, thanks to rich genomic resources and a high quality reference genome sequence (Paterson et al. 2009). In addition to genomic tools, the flexibility to make crosses between the five main sorghum races (bicolor, guinea, caudatum, durra and kafir), and with wild relatives such as *S. propinquum* and *S. halepense*, makes it particularly attractive to dissect and compare genetic components of plant architecture. In this paper, we describe a quantitative trait study of key components of plant architecture, specifically plant height and flowering time, in two half-sib tetraploid BC_1F_2 populations derived from crossing *Sorghum bicolor* BTx623 and *Sorghum halepense* Gypsum 9E. A two-year, two-environment phenotypic evaluation permits us to identify major effect and environment specific QTLs. Quantitative trait loci (QTLs) discovered in these two novel populations are compared to those from two other diploid sorghum recombinant inbred line (RIL) populations sharing BTx623 as a common parent but sampling the breadth of the Sorghum genus, one a cross to *S. bicolor* IS3620C, and the other to *S. propinquum* (Kong et al. 2013). QTLs identified in this study and their comparison provide insight into evolution of morphological diversity in the Sorghum genus, are of practical use for marker-assisted breeding, and provide a foundation for molecular cloning and functional analysis.

Materials and Methods

Genetic Stocks

S. bicolor BTx623 × *S. halepense* G9E (SH-BC_1F_1 and SH-BC_1F_2): Two tetraploid F_1 hybrids, H4 and H6, derived from crossing *S. bicolor* BTx623 × *S. halepense* Gypsum 9E (G9E) were backcrossed to the tetraploidized recurrent parent, *S. bicolor* BTx623 to develop H4 and H6-derived BC_1F_1 mapping populations consisting of 141 and 105 genotyped individuals, respectively. BC_1F_2 rows derived from selfing single BC_1F_1 plants were planted on May 28th

2013 and May 9th 2014 at the University of Georgia Plant Science Farm, Watkinsville, GA, USA (Athens 2013 and Athens 2014 hereafter), and on Jun 3rd 2013, and Jun 17th 2014 at The Land Institute, Salina, KS, USA (Salina 2013 and Salina 2014 hereafter). Within each environment, there were two blocks and three replications for each genotype, in a completely randomized design.

RIL population of *S. bicolor* BTx623 × IS3620C (IS-RIL): This population comprised 393 $F_{7\text{-}8}$ RILs derived by selfing of a single F_2 plant from *S. bicolor* BTx623 × IS3620C, expanding a population previously described (Hart et al. 2001; Kong et al. 2000). This population was planted at the University of Georgia Plant Science Farm, Watkinsville, GA, USA on 10 May 2011 and 18 May 2012. Single 1.5-m plots of each RIL were machine planted in a completely randomized design. For each progeny line, we phenotyped two plants for replication.

RIL population of *S. bicolor* BTx623 and *S. propinquum* (PQ-RIL): This population comprised 161 RILs derived by selfing a single F_2 plant from S. bicolor BTx623 and S. propinquum, as described in Kong et al. (2013). The population was planted on May 20th 2009, May 28th 2010 and May 16th 2011 in a completely randomized design, transplanting (2009, 2011) or direct seeding (2010) five plants in a 1.5 m plot. For each progeny line, we phenotyped two plants for replication. Genetic components of vegetative branching patterns were elaborated in Kong et al. (2014).

Genotyping

Leaf samples of the SH-BC_1F_1 population were frozen at -80 C and lyophilized for 48 hours. Genomic DNA was extracted from the lyophilized leaf sample based on Aljanabi et al. (1999).

The GBS platform is a slightly modified version of Multiplex Shotgun Genotyping (MSG) (Andolfatto et al. 2011) combined with the Tassel GBS 5 analysis pipeline. Genotyping of the two SH-BC_1F_1 populations used Illumina HiSeq 2500, Rapid V2 kits that generate about 150 million reads of 100 base pair (bp) fragments per run with single-end sequencing. Details of SNP calling, marker filtering, and genetic map construction can be found in Kong et al., 2017. Genotyping of the IS-RIL population used an in-house Illumina MiSeq that generates up to 25 million reads of 150 base pair (bp) fragments per run with single-end sequencing. Description of the genetic map of the IS-RIL population can be found in Kong, 2017. The genetic map of the RIL population derived from *S. bicolor* BTx623 × *S. propinquum* was published in Kong et al. (2013).

Phenotyping

We evaluated plant height and flowering time in the SH-BC_1F_2 families with three replications in two fields in two years and at two locations. Flowering time (**FL**) was measured by recording when flower heads emerged for about 50% of the plants within a plot. Plant height (**PH**) was measured from the base to the tip of the main flower head. Phenotyping of plant height and flowering time in the IS-RIL population was consistent with our system applied to the PQ-RIL as described in Kong et al. (2014). The variance component method was used to calculate broad-sense heritability [$H= V_g/(V_g+V_g*V_E/e +V_{residual}/er)$] where V_g is the variance estimate for genotype, V_E is the variance estimate for environment, e is the number of environments and r is the number of replications.

Single Marker Analysis

We screened informative markers to identify an appropriate subset for conducting single marker analysis based on the following procedure:

a. Select bi-allelic markers.
b. Select markers that are at least 100 bp apart since SNP markers within 100 bp are highly correlated and provide little additional information in genetic mapping populations in strong linkage disequilibrium.
c. Keep markers with missing data less than 60%.
d. Delete markers with minor allele frequency less than 0.02.
e. Calculate pair-wise recombination frequency, and remove markers that fail to show linkage to any other markers of $\leq$20% recombination.

After filtering the raw dataset, we conducted one-way analysis of variance (ANOVA) for each phenotypic trait as the response variable with respect to each marker genotype for each population. A significance threshold of a p-value of 10e (-3) was used. We also conducted ANOVA by pooling both SH-BC_1F_2 populations while controlling population as a blocking factor. Statistical analyses used R (R Core Team 2016). The significant lists of SNPs for each trait in each sub-population were further condensed using hierarchical clustering with pair-wise recombination frequency as the distance matrix. We visualized the clustering of SNPs using heat maps with the R 'gplots' function (Warnes et al. 2016). Potential QTLs were determined if more than 4 SNPs were found within a cluster cut at height of 0.3 (30% recombination). Peak SNPs were chosen based on the smallest p-values.

Map-based QTL Mapping

The interval mapping method (Lander and Botstein 1989) was conducted for each trait of interest in each population (H4 or H6) in each environment using R/qtl (Broman et al. 2003). A LOD score of 2.5 was used as the significance threshold to infer QTLs. For QTLs showing multiple peaks or covering large genomic regions, we used the 'scantwo' function in R/qtl to test

the hypothesis of two QTL models. A multiple QTL model (MQM) was employed to calculate the percentage of variance explained for each trait of each population in each environment. A mixed model was fitted using all QTLs as fixed effects and the environment as a random effect to understand and partition the contribution of different QTL effects to the phenotype. We used a modified method to calculate R-squared for the fixed and model effects (Nakagawa and Schielzeth 2013) for the mixed effect modeling.

Results

Summary statistics and heritability analysis

The average height of *S. halepense* G9E was 157 cm, much taller than the 98.7 cm of *S. bicolor*, albeit measured in a separate experiment in 2012 (2013-2014 data for G9E showed evidence of contamination). The tetraploid BTx623 parent was 36.7 cm taller than its diploid counterpart in 2013 ($t=2.96$, $p= 0.0050$, Table 4.1), and 28.1 cm taller in 2014 ($t=4.91$, $p<0.001$). The average height of the SH-BC_1F_2 progeny lines across all four environments is 250.2 cm, much taller than both parents, with plants grown in Salina averaging 42.7 cm taller than in Athens ($t=17.27$, $p<0.001$). The respective locations differed in opposite ways across years – average plant height in Athens was 31.3 cm shorter in 2013 than 2014 ($t=-9.38$, $p<0.001$), but in Salina was 32.9 cm taller in 2013 than 2014 ($t=10.88$, $p<0.001$). The average height of the SH-BC_1F_2 population is the largest among the three sorghum populations, at 151.2 cm taller than the IS-RILs ($t=92.46$, $p<0.001$) and 100.6 cm taller than the PQ-RILs [($t=40.49$, $p<0.001$), (Table 4.1Table 4.2 Table 4.3)]. Broad-sense heritability estimates of plant height are relatively high and consistent among the three populations, at 72%, 78% and 77% for the SH-BC_1F_2, IS-RIL and PQ-RIL populations, respectively.

Days to flowering (**FL**) of *Sorghum halepense* G9E averaged 14 days earlier than that of BTx623 in 2012 (see explanation above). Tetraploid BTx623 plants flowered about 6.9 days later on average than diploid BTx623 (t=2.49, t=0.019), and this difference was larger in 2013 (8.2 days, t=3.25, p=0.0061) than 2014 (4.0 days, t=3.35, p=0.0053) (Table 4.1). Progeny lines of SH-BC_1F_2 display large genetic variation, but their average **FL** was about 75.7 days, near the 76.4 day average of tetraploid BTx623. The average **FL** in Salina was about 4 days longer than in Athens (t=5.38, p<0.001). Within each environment, average **FL** in Athens was about 2.8 days later in 2013 than 2014 (t=2.70, p=0.0071); and in Salina was about 12.9 days later in 2013 than 2014 (t=14.42, p<0.001), a much larger difference than in Athens. Progenies of the BC_1F_2 population flowered an average of 16.9 days later than the IS-RIL (t=30.34, p<0.001), and 8.7 days later than the PQ-RIL (Table 4.2 and Table 4.3). Broad sense heritability estimates of days to flowering are relatively high, at 83.59%, 63.66% and 83.65% in the SH-BC_1F_2, PQ-RIL and IS-RIL populations, respectively.

Phenotypic values for **FL** are significantly correlated across environments, as are phenotypic values for **PH**. Further, **FL** and **PH** are also positively and significantly correlated with one another (Figure 4.1), indicating that late flowering individuals are generally taller than early flowering ones.

QTL analysis

In addition to the conventional interval mapping method (Lander and Botstein 1989), we performed single marker analysis for each trait with respect to each of the 5148 filtered SNP markers, with the order of the markers based on the published sorghum genome sequence version 1.4 (Paterson et al. 2009). An example of a heat map with hierarchical clustering was provided to visualize the relationship among significant markers and to determine the number of

potential QTLs (Figure 4.2). This is a greedy method but nevertheless permits us to analyze pooled data from the two BC_1F_2 populations and to visualize the relationship between blocks of potential QTLs without constructing genetic maps

Plant height

We detected a total of 12 QTLs for **PH** in the H4-derived population by interval mapping, with six, qPH3C.H4.1, qPH6B.H4.1, qPH7C.H4.1, qPH7C.H4.2, qPH7D.H4.1 and qPH9B.H4.1 significant in more than one environment (Table 4.4). Allele effects of the detected QTLs are all positive, indicating that *S. halepense* alleles increase **PH**. A total of five QTLs, qPH3C.H4.1, qPH6B.H4.1, qPH6B.H4.2, qPH7C.H4.2 and qPH7C.H4.3, were selected using backward selection ($p<0.05$) to fit a mixed main effect model for **PH** together with the environmental factor as the random effect (Table 4.5). The model with five QTLs and one environmental factor explains 63.2% of the total variance, with the QTL effects (fixed) explaining 24.3% of the total variance based on a modified method for calculating R-squared in the mixed model (Nakagawa and Schielzeth 2013). The random effects of the four environments from the mixed model are -34.20, -3.29, 35.47, 2.01 cm for Athens 2013, Athens 2014, Salina 2013 and Salina 2014 respectively, suggesting that the progeny plants were tallest in Salina 2013 and shortest in Athens 2013 among the four environments.

In the H6-derived population, we detected a total of 21 **PH** QTLs by interval mapping, with 11 significant in more than one environment (Table 4.4). Sixteen QTLs show positive additive effects, with five, qPH4A.H6.1, qPH.4C.H6.1, qPH.4C.H6.2, qPH.8A.H6.1 and qPH.9C.H6.1 showing negative allele substitution effects, indicating that *S. halepense* alleles decrease plant height at these loci. qPH4C.H6.1, qPH8A.H6.1 and qPH9C.H6.1 might be reciprocal alleles (i.e. different homologs) to qPH4D.H6.1, qPH8B.H6.1 and qPH9A.H6.1 or

qPH9B.H6.1 based on comparing their physical positions. A mixed effect model using backward elimination (p=0.05) with the environment as a random effect selected a total of nine QTLs, qPH1A.H6.1, qPH.1C.H6.1, qPH.2B.H6.1, qPH.3B.H6.1, qPH.4A.H6.1, qPH.4D.H6.1, qPH.6B.H6.1, qPH.7D.H6.1 and qPH.9A.H6.1 (Table 4.6), with only one QTL (qPH.4D.H6.1) showing negative allele effect of *S. halepense*. This provides some evidence for our hypothesis of reciprocal QTLs, since only one QTL in each pair is significant in the mixed effect model. QTL effects (fixed) of this model explain about 22.4% of the total variance, while inclusion of the environmental factor (random) explains about 71.9% of the total variance. The random effects of the four environments from the mixed model are -28.73, -5.86, 28.77 and 5.82 for Athens 2013, Athens 2014, Salina 2013 and Salina 2014 respectively, suggesting again that the progeny plants were tallest in Salina 2013 and shortest in Athens 2013.

We detected a total of 243, 239, 89 and 78 significant SNP markers ($p<10^{-3}$) for **PH** in Athens 2013, Athens 2014, Salina 2013 and Salina 2014, respectively, with 34 markers significant in all four environments (Figure 4.3). In the H4 derived population, we inferred a total of eight QTLs on chromosomes 2, 3, 5, 6 (2), 7 (2), 9, all increasing **PH,** with one QTL, qPH2.1.H4 newly detected. In the H6-derived population, we detected a total of 14 QTLs on all chromosomes but chromosome 5, with five pairs of QTLs (reciprocal QTLs) on the same chromosome but with different effects (Table 4.7), possibly coming from homologs in this population. A total of two QTLs, qPH10.1.H6 and qPH10.2.H6 were newly detected from the single-marker analysis in the H6-derived population. Three previously mapped genes, *Dw1*, *Dw2* and *Dw3* (Yamaguchi et al. 2016; Multani et al. 2003; Morris et al. 2013) were re-identified in our single marker analysis, corresponding to the 'skyline' signals on the long arms of chromosomes 9, 6 and 7, indicating that *S. halepense* has wild-type alleles that increase plant

height at those loci (Figure 4.4). Many small effect signals were found across the genome (Table 4.7), which support recent findings that the genetic control of plant height in sorghum is quantitative with many more loci involved (Zhang et al. 2015; Lin et al. 1995; Hart et al. 2001; Brown et al. 2008; Murray et al. 2008; Ritter et al. 2008; Upadhyaya et al. 2012) than the canonical model based on four loci, *DW1-DW4* (Quinby and Karper 1954).

A total of four general **PH** QTL regions for the two SH-BC_1F_2 populations overlap, based on aligning their physical positions on the sorghum genome, including those regions near the *DW1* (qPH9B.H4.1 and qPH9B.H6.1) (Yamaguchi et al. 2016; Hilley et al. 2016b), *DW2* (qPH6B.H4.2 and qPH.6B.H6.1) and *DW3* loci (qPH.7C.H4.2, qPH7D.H4.1, qPH7B.H6.1 and qPH7D.H6.1) (Multani et al. 2003) and a fourth region on chromosome 3 (qPH3C.H4.1 and qPH3B.H6.1) at approximately 68-69 Mb. More QTLs are detected in the H6 than the H4 population, possibly because the short-day alleles on chromosome 6 have much less effect in the H6 than the H4 population.

Days to flowering

A total of three flowering time QTLs, qFL.1A.H4.1, qFL6B.H4.1 and qFL6B.H4.2, were detected by the interval mapping method from the H4-derived population, all significant in more than one environment (Table 4.8). *Sorghum halepense* alleles of both QTLs from chromosome 6B (qFL6B.H4.1 and qFL6B.H4.2) delay flowering while *S. halepense* alleles of qFL.1A.H4.1 accelerate flowering. A log likelihood plot (Figure 4.7) suggests two possible QTLs on chromosome 6B, and a 'scantwo' analysis in R/qtl for all four environments favors two QTLs on chromosome 6B, with likelihood peaks at 6-8 cM and 46-49 cM, corresponding to 0.9 and 40-43Mb in physical location. An average LOD score is 3.58 when comparing the full model to the single QTL model, and 3.02 when comparing the additive model to the single QTL model (data

not shown). We obtained a mixed effect model adding all three QTLs, qFL.1A.H4.1, qFL6B.H4.1 and qFL6B.H4.2, as the fixed effects significant at an alpha level of 0.05 and the environment as a random effect (Table 4.9). QTL by environment interactions are not significant for **FL** in H4 for this analysis. The fixed effect explains about 28.94% of the total variance, while the model including both fixed and random effects explains about 49.74% of the variance. This indicates that flowering time varies substantially between years, and the random effects of environments are -1.86, -3.97, 13.29, -7.45 days, suggesting that **FL** in Salina 2013 took much longer than any other environment (also see Table 4.1).

We detected a total of 12 QTLs for **FL** by interval mapping in the H6 population with 5 significant in more than one environment (Table 4.8). *Sorghum halepense* alleles accelerate flowering at qFL.4A.H6.1, qFL.4A.H6.2, qFL.8A.H6.1 and qFL.10A.H6.1, and delay flowering at the rest of the QTLs. qFL8A.H6.1 and qFL8B.H6.1 might be reciprocal QTLs. We suspect that there are more than one QTL on chromosome 6B (as was found in the H6 population), however the 'scantwo' result does not formally support a two-QTL model (data not shown). We obtained a mixed effect model with a total of 8 QTLs, qFL.1C.H6.1, qFL.2C.H6.1, qFL.4A.H6.1, qFL.4A.H6.2, qFL.6B.H6.1, qFL.6B.H6.2, qFL.8A.H6.1, qFL.10A.H6.1 (Table 4.10), as the fixed effects significant at an alpha level of 0.05 and environment as the random effect. The QTL (fixed) effects explain about 25.77% of the total phenotypic variance, while the QTLs (fixed) and the environment (random) effect collectively explain about 52.34% of the total phenotypic variance. The random effect estimates from the mixed model are 0.49, -5.85, 10.71, -5.35 for the four environments, suggesting again that **FL** is much later in Salina 2013 than any other environment.

We detected totals of 118, 84, 94 and 94 significant SNP markers ($p<10^{-3}$) for **FL** in Athens 2013, Athens 2014, Salina 2013 and Salina 2014, respectively, with 54 SNP markers significant in all four environments (

Figure 4.5). The overall distribution of pooled data for **FL** suggested two major peaks on sorghum chromosome 6, with peak markers being S6_941772 and S6_42153422; and one peak on sorghum chromosome 1 at SNP marker S1_20362820 (Figure 4.6). Average *S. halepense* allele effects delay flowering on chromosome 6 but accelerate flowering on chromosome 1. The hierarchical clustering of significant markers suggested a total of four possible QTLs, one on chromosome 1 and three on chromosome 6 in the H4-derived population (Table 4.11), all overlapping with QTLs detected from the interval mapping. Allele substitution effects are negative on chromosome 1 but positive on chromosome 6. The three QTLs on chromosome 6 overlap based on their physical positions, therefore might be reciprocal QTLs. A total of seven possible QTLs were significant in the H6-derived population, distributed on chromosomes 1 (2), 4 (2), 6 (2), and 10, with two QTLs, qFL1.1.H6 and qFL10.1.H6 newly detected for the single marker analsyis. It is interesting that chromosomes 1 and 4 contain QTLs both accelerating and delaying flowering time, and these are probably not reciprocal QTLs based on their distant physical positions.

In partial summary, QTL results for **FL** reveal large differences between the two SH-BC_1F_2 populations. The H4 population seems to be dominated by two QTLs on linkage group 6B with only one other QTL detected on linkage group 1A. The QTLs on linkage group 6B in the H6-derived population show more subtle effects and explain less phenotypic variance, perhaps contributing to our ability to detect nine more QTLs on other linkage groups. Both populations showed QTL effects in both directions, which might explain the transgressive

segregation for flowering time observed in the SH-BC_1F_2 population. It also suggests that while *S. halepense* itself flowers rapidly, it can nonetheless contain late flowering alleles that are unmasked in segregating populations. However, the *S. halepense* QTL allele contributing to early flowering in the H4 population, qFL.1A.H4.1, is different from those in the H6 population, which are on chromosomes 4 and 8. We detected multiple QTLs on sorghum chromosome 6 which presumably harbors two flowering genes, with qFL.6B.H4.2 and qFL6B.H6.1 mapping near the location of *Ma1* (Cuevas et al. 2016; Murphy et al. 2011) and qFL6B.H4.1 mapping near *Ma6* (Murphy et al. 2014).

QTL correspondence across traits in the BC_1F_2 population

QTL correspondence among traits may be related to genes with pleiotropic effects, or multiple closely-linked QTLs responsible for different traits. **PH** and **FL** in the two BC_1F_2 populations are significantly correlated (Figure 4.1), indicating that we might discover genomic regions affecting both traits. In the H4-derived population, two QTLs conferring **FL** on chromosome 6B (qFL6B.H4.1 and qFL6B.H4.2) overlap with two plant height QTLs (qPH6B.H4.1 and qPH6B.H4.2). The correspondence of qFL6B.H4.2 and qPH6B.H4.2 may reflect the well-known close linkage between the *Ma1* and *Dw2* genes (Morris et al. 2013; Cuevas et al. 2016; Lin et al. 1995), and the other pair (qFL6B.H4.1 and qPH.H4.1) might suggest a similar case, possibly associated with *Ma6* and *Dw4*. In the H6-derived population, we detected a total of eight regions showing QTL correspondence between days to flowering and plant height on linkage groups 1C, 4A, 4D, 6B, 6E, 7D and 8A based on the interval based mapping (Figure 4.8), and one additional pair chromosome 10 based on single marker analysis (Table 4.7 and Table 4.11). Interestingly, one pair of overlapping QTLs qFL.4A.H6.1 and qPH.4A.H6.1 show opposite effects compared to qFL.4D.H6.1 and qPH.4D.H6.1, with the first

pair of **FL** and **PH** QTLs delaying flowering time and decreasing plant height and the second pair of **FL** and **PH** QTLs accelerating flowering time and increasing plant height.

Comparison to two other sorghum populations

We compared the QTL results from the SH-BC_1F_2 populations with two other sorghum populations sharing BTx623 as a common parent, specifically one RIL population (IS-RIL) derived from crossing *S. bicolor* BTx623 × IS3620C, and another RIL population derived from crossing *S. bicolor* BTx623 × *S. propinquum* (PQ-RIL) (Table 4.14 Table 4.15). For **PH**, three general genomic regions conferring *Dw1, Dw2* and *Dw3* on chromosomes 7, 6, and 9 respectively, overlap between at least two populations with all three showing significant signals at the *Dw1* region on chromosome 7 and regions conferring the *Dw2* and the *Dw3* significant in both SH-BC_1F_2 and IS-RIL populations. Two additional PH QTL regions overlap on chromosomes 2 and 10 between the SBSH-BC_1F_2 population and IS-RIL population (Table 4.14).

A total of five genomic regions, on sorghum chromosome 1, 4 (2), 6 and 10, showed significant association with **FL** in more than one population. Both IS-RIL and PQ-RIL populations lack the short-day *Ma1* allele on chromosome 6 (IS3620c is a 'converted' sorghum, and the PQ-RIL population was advanced under day-neutral photoperiod in Lubbock TX), and accordingly we find no QTL correspondence in this region. Overlapping regions on chromosome 1, qFL.1A.H4.1 and qFL1.1 from ISRIL, might correspond to the sorghum *Ehd1* gene, which is thought to activate *Ft* expression (Doi et al. 2004; Murphy et al. 2011). Interestingly, the BTx623 allele for qFL.1A.H4.1 delays flowering in the BC_1F_2 population while that for qFL1.1 accelerates flowering in the IS-RIL population, again suggesting three alleles with *S. halepense* conferring the most rapid flowering. In addition, qFL4A.H6.1 and qFL4D.H6.1 on sorghum chromosome 4 express different allele effects, corresponding to qFL4.1

from PQ-RIL and qFL4.1 from IS-RIL, respectively. BTx623 alleles of qFL4A.H6.1 delay flowering time in the BC_1F_2 population but of qFL4.1 from PQ-RIL accelerate PQRIL flowering, again consistent with three alleles with *S. halepense* conferring the most rapid flowering. However, for qFL4D.H6.1 and qFL4.1; and qFL10.H6.1 and qFL10.1 from SBSH-BC_1F_2 and IS-RIL respectively, BTx623 alleles confer early flowering (Table 4.15 and Figure 4.8) in both cases.

A total of 7 and 13 **PH** QTLs were unique to the H4 and H6 derived SBSH BC_1F_2 populations, respectively (Table 4.14). In addition, a total of six flowering QTLs, qFL1C.H6.1, qFL2C.H6.1, qFL7D.H6.1, qFL8A.H6.1, qFL8B.H6.1 and qFL10A.H6.1, were unique to the H6-derived BC_1F_2 population, indicating that those QTL alleles may have arisen during the radiation of *S. halepense* (Table 4.15). The fact that we detect more **PH** than **FL** QTL, and that more **PH** than **FL** QTL are unique to the SBSH cross, may suggest that the genetic control of sorghum **FL** may be more conserved than that of **PH**. A total of three flowering QTLs, qFL1.2, qFL3.1, qFL8.1, and two plant height QTLs, qPH3.1 and qPH8.1 are unique to the IS3620C population, (Table 4.12 and Figure 4.8), while none of the PQ-RIL QTLs are unique.

Discussion

The present study offers several new dimensions to knowledge of the *Sorghum* genus. First, it provides early insight into trait control and QTL polymorphism in *S. halepense*, one of the world's most important agricultural weeds and also an invasive plant now distributed over six continents. Second, the comparison of *S. halepense*, its progenitor *S. propinquum*, and a divergent form of *S. bicolor*, each crossed to the *S. bicolor* genotype from which the reference genome is derived, provides insight into the extent of genetic novelty that may have been

associated with the evolution of polyploid *S. halepense* following an estimated 96 million years of abstinence from polyploidy in the sorghum lineage (Wang et al. 2015).

QTL allele polymorphism in S. halepense

The properties of the two SH-BC_1F_2 populations are quite different, reflecting QTL allele polymorphism differentiating the two F_1 source plants obtained by crossing *S. bicolor* inbred line BTx623 to *S. halepense* accession Gypsum 9E (Figure 4.8), indicating a high level of polymorphism in *S. halepense*. Much of this difference appears to relate to the very strong effect of the chromosome 6 flowering genes in the H4-derived BC_1F_2 population, putatively *Ma1* and *Ma6* (see below). The striking effect of the chromosome 6 flowering genes in the H4-derived BC_1F_2 population might mask small effect flowering QTLs, accounting for the much larger number we found in the H6-derived BC_1F_2 population.

SMA vs QTL mapping

We proposed a novel and fast method to visualize and characterize relationships of significant SNPs with heat maps and hierarchical clustering. Constructing genetic maps in these two largely auto-tetraploid populations derived from a heterozygous parent is relatively labor-intensive and challenging in accurate genotyping calling, requiring a high depth of coverage to call heterozygosity and separate linkage groups. The single-marker analysis method is an attractive alternative to identify significant SNPs associated with traits of interest. However, SNP signals tend to be dispersed along the 10 sorghum chromosomes in this study due to doubled-ploidy of *S. halepense*. Re-grouping the significant SNP signals is especially useful in categorizing SNPs with different signs of effect, separating different QTLs from different homologs, and visualizing some small chromosomal differences between *S. bicolor* and *S.*

halepense. Single marker analysis may also detect additional potential QTL signals that eluded interval mapping, especially where portions of chromosomes are not included in the genetic map.

We detected a total of three QTLs for PH and two for FL by single marker analysis in addition to those found by interval based mapping, and the majority of QTLs detected by the GWAS-based method found their counterparts in interval mapping. This result suggests that our newly developed method functioned reasonably well in detecting and resolving the relationships of QTLs on different homologs.

QTL mapping

In this experiment, we conducted QTL analysis of two traits with relatively high heritability estimates, **PH** and **FL,** finding many QTLs significant in multiple environments (Table 4.7 and Table 4.11) . Despite high heritability estimates, environmental factors explain large portions of phenotypic variance in mixed effect models (Table 4.5,Table 4.6,Table 4.9 and Table 4.10).

The largest effect QTL for flowering time is on chromosome 6, and is especially pronounced in the H4 population where each of two chr. 6 QTLs have LOD scores greater than 10. Those two QTLs, qFL6B.H4.1 and qFL6B.H4.2 presumably correspond to the sorghum *Ma6* and *Ma1* genes (Cuevas et al. 2016; Murphy et al. 2011; Brady 2006). *Ma1* in particular is of great interest – tacitly assuming that *S. halepense* formed from progenitors resembling the wild sorghums of today, one would postulate that it had four(!) copies of the dominant *Ma1* allele conferring short-day flowering. In principle, this would delay flowering until day lengths drop below 12.5 hours, during September in the latitude of the study site. However, Gypsum 9E, and indeed an entire diversity panel of 599 *S. halepense* accessions from across the USA, flowered throughout the long days of the temperate summer. Moreover, *S. halepense* flowered

~14 days earlier than day-neutral *S. bicolor* BTx623. The difference in flowering time between the two SH-BC_1F_2 populations appears to suggest either polymorphism for *Ma1* itself, or for some interacting factor that permits day-neutral flowering, a question for further study.

Other than the two flowering QTLs on chromosome 6, only one QTL, qFL1A.H4.1, was significant in the H4-derived population, suggesting that the large effect QTLs mask the effects of QTLs on other chromosomes. QTLs on linkage group 6B have smaller LOD values in the H6 than the H4-derived population (Table 4.8), in which an additional 12 flowering QTLs were found. Across the two populations, a total of five QTLs, qFL.1A.H4.1 (overlapping with qFL1.H6.1 from single marker analysis), qFL.4A.H6.1, qFL.4A.H6.2, qFL8A.H6.1 and qFL.10A.H6.1 have *S. halepense* alleles accelerating flowering time. Thus, *S. halepense* harbors both alleles delaying flowering and alleles that accelerate it, accounting for the transgressive segregation of **FL** in progeny populations and potentially offering alleles that may be useful in breeding sorghum for specific environments.

In addition to re-identification of previously detected height genes *Dw1-Dw3*, we detected 7 and 17 more QTLs for plant height in the H4 and H6-derived populations, respectively, supporting a recent model indicating many genes with small effects (Zhang et al. 2015) rather than the classical model with only four genes controlling plant height (Quinby and Karper (1954)). The progeny lines were much taller than either parent, suggesting that *S. bicolor* and *S. halepense* might contribute different sets of genes for plant height to their progenies, or that height in early generations may be a result of heterosis.

QTL correspondence

QTL co-localization may suggest a single gene with pleiotropic effect influencing multiple traits, or different genes that are tightly linked either by chance or due to selection

during sorghum evolution. A well-known example is the *Ma1* and *Dw2* genes linked in the large heterochromatin region on chromosome 6 (Morris et al. 2013; Zhang et al. 2015; Cuevas et al. 2016). Two additional genomic locations suggest correlations between plant height and flowering time, one in the region of 6-46 Mb on chromosome 6, presumably reflecting *Ma6* and *Dw4*; and the other in the region of 56-59.5 Mb on chromosome 9 (Zhang et al. 2015). In our study, we validated these two regions on chromosome 6 and discovered six additional regions with QTL likelihood peaks at ~66.7Mb on chromosome 1C, ~4.3Mb and 61-64Mb on chromosome 4A and 4D respectively, ~50.2Mb on chromosome 6E, 57-58Mb on chromosome 7D and ~44.7Mb on chromosomes 8A, conferring QTLs for **PH** and **FL** in the H6 population. High correlation between PH and FL (Figure 4.1) may also contribute to the discovery of co-localized QTLs.

Correspondence of QTL regions between three populations sharing *S. bicolor* BTx623 as a common parent, with the other parents being morphologically and genetically distinct genotypes that represent cultivated (IS3620C), wild diploid (*S. propinquum*) and wild polyploid (*S halepense*) sorghums, provides information about common QTLs shared between or among populations and taxon-specific QTLs that contribute to divergence. Genomic regions conferring previously characterized plant height genes, *Dw1*, *Dw2* and *Dw3*, have been validated in the three sorghum populations, with both the SH-BC_1F_2 populations and the IS-RIL population segregating for all three genes, and the PQ-RIL population detecting *Dw3* (and with *Dw2* probably largely eliminated due to its close linkage to *Ma1*). Many additional QTLs for **PH** and **FL** significant in the BC_1F_2 but not in the IS-RIL or PQ-RIL populations may suggest the divergent control of plant height or inadequate statistical power to detect some QTLs in some populations. Similarly, five regions controlling **FL** corresponded in the BC_1F_2, IS-RIL and PQ-

RIL populations. Six regions might harbor the known genes controlling flowering in sorghum (Table 4.8; (Murphy et al. 2011; Cuevas et al. 2016; Yang et al. 2014a; Yang et al. 2014b; Wolabu et al. 2016). Despite of the fact that IS-RIL population has higher resolution and power to detect QTLs due to larger sample size, we detected more novel QTLs in the two SBSH- BC_1F_2 populations than any other two populations (Figure 4.8, Table 4.12 Table 4.15), demonstrating the genetic novelty arisen during the diversification of *S. halepense*. Lack of unique QTLs in the PQ-RIL population may be mainly due to a relatively low resolution genetic map with simple sequence repeat (SSR) markers and a smaller sample size compared to the IS-RIL.

Acknowledgments

We appreciate the support of the USDA Biotechnology Risk Assessment Program (2012-01658 to AHP and TSC), USAID Feed The Future (AID-OAA-A-13-00044 to AHP, TSC) program, and NIFA Global Food Security CAP (2015-68004-23492 to AHP, JNB). We thank members of the Plant Genome Mapping Laboratory (PGML) for help with field work.

References

Aljanabi, S.M., L. Forget, and A. Dookun, 1999 An improved and rapid protocol for the isolation of polysaccharide- and polyphenol-free sugarcane DNA. *Plant Molecular Biology Reporter* 17 (3):281-281.

Andolfatto, P., D. Davison, D. Erezyilmaz, T.T. Hu, J. Mast *et al.*, 2011 Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Research* 21 (4):610-617.

Bishop, G.J., 2003 Brassinosteroid Mutants of Crops. *J Plant Growth Regul* 22 (4):325-335.

Brady, J.A., 2006 Sorghum Ma5 and Ma6 maturity genes. Texas A&M University.

Broman, K.W., H. Wu, S. Sen, and G.A. Churchill, 2003 R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19 (7):889-890.

Brown, P.J., W.L. Rooney, C. Franks, and S. Kresovich, 2008 Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes. *Genetics* 180 (1):629-637.

Childs, K.L., F.R. Miller, M.M. Cordonnier-Pratt, L.H. Pratt, P.W. Morgan *et al.*, 1997 The sorghum photoperiod sensitivity gene, Ma3, encodes a phytochrome B. *Plant Physiol* 113 (2):611-619.

Cuevas, H.E., C. Zhou, H. Tang, P.P. Khadke, S. Das *et al.*, 2016 The Evolution of Photoperiod-Insensitive Flowering in Sorghum, A Genomic Model for Panicoid Grasses. *Molecular Biology and Evolution* 33 (9):2417-2428.

Doi, K., T. Izawa, T. Fuse, U. Yamanouchi, T. Kubo *et al.*, 2004 Ehd1, a B-type response regulator in rice, confers short-day promotion of flowering and controls FT-like gene expression independently of Hd1. *Genes Dev* 18 (8):926-936.

Fernandez, M.G.S., P.W. Becraft, Y. Yin, and T. Lübberstedt, 2009 From dwarves to giants? Plant height manipulation for biomass yield. *Trends in Plant Science* 14 (8):454-461.

Hart, G.E., K.F. Schertz, Y. Peng, and N.H. Syed, 2001 Genetic mapping of Sorghum bicolor (L.) Moench QTLs that control variation in tillering and other morphological characters. *Theoretical and Applied Genetics* 103 (8):1232-1242.

Higgins, R.H., C.S. Thurber, I. Assaranurak, and P.J. Brown, 2014 Multiparental Mapping of Plant Height and Flowering Time QTL in Partially Isogenic Sorghum Families. *G3-Genes Genomes Genetics* 4 (9):1593-1602.

Hill, C.B., and C. Li, 2016 Genetic Architecture of Flowering Phenology in Cereals and Opportunities for Crop Improvement. *Frontiers in Plant Science* 7:1906.

Hilley, J., S. Truong, S. Olson, D. Morishige, and J. Mullet, 2016a Identification of Dw1, a Regulator of Sorghum Stem Internode Length. *Plos One* 11 (3).

Hilley, J., S. Truong, S. Olson, D. Morishige, and J. Mullet, 2016b Identification of Dw1, a Regulator of Sorghum Stem Internode Length. *Plos One* 11 (3):e0151271.

Hilley, J.L., B.D. Weers, S.K. Truong, R.F. McCormick, A.J. Mattison *et al.*, 2017 Sorghum Dw2 Encodes a Protein Kinase Regulator of Stem Internode Length. *Sci Rep* 7 (1):4616.

Jung, C., and A.E. Müller, 2009 Review: Flowering time control and applications in plant breeding. *Trends in Plant Science* 14:563-573.

Kong, L., J. Dong, and G.E. Hart, 2000 Characteristics, linkage-map positions, and allelic differentiation of Sorghum bicolor (L.) Moench DNA simple-sequence repeats (SSRs). *Theoretical and Applied Genetics* 101 (3):438-448.

Kong, W., H. Guo, V. Goff, T.-H. Lee, C. Kim *et al.*, 2014 Genetic analysis of vegetative branching in sorghum. *Theoretical and Applied Genetics* 127 (11):2387-2403.

Kong, W., H. Jin, C.D. Franks, C. Kim, R. Bandopadhyay *et al.*, 2013 Genetic Analysis of Recombinant Inbred Lines for *Sorghum bicolor* × *Sorghum propinquum*. *G3: Genes*| *Genomes*| *Genetics* 3:101-108.

Ku, L.X., L.K. Zhang, Z.Q. Tian, S.L. Guo, H.H. Su *et al.*, 2015 Dissection of the genetic architecture underlying the plant density response by mapping plant height-related traits in maize (Zea mays L.). *Molecular Genetics and Genomics* 290 (4):1223-1233.

Kutschera, U., and Z.Y. Wang, 2012 Brassinosteroid action in flowering plants: a Darwinian perspective. *J Exp Bot* 63 (10):3511-3522.

Lander, E.S., and D. Botstein, 1989 Mapping Mendelian Factors Underlying Quantitative Traits Using Rflp Linkage Maps. *Genetics* 121 (1):185-199.

Li, X., X.R. Li, E. Fridman, T.T. Tesso, and J.M. Yu, 2015 Dissecting repulsion linkage in the dwarfing gene Dw3 region for sorghum plant height provides insights into heterosis. *Proceedings of the National Academy of Sciences of the United States of America* 112 (38):11823-11828.

Lin, Y.R., K.F. Schertz, and A.H. Paterson, 1995 Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population. *Genetics* 141 (1):391-411.

Morris, G.P., P. Ramu, S.P. Deshpande, C.T. Hash, T. Shah *et al.*, 2013 Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences of the United States of America* 110 (2):453-458.

Multani, D.S., S.P. Briggs, M.A. Chamberlin, J.J. Blakeslee, A.S. Murphy *et al.*, 2003 Loss of an MDR transporter in compact stalks of maize br2 and sorghum dw3 mutants. *Science* 302 (5642):81-84.

Murphy, R.L., R.R. Klein, D.T. Morishige, J.A. Brady, W.L. Rooney *et al.*, 2011 Coincident light and clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum. *Proceedings of the National Academy of Sciences of the United States of America* 108 (39):16469-16474.

Murphy, R.L., D.T. Morishige, J.A. Brady, W.L. Rooney, S.S. Yang *et al.*, 2014 Ghd7 (Ma(6)) Represses Sorghum Flowering in Long Days: Ghd7 Alleles Enhance Biomass Accumulation and Grain Production. *Plant Genome* 7 (2).

Murray, S.C., W.L. Rooney, M.T. Hamblin, S.E. Mitchell, and S. Kresovich, 2009 Sweet Sorghum Genetic Diversity and Association Mapping for Brix and Height. *Plant Genome* 2 (1):48-62.

Murray, S.C., W.L. Rooney, S.E. Mitchell, A. Sharma, P.E. Klein *et al.*, 2008 Genetic Improvement of Sorghum as a Biofuel Feedstock: II. QTL for Stem and Leaf Structural Carbohydrates. *Crop Science* 48 (6):2180-2193.

Nakagawa, S., and H. Schielzeth, 2013 A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4 (2):133-142.

Paterson, A., H., J.E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood *et al.*, 2009 The Sorghum bicolor genome and the diversification of grasses *Nature* 457:551-556.

Peiffer, J.A., M.C. Romay, M.A. Gore, S.A. Flint-Garcia, Z.W. Zhang *et al.*, 2014 The Genetic Architecture Of Maize Height. *Genetics* 196 (4):1337-+.

Peng, J., D.E. Richards, N.M. Hartley, G.P. Murphy, K.M. Devos *et al.*, 1999 'Green revolution' genes encode mutant gibberellin response modulators. *Nature* 400 (6741):256-261.

Quinby, J., 1966 Fourth maturity gene locus in sorghum. *Crop Science* 6 (6):516-518.

Quinby, J., and R. Karper, 1945a Inheritance of three genes that influence time of floral initiation and maturity date in milo. *Journal of the American Society of Agronomy*.

Quinby, J., and R. Karper, 1945b The inheritance of three genes that influence time of floral initiation and maturity date in milo. *Jour. Amer. Soc. Agron.* 37 (916).

Quinby, J.R., and R.E. Karper, 1954 Inheritance of Height in Sorghum. *Agronomy Journal* 46 (5):211-216.

R Core Team, 2016 R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ritter, K.B., D.R. Jordan, S.C. Chapman, I.D. Godwin, E.S. Mace *et al.*, 2008 Identification of QTL for sugar-related traits in a sweet x grain sorghum (Sorghum bicolor L. Moench) recombinant inbred population. *Molecular Breeding* 22 (3):367-384.

Salas Fernandez, M.G., P.W. Becraft, Y. Yin, and T. Lubberstedt, 2009 From dwarves to giants? Plant height manipulation for biomass yield. *Trends Plant Sci* 14 (8):454-461.

Sasaki, A., M. Ashikari, M. Ueguchi-Tanaka, H. Itoh, A. Nishimura *et al.*, 2002 Green revolution: a mutant gibberellin-synthesis gene in rice. *Nature* 416 (6882):701-702.

Upadhyaya, H.D., Y.-H. Wang, S. Sharma, and S. Singh, 2012 Association mapping of height and maturity across five environments using the sorghum mini core collection. *Genome* 55 (6):471-479.

Wang, X., J. Wang, D. Jin, H. Guo, T.H. Lee *et al.*, 2015 Genome Alignment Spanning Major Poaceae Lineages Reveals Heterogeneous Evolutionary Rates and Alters Inferred Dates for Key Evolutionary Events. *Mol Plant* 8 (6):885-898.

Wang, Y., J. Zhao, W. Lu, and D. Deng, 2017 Gibberellin in plant height control: old player, new story. *Plant Cell Rep* 36 (3):391-398.

Wang, Y.H., and J.Y. Li, 2006 Genes controlling plant architecture. *Current Opinion in Biotechnology* 17 (2):123-129.

Warnes, G.R., B. Bolker, L. Bonebakker, R. Gentleman, W. LIaw *et al.*, 2016 gplots: Various R Programming Tools for Plotting Data.

Wolabu, T.W., F. Zhang, L.F. Niu, S. Kalve, P. Bhatnagar-Mathur *et al.*, 2016 Three FLOWERING LOCUS T-like genes function as potential florigens and mediate photoperiod response in sorghum. *New Phytologist* 210 (3):946-959.

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

Yang, S.S., R.L. Murphy, D.T. Morishige, P.E. Klein, W.L. Rooney *et al.*, 2014a Sorghum Phytochrome B Inhibits Flowering in Long Days by Activating Expression of SbPRR37 and SbGHD7, Repressors of SbEHD1, SbCN8 and SbCN12. *Plos One* 9 (8).

Yang, S.S., B.D. Weers, D.T. Morishige, and J.E. Mullet, 2014b CONSTANS is a photoperiod regulated activator of flowering in sorghum. *BMC Plant Biology* 14.

Zhang, D., H. Guo, C. Kim, T.H. Lee, J.P. Li *et al.*, 2013 CSGRqtl, a Comparative Quantitative Trait Locus Database for Saccharinae Grasses. *Plant Physiology* 161 (2):594-599.

Zhang, D., W. Kong, J. Robertson, V.H. Goff, E. Epps *et al.*, 2015 Genetic analysis of inflorescence and plant height components in sorghum (Panicoidae) and comparative genetics with rice (Oryzoidae). *BMC Plant Biol* 15:107.

Chapter 4 Tables and Figures

Table 4.1 Summary statistics for flowering time (FL), plant height (PH), number of mature tillers (TL) and number of secondary branches (BRCH) in the SH-BC_1F_2 [*S. halepense* derived (*S. bicolor* BTx623× *S. halepense* G9E) backcross] population and parents.

			SH-BC1F2							BTx623 (2x)			BTx623 (4x)			
Trait	Location	Year	Population	N	Mean	Median	SD	Min	Max	N	Mean	SD	N	Mean	SD	Heritability (%)
PH	Athens	2013	Pooled	246	213.1	207.6	34.27	128.6	329.9	12	121.1	18.42	18	157.8	47.79	71.66
PH	Athens	2013	H4	141	212.4	207.0	34.08	144.6	300.7							
PH	Athens	2013	H6	105	214.0	212.4	34.67	128.6	329.9							
PH	Athens	2014	Pooled	236	244.4	242.3	38.78	155.8	342.7	8	113.6	9.58	12	134.6	20.11	
PH	Athens	2014	H4	134	243.7	246.8	41.11	155.8	342.7							
PH	Athens	2014	H6	102	245.3	239.8	35.65	180.8	336.7							
PH	Salina	2013	Pooled	246	287.2	288.3	34.08	186.7	360.0	12	144.2	39.19	12	193.3	52.80	
PH	Salina	2013	H4	141	283.2	285.0	36.36	186.7	360.0							
PH	Salina	2013	H6	105	292.6	291.7	30.10	223.3	358.3							
PH	Salina	2014	Pooled	236	254.4	255.0	32.21	178.0	335.0	12	118.3	7.99	12	155.0	24.32	
PH	Salina	2014	H4	134	249.1	249.2	32.72	178.0	335.0							
PH	Salina	2014	H6	102	261.3	261.7	30.34	185.0	323.3							
FL	Athens	2013	Pooled	245	75.1	74.0	11.45	51.0	115.0	3	73.0	4.00	6	81.3	8.76	83.59
FL	Athens	2013	H4	140	75.2	74.0	11.95	53.5	115.0							
FL	Athens	2013	H6	105	75.0	73.0	10.79	51.0	103.5							
FL	Athens	2014	Pooled	241	72.3	70.0	11.72	52.0	107.5	3	65.3	2.08	4	68.5	3.42	
FL	Athens	2014	H4	138	73.8	71.5	12.74	52.5	107.5							
FL	Athens	2014	H6	103	70.2	69.0	9.90	52.0	106.5							
FL	Salina	2013	Pooled	246	84.1	83.0	9.77	63.5	121.5	4	76.5	1.73	4	86.0	1.41	
FL	Salina	2013	H4	141	85.1	84.5	10.61	63.5	121.5							
FL	Salina	2013	H6	105	82.7	82.0	8.36	67.5	119.0							

FL	Salina	2014	Pooled	236	71.1	70.5	9.90	49.0	94.0	4	62.0	1.41	4	67.5	1.91	
FL	Salina	2014	H4	134	71.5	70.3	10.60	49.0	94.0							
FL	Salina	2014	H6	102	70.6	70.5	8.92	51.0	93.5							

Table 4.2 Summary statistics for flowering time (FL), plant height (PH), number of mature tillers (TL) and number of secondary branches (BRCH) in the PQ-RIL [propinquum derived (*S. bicolor* BTx623× *S. propinquum*) recombinant inbred line] population and parents.

		PQ-RIL						BTx623			
Trait	Year	N	Mean	Median	SD	Min	Max	N	Mean	SD	Heritability (%)
FL	2009	157	71.8	68.8	16.56	39.0	124.7	10	65.8	2.54	63.66
FL	2010	134	64.7	63.3	9.72	46.2	97.0	7	59.4	4.35	
FL	2011	130	64.1	62.3	9.75	40.4	86.5	8	64.6	3.10	
PH	2009	157	155.8	153.0	38.09	29.0	246.0	10	106.3	1.95	77.03
PH	2010	121	156.0	145.0	49.60	74.5	275.0	14	112.2	10.63	
PH	2011	140	135.8	131.5	34.76	64.3	228.0	20	100.9	8.27	

Table 4.3 Summary statistics for flowering time (FL), plant height (PH), number of mature tillers (TL) and number of secondary branches (BRCH) in the IS-RIL [IS3620C derived (*S. bicolor* BTx623× *S. bicolor* IS3620C) recombinant inbred line] population and parents.

	IS-RIL							BTx623			IS3620C			
Trait	Year	N	Mean	Median	SD	Min	Max	N	Mean	SD	N	Mean	SD	Heritability (%)
FL	2011	393	61.9	59.8	9.47	44.8	99.2	26	72.1	5.37	15	58.3	6.13	83.65
FL	2012	383	56.0	54.4	8.02	40.8	87.4	14	70.1	5.01	8	54.5	5.36	
PH	2011	388	97.1	93.2	20.99	45.0	175.3	16	95.5	7.24	13	80.8	8.03	78.25
PH	2012	384	100.0	95.0	24.50	54.2	181.0	12	92.9	8.83	16	87.5	9.11	

Table 4.4 Parameters of plant height QTLs from interval mapping of the H4 and H6 BC_1F_2 populations

QTL	Pos (cM)	Pos (Mb)	LOD	% of Variance Explained	Effect	Left(Mb)	Right(Mb)	Env
qPH.3C.H4.1	147.0	68.7	3.7	11.03	23.38	60.7	68.7	AT13
qPH.3C.H4.1	156.0	68.7	2.7	8.10	24.89	60.7	66.1	AT14
qPH.3C.H4.1	165.0	68.7	3.0	9.47	20.78	64.7	68.7	SL14
qPH.3D.H4.1	47.5	69.5	3.1	9.42	22.07	56.3	73.8	AT13
qPH.5C.H4.1	16.8	20.6	2.9	9.08	24.65	2.5	61.9	AT13
qPH.6B.H4.1	13.8	4.3	10.2	28.17	37.32	3.6	37.3	AT13
qPH.6B.H4.1	26.0	37.2	12.9	32.96	48.65	1.9	37.3	AT14
qPH.6B.H4.1	25.0	37.2	8.2	20.86	37.52	0.9	2.0	SL13
qPH.6B.H4.2	44.0	40.8-42.2	6.3	19.84	34.09	1.9	45.3	SL14
qPH.6C.H4.1	14.9	53.9	4.5	13.53	27.26	42.6	58.1	AT13
qPH.7A.H4.1	81.0	11.9-14.1	3.8	12.26	40.21	1.1	53.5	AT14
qPH.7A.H4.2	133.0	57.5	3.0	7.07	20.96	55.0	57.7	SL14
qPH.7C.H4.1	11.6	49.5	4.4	13.46	25.50	6.7	57.9	AT13
qPH.7C.H4.1	18.0	57.8-58.0	6.5	19.47	37.84	36.9	57.9	AT14
qPH.7C.H4.1	18.0	57.8-58.0	4.8	13.86	27.92	36.9	57.9	SL13
qPH.7C.H4.1	18.0	57.8-58.0	3.8	11.24	22.89	36.9	57.9	SL14
qPH.7C.H4.2	61.0	59.2	3.3	10.18	22.17	58.0	60.7	AT13
qPH.7C.H4.2	61.0	58.6	4.6	14.27	31.56	57.7	60.7	AT14
qPH.7C.H4.2	60.0	58.6	4.9	14.02	27.77	58.0	60.7	SL13
qPH.7C.H4.2	55.5	58.1	3.9	12.47	23.41	56.4	60.7	SL14
qPH.7D.H4.1	10.0	58.8	5.0	15.18	27.86	4.9	58.8	AT13
qPH.7D.H4.1	9.1	58.8	3.3	10.72	27.74	4.9	58.8	AT14
qPH.9B.H4.1	63.5	56.0	3.3	10.14	22.24	53.7	57.1	AT13
qPH.9B.H4.1	63.5	56.0	2.9	9.57	25.80	53.7	57.1	AT14
qPH.1A.H6.1	104.8	1.0-1.5	3.3	13.67	33.18	0.0	2.1	AT14
qPH.1A.H6.1	103.7	0.7-0.8	3.1	13.30	27.78	0.0	2.1	SL14
qPH.1B.H6.1	29.0	60.8	3.7	14.94	29.30	58.8	61.2	AT14

qPH.1C.H6.1	131.4	67.0	4.4	17.32	29.61	66.9	69.9	AT13
qPH.1C.H6.1	131.4	67.0	9.0	33.07	41.93	66.9	69.2	AT14
qPH.1C.H6.1	131.4	67.0	5.1	19.97	27.84	66.9	70.2	SL13
qPH.1C.H6.1	131.4	67.0	5.8	22.84	30.01	66.9	69.2	SL14
qPH.1D.H6.1	147.8	54.8-57.9	3.7	15.55	-30.05	51.1	65.9	SL14
qPH.2B.H6.1	93.6	69.7	3.1	12.32	26.50	66.4	69.7	AT13
qPH.2B.H6.2	10.0	61.9	4.5	16.72	30.50	59.8	62.1	AT14
qPH.2B.H6.2	29.0	62.1	2.5	10.08	20.15	61.9	65.7	SL13
qPH.3B.H6.1	139.0	69.1	2.7	11.58	26.20	68.2	69.5	AT13
qPH.3B.H6.1	139.0	69.1	3.6	15.04	30.13	68.2	69.5	AT14
qPH.3B.H6.1	143.0	69.1	2.9	12.18	22.75	68.7	69.5	SL13
qPH.3B.H6.2	91.0	60.9	4.2	16.68	30.61	53.5	63.0	SL14
qPH.4A.H6.1	85.3	4.3	3.5	13.78	-35.84	4.3	34.3	AT14
qPH.4C.H6.1	111.4	64.8	5.3	21.17	-35.14	11.3	60.8	AT14
qPH.4C.H6.2	49.0	20.7	5.0	18.82	-27.74	11.3	64.8	SL14
qPH.4D.H6.1	281.0	63.4	4.1	16.06	32.84	58.7	65.1	AT13
qPH.4D.H6.1	240.0	62.8	5.2	12.89	33.48	58.7	65.1	AT14
qPH.4D.H6.1	281.0	63.4	5.8	21.97	32.87	58.7	64.0	SL13
qPH.4D.H6.1	171.6	61.1	3.4	14.37	-33.92	1.3	61.1	SL14
qPH.6B.H6.1	79.0	40.8	7.7	27.13	37.57	37.2	47.0	AT13
qPH.6B.H6.1	70.7	41.8-42.2	11.0	39.28	45.00	37.2	42.2	AT14
qPH.6B.H6.1	71.0	41.8-42.1	6.5	25.17	30.71	37.2	42.2	SL13
qPH.6B.H6.1	70.7	41.8-42.1	5.5	22.10	28.69	37.2	42.2	SL14
qPH.6E.H6.1	187.0	50.2	3.7	14.49	34.57	48.7	58.7	AT14
qPH.7B.H6.1	0.0	58.1	3.9	16.07	29.00	58.0	59.2	AT14
qPH.7D.H6.1	137.7	57.9	4.9	20.02	34.16	37.9	58.6	AT14
qPH.7D.H6.1	137.7	57.9	2.7	11.35	21.66	11.7	58.6	SL13
qPH.8A.H6.1	257.0	44.7	5.1	15.39	-35.47	1.8	45.3	AT13
qPH.8A.H6.1	269.0	1.8	3.1	12.36	-28.60	1.8	45.3	AT14
qPH.8A.H6.1	256.0	1.8	4.1	12.88	-27.19	1.8	45.3	SL13

qPH.8A.H6.1	269.0	1.8	2.6	10.42	-22.38	1.8	45.3	SL14
qPH.8B.H6.1	43.0	6.1	4.9	14.83	29.03	5.2	6.1	AT13
qPH.8B.H6.1	45.0	5.3	5.0	18.11	31.24	1.0	45.7	AT14
qPH.8B.H6.1	37.0	6.1	3.3	11.74	21.72	6.0	45.7	SL14
qPH.9A.H6.1	168.8	53.9	3.0	12.27	31.66	49.4	53.7	AT13
qPH.9A.H6.1	168.8	53.9	3.8	15.68	36.32	49.4	53.7	AT14
qPH.9A.H6.1	170.0	52.9	4.6	18.91	33.91	49.4	53.7	SL14
qPH.9B.H6.1	95.5	53.6	2.6	10.16	19.94	53.5	55.1	SL13
qPH.9B.H6.1	94.0	53.6	3.7	14.51	24.43	53.5	55.1	SL14
qPH.9C.H6.1	146.5	47.0	3.5	14.30	-36.36	4.5	55.9	AT14
qPH.9C.H6.1	146.5	47.0	3.2	13.66	-30.16	2.3	55.9	SL14

Table 4.5 A mixed main effect model for plant height in the H4 derived population

	Sum Sq[1]	DF	Estimate[2]	F value	P value
qPH.3C.H4.1	22923	1	13.21	27.95	1.806e-07 ***
qPH.6B.H4.1	26783	1	16.66	32.66	1.815e-08 ***
qPH.6B.H4.2	37628	1	22.78	45.89	3.278e-11 ***
qPH.7C.H4.1	18812	1	15.02	22.94	2.161e-06 ***
qPH.7C.H4.2	10347	1	10.90	12.62	0.0004154 ***

Random effects:		
Source	Variance	SD
ENV	836.1	28.92
Residual	820.0	28.64

[1]Sum of squares of a single QTL in the mixed effect model.
[2]Estimated effects of allele substitution

Table 4.6 A mixed main effect model for plant height in the H6 population

	Sum Sq	DF	Estimates	F value	P value
qPH.1A.H6.1	8239	1	-19.53	14.28	0.0001814 ***
qPH.1C.H6.1	16008	1	16.20	27.75	2.267e-07 ***
qPH.2B.H6.2	4335	1	7.56	7.51	0.0063978 **
qPH.3B.H6.1	6518	1	9.87	11.30	0.0008510 ***
qPH.4A.H6.1	7605	1	-13.82	13.18	0.0003195 ***
qPH.4C.H6.2	3710	1	-7.91	6.43	0.0115983 *
qPH.4D.H6.1	10734	1	17.55	18.61	2.030e-05 ***
qPH.6B.H6.1	47525	1	25.83	82.38	< 2.2e-16 ***
qPH.7D.H6.1	5812	1	-12.35	10.07	0.0016203 **
qPH.8B.H6.1	1841	1	-5.32	3.19	0.0748162
qPH.9A.H6.1	5177	1	14.66	8.97	0.0029092 **

Random effects:		
Source	Variance	SD
ENV	1065.8	32.65
Residual	576.9	24.02

[1]Sum of squares of a single QTL in the mixed effect model.
[2]Estimated effects of allele substitution

Table 4.7 Possible plant height (PH) QTL detected by single marker analysis and hierarchical clustering

QTL	Peak SNP	Chr	Pos	P-vlaue	Effect	Left	Right
qPH2.1.H4	S2_41169589	2	41.2	0.0000525	30.60	S2_20718298	S2_69631146
qPH3.1.H4	S3_60878509	3	60.9	0.0000777	24.90	S3_2064439	S3_74033233
qPH5.1.H4	S5_54776509	5	54.8	0.00002451	38.31	S5_1733420	S5_56412809
qPH6.1.H4	S6_4292628	6	4.3	4.10E-07	40.70	S6_25386	S6_55609888
qPH6.1.H4	S6_42153422	6	42.2	5.67E-07	38.45	S6_1942522	S6_48598662
qPH7.1.H4	S7_57866027	7	57.9	2.32E-06	32.44	S7_6735921	S7_63484586
qPH7.2.H4	S7_58436356	7	58.4	5.89E-06	36.24	S7_7584188	S7_63881844
qPH9.1.H4	S9_55964590	9	56.0	0.000139313	30.73	S9_55444762	S9_57266873
qPH1.1.H6	S1_69985214	1	70.0	1.12E-06	-35.22	S1_8003716	S1_70832264
qPH1.2.H6	S1_71002531	1	71.0	1.30E-06	37.79	S1_732248	S1_73686219
qPH2.2.H6	S2_61673023	2	61.7	0.00000659	35.35	S2_5337548	S2_69921627
qPH3.1.H6	S3_68603744	3	68.6	6.44E-06	39.37	S3_2787432	S3_72459542
qPH4.1.H6	S4_66390904	4	66.4	3.30E-05	30.27	S4_59149048	S4_66947707
qPH4.2.H6	S4_65038519	4	65.0	2.24E-05	-32.64	S4_3893501	S4_67715604
qPH6.1.H6	S6_3185305	6	3.2	4.47E-07	37.65	S6_3185305	S6_58868689
qPH7.1.H6	S7_5178515	7	5.2	0.000019045	36.53	S7_439052	S7_58620694
qPH8.1.H6	S8_1751335	8	1.8	0.000051	36.98	S8_1293001	S8_6056426
qPH8.2.H6	S8_44723812	8	44.7	0.000021495	-38.43	S8_1567708	S8_45322860
qPH9.1.H6	S9_55760094	9	55.8	0.0000298	-41.70	S9_2962982	S9_56423964
qPH9.2.H6	S9_53579966	9	53.6	0.0000432	25.06	S9_47399246	S9_55968872
qPH10.1.H6	S10_51186443	10	51.2	0.000000471	54.10	S10_50060082	S10_56034268
qPH10.2.H6	S10_53649131	10	53.6	0.0000789	-30.74	S10_6265233	S10_60321156

Table 4.8 Parameters of days to flowering (FL) QTLs from interval mapping of the H4 and H6 BC_1F_2 populations

QTL name	Pos (cM)	Pos (Mb)	LOD	% Phenotyp Explaine d (Sginle QTL)	Effect	Left SNP*	Right SNP	Env
qFL.1A.H4.1	145.0	22.0	2.6	8.11	-9.27	17.27	57.61	AT13
qFL.1A.H4.1	146.0	22.0	3.1	8.89	-10.70	17.59	57.61	AT14
qFL.6B.H4.1	12.0	4.3	12.2	32.09	13.88	0.94	37.25	AT13
qFL.6B.H4.1	13.8	4.3	11.5	31.90	14.76	0.94	37.25	AT14
qFL.6B.H4.2	45.0	40.8-42.1	10.3	28.26	13.26	1.97	43.16	SL13
qFL.6B.H4.2	44.0	14.5-37.2	12.5	34.64	14.41	0.94	43.16	SL14
qFL.1C.H6.1	101.8	66.7	3.0	12.41	7.51	65.49	68.79	AT14
qFL.2C.H6.1	90.0	75.5	3.4	6.14	8.84	73.66	75.54	SL13
qFL.4A.H6.1	51.0	4.3	4.0	13.71	-9.60	4.10	5.07	AT13
qFL.4A.H6.1	55.0	4.3	3.1	11.44	-7.98	4.10	4.32	AT14
qFL.4A.H6.2	85.3	4.3	3.8	15.03	-9.38	5.07	34.25	SL13
qFL.4A.H6.1	51.0	4.3	6.0	20.14	-9.59	4.10	4.32	SL14
qFL.4D.H6.1	238.0	62.8	5.4	11.31	9.78	62.80	65.05	AT13
qFL.4D.H6.1	234.0	62.8	3.1	11.08	8.57	62.80	66.86	SL13
qFL.4D.H6.1	236.0	62.8	2.6	8.51	7.75	62.80	66.86	SL14
qFL.6B.H6.1	59.6	14.5-36.8	6.2	23.93	10.82	3.19	41.89	AT13
qFL.6B.H6.1	60.6	37.3	2.9	12.04	7.09	3.33	50.89	AT14
qFL.6B.H6.1	60.6	37.3	3.5	14.40	6.54	3.19	50.89	SL13
qFL.6B.H6.2	104.9	47.0	5.0	20.29	8.19	3.19	50.89	SL14
qFL.6E.H6.1	179.0	50.2	3.7	11.83	9.51	48.74	58.87	AT13
qFL.6E.H6.1	182.0	57.4	3.4	12.92	7.65	48.74	58.87	SL13
qFL.7D.H6.1	133.0	57.7	3.3	13.74	8.65	55.64	60.84	AT13
qFL.8A.H6.1	229.0	14.5-41.7	4.1	8.54	-9.75	14.56	45.32	SL13
qFL.8A.H6.1	258.0	44.7	3.0	9.34	-7.36	1.81	45.32	SL14

qFL.8B.H6.1	85.7	45.7	2.7	11.21	8.38	1.07	45.74	AT13
qFL.10A.H6.1	11.0	8.6-36.6	4.4	3.44	-6.73	8.60	36.58	SL13

*Corresponding to 1-lod interval

Table 4.9 A mixed effect model for days to flowering in the H4 derived population.

	Sum Sq	DF	Estimates	F value	P value
qFL.1A.H4.1	1162.8	1	-4.1322	13.642	0.0002435 ***
qFL.6B.H4.1	5320.1	1	7.3288	62.416	1.554e-14 ***
qFL.6B.H4.2	6281.1	1	9.1940	73.69	< 2.2e-16 ***

Random effects:		
Source	Variance	SD
Environment	34.94	5.911
Residual	84.24	9.232

[1]Sum of squares of a single QTL in the mixed effect model.
[2]Estimated effects of allele substitution

Table 4.10 A mixed effect model for days to flowering in the H6 derived population

	Sum Sq	DF	Estimates	F value	P value
qFL.1C.H6.1	502.20	1	2.5103	8.334	0.004100 **
qFL.2C.H6.1	643.56	1	5.6277	10.680	0.001176 **
qFL.4A.H6.1	1038.64	1	-3.9864	17.236	4.030e-05 ***
qFL.4A.H6.2	405.95	1	-3.2681	6.737	0.009789 **
qFL.6B.H6.1	1230.46	1	4.9764	20.420	8.184e-06 ***
qFL.6B.H6.2	355.87	1	2.5166	5.906	0.015527 *
qFL.8A.H6.1	589.51	1	-3.2735	9.783	0.001889 **
qFL.10A.H6.1	178.79	1	-2.1279	2.967	0.085747 .

Random effects:		
Source	Variance	SD
ENV	33.27	5.768
Residual	60.26	7.763

Table 4.11 Possible flowering time (**FL**) QTLs from single marker analysis and hierarchical clustering

QTL	Peak SNP	Chr	Pos	P-vlaue	Effect	Left	Right
qFL1.1.H4	S1_43188994	1	43.2	0.000014	-15.49	S1_11475519	S1_50276681
qFL6.1.H4	S6_45033138	6	45.0	2.11E-06	9.41	S6_5554081	S6_54550973
qFL6.2.H4	S6_36770500	6	36.8	4.08E-08	12.08	S6_1942522	S6_47239380
qFL6.3.H4	S6_42153422	6	42.2	8.18E-10	14.34	S6_25386	S6_49819552
qFL1.1.H6	S1_20949893	1	20.9	0.000133967	-8.67	S1_8003716	S1_54606054
qFL1.2.H6	S1_68442832	1	68.4	0.000192052	18.18	S1_25198125	S1_72932488
qFL4.1.H6	S4_66692086	4	66.7	0.0000514	8.57	S4_61046019	S4_66947707
qFL4.2.H6	S4_4207328	4	4.2	0.000003475	-12.58	S4_1404222	S4_8893237
qFL6.1.H6	S6_47106747	6	47.1	3.02E-06	10.97	S6_941772	S6_49964543
qFL6.2.H6	S6_45819310	6	45.8	2.54E-06	13.68	S6_885326	S6_59263777
qFL10.1.H6	S10_55779157	10	55.8	0.00008612	12.75	S10_14550404	S10_55779157

Table 4.12 Parameters of plant architecture related QTLs from interval mapping of the IS-RIL population (Supplimentary)

QTL	Peak (cM)	Peak (Mb)	LOD	% Var	Additive Effect	Left Flanking SNP Positions[1]	Right Flanking SNP Positions[1]
qFL1.1	70.0	23.1	9.6	6.37	1.56	22.0	26.2
qFL1.2[2]	128.0	58.2	8.1	5.35	-1.13	55.7	59.4
qFL3.1	116.6	62.5	8.9	5.88	-1.15	59.3	63.4
qFL4.1	151.0	64.5	5.4	3.51	0.92	63.7	65.8
qFL6.1	81.0	56.0	4.2	2.71	-0.78	55.5	57.9
qFL8.1	31.0	4.3	4.4	2.86	-0.79	2.7	6.2
qFL8.2	84.6	52.4	16.1	11.16	-1.70	52.3	52.8
qFL9.1	122.4	59.3	11.4	7.69	1.37	56.1	59.5
qFL10.1	91.0	55.9	5.5	3.60	0.89	55.2	56.5
qPH2.1	121.8	65.5	3.1	2.20	2.37	63.7	68.8
qPH3.1	19.0	3.8	4.4	3.19	-2.86	3.2	5.6
qPH6.1	21.0	42.4	22.1	17.68	-6.83	39.6	44.9
qPH7.1	89.0	58.4	14.5	11.12	5.22	57.7	59.5
qPH8.1	69.0	50.2	3.1	2.21	2.53	47.5	51.8
qPH9.1	93.0	52.6	3.3	2.35	2.36	52.0	55.8
qPH10.1	55.0	12.3	4.3	3.06	-2.87	8.1	52.2

[1] Flanking SNP positions correspond to 1-LOD interval in genetic distance

[2] QTLs in bold are unique to the ISRIL population

Table 4.13 Parameters of plant architecture related QTLs from interval mapping of the PQ-RIL population (Supplimentary)

QTL	Peak (cM)	Peak (Mb)	LOD	% Var	Additive Effect	Left Flanking Marker (Mb)	Right Flanking Marker (Mb)
qFL4.1	34.0	5.4	3.5	6.03	1.21	1.5	10.0
qFL8.1	71.0	51.9	2.8	7.00	-1.29	49.1	51.9
qFL9.1	35.0	50.2	4.1	8.16	1.66	50.2	54.5
qPH7.1	42.0	57.3	2.7	7.18	6.81	51.1	59.4

Table 4.14 Plant height (PH) QTL comparisons in SBSH BC1F2, IS-RIL and PQ-RIL population

QTL	IS-RIL	PQ-RIL	Inclusion of Genes
qPH2.H4.1 (+)			
qPH.3C.H4.1 (+)			
qPH.3D.H4.1 (+)			
qPH.5C.H4.1 (+)			
qPH.6B.H4.1 (+)			
qPH.6B.H4.2 (+)	qPH6.1 (-)		*Dw2*
qPH.6C.H4.1 (+)			
qPH.7A.H4.1 (+)			
qPH.7A.H4.2 (+)	qPH7.1 (+)	qPH7.1 (+)	*Dw3*
qPH.7C.H4.1 (+)			
qPH.7C.H4.2 (+)	qPH7.1 (+)	qPH7.1 (+)	*Dw3*
qPH.7D.H4.1 (+)	qPH7.1 (+)	qPH7.1 (+)	*Dw3*
qPH.9B.H4.1 (+)	qPH9.1 (+)		*Dw1*
qPH.1A.H6.1 (+)			
qPH.1B.H6.1 (+)			
qPH.1C.H6.1 (+)			
qPH.1D.H6.1 (-)			
qPH.2B.H6.1 (+)	qPH2.1 (+)		
qPH.2B.H6.2 (+)	qPH2.1 (+)		
qPH.3B.H6.1 (+)			
qPH.3B.H6.2 (+)			
qPH.4A.H6.1 (-)			
qPH.4C.H6.1 (-)			
qPH.4C.H6.2 (-)			
qPH.4D.H6.1 (+)			
qPH.6B.H6.1 (+)	qPH6.1 (-)		*Dw2*
qPH.6E.H6.1 (+)			
qPH.7B.H6.1 (+)	qPH7.1 (+)	qPH7.1 (+)	*Dw3*

qPH.7D.H6.1 (+)	qPH7.1 (+)	qPH7.1 (+)	*Dw3*
qPH.8A.H6.1 (-)			
qPH.8B.H6.1 (+)			
qPH.9A.H6.1 (+)	qPH9.1 (+)		*Dw1*
qPH.9B.H6.1 (+)	qPH9.1 (+)		*Dw1*
qPH.9C.H6.1 (-)	qPH9.1 (+)		*Dw1*
qPH10.1.H6 (+)	qPH10.1 (-)		
qPH10.2.H6 (-)	qPH10.1 (-)		

Table 4.15 Flowering time (FL) comparisons in SBSH-BC1F2, IS-RIL and PQ-RIL population

QTLname	ISRIL	PQRIL	Inclusion of genes
qFL.1A.H4.1 (-)	qFL1.1 (+)		*Ehd1*
qFL.6B.H4.1 (+)			*Ma6*
qFL.6B.H4.1 (+)			*Ma6*
qFL.6B.H4.2 (+)			*Ma1*
qFL1.H6.1 (-)	qFL1.1 (+)		*Ehd1*
qFL.1C.H6.1 (+)			
qFL.2C.H6.1 (+)			
qFL.4A.H6.1 (-)		qFL4.1 (+)	
qFL.4A.H6.2 (-)		qFL4.1 (+)	
qFL.4D.H6.1 (+)	qFL4.1 (+)		
qFL.6B.H6.1 (+)			*Ma1*
qFL.6B.H6.2 (+)			*SbTFL1-1*
qFL.6E.H6.1 (+)	qFL6.1 (-)		*SbFT3*
qFL.7D.H6.1 (+)			
qFL.8A.H6.1 (-)			
qFL.8B.H6.1 (+)			
qFL.10A.H6.1 (-)			*CO/SbMFT2*
qFL10.H6.1 (+)	qFL10.1 (+)		

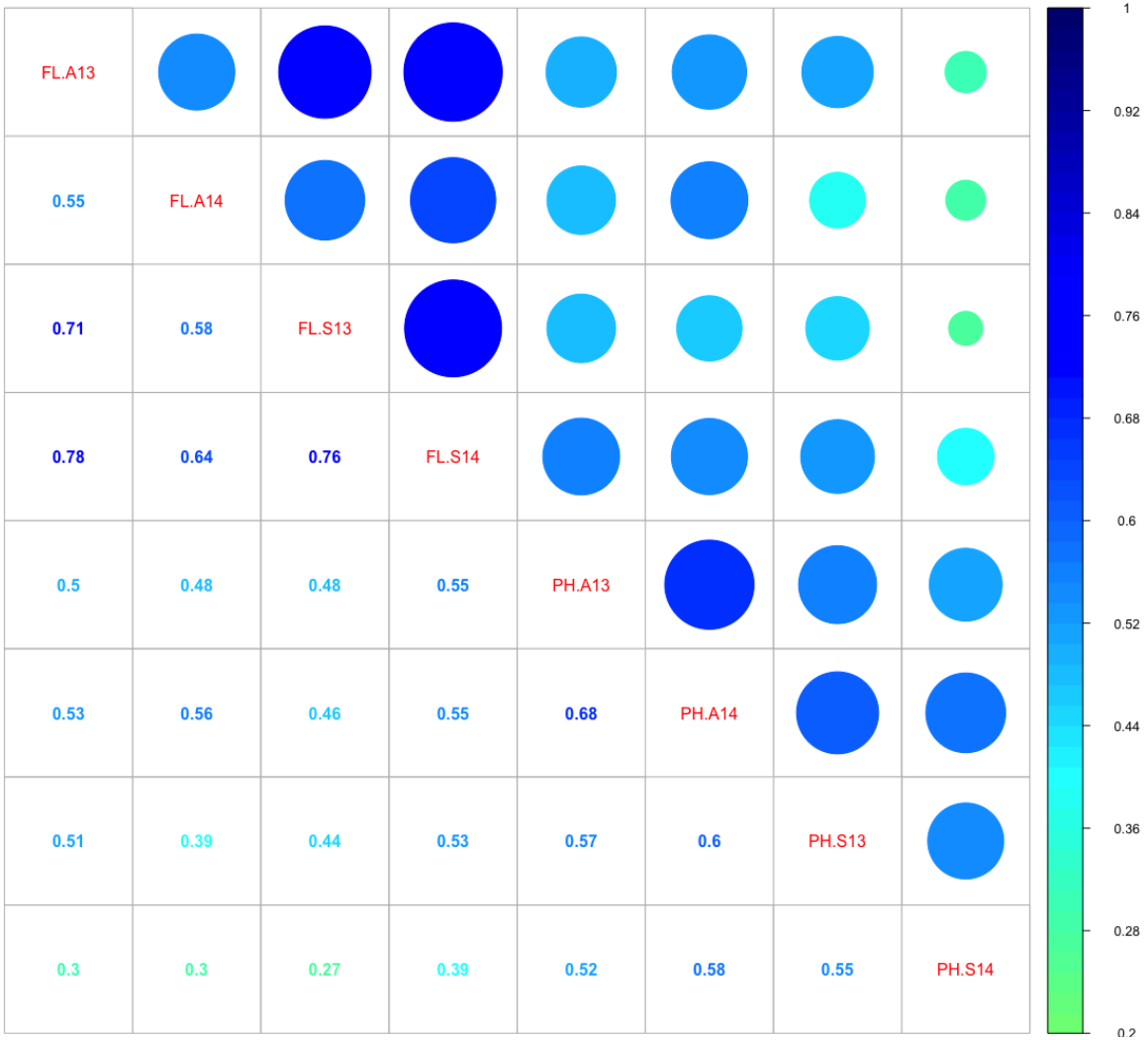


Figure 4.1Correlation coefficients between days to flowering (FL) and plant height (PH) in the BC_1F_2 populations in four environments.

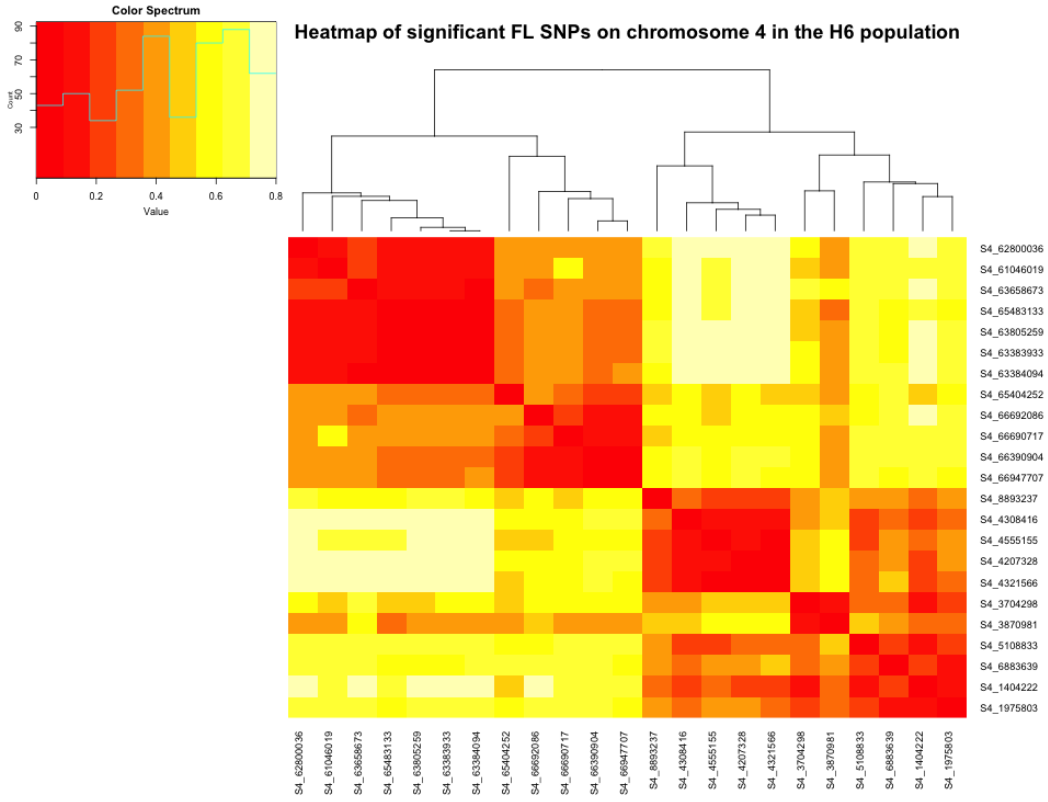


Figure 4.2 A heat map with hierarchical clustering of significant chromosome 4 SNP markers from single marker analysis for flowering time.

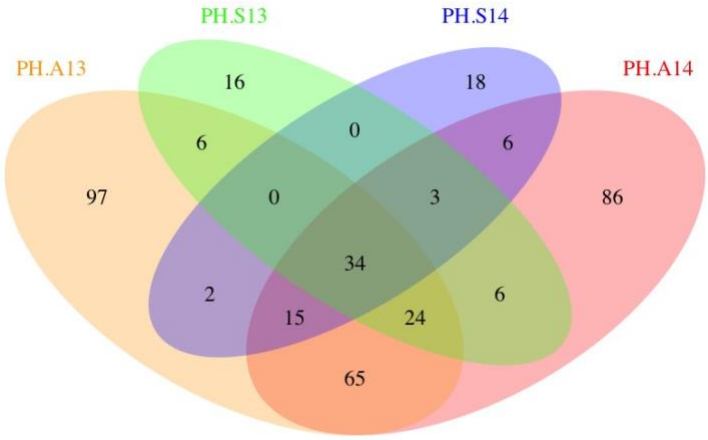


Figure 4.3 Venn diagram of the number of SNP markers for plant height significant at a p value<10^{-3} in different environments for pooled populations

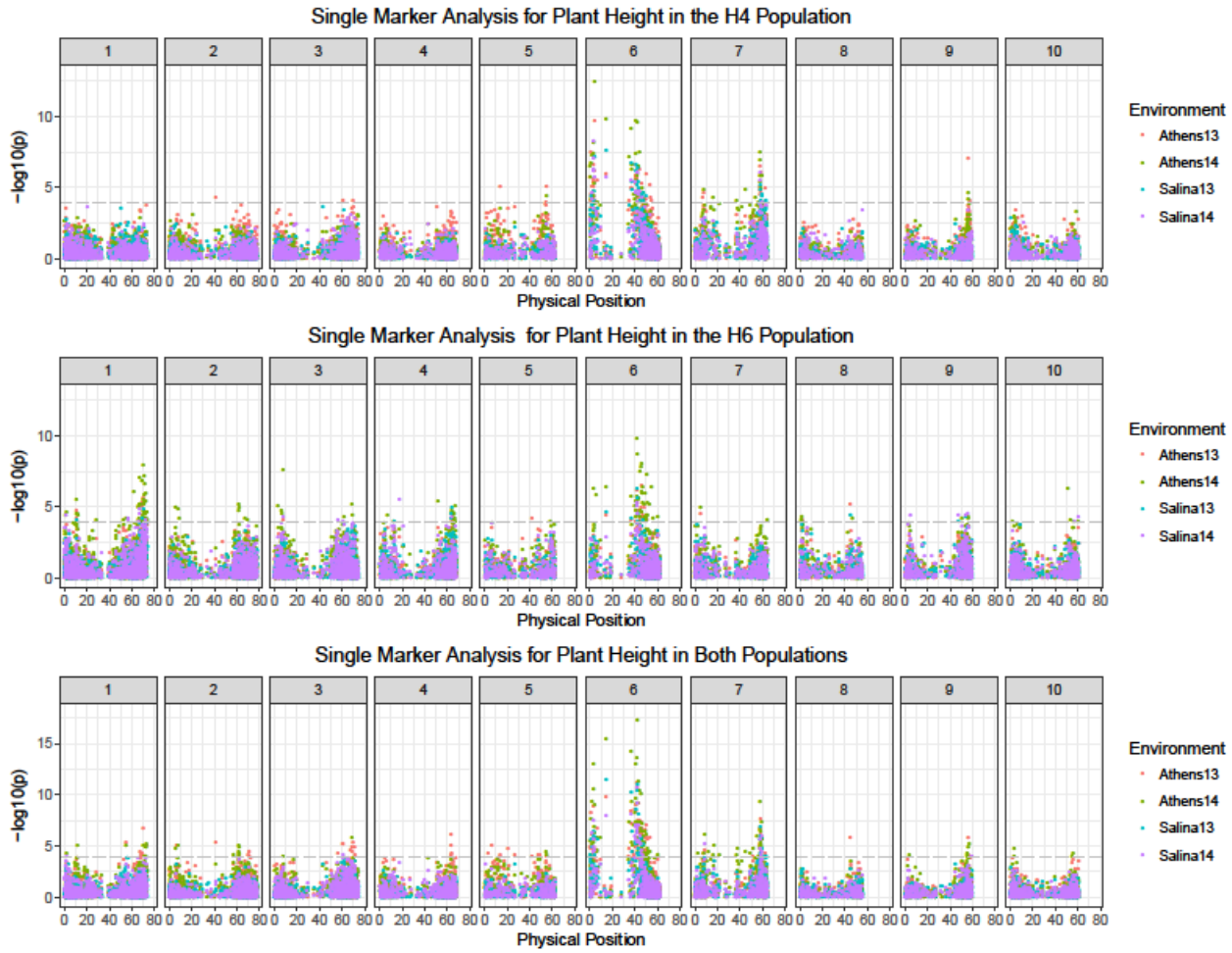


Figure 4.4 Single marker analysis of plant height in the H4, H6 -derived and pooled BC_1F_2 populations

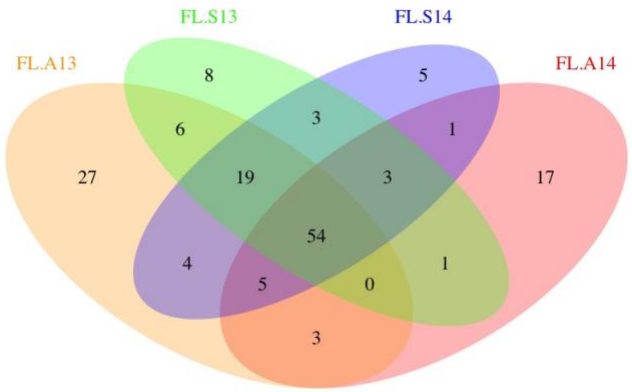


Figure 4.5 Venn diagram of the number of SNP markers for days to flowering significant at a pvalue<10^{-3} in different environments for pooled populations

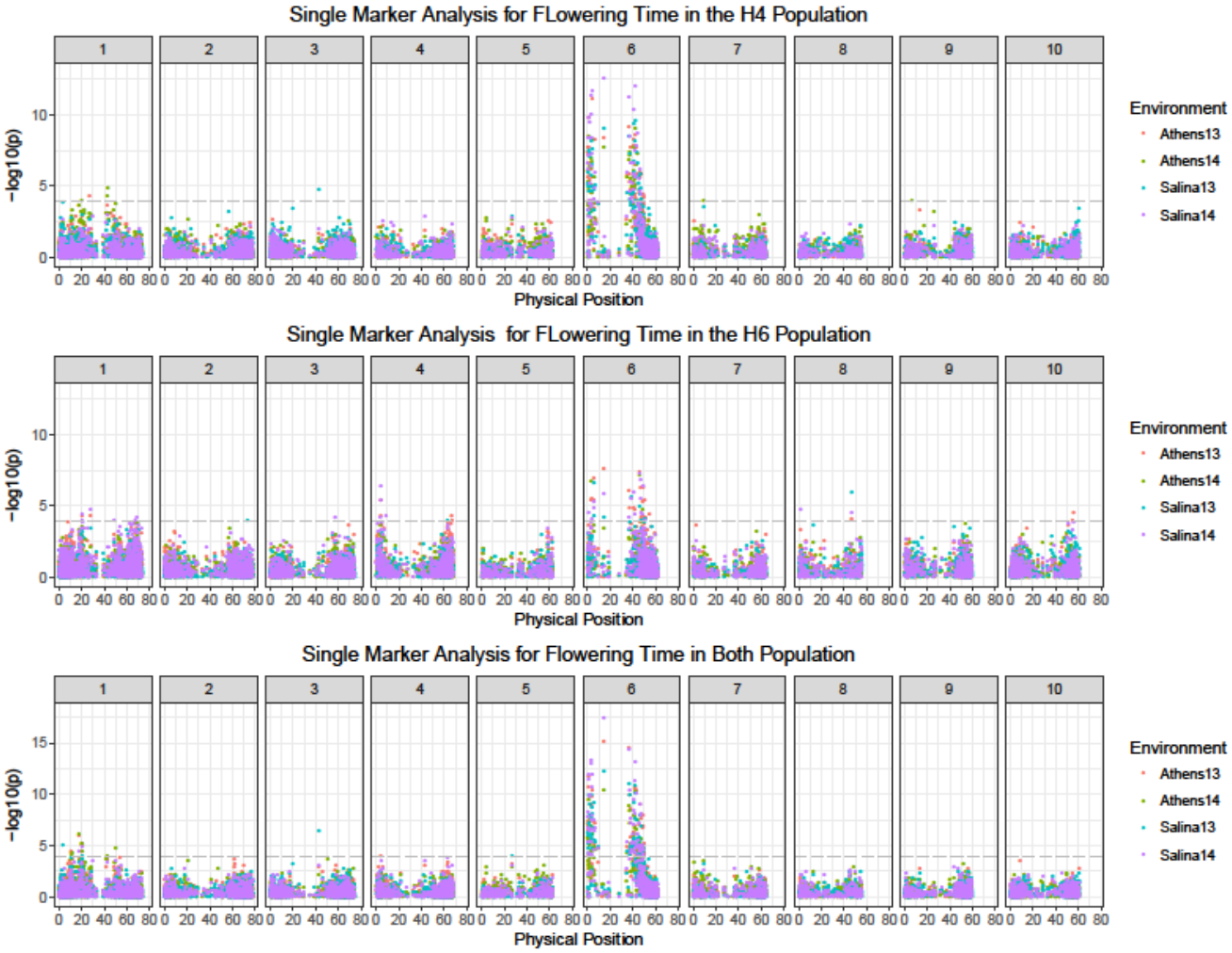


Figure 4.6 Single marker analysis for days to flowering in the H4-, H6-derived and pooled BC_1F_2 populations

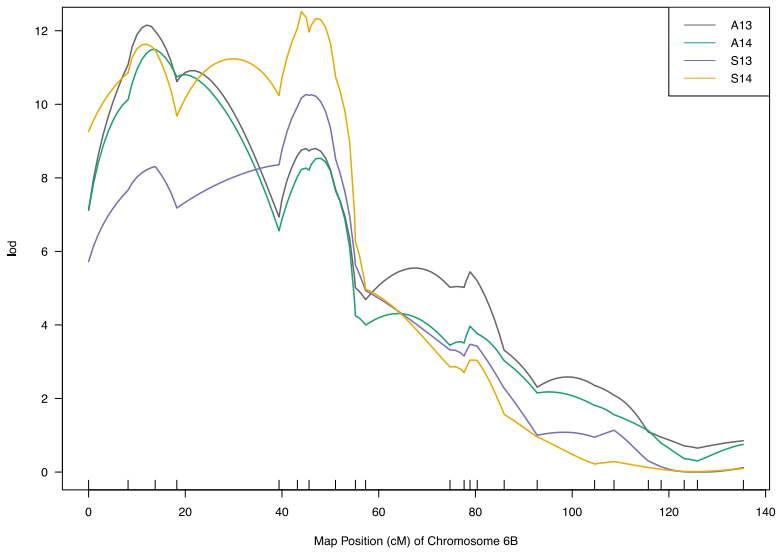


Figure 4.7 LOD plot for days to flowering (FL) in the H4-derived population

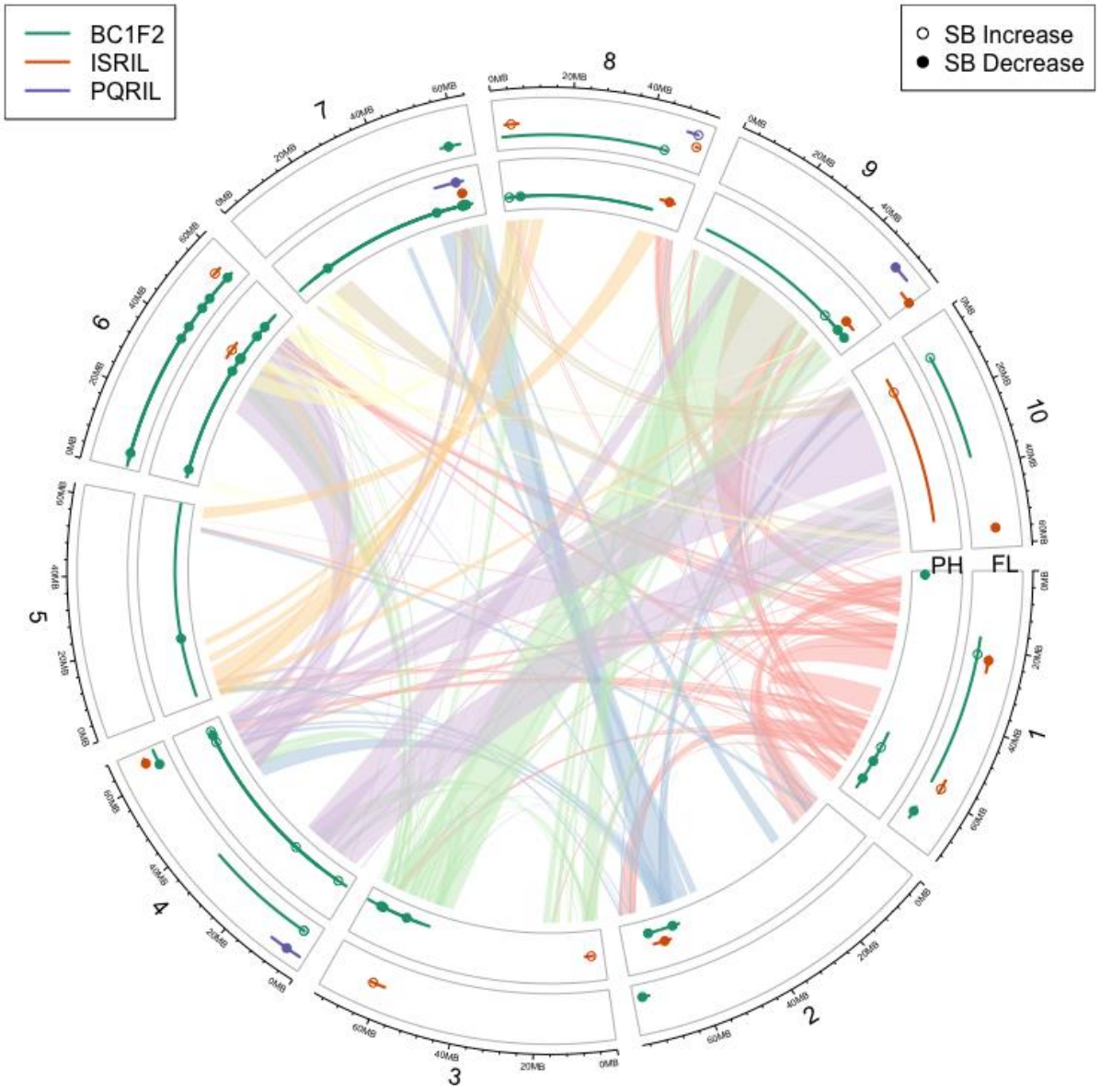


Figure 4.8 QTL correspondence plot. QTL locations were converted into their physical positions. Links are genome duplication event in sorghum (Lee et al. 2013)

CHAPTER 5

COMPARATIVE EVOLUTION OF VEGETATIVE BRANCHING IN SORGHUM[4]

[4] Kong, WQ., Nabukalu, P., Robertson, J., Goff, V., Pierce, G., Lemke, C., Compton, R., Cox, S. Paterson, AH.. To be submitted to Theoretical and Applied Genetics.

Abstract

Tillering and secondary branching are two plastic traits with high agronomic importance, especially in terms of the ability of plants to adapt to changing environments. We describe a quantitative trait analysis of tillering and secondary branching in two novel BC_1F_2 populations totaling 246 genotypes derived from backcrossing two *Sorghum bicolor* x *S. halepense* F_1 plants to a tetraploidized *S. bicolor*. A two-year, two-environment phenotypic evaluation in Athens, GA and Salina, KS permitted us to identify major effect and environment specific QTLs. Significant correlation between tillering and secondary branching followed by discovery of overlapping sets of QTLs continue to support the developmental relationship between these two organs and suggest the possibility of pleiotropy. Comparisons with two other populations *S. bicolor* BTx623 as a common parent but sampling the breadth of the Sorghum genus, increase confidence in QTL detected for these two plastic traits and provide insight into the evolution of morphological diversity in the Eusorghum clade. Correspondence between flowering time and vegetative branching supports other evidence in suggesting a pleiotropic effect of flowering genes. We propose a model to predict biomass weight from plant architecture related traits, quantifying contribution of each trait to biomass and providing guidance for future breeding experiments.

Introduction

Plant architecture is the three-dimensional organization of a plant body. The above-ground architecture includes the pattern of vegetative branching, sizes and shapes of stalks, leaves and floral organs, and plant height. The expression of plant architecture varies during different developmental stages, by a series of highly regulated endogenous genetic factors (Rameau et al. 2014; McSteen 2009; Kebrom et al. 2013) and exogenous constraints exerted by environments. Genetic factors impart the biodiversity of plant architecture, contributing to adaptation to different niches, and often utilized in classification of taxa. On the other hand, responsiveness to biotic and abiotic stresses tailors plant architecture to fitness under different environments (Krishna Reddy and Finlayson 2014; Kebrom et al. 2010).

Important aspects of plant architecture are tillering and vegetative branching, that are considered to be medium to low heritability traits with a high degree of plasticity (Kong et al. 2014; Doust 2007). The complexity of these traits is due in part to their non-deterministic development and responsiveness to environmental changes. Voluminous studies describe genes and genetic pathways controlling axillary meristem initiation and outgrowth that affect the number of tillers and patterns of vegetative branching (Wang and Li 2011; Kebrom et al. 2013; McSteen 2009). Many of these genes are involved in the production, signal transduction, transport, degradation and interactions of hormones such as auxin, cytokinin and strigolactone which act directly and locally to promote or repress axillary meristem activity (Mueller and Leyser 2011; Gallavotti 2013; Waldie et al. 2010; Doust 2007; Waldie et al. 2014).

Recent studies have also suggested that genes involved in controlling flowering time also influence the activity of axillary meristems and thus influence tillering and vegetative branching. For example, the flowering locus T (*Ft*) gene that regulates flowering time in many species, has

recently been found to trigger storage organ formation through direct interaction with the TCP factors (Navarro et al. 2015). The rice homolog of *Leafy* (*Lfy*) from Arabidopsis, expressed during the development of axillary bud and inflorescence branch primordia, is also required to produce tillers and panicle branches (Rao, 2008).

As a C4 model plant, sorghum has a relatively small genome (~730 Mb) and can provide many avenues to study traits related to plant architecture, thanks to a high quality reference genome sequence (Paterson et al. 2009). Using colinearity, results from sorghum may be extrapolated to many other C4 plants with large genomes, such as sugarcane. The flexibility to make crosses between the five main sorghum races (bicolor, guinea, caudatum, durra and kafir), and with wild relatives such as *S. propinquum* and *S. halepense* which vary widely in plant architecture, makes sorghum particularly attractive to dissect and compare genetic components of plant architecture. Compared to voluminous studies of plant height and flowering (Morris et al. 2013; Yang et al. 2014a; Yang et al. 2014b; Cuevas et al. 2016; Zhang et al. 2015; Murphy et al. 2011; Lin et al. 1995; Murray et al. 2009; Brown et al. 2008; Upadhyaya et al. 2012a), understanding of genetic components for tiller number and vegetative branching in sorghum has been relatively limited (Kong et al. 2014; Upadhyaya et al. 2012b; Hart et al. 2001; Paterson et al. 1995; Alam et al. 2014), possibly due to difficulties in phenotyping and the lower heritability of these traits.

In this paper, we describe a quantitative trait study of two important components of plant architecture, tillering and vegetative branching, in two half-sib tetraploid BC_1F_2 populations derived from crossing *Sorghum bicolor* BTx623 and *Sorghum halepense* Gypsum 9E. A two-year, two-environment phenotypic evaluation in Athens, GA and Salina, KS permitted us to identify major effect and environment specific QTLs. Quantitative trait loci (QTLs) discovered

in these two populations will be compared to those from two diploid sorghum recombinant inbred line (RIL) populations sharing BTx623 as a common parent but sampling the breadth of the Sorghum genus, one a cross to *S. bicolor* IS3620C, and the other to *S. propinquum* (Kong et al. 2013). QTLs identified in this study and their comparison provides insight into evolution of morphological diversity in the Sorghum genus, are of practical use for marker-assisted breeding, and provide a foundation for molecular cloning and functional analysis.

Materials and Methods

Population development, details of genotyping methods and methods for QTL analysis can be found in Kong, et al (2017).

Phenotyping

We evaluated tillering (**TL**) and secondary branching per tiller (**BRCH**) in the BC_1F_2 families with three replications in two fields in two years, 2013 and 2014 and at two locations, Athens GA and Salina KS. Tillering (**TL**) was measured by counting the number of tillers with mature seed heads before plants senesced. Secondary branches per tiller (**BRCH**) was calculated by taking the average of the number of the secondary branches from two representative tillers.

Phenotyping of vegetative branching in the IS-RIL population was consistent with our system applied to the *S. bicolor* × *S. propinquum* RILs described in Kong et al. (2014). To compare secondary branching across population and environments, we used the number of mature tillers (**TL**), and calculated the average number of secondary branches per mature tiller (**BRCH**) in the IS-RIL and PQ-RIL population. The variance component method was used to calculate broad-sense heritability [$H= V_g/(V_g+V_g*V_E/e +V_{residual}/er)$] where V_g is the variance estimate for genotype, V_E is the variance estimate for environment, e is the number of environments and r is the number of replications.

Genetic Analysis

To fully utilize the available data while protecting against false-positive results, genetic analysis employed two approaches. Using genetic maps that were constructed as described (Kong, 2017) from selected well-groomed SNP segregation data for each of the two SBSH-BC_1F_2 populations, interval mapping was conducted as described below. In addition, single marker analysis was conducted using each SNP marker that met quality standards described (whether in the genetic map or not), using hierarchical clustering to separate SNP markers on potentially different homologous chromosomes and inferring QTLs only if more than 4 SNPs were found within a cluster cut at height of 0.3 in recombination frequency to mitigate spurious associations. Similarities and differences in the results of these analyses were addressed in results.

Mixed modeling for biomass

We constructed a mixed effect model with **Biomass** as the response variable; **FL**, **PH**, **TL**, **BRCH**, mid-stalk diameter (**MD**), the number of nodes (**ND**), and population (H4 or H6) as fixed explanatory variables; and the environment (**ENV**) as a random effect. **MD** was the diameter of at the middle of a plant. The average of the six phenotypes from two blocks and three replications was taken for the mixed effect modeling. A natural log transformation was used for **Biomass** to normalize the data. Mixed effect modeling and model selection used the lme4 and lmerTest packages in R (Kuznetsova et al. 2015; Bates et al. 2015).

Results

Summary statistics and heritability analysis

The average number of mature tillers of *S. halepense* G9E was 16, higher than the 2.6 of diploid *S. bicolor* BTx623 (Table 5.1). Tetraploid BTx623 had an average of 0.77 more tillers

than diploid BTx623 in 2013 (t=2.91, p=0.006) and 1.58 more in 2014 (t=3.82, p=0.0005). In the BC_1F_2 population, average **TL** for most lines fell within the range of that of their parents, showing less transgressive segregation than plant height (**PH**) and flowering time (**FL**) (Kong et al, 2017). The average **TL** was 1.46 more in Salina than Athens (t=14.07, p<0.001). Average **TL** in Athens was 2.24 (t=-21.87, p<0.001) fewer in 2013 than 2014; and in Salina 2013 was 2.14 fewer in 2013 than 2014 (t=19.07, p<0.001). Average **TL** of the BC_1F_2 population is 0.30 greater than that of the PQ-RILs (t=2.52, p=0.020, Table 5.2), and 2.83 greater than that of the IS-RILs (t=36.19, p<0.001, Table 5.3). Broad sense heritability estimates for **TL** were intermediate for all three populations, at 35%, 36%, and 30% for the PQRIL, ISRIL and SH-BC_1F_2 populations, respectively (Table 5.1-5.3).

The number of secondary branches per primary tiller (**BRCH**) is sensitive to environmental changes and is also a fail-safe for a plant in case the primary seed head is damaged. Average **BRCH** of *S. halepense* is 13, dramatically higher than the 0.286 of *S. bicolor* BTx623 (Table 5.1). There were no statistically significant differences for **BRCH** between diploid and tetraploid BTx623 in Athens 2014, Salina 2013 and 2014, while there was 2.1 more **BRCH** in tetraploid BTx623 than in diploid BTx623 (t=4.16, p=0.0011) in Athens 2013. The average number of **BRCH** of most progeny lines fell within the range of the respective parents. For the SH-BC_1F_2 progeny lines, the average number of **BRCH** in Athens was 1.29 more than in Salina (t=25.50, p<0.001). Average **BRCH** in Athens was 0.45 more in 2013 than 2014 (t=7.70, p<0.001); and in Salina was 0.60 more in 2013 than 2014 (t=7.98, p<0.001). The average number of **BRCH** of the SH-BC_1F_2 population is 2.28 smaller than that of the PQ-RIL population (t=-14.38, p<0.001,Table 5.2), and 0.99 smaller than that of the ISRIL population (t=-0.99, p<0.001,Table 5.3). Broad-sense heritability estimates for **BRCH** are relatively low, 7%

and 10% for the PQ-RIL and SH-BC_1F_2 populations, respectively, but intermediate for the ISRIL population, 40.9% (Table 5.1Table 5.2Table 5.3).

Trait correlations

In all four environments, **FL** is positively and significantly correlated with **PH** (Figure 5.1), i.e., late flowering individuals are generally taller than early flowering ones. **FL** and **TL** are negatively correlated in both the H4 ($p=0.034$) and H6- derived populations ($p=0.032$) in Athens in 2013, and positive in other three environments, although not significant ($p>0.05$) for H4-derived populations in Athens 2014 or Salina 2013 and the H6-derived population in Salina 2014. In three out of four environments, Athens 2013, Salina 2013 and 2014, **FL** and **BRCH** are negatively correlated, with a non-significant positive correlation in Athens 2014. Correlations between **TL** and **BRCH** are generally positive, except for the H6 population in Athens 2013 where the correlation is negative but not significant.

Genetic analysis

Number of tillers

We detected a total of two QTLs, qTL.4A.H4.1 and qTL.4D.H4.1, for **TL** in the H4-derived population (Table 5.4). qTL.4A.H4.1 is significant in both Athens 2013 and Salina 2014, and qTL.4D.H4.1 is significant in Salina 2013 and Salina 2014. An additive model of the two QTLs, qTL.4A.H4.1 and qTL.4D.H4.1 explains 13.9% of the total phenotypic variance in Salina 2014. Although the peaks of qTL.4A.H4.1 are ~26 cM apart in Athens 2013 and Salina 2014, their physical locations corresponding to the one-lod interval in genetic distance overlap. No QTLs for **TL** were detected in Athens 2014. Both QTLs have positive allele effects, indicating that *S. halepense* alleles increase **TL**.

We detected a total of seven QTLs for the number of **TL** in the H6-derived population with only qTL.2C.H6.1 significant in both Athens 2013 and Salina 2013 (Table 5.4). Five QTLs detected in Athens 2013, qTL.2C.H6.1, qTL.6A.H6.1, qTL.6B.H6.1, qTL.9B.H6.1 and qTL.10C.H6.1, collectively explain 34% of the total phenotypic variance, one QTL detected in Athens 2014 explains 11.42% of the total phenotypic variance, and two QTLs detected in Salina 2013 explain 13.9% of the total phenotypic variance. No QTLs were found in Salina 2014. Two QTLs, qTL.6B.H6.1 and qTL.10C.H6.1 detected in Athens 2013, have negative allele effects, suggesting that *S. halepense* alleles decrease the number of tillers.

Using single marker analysis, we detected a total of 63, 46, 26 and 48 significant SNP markers ($p<10^{-3}$) for **TL** for pooled (H4 and H6) data in Athens 2013, Athens 2014, Salina 2013 and Salina 2014, respectively, with only one SNP marker, S4_58879601, significant in all environments (Figure 5.2 and Figure 5.3). Fewer signals detected for **TL** in multiple environments reflects lower heritability and large genotype by environment interactions. In the H4 population, we detected two QTLs for **TL**, qTL2.H4.1 and qTL.H7.1 in addition to the two QTLs on chromosome 4 detected by interval mapping. As was true for H4 QTLs found by interval mapping, *S. halepense* alleles increase the number of **TL**. In the H6 population, we detected a total of 14 QTLs for **TL** on chromosomes 1 (2), 2, 3(2), 4 (2), 6 (3), 9 (2), 10 (2) with three new QTLs not overlap with any QTLs detected in the interval mapping, qTL4.H6.1, qTL4.H6.2 and qTL10.H6.2, all with *S. halepense* alleles increasing the number of **TL** (Table 5.5). The other 11 QTLs form the single-marker analysis overlap with seven QTLs from interval mapping based on their physical positions

Number of Secondary Branches per Primary branch (**BRCH**)

We detected a total of seven QTLs for **BRCH** in the H4-derived population, including six from Athens 2014 and one from Salina 2013 (Table 5.6). No QTLs were found in Athens 2013 or Salina 2014. The six QTLs detected in Athens 2014 together explain 22.0% of the total phenotypic variance, while the one QTL detected from Salina 2013 explains about 8.28% of the phenotypic variance. It is interesting that six out of seven QTLs show negative allele effects Table 5.6, suggesting that *S. halepense* alleles contribute to decreased **BRCH**, which is unexpected and contrary to the difference between parents. Those QTLs with negative additive effect might reflect late release of apical dominance from *S. halepense*, which is associated with fewer **BRCH**.

We detected a total of seven QTLs for **BRCH** in the H6-derived population, with one QTL, qBRCH.3E.H6.1, significant in two environments, Salina 2013 and 2014 (Table 5.6). Two **BRCH** QTLs found in Athens 2014, qBRCH.1C.H6.1 and qBRCH.10C.H6.2, three BRCH QTLs found in Salina 2013, qBRCH3E.H6.1, qBRCH6B.H6.1,qBRCH10C.H6.1, and two BRCH QTLs found in Salina 2014, collectively explain 19.3% , 19.5% and 26.4 % of the total phenotypic variance, respectively. For four QTLs, qBRCH.1C.H6.1, qBRCH.3E.H6.1, qBRCH.5C.H6.1 and qRBCH10C.H6.1, *S. halepense* alleles increase **BRCH** as predicted based on the parental phenotypes, while *S. halepense* alleles decrease **BRCH** for the other three QTLs, qBRCH6b.H6.1, qBRCH.6B.H6.2 and qBRCH.10C.H6.2.

We detected a total of 4, 110, 65 and 20 significant SNP markers ($p<10^{-3}$) for **BRCH** in Athens 2013, Athens 2014, Salina 2013 and Salina 2014 for pooled data with very little correspondence among different environments, consistent with low heritability estimates and large genotype by environment interactions (Figure 5.4

Figure 5.5). In the H4-derived population, we detected a total of 11 QTLs for **BRCH** on chromosomes 1 (2), 3, 4 (2), 5, 6(2), 7, 9, 10, with three negative effect QTLs, suggesting that *S. halepense* may decrease **BRCH** (Table 5.7). A total of four QTLs, qBRCH1.H4.2, qBRCH3.H4.1, qRBCH9.H4.1 and qBRCH10.H4.1 were newly detected only in the single marker analysis, all with *S. halepense* alleles increasing **BRCH**. In the H6 population, we detected a total of 11 QTLs on chromosomes 1, 3 (2), 4, 5, 6(3), 7, 9 10 with only one negative effect QTL, qBRCH.H6.2 (Table 5.7). A total of three QTLs, qBRCH4.H6.1, qBRCH7.H6.1 qBRCH9.H6.1, were newly detected in the single marker analysis, with all three increasing **BRCH**. The other 8 QTLs detected in single markers analysis overlap with the seven BRCH QTLs from interval mapping by comparing their physical positions.

QTL correspondence across traits in the BC_1F_2 population

In most environments, **TL** and **BRCH** are significantly and positively correlated (Figure 5.1), therefore some QTL regions are expected to overlap due to their developmental relationship (Kong et al. 2014). Indeed, we found two **TL** QTLs, qTL2.H4.1 and qTL.4D.H4.1 overlapping with qBRCH.2D.H4.1 and qBRCH.4D.H4.1 in the H4-derived population based on their physical positions. Four QTLs, qTL.3E.H6.1, qTL.6B.H6.1, qTL6A.H6.1 and qTL.10C.H6.1 overlap with qBRCH.3E.H6.1, qBRCH.6B.H6.2, qBRCH6.H6.3 and qBRCH.10C.H6.2 in the H6-derived population, respectively. Interestingly, *S. halepense* contributed opposite allele effects for the two pairs of overlapping QTLs in the H4-derived population, but the same effect for all overlapping pairs in the H6-derived population.

Recent studies have suggested that genes controlling days to flowering might also influence tillering and vegetative branching (Navarro et al. 2015; Rameau et al. 2014; Rao et al. 2008; Peng et al. 2006). We found a total of six **TL** QTLs overlapping with **FL** QTLs in the H6-

derived population, with two pairs of QTLs, qTL.4.H6.1 with qFL4A.H6.1 and qTL6B.H6.1 with qFL6B.H6.2, showing opposite effects from *S. halepense* (Table 5.10).

Similarly, a total of two and five QTLs for **BRCH** show possible correspondence to **FL** in the H4 and H6-derived BC_1F_2 populations, respectively (Table 5.11), with four pairs of overlapping QTLs showing opposite allele effect from *S. halepense*. Additional QTLs that overlap but are not limited to the same linkage group and population are qBRCH.4C.H4.1 and qFL4A H6.1, peaking at 4.8 and 4.3 Mb respectively, qBRCH4D.H4.1 and qFL4D.H6.1, peaking at 61.2 and 62.8 Mb, respectively. QTLs qBRCH10C.H6.1/2 and qFL10A.H6.1 might be loosely associated, since they both cover a large genomic region.

Comparison to QTLs found in IS-RIL and PQ-RIL

We found a total of five **TL** QTLs in the SH-BC_1F_2, qTL.1D.H6.1, qTL.3E.H6.1, qTL.6A.H6.1, qTL.6B.H6.1, and qTL7.H4.1, corresponding in physical position on the sorghum genome sequence to four IS-RIL **TL** QTLs, qTL_1.1, qTL_3.1, qTL_6.1 and qTL7.1; and a total of seven SBSH-BC_1F_2 **TL** QTLs correspond to five PQ-RIL QTLs (Table 5.10). Curiously, qTL.6A.H6.1 and qTL.6B.H6.1, both overlapping with QTLs found in the IS-RIL and PQ-RIL populations, display opposite allele effects. qTL.6B.H6.1 from the SH-BC_1F_2 population shows a negative effect, indicating that fewer tillers may be associated with late flowering, which might be associated with the *Ma1* (Cuevas et al. 2016) gene on chromosome 6. However, there appears to be another QTL region on chromosome 6 significant in all three populations, roughly spanning 50 - 60 Mb, and suggesting that *S. bicolor* alleles reduce the number of tillers.

Despite that **BRCH** is a plastic trait with low heritability, we still found two **BRCH** QTLs in the H4-derived SBSH-BC_1F_2, qBRCH4D.H4.1 and qBRCH5C.H4.1 overlapping with two IS-RIL QTLs, qRBCH4.1 and qBRCH5.1; and three H6-derived SBSH-BC_1F_2 QTLs

overlapping with two IS-RIL QTLs, qBRCH4.2 and qBRCH10.1 (Table 5.11). Two pairs of QTLs, qBRCH.5C.H4.1 and qBRCH5.1 from IS-RIL and qBRCH.10C.H6.1 and qBRCH10.1 from IS-RIL show opposite allele effects from *S. halepense*, suggesting that *S. bicolor* alleles increase BRCH in the SBSH-BC_1F_2 population but decrease it in the ISRIL. In addition, a total of five and four H4 and H6–derived SBSH-BC_1F_2 QTLs for **BRCH** overlap with QTLs for various degrees (primary, secondary or tertiary) of vegetative branching described in Kong et al. (2014). Most overlapping pairs of QTLs of SBSH-BC_1F_2 and the PQRIL show the same direction of effect, from *S. halepense* and *S. propinquum*, respectively, except one case on chromosome 7 where QTLs within PQRIL shows different directions of effects.

A regression model to predict biomass

We performed a regression analysis to predict biomass weight (**Biomass**, using natural log transformation) with respect to traits related to plant architecture while controlling for population structure and environmental factors. Our final model consists of a total of seven variables, with plant height (**PH**), mid-stalk diameter (**MD**), number of mature tillers (**TL**), number of secondary branches (**BRCH**), flowering time (**FL**), and population (**H4** or **H6**) as fixed effects and environmental factors as a random effect (Table 5.12 Table 5.13). Fixed effects in this model collectively explain about 71.76% of the total variance using a modified method for estimating R-squared in mixed models (Nakagawa and Schielzeth 2013). The typical log error in this model is about 0.3148, and can be decomposed into environmental error that is estimated to be normally distributed with a mean of zero and standard deviation of 0.1260; and the inherent error that is estimated to be normally distributed with a mean of zero and standard deviation of 0.2885. The model suggests that **PH**, **TL** and **MD** are the three most important variables in predicting **Biomass**, followed by **FL** and **BRCH**. For example, a 10 cm increase in plant height

leads to 6.4% increase in **Biomass** weight, keeping other variables constant, while an increase of one TL leads to a 15.1% increase in Biomass weight, keeping other variables constant.

Discussion

The present study offers several new insights into the genetic control of tillering and vegetative branching. First, it adds more information to current knowledge of vegetative branching in sorghum, an under-studied trait, especially providing early insights into QTL polymorphism in *S. halepense*. Correspondence of QTL regions between three populations sharing *S. bicolor* BTx623 as a common parent, with the other parents being morphologically and genetically distinct genotypes that represent cultivated (IS3620C), wild diploid (*S. propinquum*) and wild polyploid (*S halepense*) sorghums, provides information about common QTLs shared between or among populations and taxon-specific QTLs that contribute to divergence. Finally, constructing a mixed model to predict dry biomass with respect to various traits associated with plant architecture and the environmental factors provides a framework to understand the contribution of each trait to biomass as well as environmental influences.

QTL mapping

QTL mapping results for two relatively plastic traits, **TL** and **BRCH**, suggest high genotype by environment interactions and population differences. We only found three and one QTLs significant in multiple environments for **TL** and **BRCH** with interval mapping, respectively, with 6 and 13 significant in only single environments. Overlapping SNP sets from single marker analysis are much lower for these traits than for highly heritable traits such as plant height and flowering time (Figure 5.2, Figure 5.4).

QTL results are very different in the two populations derived from two different sibling F_1 plants, possibly due to *Ma* and *Dw* genes on chromosome 6. We detected fewer **TL** QTLs in

the H4 than the H6-derived populations (Table 5.5 Table 5.4), similar with the population differences of **FL** and **PH** QTLs. The number of **BRCH** QTLs for the two populations does not follow this pattern, but most H4-derived QTLs had negative effects in interval mapping. The opposite effect **TL** QTLs associated with the **FL** QTLs, qTL4.H6.1 and qFL4A.H6.1, qTL.6B.H6.1 and qFL6B.H6.2, and **BRCH** QTLs with **FL** QTLs, qBRCH1.H4.1 and qFL.1A.H4.1 , qBRCH.6B.H4.1, and qFL.6B.H4.1, qBRCH.6B.H6.1 and qFL.6B.H6.1, and qBRCH.6B.H6.2 and qFL.6B.H6.2 (Table 5.10 and Table 5.11), suggest that delaying flowering might reduce tillers and branching, perhaps due to late release of apical dominance.

QTL correspondence

Two **TL** QTLs and one **BRCH** QTL overlapped in all three populations with the same direction of allele effect (Table 5.10 andTable 5.11), suggesting a parsimonious hypothesis that *S. halepense*, *S. propinquum* and *S. bicolor* IS3620C share an ancestral allele, while a different recently-derived allele has been selected in the elite cultivar *S. bicolor* BTx623. Cases in which overlapping QTLs have different directions of allele effect are more complex, possibly suggesting more than two alleles, or perhaps representing spurious correspondence due to relatively large QTL intervals.

The *S. halepense* data continue to support the hypothesis that **TL** and **BRCH** are developmentally related (Kong et al. 2014)—six QTL pairs (qTL2.H4.1 and qBRCH.2D.H4.1, qTL.4D.H4.1 and qBRCH.4D.H4.1,qTL.3E.H6.1 and qBRCH.3E.H6.1, qTL.6B.H6.1 and qRBCH.6B.H6.2, qTL6A.H6.1 and qBRCH6.H6.3, qTL.10C.H6.1 and qBRCH.10C.H6.2) overlapped, perhaps harboring genes influencing axillary meristem development at early stages.

A surprising number of genomic regions were significant for **FL** and **TL** or **FL** and **BRCH**, perhaps suggesting pleiotropic relationships with genes controlling **FL** (Table 5.10Table

5.11, Figure 5.6). For example, genes regulating flowering such as MADS box proteins also influence determinacy of all meristems (Melzer et al. 2008). Further, the flowering locus T (*FT*) gene that regulates flowering time in many species, has recently been found to trigger storage organ formation through direct interaction with the TCP factors (Navarro et al. 2015). We found a total of six genomic regions harboring QTLs responsible for both **FL** and **TL**, and four regions for both **FL** and **BRCH** in their respective populations. Previous study (Feltus et al. 2006; Paterson et al. 1995; Lin et al. 1995) has suggested that regions on chromosome 6 that harbor *Ma1* also contain QTLs for tiller number. One explanation might be that *Ma1*, which appears to be a homolog of the Arabidopsis *Ft* and Rice *Hd3a* genes (Cuevas et al. 2016), influences organ formation. The *Ma1* associated region in this study affected both **TL** and **BRCH**, while another QTL region at ~47.2Mb on chromosome 6 affecting all three traits, **FL**, **TL** and **BRCH**, might be related to the Sb06g019010 gene encoding the 'number of apical meristem' (NAM) protein (Kong et al. 2014; Finn et al. 2014).

Regression model for predicting biomass

A mixed model for predicting dry biomass weight (**Biomass**) retained a total of five traits, plant height (**PH**), mid-stalk diameter (**MD**), mature tillers (**TL**), number of secondary branches (**BRCH**), and flowering time (**FL**) as significant predictors of dry biomass, with all fixed effects explaining 71.76% of the total variance, and a log error of 0.3148. This model indicates that plant architecture related traits can predict **Biomass** with relatively high accuracy. Application of this model might be a cost-efficient method for predicting **Biomass** for future experiments, quantifying the contribution of individual traits to **Biomass** and providing guidance for improving genotypes aimed at biomass production.

Acknowledgments

We appreciate the support of the AFRI Plant Growth and Development Program (2016-67013-24608 to AHP), USDA Biotechnology Risk Assessment Program (2012-01658 to AHP and TSC), and NIFA Global Food Security CAP (2015-68004-23492 to AHP).

References

Alam, M.M., E.S. Mace, E.J. van Oosterom, A. Cruickshank, C.H. Hunt *et al.*, 2014 QTL analysis in multiple sorghum populations facilitates the dissection of the genetic and physiological control of tillering. *Theoretical and Applied Genetics* 127 (10):2253-2266.

Bates, D., M. Machler, B.M. Bolker, and S.C. Walker, 2015 Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67 (1):1-48.

Brown, P.J., W.L. Rooney, C. Franks, and S. Kresovich, 2008 Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes. *Genetics* 180 (1):629-637.

Cuevas, H.E., C. Zhou, H. Tang, P.P. Khadke, S. Das *et al.*, 2016 The Evolution of Photoperiod-Insensitive Flowering in Sorghum, A Genomic Model for Panicoid Grasses. *Molecular Biology and Evolution* 33 (9):2417-2428.

Doust, A.N., 2007 Grass architecture: genetic and environmental control of branching. *Current Opinion in Plant Biology* 10 (1):21-25.

Feltus, F.A., G.E. Hart, K.F. Schertz, A.M. Casa, P. Brown *et al.*, 2006 Genetic map alignment and QTL correspondence between inter- and intra-specific sorghum populations. *Theoretical and Applied Genetics* 112:1295-1305.

Finn, R.D., A. Bateman, J. Clements, P. Coggill, R.Y. Eberhardt *et al.*, 2014 Pfam: the protein families database. *Nucleic Acids Research* 42 (D1):D222-D230.

Gallavotti, A., 2013 The role of auxin in shaping shoot architecture. *Journal of Experimental Botany* 64 (9):2593-2608.

Hart, G.E., K.F. Schertz, Y. Peng, and N.H. Syed, 2001 Genetic mapping of Sorghum bicolor (L.) Moench QTLs that control variation in tillering and other morphological characters. *Theoretical and Applied Genetics* 103 (8):1232-1242.

Kebrom, T.H., T.P. Brutnell, and S.A. Finlayson, 2010 Suppression of sorghum axillary bud outgrowth by shade, phyB and defoliation signalling pathways. *Plant Cell and Environment* 33 (1):48-58.

Kebrom, T.H., W. Spielmeyer, and E.J. Finnegan, 2013 Grasses provide new insights into regulation of shoot branching. *Trends in Plant Science* 18 (1):41-48.

Kong, W., H. Guo, V. Goff, T.-H. Lee, C. Kim *et al.*, 2014 Genetic analysis of vegetative branching in sorghum. *Theoretical and Applied Genetics* 127 (11):2387-2403.

Kong, W., H. Jin, C.D. Franks, C. Kim, R. Bandopadhyay *et al.*, 2013 Genetic Analysis of Recombinant Inbred Lines for *Sorghum bicolor* × *Sorghum propinquum*. *G3: Genes*| *Genomes*| *Genetics* 3:101-108.

Krishna Reddy, S., and S.A. Finlayson, 2014 Phytochrome B promotes branching in Arabidopsis by suppressing auxin signaling. *Plant Physiol* 164 (3):1542-1550.

Kuznetsova, A., P.B. Brockhoff, and R.H.B. Christensen, 2015 Package 'lmerTest'. *R package version* 2.

Lin, Y.R., K.F. Schertz, and A.H. Paterson, 1995 Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population. *Genetics* 141 (1):391-411.

McSteen, P., 2009 Hormonal Regulation of Branching in Grasses. *Plant Physiology* 149 (1):46-55.

Melzer, S., F. Lens, J. Gennen, S. Vanneste, A. Rohde *et al.*, 2008 Flowering-time genes modulate meristem determinacy and growth form in Arabidopsis thaliana. *Nat Genet* 40 (12):1489-1492.

Morris, G.P., P. Ramu, S.P. Deshpande, C.T. Hash, T. Shah *et al.*, 2013 Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences of the United States of America* 110 (2):453-458.

Mueller, D., and O. Leyser, 2011 Auxin, cytokinin and the control of shoot branching. *Annals of Botany* 107 (7):1203-1212.

Murphy, R.L., R.R. Klein, D.T. Morishige, J.A. Brady, W.L. Rooney *et al.*, 2011 Coincident light and clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum. *Proceedings of the National Academy of Sciences of the United States of America* 108 (39):16469-16474.

Murray, S.C., W.L. Rooney, M.T. Hamblin, S.E. Mitchell, and S. Kresovich, 2009 Sweet Sorghum Genetic Diversity and Association Mapping for Brix and Height. *Plant Genome* 2 (1):48-62.

Nakagawa, S., and H. Schielzeth, 2013 A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4 (2):133-142.

Navarro, C., E. Cruz-Oro, and S. Prat, 2015 Conserved function of FLOWERING LOCUS T (FT) homologues as signals for storage organ differentiation. *Current Opinion in Plant Biology* 23:45-53.

Paterson, A., H., J.E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood *et al.*, 2009 The Sorghum bicolor genome and the diversification of grasses *Nature* 457:551-556.

Paterson, A.H., K.F. Schertz, Y.R. Lin, S.C. Liu, and Y.L. Chang, 1995 The Weediness of Wild Plants - Molecular Analysis of Genes Influencing Dispersal and Persistence of Johnsongrass, Sorghum-Halepense (L) Pers. *Proceedings of the National Academy of Sciences of the United States of America* 92 (13):6127-6131.

Peng, M., Y. Cui, Y.M. Bi, and S.J. Rothstein, 2006 AtMBD9: a protein with a methyl-CpG-binding domain regulates flowering time and shoot branching in Arabidopsis. *Plant Journal* 46 (2):282-296.

Rameau, C., J. Bertheloot, N. Leduc, B. Andrieu, F. Foucher *et al.*, 2014 Multiple pathways regulate shoot branching. *Frontiers in Plant Science* 5:741.

Rao, N.N., K. Prasad, P.R. Kumar, and U. Vijayraghavan, 2008 Distinct regulatory role for RFL, the rice LFY homolog, in determining flowering time and plant architecture. *Proceedings of the National Academy of Sciences of the United States of America* 105 (9):3646-3651.

Upadhyaya, H.D., Y.-H. Wang, S. Sharma, and S. Singh, 2012a Association mapping of height and maturity across five environments using the sorghum mini core collection. *Genome* 55 (6):471-479.

Upadhyaya, H.D., Y.H. Wang, S. Sharma, S. Singh, and K.H. Hasenstein, 2012b SSR markers linked to kernel weight and tiller number in sorghum identified by association mapping. *Euphytica* 187 (3):401-410.

Waldie, T., A. Hayward, and C.A. Beveridge, 2010 Axillary bud outgrowth in herbaceous shoots: how do strigolactones fit into the picture? *Plant Molecular Biology* 73 (1-2):27-36.

Waldie, T., H. McCulloch, and O. Leyser, 2014 Strigolactones and the control of plant development: lessons from shoot branching. *Plant Journal* 79 (4):607-622.

Wang, Y.H., and J.Y. Li, 2011 Branching in rice. *Current Opinion in Plant Biology* 14 (1):94-99.

Yang, S.S., R.L. Murphy, D.T. Morishige, P.E. Klein, W.L. Rooney *et al.*, 2014a Sorghum Phytochrome B Inhibits Flowering in Long Days by Activating Expression of SbPRR37 and SbGHD7, Repressors of SbEHD1, SbCN8 and SbCN12. *Plos One* 9 (8).

Yang, S.S., B.D. Weers, D.T. Morishige, and J.E. Mullet, 2014b CONSTANS is a photoperiod regulated activator of flowering in sorghum. *BMC Plant Biology* 14.

Zhang, D., W. Kong, J. Robertson, V.H. Goff, E. Epps *et al.*, 2015 Genetic analysis of inflorescence and plant height components in sorghum (Panicoidae) and comparative genetics with rice (Oryzoidae). *BMC Plant Biol* 15:107.

Chapter 5 Tables and Figures

Table 5.1 Summary statistics for number of mature tillers (TL) and number of secondary branches (BRCH) in the SBSH-BC_1F_2 [*S. halepense* derived (*S. bicolor* BTx623× *S. halepense* G9E) backcross] population and parents.

			SH-BC1F2							BTx623 (2x)			BTx623 (4x)			
Trait	Location	Year	Population	N	Mean	Median	SD	Min	Max	N	Mean	SD	N	Mean	SD	Heritability (%)
TL	Athens	2013	Pooled	246	2.43	2.17	0.9303	0.83	5.33	12	1.3	0.49	18	1.9	1.39	30.46
TL	Athens	2013	H4	141	2.29	2.17	0.8410	0.83	5.33							
TL	Athens	2013	H6	105	2.61	2.50	1.0140	1.17	5.17							
TL	Athens	2014	Pooled	236	4.66	4.67	1.2782	2.00	8.75	8	2.0	0.87	12	3.7	1.03	
TL	Athens	2014	H4	134	4.63	4.67	1.1948	2.00	8.67							
TL	Athens	2014	H6	102	4.71	4.58	1.3849	2.17	8.75							
TL	Salina	2013	Pooled	246	3.92	3.83	1.0259	1.67	7.50	12	1.3	0.45	12	2.2	1.19	
TL	Salina	2013	H4	141	3.83	3.67	1.0840	2.00	7.50							
TL	Salina	2013	H6	105	4.04	4.00	0.9346	1.67	6.67							
TL	Salina	2014	Pooled	236	6.09	5.75	1.4246	3.50	12.00	12	3.3	1.48	12	5.0	1.08	
TL	Salina	2014	H4	134	5.96	5.67	1.3826	4.00	12.00							
TL	Salina	2014	H6	102	6.26	6.00	1.4672	3.50	11.00							
BRCH	Athens	2013	Pooled	246	2.79	2.71	0.7409	1.25	5.42	12	1.6	1.03	18	3.7	1.15	10.38
BRCH	Athens	2013	H4	141	2.74	2.67	0.6851	1.25	4.70							
BRCH	Athens	2013	H6	105	2.87	2.75	0.8072	1.25	5.42							
BRCH	Athens	2014	Pooled	236	2.34	2.33	0.5415	0.78	3.83	12	2.4	0.79	12	1.1	0.83	
BRCH	Athens	2014	H4	134	2.27	2.25	0.5517	0.78	3.67							
BRCH	Athens	2014	H6	102	2.42	2.42	0.5177	1.33	3.83							
BRCH	Salina	2013	Pooled	246	1.57	1.50	0.8560	0.00	4.67	12	0.1	0.29	12	0.1	0.29	
BRCH	Salina	2013	H4	141	1.31	1.17	0.8050	0.00	4.67							
BRCH	Salina	2013	H6	105	1.91	1.83	0.8036	0.50	4.00							

BRCH	Salina	2014	Pooled	236	0.97	0.83	0.7857	0.00	3.50	12	0.3	1.11	12	0.0	0.00	
BRCH	Salina	2014	H4	134	0.71	0.63	0.5858	0.00	2.67							
BRCH	Salina	2014	H6	102	1.32	1.17	0.8800	0.00	3.50							

Table 5.2 Summary statistics for flowering time (**FL**), plant height (**PH**), number of mature tillers (**TL**) and number of secondary branches (**BRCH**) in the PQ-RIL [propinquum derived (*S. bicolor* BTx623× *S. propinquum*) recombinant inbred line] population and parents.

		PQ-RIL						BTx623			
Trait	Year	N	Mean	Median	SD	Min	Max	N	Mean	SD	Heritability (%)
TL	2009	155	3.11	3.00	1.8495	1.0	11.0	10	1.20	0.4216	35.17
TL	2010	132	4.79	4.50	2.8673	1.0	14.5	14	1.21	0.4472	
TL	2011	141	4.08	3.50	2.1624	1.0	13.5	20	1.65	0.8127	
BRCH	2009	155	4.34	3.67	2.6632	0.2	16.0	10	3.25	1.0341	10.38
BRCH	2010	132	4.70	4.00	2.5093	0.0	15.6	14	3.47	1.2175	
BRCH	2011	141	3.60	3.43	1.7645	0.0	10.7	20	1.83	1.4754	

Table 5.3 Summary statistics for flowering time (**FL**), plant height (**PH**), number of mature tillers (**TL**) and number of secondary branches (**BRCH**) in the IS-RIL [IS3620C derived (*S. bicolor* BTx623× *S. bicolor* IS3620C) recombinant inbred line] population and parents.

	IS-RIL							BTx623			IS3620C			
Trait	Year	N	Mean	Median	SD	Min	Max	N	Mean	SD	N	Mean	SD	Heritability (%)
TL	2011	388	2.29	2.00	1.2250	1.00	10.00	16	1.56	0.8638	13	5.23	1.9277	36.09
TL	2012	384	2.58	2.50	1.2970	1.00	10.00	12	1.25	0.4330	16	4.00	1.7678	
BRCH	2011	388	2.63	2.50	1.7218	0.00	10.50	16	3.31	1.7399	13	5.03	3.2963	40.92
BRCH	2012	383	2.90	2.80	1.5127	0.00	13.50	12	4.63	1.9378	16	2.89	0.8393	

Table 5.4 Parameters of **TL** (mature tillers) QTLs from interval mapping of the H4 and H6 SBSH-BC_1F_2 populations

QTL Name	Pos (CM)	Pos (Mb)	LOD	% of Variance explained	Effect	Left (Mb)	Right (Mb)	Env
qTL.4A.H4.1	123.2	53.4	2.6	7.94	0.58	52.5	61.2	AT13
qTL.4A.H4.1	149.0	58.9	2.6	8.53	1.01	57.4	61.2	SL14
qTL.4D.H4.1	28.0	61.8-62.5	3.2	9.54	0.87	53.1	65.8	SL13
qTL.4D.H4.1	29.4	61.8-62.5	2.5	8.21	1.03	53.1	65.8	SL14
qTL.1D.H6.1	100.4	65.3	2.8	11.42	1.38	19.2	65.3	AT14
qTL.2C.H6.1	108.0	9.0	3.0	8.34	1.21	8.4	65.8	AT13
qTL.2C.H6.1	110.0	9.0	2.5	10.68	1.19	8.4	65.8	SL13
qTL.3E.H6.1	205.0	59.7	2.9	11.15	1.05	4.5	59.7	SL13
qTL.6A.H6.1	186.0	57.5	3.0	11.81	0.73	56.5	60.5	AT13
qTL.6B.H6.1	121.0	47.2	6.7	24.58	-1.09	47.2	50.9	AT13
qTL.9B.H6.1	55.0	53.6	3.2	12.57	0.78	47.9	55.8	AT13
qTL.10C.H6.1	87.7	6.0	2.8	11.69	-0.74	1.2	12.8	AT13

Table 5.5 Parameters of tillering (**TL**) QTLs from single marker analysis of the H4 and H6 SBSH-BC_1F_2 populations

QTL	Peak SNP	Chr	Pos	P vlaue	Effect	Left [1]	Right[1]
qTL2.H4.1	S2_56654701	2	56.7	3.16E-06	1.44	S2_56654701	S2_77026507
qTL4.H4.1	S4_53529686	4	53.5	0.000103577	0.95	S4_52973418	S4_61151050
qTL4.H4.2	S4_57873900	4	57.9	6.20E-05	1.24	S4_57041432	S4_62529365
qTL7.H4.1	S7_9441761	7	9.4	0.000282094	1.09	S7_7725211	S7_53456010
qTL1.H6.1	S1_62664534	1	62.7	3.78E-05	2.10	S1_5081943	S1_70788904
qTL1.H6.2	S1_67465078	1	67.5	5.19E-05	1.18	S1_1265914	S1_69679431
qTL2.H6.1	S2_7106903	2	7.1	2.61E-05	0.90	S2_6849865	S2_69967980
qTL3.H6.1	S3_15352565	3	15.4	0.000180592	1.07	S3_6850696	S3_66467840
qTL3.H6.2	S3_57375930	3	57.4	3.12E-05	1.30	S3_6871352	S3_72811253
qTL4.H6.1	S4_1267007	4	1.3	0.000120412	1.58	S4_1051006	S4_67106558
qTL4.H6.2	S4_59173211	4	59.2	4.34E-06	1.53	S4_55435164	S4_62067548
qTL6.H6.1	S6_50834320	6	50.8	0.000142023	1.91	S6_45819310	S6_61351403
qTL6.H6.2	S6_50892527	6	50.9	8.30E-07	-1.06	S6_941772	S6_51751708
qTL6.H6.3	S6_61374763	6	61.4	7.32E-05	0.93	S6_56308299	S6_61939440
qTL9.H6.1	S9_53370872	9	53.4	1.71E-05	0.99	S9_8217569	S9_58139966
qTL9.H6.2	S9_56158950	9	56.2	0.000119584	1.52	S9_51954880	S9_57982011
qTL10.H6.1	S10_6764501	10	6.8	0.000158086	-1.26	S10_551695	S10_47850983
qTL10.H6.2	S10_45853249	10	45.9	8.56E-05	1.48	S10_5531584	S10_50040659

[1] Markers with smallest/largest (left/right) physical distances within in an interval

Table 5.6 Parameters of branching (**BRCH**) QTLs from interval mapping of the H4 and H6 SBSH-BC_1F_2 populations

QTL Name	Pos (cM)	Pos (Mb)	LOD	%Var Explained	Effect	Left (Mb)	Right (Mb)	Env
qBRCH.1F.H4.1	4.0	1.6-3.2	3.7	11.56	-0.47	1.6	8.6	AT14
qBRCH.2D.H4.1	122.0	74.5	3.0	9.98	-0.38	66.1	75.5	AT14
qBRCH.4C.H4.1	8.0	4.8	2.6	8.48	-0.34	3.7	6.9	AT14
qBRCH.4D.H4.1	102.7	61.2	3.2	10.22	-0.38	20.7	61.8	AT14
qBRCH.5C.H4.1	59.8	11.6	3.6	11.46	-0.38	1.7	57.9	AT14
qBRCH.6B.H4.1	8.2	0.9	2.6	8.28	-0.48	0.9	37.2	SL13
qBRCH.7C.H4.1	86.0	61.6	3.1	9.96	0.36	56.5	62.8	AT14
qBRCH.1C.H6.1	142.0	70.2	3.5	13.82	0.41	69.1	72.5	AT14
qBRCH.3E.H6.1	203.0	59.7	3.8	13.72	0.95	4.5	59.7	SL13
qBRCH.3E.H6.1	218.0	59.7	3.9	7.91	0.88	2.7	59.7	SL14
qBRCH.5C.H6.1	6.0	54.5	6.1	19.19	1.37	2.6	3.1	AT13
qBRCH.6B.H6.1	20.0	3.1	3.5	13.72	-0.63	2.0	42.2	SL13
qBRCH.6B.H6.2	95.0	47.0	2.9	11.66	-0.65	3.3	50.9	SL14
qBRCH.10C.H6.1	19.2	2.4	4.1	16.55	1.11	2.4	58.3	SL13
qBRCH.10C.H6.2	91.0	6.0	2.8	11.76	-0.38	1.2	53.4	AT14

Table 5.7 Parameters of number of secondary branches per tiller (**BRCH**) QTLs from single marker analysis of the H4 and H6 SBSH-BC_1F_2 populations

QTL	Peak SNP	Chr	Pos	P vlaue	Effect	Left	Right
qBRCH1.H4.1	S1_2170283	1	2.2	1.96E-05	0.44	S1_738323	S1_70832264
qBRCH1.H4.2	S1_25951730	1	26.0	8.89E-06	0.83	S1_21843308	S1_58839105
qBRCH3.H4.1	S3_13576771	3	13.6	4.72E-06	0.55	S3_2231352	S3_69864145
qBRCH4.H4.1	S4_5197267	4	5.2	4.09E-05	0.66	S4_2665334	S4_65162214
qBRCH4.H4.2	S4_68004088	4	68.0	8.23E-06	-0.58	S4_4826988	S4_68004088
qBRCH5.H4.1	S5_11626618	5	11.6	9.53E-08	-0.55	S5_2287684	S5_56190491
qBRCH6.H4.1	S6_941772	6	0.9	4.45E-05	-0.59	S6_941772	S6_58837002
qBRCH6.H4.2	S6_50891801	6	50.9	6.19E-05	0.48	S6_48190030	S6_59129413
qBRCH7.H4.1	S7_61629178	7	61.6	0.000126756	0.39	S7_887957	S7_62553133
qBRCH9.H4.1	S9_50415229	9	50.4	0.00011769	0.62	S9_4506578	S9_50415229
qBRCH10.H4.1	S10_11562619	10	11.6	4.03E-05	0.71	S10_2754555	S10_53317886
qBRCH1.H6.1	S1_70199577	1	70.2	1.72E-05	0.44	S1_69679431	S1_72932488
qBRCH3.H6.1	S3_61786995	3	61.8	4.93E-05	1.10	S3_4015390	S3_71069563
qBRCH3.H6.2	S3_65327695	3	65.3	3.28E-05	0.55	S3_61695348	S3_70980529
qBRCH4.H6.1	S4_66003764	4	66.0	0.000129728	1.42	S4_1267007	S4_66003764
qBRCH5.H6.1	S5_9191183	5	9.2	1.82E-05	1.25	S5_791861	S5_54509359
qBRCH6.H6.1	S6_48740921	6	48.7	2.01E-05	1.40	S6_19018634	S6_58002893
qBRCH6.H6.2	S6_49144952	6	49.1	0.000135526	-0.67	S6_3185305	S6_51252650
qBRCH6.H6.2	S6_56567550	6	56.6	0.00059907	0.76	S6_56567550	S6_61939440
qBRCH7.H6.1	S7_63121928	7	63.1	4.14E-05	1.08	S7_42048	S7_63121928
qBRCH9.H6.1	S9_17340892	9	17.3	4.44E-05	0.99	S9_825289	S9_39917345
qBRCH10.H6.1	S10_15907044	10	15.9	5.11E-06	1.24	S10_2406347	S10_57704601

Table 5.8 Parameters of tillering and vegetative branching related QTLs from interval mapping of the IS-RIL population

QTL	Year	Peak (cM)	Peak (Mb)	LOD	% Var	Additive Effect	Left Flanking SNP Positions[1]	Right Flanking SNP Positions
qTL_1.1	2012	171.0	69.2	1.9	2.96	0.0695	67.7	71.3
qTL_2.1	2011	6.0	3.2	5.1	5.98	-0.0946	3.1	3.8
qTL_3.1	2011	98.0	58.1	3.8	3.65	0.0758	52.0	60.9
qTL_3.2	2012	143.0	69.2	2.7	3.40	0.0787	58.1	72.7
qTL_6.1	2012	55.0	50.6	3.2	2.47	0.0638	49.6	53.0
qTL_6.1	2011	102.0	59.8	3.5	2.16	0.0593	49.6	60.9
qTL_7.1	2012	27.0	43.8	3.5	3.55	0.0764	3.1	49.0
qBRCH4.1	2012	121.6	60.0	3.0	4.04	-0.1049	58.9	65.8
qBRCH4.2	2012	153.6	65.2	1.1	2.77	0.0916	63.1	65.8
qBRCH4.2	2011	162.0	66.9	3.5	3.53	0.1203	65.8	67.6
qBRCH5.1	2011	30.6	4.8	4.6	4.70	0.1350	1.5	7.4
qBRCH5.1	2012	30.0	4.8	4.3	5.37	0.1241	1.5	6.1
qBRCH10.1	2012	57.8	17.0	3.3	3.71	-0.0982	9.8	49.4

[1] SNP positions correspond to 1-lod QTL interval

Table 5.9 Parameters of plant architecture related QTLs from interval mapping of the PQ-RIL population

QTL	Env	Peak (cM)	Peak (Mb)	LOD	% Var	Additive Effect	Left Flanking Marker (Mb)	Right Flanking Marker (Mb)
qTL_2.1	2010	55.9	63.2	3.9	12.66	1.30	59.1	60.5
qTL_5.1	2011	6.0	0.2-1.9	2.7	5.77	0.70	0.2	4.5
qTL_7.1	2009	40.0	57.3	2.5	7.12	0.14	0.9	59.4
qBRCH4.1	2011	48.0	51.2	2.6	7.48	-0.49	10.0	58.0
qBRCH8.1	2011	7.5	0-2.9	2.6	7.85	0.50	0.2	3.0

Table 5.10 Comparisons of **TL** and **FL** QTL in the SBSH-BC_1F_2, IS-RIL and PQ-RIL population

QTL Name	ISRIL	PQRIL	FLQTL
qTL2.H4.1 (+)		qM1_2.1 (+)	
qTL.4A.H4.1 (+)		qTL4.1 (-)[1]	
qTL.4D.H4.1 (+)		qTL4.1 (-)[1]	
qTL7.H4.1 (+)	qTL7.1 (+)	qTL7.1 (+)	
qTL.1D.H6.1 (+)	qTL_1.1 (+)		qFL1C.H6.1 (+)
qTL.2C.H6.1 (+)		qM1_2.1 (+)	
qTL.3E.H6.1 (+)	qTL_3.1 (+)		
qTL4.H6.1 (+)			qFL4A.H6.1 (-)
qTL4.H6.2 (+)			qFL4D.H6.1 (+)
qTL.6A.H6.1 (+)	qTL_6.1 (+)	qM1_6.1 (+) [1]	
qTL.6B.H6.1 (-)	qTL_6.1 (+)	qM1_6.1 (+)	qFL6B.H6.2 (+)
qTL.9B.H6.1 (+)			
qTL.10C.H6.1 (-)			qFL10A.H6.1 (-)
qTL10.H6.2 (+)			qFL10.H6.1 (+)

[1] From Kong et al, (2014).

Table 5.11 Comparisons of BRCH QTL in SBSH-BC_1F_2, IS-RIL and PQ-RIL populations

QTL Name	ISRIL	PQRIL	FLQTL
qBRCH1.H4.2 (+)		qAX1.1 (+)	
qBRCH1.H4.1 (+)		qAX1.1 (+)	qFL.1A.H4.1 (-)
qBRCH.1F.H4.1 (-)			
qBRCH.2D.H4.1 (-)			
qBRCH3.H4.1 (+)		qAX3.1, qIM3.1 qVG3.1, qSR3.1, qTR3.1 (+)	
qBRCH.4C.H4.1 (-)			
qBRCH4.H4.1 (+)			
qBRCH.4D.H4.1 (-)	qBRCH4.1 (-)	qBRCH4.1 (-)	
qBRCH.5C.H4.1 (-)	qBRCH5.1 (+)	qTR5.1 (-)	
qBRCH.6B.H4.1 (-)			qFL.6B.H4.1 (+)
qBRCH6.H4.2 (+)			
qBRCH.7C.H4.1 (+)			
qBRCH9.H4.1 (+)			
qBRCH10.H4.1 (+)			
qBRCH.1C.H6.1 (+)		qIM2_1.1 (+)	
qBRCH.3E.H6.1 (+)		qAX3.1, qIM3.1, qVG3.1, qSR3.1, qTR3.1 (+)	
qBRCH3.H6.2 (+)			
qBRCH4.H6.1 (+)	qBRCH4.2 (+)		qFL.4D.H6.1 (+)
qBRCH5.H6.1 (+)			
qBRCH.5C.H6.1 (+)			
qBRCH.6B.H6.1 (-)			qFL.6B.H6.1 (+)
qBRCH.6B.H6.2 (-)			qFL.6B.H6.2 (+)
qBRCH6.H6.1 (+)			qFL.6B.H6.2 (+)
qBRCH6.H6.3 (+)			qFL.6E.H6.1 (+)
qBRCH7.H6.1 (+)		qVG7.1 (+), qIM2_7.1 (+), qSR_7.1 (-)	
qBRCH9.H6.1 (+)		qTR9.1 (+)	
qBRCH.10C.H6.1 (+)	qBRCH10.1 (-)		
qBRCH.10C.H6.2 (-)	qBRCH10.1 (-)		

Table 5.12 A mixed-effect model for predicting **Biomass** (natural log transformation) in the SBSH-BC_1F_2 population.

Groups	Variance	Std. dev.
Env	0.01589	0.126
Residual	0.08324	0.2885

	Sum Sq	Mean Sq	DF	F-stat	P value
PH	32.561	32.561	1	391.19	< 2.2e-16 ***
MD	10.609	10.609	1	127.46	< 2.2e-16 ***
TL	25.421	25.421	1	305.42	< 2.2e-16 ***
BRCH	2.303	2.303	1	27.67	1.901e-07 ***
FL	4.892	4.892	1	58.77	4.396e-14 ***
Population	1.944	1.944	1	23.36	1.564e-06 ***

PH: plant height

MD: mid-stalk diameter

TL: number of mature tillers

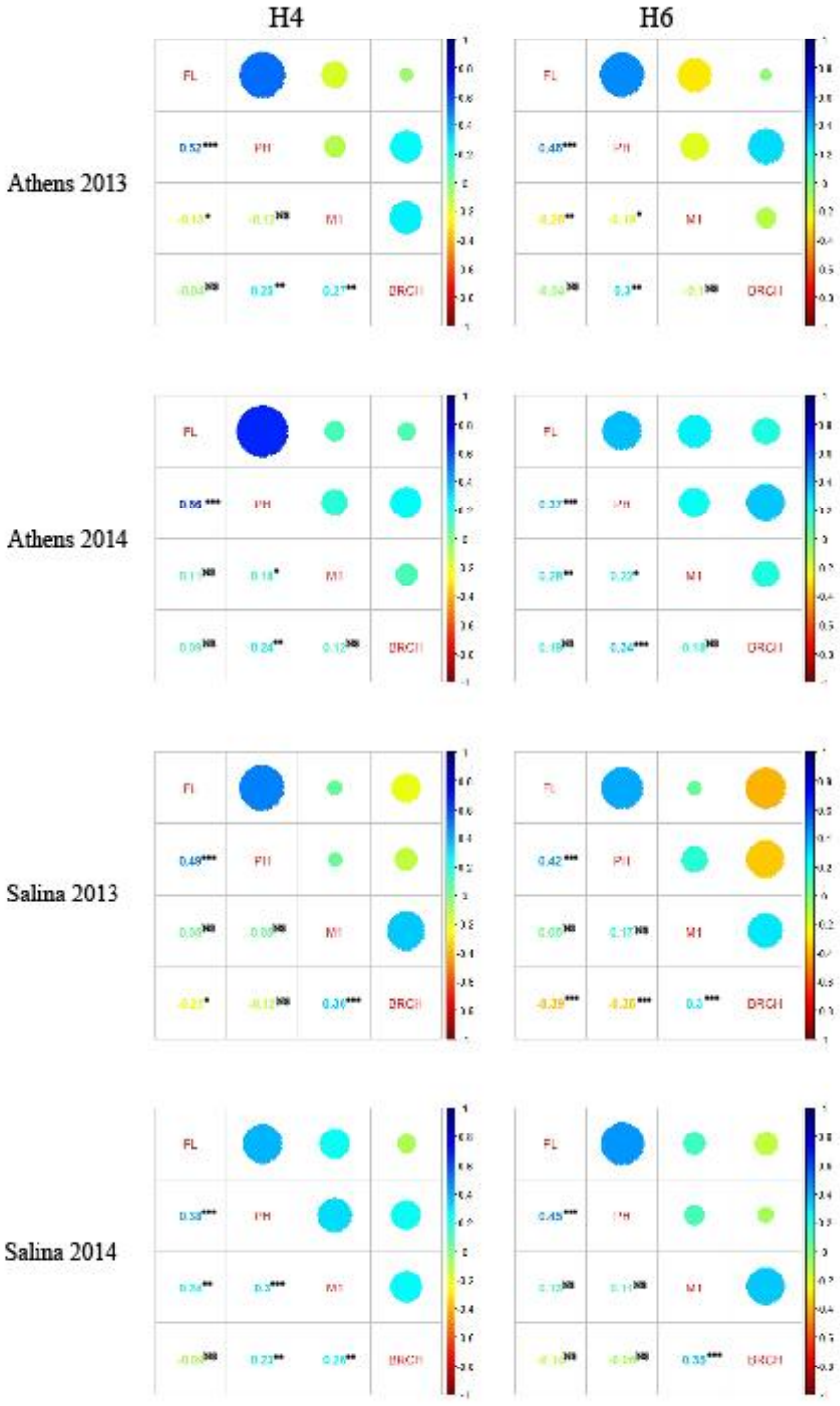
BRCH: number of secondary branches per tiller

Env: environmental effects

Table 5.13 Parameter estimation for a mixed-effect model for predicting **Biomass** in the SBSH-BC_1F_2 population

	Estimate	Std. Error	df	t-stat	P-value
(Intercept)	2.6460	0.1066	19.8	24.815	2.22E-16***
PH	0.006197	0.000313	936.3	19.779	< 2e-16***
MD	0.02962	0.002624	953.4	11.290	< 2e-16***
TL	0.1409	0.008061	675.1	17.476	< 2e-16***
BRCH	0.06978	0.01327	720.6	5.260	1.9E-07***
FL	0.007882	0.001028	942.7	7.666	4.4E-14***
Population H6	-0.09472	0.01960	954.8	-4.833	1.56E-06***

Figure 5.1: Correlation coefficients among days to flowering (**FL**), plant height (**PH**), number of mature tillers (**TL**) and number of secondary branches (**BRCH**) in the H4 and H6- derived SBSH-BC_1F_2 populations in four environments.


H4
H6
Athens 2013
Athens 2014
Salina 2013
Salina 2014
FL
PH
MI
DRGI

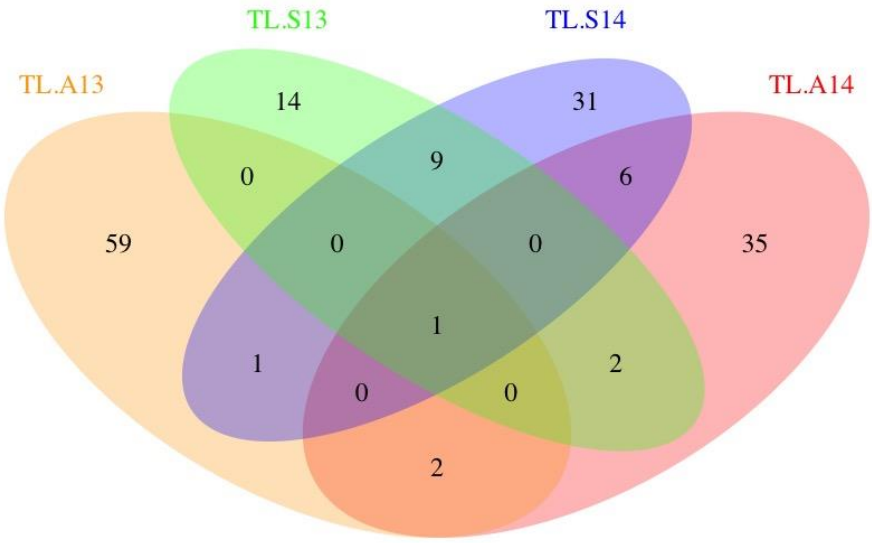
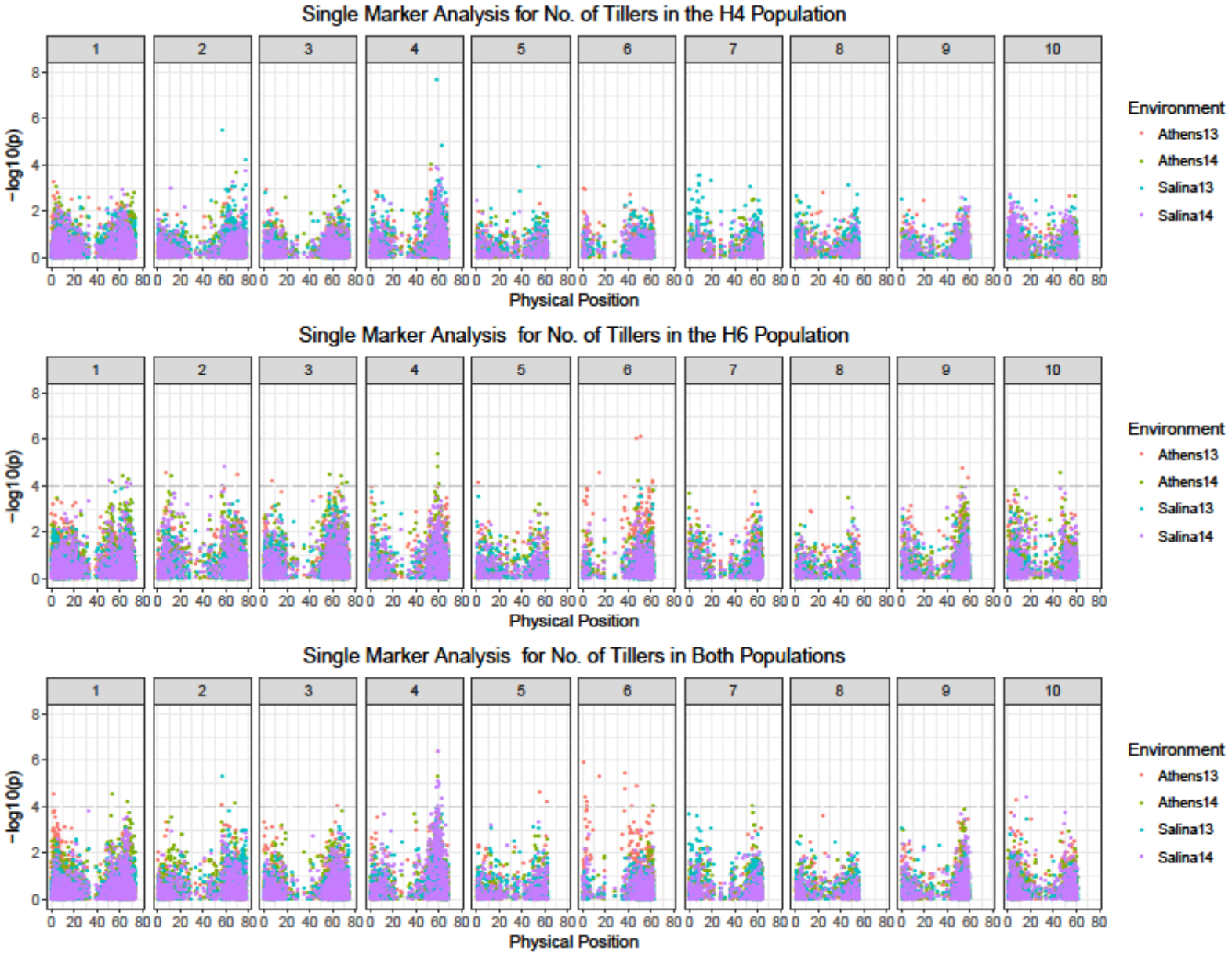


Figure 5.2 Venn diagram of the number of SNP markers for tillering (TL) significant at a p value<10^{-3} in different environments for pooled SBSH BC_1F_2 populations

Figure 5.3: Single marker analysis of the number of mature tillers in the H4, H6 and pooled SBSH BC_1F_2 populations


Single Marker Analysis for No. of Tillers in the H4 Population
Single Marker Analysis for No. of Tillers in the H6 Population
Single Marker Analysis for No. of Tillers in Both Populations
1
2
3
4
5
6
7
8
9
10
−log10(p)
Physical Position
Environment
Athens13
Athens14
Salina13
Salina14

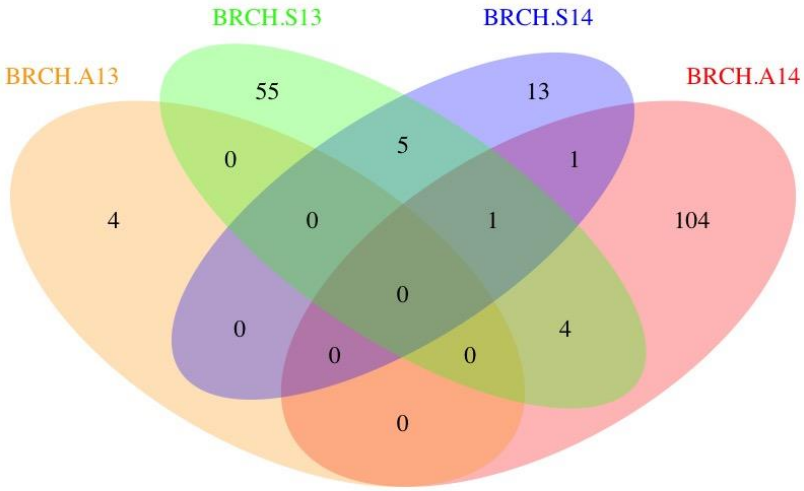
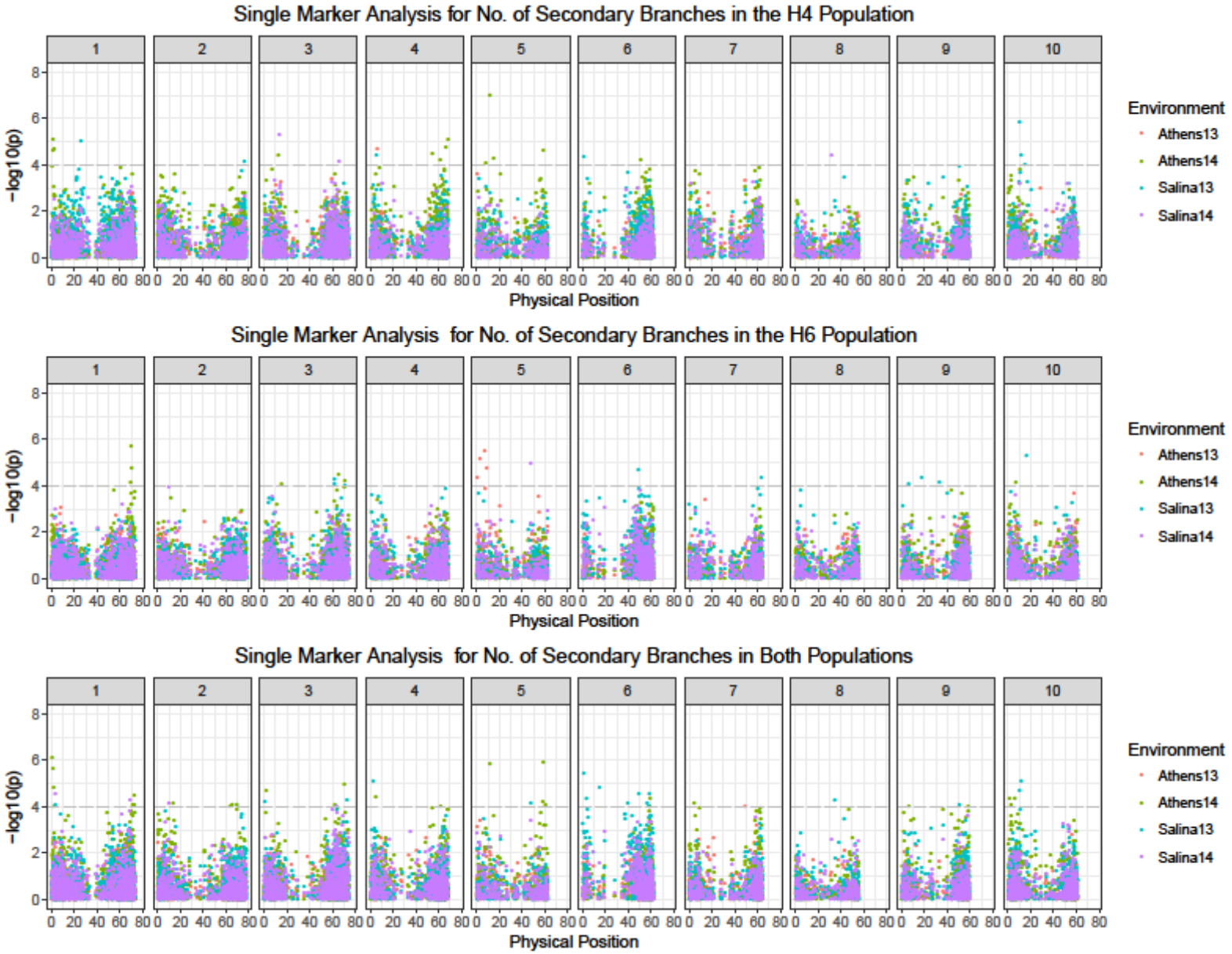


Figure 5.4: Venn diagram of the number of SNP markers for secondary branches per tiller (BRCH) significant at a pvalue<10^{-4} in different environments for SBSH BC_1F_2 pooled populations

Figure 5.5 Single marker analysis for the number of secondary branches per tiller in the H4, H6-derived and the pooled BC_1F_2 populations


Single Marker Analysis for No. of Secondary Branches in the H4 Population
Single Marker Analysis for No. of Secondary Branches in the H6 Population
Single Marker Analysis for No. of Secondary Branches in Both Populations
1
2
3
4
5
6
7
8
9
10
−log10(p)
Physical Position
Environment
Athens13
Athens14
Salina13
Salina14

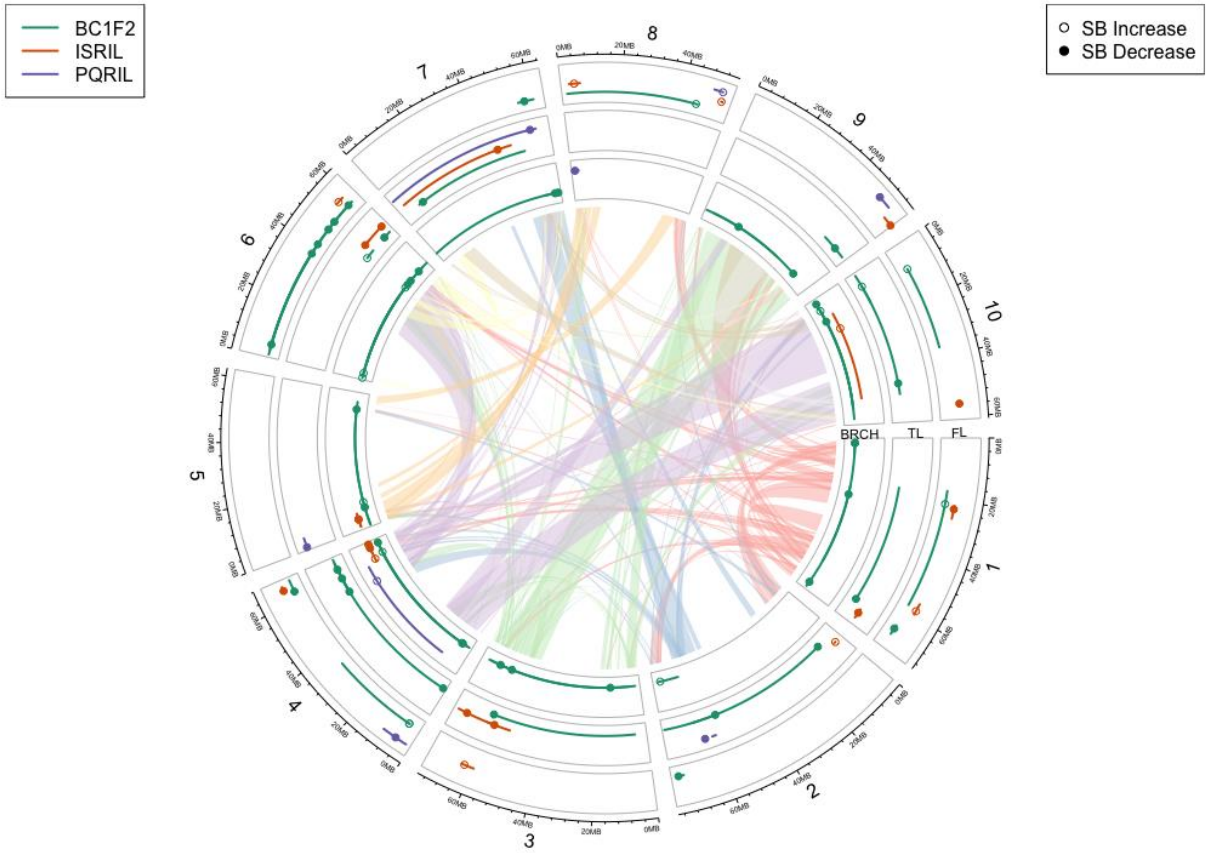


Figure 5.6 QTL correspondence for flowering time (**FL**), tillering (**TL**), and secondary branching (**BRCH**) in SBSH-BC_1F_2, ISRIL and PQRIL populations in physical distance. Links are the duplication events in sorghum (Lee et al. 2013)

CHAPTER 6

UNRAVELING THE GENETIC COMPONENTS OF PERENNIALITY TOWARD BREEDING FOR PERENNIAL SORGHUM[5]

[5] Kong, WQ., Nabukalu, P., Robertson, J., Goff, V., Pierce, G., Lemke, C., Compton, R., Cox, S. Paterson, AH.. To be submitted to Theoretical and Applied Genetics.

Abstract

Perennial crops may provide consistent food and biomass supplies while preserving ecological capital and reducing water and energy inputs. Sorghum, one of the few multi-purpose crops that provide both grain and biomass in some of the world's most adverse conditions, has two perennial relatives as well as rich morphological diversity occurring during divergent evolution both in the wild and under domestication, making it a good candidate for breeding for perenniality. We describe a quantitative study to elucidate the genetic components conferring perenniality related traits, mainly rhizomatousness and winter survival, in two novel BC_1F_2 populations totaling 246 genotypes derived from backcrossing two *Sorghum bicolor* x *S. halepense* F_1 plants to a tetraploidized *S. bicolor*. Phenotyping for two years in Bogart, GA and Salina, KS permits understanding of the relationship between rhizomatousness and winter survival, providing materials for future plant breeding. Correspondence of rhizomatousness quantitative trait loci (QTLs) with two populations derived from its progenitors, *S. bicolor* × *S. propinquum*, suggests the set of QTLs inherited and genetic novelty arisen after the divergence from its progenitors, while comparisons to tillering and branching QTLs further support their developmental relationship of these two organs. An unexpected finding supported from both the BC_1F_2 and PQ-F_2 populations suggests alleles contributing to late flowering are related to fewer rhizomatousness. Interestingly, twelve out of sixteen QTL regions conferring rhizomatousness fall in genome-duplicated regions, indicating that those genes retain their function after the duplication event 96 million years ago. This study also assists in narrowing down candidate genes for rhizomatousness from expression profiling, provide diagnostic DNA markers, and facilitate breeding for perenniality in sorghum.

Introduction

The global population is projected to reach 8.5 billion by 2030 and 9.7 billion by 2050, with demand for food predicted to double by 2030. Meanwhile, the area of agricultural land and arable land per capita has been decreasing (www.fao.org). Annual cropping systems that currently dominate production of row-crops require excessive water, extensive tillage and considerable amounts of synthetic fertilizer which disturb natural biological processes and cause soil degradation and erosion (Montgomery 2007). Cropping systems that provide multiple harvests from single plantings, by ratooning or perenniality, have recently gained recognition for their ability to maintain or improve agroecological conditions on marginal land while also substantially improving water infiltration of soils, preventing soil erosion, reducing flooding and increasing runoff water quality (Glover et al. 2010). Perennial systems may also benefit smallholder farmers by amortizing seed costs over multiple cropping seasons and reducing labor for tilling and weeding where equipment is limited.

Rhizomes are subterranean storage organs that develop from axillary buds on the basal portion of the stem or specialized creeping roots and grow diageotropically. Rhizomes are a major means for perennial grasses, including some invasive species such as johnsongrass [*Sorghum halepense* (L.) Pers.], bermudagrass [*Cynodon dactylon* (L.) Pers.], and reed (*Phragmites australis*) to overwinter and store nutrients to support spring regrowth. In addition to rhizomatousness, ratooning, in which crops are harvested and allow to regrow from stubble to produce another crop, is widely practiced in sugarcane (Hunsigi 1989), a close relative of sorghum. The benefits of ratooning conventional rice cultivars motivate its practice on 68-77% of the rice area in Texas (Wilson et al. 2014), and 25-32% in Louisiana (https://www.lsuagcenter.com/en/crops_livestock/crops/rice/Statistics/Rice-Maps.htm#16), with

progress being made in breeding for perenniality (Hu et al. 2003). Ratooning is occasionally practiced in sorghum (McCormick et al. 1995; Paterson et al. 1995b) and there is variation among sorghum cultivars for ratoon crop productivity (Duncan et al. 1980; Duncan and Moss 1987), however breeding efforts are needed to stabilize ratoon-crop grain yields (Duncan and Moss 1987).

Perenniality through ratooning may be highly correlated with the ability to generate tillers, which are developmentally related to, but functionally distinct from, rhizomes (Kong et al. 2014). Both tillers and rhizomes are initiated from axillary meristems at the base of a plant, but experience differentiation along a positional gradient, for the former develops in an upward direction and produces flower and seeds, and the latter grow horizontally and stores nutrients for spring regrowth (Gizmawy et al. 1985).

Study of the molecular control of perenniality has been limited, with some quantitative and comparative studies initiated in rice, sorghum and maize (Paterson et al. 1995b; Hu et al. 2003; Westerbergh and Doebley 2004; Sacks et al. 2006; Hu et al. 2011; Kong et al. 2015). Comparative studies reveal that QTLs responsible for rhizomatousness in rice and sorghum fall largely in corresponding genomic regions (Hu et al. 2003), while some of these also correspond to maize (Westerbergh and Doebley 2004), supporting the notion that independent but convergent mutations at corresponding loci might exert large effects during grass domestication (Paterson et al. 1995a). Correspondence between divergent grasses such as panicoid sorghum and oryzoid rice also suggests that locations genes controlling rhizomatousness may extrapolate to many other species. However, it still remains unclear which specific genes are responsible for development and metabolism of the rhizomes.

Sorghum has become an excellent model in which to dissect the genetic basis of 'ratooning', rhizomatousness and overwintering, largely due to crosses made with *S. propinquum*, a diploid tropical plant that produces abundant tillers and rhizomes (Paterson et al. 1995b; Kong 2013; Kong et al. 2015). A segregating population derived from crossing *S. bicolor* with *S. propinquum* has permitted unraveling of genomic regions conferring rhizomatousness and other traits related to perenniality (Jang et al. 2006). Genes associated with rhizome expression are somewhat enriched in the QTL intervals and cover a wide range of functional categories (Jang et al. 2009). *S. halepense*, a naturally-formed polypoid derived from interspecific hybridization of *S. bicolor* and *S. propinquum*, has more extensive rhizomes than its rhizomatous progenitor, perhaps due in part to some 'recruitment' of genes from *S. bicolor* to expression in rhizomes (Jang et al. 2009).

In this project, we describe a quantitative trait locus (QTL) study for traits related to perenniality in two previously described BC_1F_1 populations (Kong, 2017 in preperation) derived from crosses of tetraploid *Sorghum bicolor* breeding line 'BTx623' × *S. halepense* accession 'Gypsum 9E'. Phenotypes of their progeny lines have been evaluated in two years, 2013 and 2014, both in Salina, KS and Athens, GA. QTLs associated with perenniality from this study provide for comparison of rhizome growth to vegetative branching patterns in the same population, further advancing knowledge of the developmental relationships between these two similar but functionally distinct organs (Paterson et al. 1995b; Kong et al. 2015). Comparison of rhizome and flowering time QTLs from the same populations may provide insight into the relationship of apical dominance to rhizomatousness. Because *S. propinquum* is thought to have been the progenitor that contributed rhizomatousness to *S. halepense*, comparing the results of this study to those derived from *S. bicolor* × *S. propinquum* (Paterson et al. 1995b; Kong et al.

2015) may provide new insight into the evolution of genetic novelty by polyploid *S. halepense*, facilitating understanding of its competitiveness and aggressiveness.

Materials and Methods

Genetic Stocks

Details of development of the two *S. bicolor* × *S. halepense* BC_1F_2 populations (SBSH-BC_1F_2) can be found in Kong et al (2017). Results from these two populations were compared *S. bicolor* × *S. propinquum* F_2 (PQ-F_2) and RIL (PQ-RIL) populations (Paterson et al. 1995b; Kong et al. 2015).

Genetic Mapping

Details of genotyping and construction of genetic maps can be found in Kong, et al. (2017).

Phenotype

Aboveground vegetative shoots emanating from rhizomes were quantified as an indicator for rhizomes. In both Athens 2013 and 2014, the number of rhizome-derived shoots and the distance of the rhizome-derived shoots ('rhizome distance') from the middle of the source crown of three representative plants (reps) was recorded for each genotype in two fields (block), while in Salina 2013 and 2014, we recorded the total number of rhizome-derived shoots and the rhizome distance for each genotype in each field. Using an indicator variable for presence (1) or absence (0) of at least one rhizome for each plant in Athens 2013 and 2014, we calculated the proportion of rhizomatousness **(Rhiz**) for each genotype by dividing the number of plants with rhizomes by the total number of plants (6, if there were no missing data). For Salina 2013 and 2014, **Rhiz** is 0 for non-rhizomatousness for both fields, 0.5 for rhizome occurrence in one out of two fields and 1 for rhizome occurrence in both fields. We created another categorical variable,

rhizome groups (**RG)**, to collapse **Rhiz** into five groups to smooth the distribution: 0 for **Rhiz**=0, 1 for **Rhiz**>0 and **Rhiz** <=0.3, 2 for **Rhiz** >0.3 and **Rhiz** <=0.6, 3 for **Rhiz** >0.6 and **Rhiz** <0.9 and 4 for **Rhiz** >0.9. The number of rhizomes for each plant (**RN**) were recorded and the average number of rhizomes for each genotype were calculated for QTL analysis. In addition, the average rhizome distance (**RD**) for each plant were recorded and the average **RD** for each genotype were calculated for QTL mapping. However, if a plant does not have rhizomes, the distance would be treated as zero. In partial summary, the three variables for QTL analysis are thus **RG**, **RN** and **RD**.

In addition to measurements of rhizomatousness, we also recorded the number of plants that survived (**Survival)** winters based on production of above ground shoots in the following spring. The number of SBSH-BC_1F_2 plants that survived was very few in Athens 2013 as well as both years in Salina due to extreme winter conditions. We created a categorical variable with three levels of survival (**ISur**) with 0 being no survival, 0.5 indicating that at least one plant for each genotype in one of two fields survived, and 1 indicating that at least one plant for each genotype survived in both fields.

Genetic Analysis

To fully utilize the available data while protecting against false-positive results, genetic analysis employed two approaches. Using genetic maps that were constructed as described (Kong, 2017) from selected well-groomed SNP segregation data for each of the two SBSH-BC_1F_2 populations, interval mapping was conducted as described below. In addition, single marker analysis was conducted using each SNP marker that met quality standards described (whether in the genetic map or not), using hierarchical clustering to separate SNP markers on potential different homologous chromosomes and inferring QTLs only if more than 4 SNPs were

found within a cluster cut at height of 0.3 in recombination frequency to mitigate spurious associations. Similarities and differences in the results of these analyses were addressed in results.

Single marker analysis

Single marker analysis was applied to both pooled data, which we fit a model with both populations while at the same time using the population as a covariate, and to each of the two population. To predict **RG**, we used analysis of variance (ANOVA) with respect to each marker genotype for Athens 2013 and 2014 data, and logistic regression for Salina 2013 and 2014 data since **RG** distribution in Salina is more suitable for logistic regression. To predict both **RN** and **RD**, we used ANOVA with the response variable transformed with square root only for data from Athens 2013 and Athens 2014, since the binomial distribution of **RN** in Salina 2013 and Salina 2014 is the same with modeling **RG**. To predict **ISur** in Athens 2014, we used ordinal logistic regression with respect to each marker genotype. SNP markers significant at a p value of 10^{-3} for each trait in each sub-population were clustered using hierarchical clustering with pair-wise recombination frequency as the distance measure (Kong et al, 2017 in preparation). To further mitigate spurious associations, QTLs were claimed only if more than 4 SNPs were found within a cluster cut at height of 0.3 in recombination frequency. Peak SNPs were chosen based on the smallest p-values. Details of single marker QTL analysis can be found in Kong et al, 2017 (in preparation). Statistical analysis used R Broman (2003).

QTL interval mapping

A total of 246 individuals were included in the mapping population, with 141 derived from the H4 parent and 105 from the H6 parent. For **RG** in Athens, interval mapping was

applied treating **RG** as normally distributed; interval mapping with a binomial model using only 0 and 1 for **RG** was applied for Salina (Arends et al. 2010; R Core Team 2016). QTL analysis for **RN** used interval mapping while treating the response variable with the square-root transformation in Athens 2013 and 2014, since the result of **RN** was similar to **RG**, both suggesting a binomial distribution. For **RD**, we fitted a two-part model described in (Warwick et al. 1986; Paterson et al. 1995b; Kong et al. 2015). The variances and effects for **RD** QTLs were obtained for both binary and normal models. For binary data, the variance and effects were obtained from a logistic regression model for presence and absence of rhizomatousness. For normal data, we replaced the non rhizome genotype with missing data and calculated the variances and effects based on the normal model. QTL effects of **RD** are inferred based on the significant model, either binary or normal. Data analyses used R and R/qtl (Paterson et al. 1995b).

Results

Summary Statistics

Phenotypic data were recorded for a total of 246 individuals in 2013, with 10 individuals missing in 2014 (Table 6.1). About 1.2% and 22.0% of progeny lines had no rhizomes in Athens 2013 and 2014 respectively, while 46.3% and 48.7% of the progeny lines have no rhizomes in Salina 2013, and 2014, respectively (Table 6.1). We observed higher **RN** in Athens 2013 than any other test environment, 0.595 more rhizomes on average than the second highest environment, Athens 2014 ($t=12.22$, $p<0.001$,Table 6.2). The difference between **RN** of the pooled data (both populations), in Athens 2014 is not significantly higher than that in Salina 2013 ($t=0.5$, $p=0.617$), but Salina 2013 is significantly higher than Salina 2014 ($t=7.094$, $p<0.001$). Average rhizome distances (**RD**), after excluding progenies without rhizomes, is the

highest in Athens 2013 and lowest in Athens 2014 (Table 6.3), and this difference is significant ($t=10.68$, $P<0.001$) for the pooled data, while the difference of the pooled data between the two years in Salina is not significant ($t=0.53$, $p=0.60$).

We observed very little winter survival in the two environments in Salina, but a modest number of surviving plants in Athens (Table 6.4). A total of 25 plots survived in Athens 2013, 10 in one field (Block) and 15 in the other field (Block), and two genotypes survived in both fields. Survival of Athens 2014 was significantly higher, with a total of 117 plots surviving in the first field, 50 in the second field, and 30 in both (Table 6.4). Low survival in Athens 2013 might be related to unusually cold conditions, exemplified by a low temperature of -13.9 °C on Jan 7th, 2014.

Rhizome spread and depth are associated with winter survival (Kong et al. 2015). In the relative severe winter of Athens 2013, only rhizomatous plants survived, while in Athens 2014 some progeny lines survived without deep rhizomes (Table 6.5 and Table 6.6).

QTL analysis

Presence and absence of rhizomatousness collapsed in groups (**RG**)

In the H4-derived population, the only QTL detected, qRG6B.H4.1 is significant in Athens 2014, with a LOD score of 2.6 and an allele substitution effect of -0.71. In the H6-derived population, we detected a total of 13 QTLs with two, qRG.6A.H6.1 and qRG.6B.H6.1, significant in multiple environments (Table 6.7). A total of 9 out of 13 QTLs show negative allele effects, indicating that *S. halepense* decreases the occurrence of rhizomatousness (Table 6.7). Two QTLs detected in Athens 2013 collectively explain about 24.94% of the phenotypic variance in that experiment; seven QTLs in Athens 2014 explain about 47.26% of the phenotypic variance; four QTLs in Salina 2013 explain about 24.79% of the total phenotypic variance; and

three QTLs in Salina 2014 explain about 20.79 % of the total phenotypic variance. It worth noting that effects from Salina are on a different scale than those from Athens, since the Salina data were fitted using a logistic regression model and the corresponding effects represented odds ratios for occurrence of rhizomatousness.

Using single-marker analysis, we detected a total of 19, 51, 10 and 33 SNP markers significant at the level of 0.001 for pooled data (Figure 6.2). Based on these data, in the H4-derived population, a total of two QTLs were significant (Table 6.8), with *S. halepense* alleles at qRG.1.H4.1 increasing the probability of rhizome occurrence and qRG.6.H4.1 decreasing it. In the H6-derived population, a total of seven QTLs were significant, on chromosomes 1, 2, 3, 4, 6 (2) and 9, with two QTLs, qRG.1.H6.1 and qRG.6.H6.1 showing negative allele effects, indicating *S. halepense* alleles decreased rhizome occurrence (Table 6.8). Single marker analysis agreed well with the interval mapping but detected additional two QTLs, qRG1.H4.1 and qRG3.H6.1.

*Number of rhizomes (**RN**)*

For **RN**, we only analyzed data for the two Athens environments, since rhizome data from Salina suggest a logistic regression model, which is the same with the result as **RG** in Salina. In the H4-derived population, we detected only one QTL, qRN.6B.H4.1, significant in Athens 2014, which explained about 11.12 % of the total phenotypic variance with a negative allele substitution effect (Table 6.9). In the H6-derived population, we detected a total of 11 QTLs for **RN**, with one QTL, qRN.6B.H6.1, significant in both environments. A total of three QTLs were found significant in Athens 2013, collectively explaining about 32.53% of the total phenotypic variance; a total of nine QTLs were found significant in Athens 2014, collectively explaining about 43.33% of the phenotypic variance. Among the 11 QTLs detected for **RN**, eight

QTLs overlapped with the **RG** QTLs, while three QTLs, qRN.1B.H6.1, qRN.6E.H6.1 and qRN.9C.H6.1 are newly detected for **RN**.

Totals of 29 and 60 SNP markers were found significant at a level of 0.001 of which 17 and 53 (Figure 6.3) show negative allele effects for pooled data in Athens 2013 and 2014 respectively. Based on these data, in the H4-derived population, we inferred totals of 21 and 12 markers to be significant with 3 and 8 showing negative allele effects in Athens 2013 and 2014, respectively. In the H6-derived population, we found a total of 35 and 124 markers significant with 10 and 62 markers showing negative allele effects.

*Rhizome Distance (**RD**)*

With a two-part model, we can deconvolute QTLs affecting the occurrence of rhizomes from those affecting the distance that rhizomes spread. In the H4-derived population, we only detected one QTL significant in Salina 2014, qRD.2D.H4.1, significantly reducing the occurrence of rhizomes by an odds ratio of -1.46 (Table 6.10). In the H6-derived population, we detected a total of 16 QTLs for **RD**, with three QTLs, qRD.4D.H6.1, qRD.6A.H6.1 and qRD.6B.H6.2 significant in multiple environments. A total of five QTLs were found significant in Athens 2013 with all affecting the distance of rhizomes suggested by the large 'LOD normal' scores, which agrees with our observation that **RD** in Athens 2013 is generally normally distributed (Figure 6.1). One QTL, qRD.6B.H6.2, showed negative allele effect in Athens 2013 (Table 6.10), suggesting that *S. halepense* alleles decrease **RD**. A total of three QTLs were found significant in Athens 2014, with all three affecting both the occurrence and the distance of rhizomes. Two QTLs, qRD.1C.H6.1and qRD.6B.H6.2 show negative effects in Athens 2014. A total of five QTLs were found significant in Salina 2013, with two QTLs, qRD.6B.H6.3 and qRD.9B.H6.1 only affecting occurrence of rhizomes and two QTLs qRD.8A.H6.2 and

qRD.9A.H6.1 only affecting rhizome distance, and one QTL, qRD.8B.H6.1, affecting both occurrence and distance of rhizomes. Three QTLs from Salina 2013, qRD.6B.H6.3, qRD.8A.H6.2 and qRD.8B.H6.1, show negative allele effects. Curiously, the LOD scores of qRD8B.H6.1 are significant in both binary and normal models, and the effects suggest that this QTL decreases the presence of rhizomes but increases the rhizome distance. A total of six QTLs were found significant in Salina 2014, with four QTLs, qRD.6A.H6.1, qRD.6B.H6.1, qRD.8A.H6.1 and qRD.10C.H6.1 affecting the occurrence and two, qRD.4A.H6.1 and qRD.4D.H6.1, affecting the distance of rhizomes. A total of four QTLs, qRD.4A.H6.1, qRD.6B.H6.1, qRD.8A.H6.1 and qRD.10C.H6.1 in Salina 2014 show negative allele effects.

Using single marker analysis, we detected totals of 17 and 37 SNP markers significant at a level of 0.001 for **RD** (Figure 6.4), with 10 and 34 showing negative allele effects for the pooled data in Athens 2013 and 2014, respectively. Based on these data, in the H4-derived population, we found totals of 15 and 7 markers significant with 8 and 4 markers showing negative allele effects in Athens 2013 and Athens 2014, respectively. In the H6-derived population, we found totals of 37 and 64 markers significant with 8 and 37 showing negative allele effects in Athens 2013 and 2014, respectively.

Survival

The map-based QTL analysis failed to discover any statistically significant QTLs for winter survival in Athens 2014 (the only test environment for which sufficient numbers of lines survived to make inferences). Using a cumulative logistic regression with single marker analysis by pooling the H4 and H6 populations, we detected a total of 28 SNP markers significant at a level of 0.001 (Figure 6.5) with 10 showing negative allele effect for the pooled data. In the H4-derived population, we only found two markers significant with one showing negative effect,

while in the H6-derived population, we found a total of six markers significant with one showing negative effect. Increasing the number of individuals or testing in milder conditions may assist in discovering significant regions of chromosomes associated with winter survival. Among the 28 significant SNP markers, 16 came from a potential genomic region from 53.5-73.2 Mb on chromosome 3. The majority of SNP effects in this region are positive, indicating that *S. halepense* increases winter survival. This QTL for winter survival may also correspond to a rhizome QTL detected from single-marker analysis (Table 6.8), qRG3.H6.

Comparisons of rhizomatousness QTLs to other sorghum studies

Both **RG** QTLs detected in the H4-derived populations, qRG.1.H4.1 and qRG,6B.H4.1, overlap with rhizome QTLs from the PQ-F_2 population, while 9 of 14 **RG** QTLs in the H6 population overlap with PQ-F_2 rhizome QTLs (Paterson et al. 1995b; Kong et al. 2015). Totals of one and five QTLs respectively detected from H4 and H6-derived populations overlap with rhizome QTLs from the PQ-RIL population (Paterson et al. 1995b). Four QTLs, qRG.1.H4.1, qRG.1C.H6.1, qRG.3.H6.1, and qRG9B.H6.1 overlap in all three populations. All **RG** QTLs from the H4-derived population and a total of six QTLs, qRG.2C.H6.1, qRG.3.H6.1, qRG.4B.H6.1, qRG.4B.H6.2, qRG.6B.H6.3 and qRG.9B.H6.1 from H6-derived population overlap between at least two studies with corresponding allele effects (Paterson et al. 1995b; Kong et al. 2015). Three of the six SBSH-BC_1F_2 QTLs, qRG.2C.H6.1, qRG.6B.H6.2 and qRG.6B.H6.3 overlapping with LSR in the PQ-F_2 population (Paterson et al. 1995b; Kong et al. 2015), show negative allele effects, indicating that alleles from *S. halepense* or *S. propinquum* decrease rhizomatousness. In addition, a total of three SBSH-BC_1F_2 QTLs, qRG.1C.H6.1, qRG.8B.H6.2 and qRG.10B.H6.1 showed negative allele effects while previous studies found

positive effects. Two rhizome QTLs, qRG8B.H6.1 and qRG10C.H6.1 are unique to the BC_1F_2 population.

Comparisons of rhizomatousness and vegetative branching QTLs

Previously, we found correspondence between rhizomatousness and vegetative branching QTLs in six general genomic regions, possibly due to developmental relationships between the two organs (Paterson et al. 1995b; Kong et al. 2015). A total of four QTLs for rhizomatousness from the H6-derived population, qRG.4B.H6.2, qRG.6A.H6.1, qRG.6B.H6.3, qRG.9B.H6.1, correspond only to H6-derived **TL** (mature tillers) QTLs comparing their physical distances; three H6-derived rhizomatousness QTLs, qRG.6B.H6.1, qRG.6B.H6.2, qRG.10C.H6.1, correspond only to **BRCH** (secondary branches per mature tiller; Table 6.11) and two H6-derived rhizomatousness QTLs, qRG.3.H6.1 and qRG.10C.H6.1, correspond to both **TL** and **BRCH**. Directions of QTL effects between the two traits are consistent within the SBSH-BC_1F_2 population: *S. halepense* alleles that increase **TL** or **BRCH** also increase **RG,** and *vice versa*. A total of five SBSH-BC_1F_2 rhizome QTLs, qRG.4B.H6.2, qRG.6B.H6.1, qRG.8B.H6.1, qRG.8B.H6.2 and qRG.9B.H6.1, overlap with various levels of vegetative branching QTLs in the PQ-RIL population with two QTLs, qRG.8B.H6.1 and qRG.8B.H6.2, suggesting opposite effects compared to the previous study (Table 6.11).

Comparisons of rhizomatous QTLs to flowering time

A total of six QTLs, qRG.1C.H6.1, qRG.6A.H6.1, qRG.6B.H6.1, qRG.6B.H6.2, qRG.6B.H6.3 and qRG.10C.H6.1 controlling rhizome presence overlap with flowering QTLs (Table 6.11). Five of six pairs of corresponding QTLs (except qRG.6A.H6.1) show negative allele effects for rhizomatousness but positive for flowering time, suggesting that late flowering is associated with reduced rhizomatousness. We observed the same directions of additive effects

for overlapping rhizome (LSR) and flowering QTLs (Flravg) on chromosomes 2 and 6 in the PQ-F_2 population. Correspondence between qRG.6A.H6.1 and qFL.6E.H6.1 suggests that the *S. halepense* alleles increase both days to flowering and rhizomatousness, while the correspondence on chromosome 10, qRG.10C H6.1 and qFL.10A.H6.1 suggest that *S. halepense* alleles decrease both traits.

Expression patterns of rhizomatousness

Among the 162 genes upregulated in rhizome buds compared to rhizome-derived shoots, 45 genes were within the SBSH-BC_1F_2 QTL intervals for **RG**. A total of nine genes were found within QTL intervals both in the PQ-F_2 and SBSH-BC_1F_2, six in both SBSH-BC_1F_2 and PQ-RIL, and three in all three populations. Many of the significant genes are involve in protein synthesis, agreeing with other transcriptomic and proteomic studies in rhizomes (He et al. 2012; Ma et al. 2016). Among the 94 genes upregulated in rhizome-derived shoots compared to rhizome buds, 26 were within the QTL intervals for **RG** in the SBSH-BC_1F_2 population. A total of nine genes were found within QTL intervals both in the PQ-F2 and SBSH-BC_1F_2, one in both SBSH-BC_1F_2 and PQ-RIL, and none in all three populations.

QTL regions correspondence to paleo-duplication

Among a total of 18 rhizome QTLs mapped to date in the three populations studied, 12 (66.7%) occur in duplicated chromosomal regions tracing to a 96 million-year old genome duplication (Wang et al. 2015) (QTL intervals have been modified as 1 Mb before the peak and 1 Mb after the peak). Noting that there are a total of 202 duplicated regions on the ten sorghum chromosomes (Lee et al. 2013), and that the 12 duplicated regions containing corresponding rhizome QTLs cover only approximately 26.0% of the total sorghum genome (Figure 6.6), the likelihood that so many QTLs would occur in corresponding duplicated regions by chance is

small. This finding suggests that both members of gene pairs duplicated during the formation of these regions 96 million years ago (Wang et al. 2015) may continue to function in rhizome development. Regions on chromosomes 1 are duplicated within the same chromosome, while the other ten pairs of regions are located on different chromosomes (Figure 6.6). For the 12 pairs of duplicated regions harboring rhizome QTLs, five contain QTLs conferring the same direction of allele effects [chrs. 1-1(+), chrs.3-3 (+), chrs.3-4 (+), chrs.3-9 (+), chrs.6-8 (-), + means the other parent (not BTx623) increase rhizome presence], while the other seven pairs contain QTLs conferring different direction of allele effects [chrs.1(-)-9(+), chrs. 4(+)-6(-), chrs. 4(+)-10 (-) (2), chrs.5(+)-6(-), chrs. 5(+)-8(-), chrs.6(-)-9(+), Figure 6.6].

Discussion

Detailed analysis of BC_1F_1 populations (Kong, 2017) derived from crosses of tetraploid *Sorghum bicolor* breeding line 'BTx623' × *S. halepense* accession 'Gypsum 9E' provide insight into the genetic control of traits key to the spread of one of the world's most widespread weedy and invasive plants, but which might be harnessed in breeding of grains suitable for producing multiple harvests from single plantings. The current study adds new information to genomic regions conferring rhizomatousness in the Sorghum genus and also provides additional support for many previous results (Jang et al. 2009), strengthening the case for extrapolation of this information to other studies and other taxa (Hu et al. 2003). Various degree of correspondence of rhizomatousness with vegetative branching and flowering time QTLs improves knowledge of relationships among these traits, toward clarification of underlying developmental mechanisms and potential trade-offs related to perenniality. Comparisons of rhizome QTLs derived from *S. halepense* populations to those derived from its putative progenitor *S. propinquum*, improves knowledge of genetic novelty that may have been associated with evolution of *S. halepense*.

Rhizome related phenotypes are qualitatively different from previously studied flowering time and plant height traits in the *Sorghum bicolor* × *S. halepense* populations (Kong, et al 2017, in preparation), with less transgressiveness and larger genotype by environment interactions. The relatively mild climate of Athens, GA was generally more effective at discriminating winter survival than Salina, KS (Table 6.1), although the severe winter of 2013 in Athens, GA provided relatively little information. Benefiting from this large winter temperature differences, we concluded that deep rhizomes are essential for plants to survive a relatively severe winter, while short rhizomes or even vegetative tillers of a plant may be adequate for survival under mild weather (Table 6.5 and Table 6.6). This hypothesis suggests that breeding for perenniality under different environments may strive for different ideotypes with different developmental features.

QTLs discovered in this study increase knowledge of rhizomatousness in many ways. First, overlap of *S. halepense* QTLs with those found in previous studies of *S. bicolor* × *S. propinquum* further support previous results (Paterson et al. 1995b; Kong et al. 2015), and indicate the possible set of rhizome QTLs in *S. halepense* inherited from its progenitors. However, rhizome growth of polyploid *Sh* transgresses that of its rhizomatous diploid progenitor. In adjacent growouts, *SbxSh* F_2 progeny had a higher frequency of rhizome-derived shoots emerging from the soil (37.6%), larger average number of rhizomes producing above-ground shoots (0.77), and greater distance of rhizome-derived shoots from the crown (11.97 cm) than *SbxSp* (30%, 0.32, 7.5). Novel QTLs and QTLs from different homologs discovered from these two *Sorghum bicolor* × *S. halepense* populations might be associated with the greater aggressiveness of *S. halepense* than its progenitors due to new *S. halepense*-specific alleles, perhaps related to polyploidization.

Correspondence among rhizome and vegetative branching (**RG**, **TL** and **BRCH)** QTLs in *Sorghum bicolor* × *S. halepense* continue to support that rhizomes and tillers are developmentally related, with many of the same genes expressed in each of these organs(Jang et al. 2009). Two rhizome QTLs, qRG.3.H6.1 and qRG.10C.H6.1, are both correlated with **TL** and **BRCH** QTLs, indicating that these two QTLs might confer genes controlling axillary meristem initiations affecting all three traits. A total of five SBSH QTLs, qRG.1.H4.1, qRG.1C.H6.1, qRG.2C.H6.1, qRG.4B.H4.1, qRG.10B.H6.1 are not associated with any **TL** or **BRCH** QTLs, suggesting those QTLs might represent allelic variation in genes that are only related to rhizome formation.

Au unexpected finding is the relationship between **RG** and **FL** QTLs. A total of four overlapping QTLs from the SBSH-BC_1F_2 and two QTLs from the PQ-F_2 population suggested that later flowering are associated with fewer rhizomes, a surprising finding in view of the fact that rhizomes are vegetative organs and might be expected to grow more during a long vegetative period. Across 4 environments, early flowering is correlated with reduced aboveground vegetative biomass ($r=-0.26$ to -0.62, $p<0.001$), but increased rhizome growth ($r=0.17$ to 0.30, $p<0.001$) in *SbxSh* progeny. A tantalizing hypothesis (Paterson et al. 2017) is that the selective advantage of rhizomes as propagules may outweigh their importance as storage organs, and that rhizome growth has become positively correlated with reproductive growth but negatively correlated with other vegetative growth. While this result is surprising, studies in other species provide evidence that genes controlling flowering are active at floral primordia and shoot apical meristems (Skipper 2002; Gamuyao et al. 2017). A parsimonious model is that 'signals' (e.g., florigens) controlling flowering have also been co-opted to control rhizomatousness, and release of apical dominance by flowering accelerates rhizome

development. This model may also explain an unexpectedly high number of cases in which *S. halepense* alleles were associated with reduced rhizomatousness, as these same alleles were frequently associated with late flowering with 1 out 2 and 9 out of 14 in the H4 and H6-derived population, respectively (Table 6.11). The subset of rhizomatousness genes that were not related to flowering time might function in early development of rhizomes in *S. halepense*. Additional functional analysis is needed to verify this hypothesis.

Identification of genes, regulatory elements, and biochemical functions that are important to rhizome development but dispensable to other plant processes would provide the foundation to apply high-throughput methods (Randall and Mulla 2001) to identify exogenous treatments (plant growth regulators) to specifically target and perturb rhizomes. Such targeting of growth regulation might provide for control of rhizomatous weeds even in closely related crops, for example Johnsongrass (*S. halepense*) in sorghum. Such weeds introduce a host of problems into agricultural systems, reducing the productivity and quality of many crops, serving as an over-wintering site for insects and diseases that subsequently attack crops, and offering opportunity for transgenes to 'escape' and make weeds more difficult to control.

The fact that the majority of **RG** QTLs are located in the duplicated regions in sorghum indicates that the duplicated rhizome genes are still functioning in similar ways despite genome duplication dated 96 million years ago (Wang et al. 2015). Retention of duplicated copies of rhizomatous genes after the polyploidization of *S. halepense* may contribute to its adeptness and invasiveness to the new areas that none of its progenitors have been to (te Beest et al. 2012; Renny-Byfield and Wendel 2014).

The competitiveness that rhizomes confer to weeds might also be turned to advantage – rhizomes contribute to the productivity of many forage and biomass crops, and play a major role

in erosion control. This feature, together with conferring perenniality, makes rhizomatous grasses a promising vehicle for bringing marginal lands into sustainable biomass production , maximizing ecosystem productivity and minimizing losses of topsoil , water, and nutrients . The present project provides DNA markers useful in support of our efforts to breed sorghums that are suitable for producing multiple crops from single plantings (Paterson et al. 2013), and prior evidence (Hu et al. 2003) suggests that these markers may also be useful in the breeding of other ratooning and/or perennial grain crops.

Acknowledgments

We appreciate the support of the USAID Feed The Future (AID-OAA-A-13-00044 to AHP, TSC), AFRI Plant Growth and Development Program (2016-67013-24608 to AHP), USDA Biotechnology Risk Assessment Program (2012-01658 to AHP and TSC), AFRI Controlling Weedy and Invasive Plants program (2013-67013-21306 to J.N. Barney and AHP), and NIFA Global Food Security CAP (2015-68004-23492 to AHP, JNB).

References

Arends, D., P. Prins, R.C. Jansen, and K.W. Broman, 2010 R/qtl: high-throughput multiple QTL mapping. *Bioinformatics* 26 (23):2990-2992.

Broman, K.W., 2003 Mapping quantitative trait loci in the case of a spike in the phenotype distribution. *Genetics* 163 (3):1169-1175.

Duncan, R.R., F.R. Miller, and A.J. Bockholt, 1980 INHERITANCE OF TILLER REGROWTH IN RATOONED SORGHUM. *Canadian Journal of Plant Science* 60 (2):473-478.

Duncan, R.R., and R.B. Moss, 1987 Comparative Yields of Ratoon Cropped Temperately and Tropically Adapted Grain-Sorghum Hybrids. *Crop Science* 27 (3):569-571.

Gamuyao, R., K. Nagai, M. Ashikari, and S. Reuscher, 2017 A new outlook on sporadic flowering of bamboo. *Plant Signal Behav* 12 (7):e1343780.

Gizmawy, I., J. Kigel, D. Koller, and M. Ofir, 1985 Initiation, Orientation and Early Development of Primary Rhizomes in Sorghum-Halepense (L) Pers. *Annals of Botany* 55 (3):343-350.

Glover, J.D., J.P. Reganold, L.W. Bell, J. Borevitz, E.C. Brummer *et al.*, 2010 Increased Food and Ecosystem Security via Perennial Grains. *Science* 328 (5986):1638-1639.

He, R., M.J. Kim, W. Nelson, T.S. Balbuena, R. Kim *et al.*, 2012 Next-generation sequencing-based transcriptomic and proteomic analysis of the common reed, Phragmites australis (Poaceae), reveals genes involved in invasiveness and rhizome specificity. *American Journal of Botany* 99 (2):232-247.

Hu, F.Y., D.Y. Tao, E. Sacks, B.Y. Fu, P. Xu *et al.*, 2003 Convergent evolution of perenniality in rice and sorghum. *Proceedings of the National Academy of Sciences of the United States of America* 100 (7):4050-4054.

Hu, F.Y., D. Wang, X.Q. Zhao, T. Zhang, H.X. Sun *et al.*, 2011 Identification of rhizome-specific genes by genome-wide differential expression Analysis in Oryza longistaminata. *BMC Plant Biology* 11.

Hunsigi, G., 1989 Ratooning in sugarcane. *Outlook on Agriculture* 18 (4):175-184.

Jang, C.S., T.L. Kamps, D.N. Skinner, S.R. Schulze, W.K. Vencill *et al.*, 2006 Functional classification, genomic organization, putatively cis-acting regulatory elements, and relationship to quantitative trait loci, of sorghum genes with rhizome-enriched expression. *Plant Physiology* 142 (3):1148-1159.

Jang, C.S., T.L. Kamps, H. Tang, J.E. Bowers, C. Lemke *et al.*, 2009 Evolutionary fate of rhizome-specific genes in a non-rhizomatous Sorghum genotype. *Heredity* 102 (3):266-273.

Kong, W., 2013 Genetic Analysis of Plant Architecture in Sorghum in *Crop and Soil Sciences*. The University of Georgia, Athens, GA.

Kong, W., H. Guo, V. Goff, T.-H. Lee, C. Kim *et al.*, 2014 Genetic analysis of vegetative branching in sorghum. *Theoretical and Applied Genetics* 127 (11):2387-2403.

Kong, W.Q., C. Kim, V.H. Goff, D. Zhang, and A.H. Paterson, 2015 Genetic Analysis of Rhizomatousness and Its Relationship with Vegetative Branching of Recombinant Inbred Lines of Sorghum Bicolor X S-Propinquum. *American Journal of Botany* 102 (5):718-724.

Lee, T.H., H.B. Tang, X.Y. Wang, and A.H. Paterson, 2013 PGDD: a database of gene and genome duplication in plants. *Nucleic Acids Research* 41 (D1):D1152-D1158.

Ma, X., Q. Xu, W.A. Meyer, and B. Huang, 2016 Hormone regulation of rhizome development in tall fescue (Festuca arundinacea) associated with proteomic changes controlling respiratory and amino acid metabolism. *Ann Bot* 118 (3):481-494.

McCormick, M., M. Morris, B. Ackerson, and D. Blouin, 1995 Ratoon cropping forage sorghum for silage: yield, fermentation, and nutrition. *Agronomy Journal* 87 (5):952-957.

Montgomery, D.R., 2007 Soil erosion and agricultural sustainability. *Proc Natl Acad Sci U S A* 104 (33):13268-13272.

Paterson, A., H., T.S. Cox, W. Kong, and M. Navarro, 2013 Viewpoint: multiple-harvest sorghums toward improved food securit in *Perennial Crops for Food Security*, edited by C. Batello, L. Wade, S. Cox, N. Pogna, A. Bozzini *et al.*

Paterson, A., H., W. Kong, R. Johnston, P. Nabukalu, G. Wu *et al.*, 2017 The evolution of an invasive plant, Sorghum halepense L. ('Johnsongrass').

Paterson, A.H., Y.R. Lin, Z.K. Li, K.F. Schertz, J.F. Doebley *et al.*, 1995a Convergent Domestication of Cereal Crops by Independent Mutations at Corresponding Genetic-Loci. *Science* 269 (5231):1714-1718.

Paterson, A.H., K.F. Schertz, Y.R. Lin, S.C. Liu, and Y.L. Chang, 1995b The Weediness of Wild Plants - Molecular Analysis of Genes Influencing Dispersal and Persistence of Johnsongrass, Sorghum-Halepense (L) Pers. *Proceedings of the National Academy of Sciences of the United States of America* 92 (13):6127-6131.

R Core Team, 2016 R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Randall, G.W., and D. Mulla, 2001 Nitrate nitrogen in surface waters as influenced by climatic conditions and agricultural practices. *Journal of Environmental Quality* 30:337-344.

Renny-Byfield, S., and J.F. Wendel, 2014 Doubling down on genomes: polyploidy and crop plants. *American Journal of Botany* 101 (10):1711-1725.

Sacks, E.J., M.P. Dhanapala, D.Y. Tao, M.T.S. Cruz, and R. Sallan, 2006 Breeding for perennial growth and fertility in an Oryza sativa O longistaminata population. *Field Crops Research* 95 (1):39-48.

Skipper, M., 2002 Genes from the APETALA3 and PISTILLATA lineages are expressed in developing vascular bundles of the tuberous rhizome, flowering stem and flower Primordia of Eranthis hyemalis. *Ann Bot* 89 (1):83-88.

te Beest, M., J.J. Le Roux, D.M. Richardson, A.K. Brysting, J. Suda *et al.*, 2012 The more the better? The role of polyploidy in facilitating plant invasions. *Ann Bot* 109 (1):19-45.

Wang, X., J. Wang, D. Jin, H. Guo, T.H. Lee *et al.*, 2015 Genome Alignment Spanning Major Poaceae Lineages Reveals Heterogeneous Evolutionary Rates and Alters Inferred Dates for Key Evolutionary Events. *Mol Plant* 8 (6):885-898.

Warwick, S.I., D. Phillips, and C. Andrews, 1986 RHIZOME DEPTH - THE CRITICAL FACTOR IN WINTER SURVIVAL OF SORGHUM-HALEPENSE (L) PERS (JOHNSON GRASS). *Weed Research* 26 (6):381-387.

Westerbergh, A., and J. Doebley, 2004 Quantitative trait loci controlling phenotypes related to the perennial versus annual habit in wild relatives of maize. *Theoretical and Applied Genetics* 109 (7):1544-1553.

Wilson, L.T., Y. Yang, J. Wang, and B.L. Morace, 2014 Texas Rice Crop Survey.

Chapter 6 Tables and Figures:

Table 6.1: Summary statistics for presence/absence of rhizomatousness in the *S. bicolor* × *S. halepense* BC_1F_2 population

	N	No Rhizome (0)†	Weak Rhizome (0-0.3)	Medium Rhizome (0.3-0.6)	Strong Rhizome (0.6-0.9)	All Rhizomes (1)
Athens13	246	3	19	65	124	35
Athens14	236	54	46	94	38	4
Salina13*	246	114	-	92	-	40
Salina14*	236	115	-	83	-	38

† Proportion of presence/absence of rhizomatousness for each genotype

* In Salina 2013, rhizome occurrence was recorded at genotype level for each replication, so the presence/absence of rhizomatousness were grouped in to three groups. We also collapse the data from Salina 2014 to match Salina 2013.

Table 6.2: Summary statistics for the number of rhizomes (**RN**) in the SBSH-BC_1F_2 population.

Env	Pop	N	Mean	Median	SD	Max.
A13	Pooled	246	1.09	1.00	0.5963	4.333
A13	H4	141	1.06	1.00	0.4917	2.500
A13	H6	105	1.14	1.17	0.7131	4.333
A14	Pooled	236	0.50	0.33	0.4682	2.500
A14	H4	134	0.45	0.33	0.4400	2.333
A14	H6	102	0.57	0.50	0.4971	2.500
S13	Pooled	246	0.47	0.17	0.6768	3.667
S13	H4	141	0.35	0.00	0.5530	3.333
S13	H6	105	0.65	0.33	0.7857	3.667
S14	Pooled	236	0.12	0.00	0.3524	2.833
S14	H4	134	0.07	0.00	0.1619	0.833
S14	H6	102	0.20	0.00	0.4942	2.833

Table 6.3: Summary statistics for Rhizome Distance (**RD**) in the SBSH-BC_1F_2 population.

Env	Pop	N	No of Genotypes with Rhizomes	Mean	Median	SD	Min	Max
A13	Pooled	246	243	5.44	5.42	2.4258	0.67	13.69
A13	H4	141	139	5.29	5.42	2.1364	0.83	10.14
A13	H6	105	104	5.65	5.58	2.7630	0.67	13.69
A14	Pooled	236	181	3.23	3.00	1.8321	0.67	10.67
A14	H4	134	101	3.01	2.56	1.8185	0.67	10.67
A14	H6	102	80	3.52	3.43	1.8200	0.67	8.50
S13 *	Pooled	246	132	4.18	3.25	2.6727	0.50	17.50
S13	H4	141	65	3.79	3.17	2.1320	1.00	10.85
S13	H6	105	67	4.55	3.86	3.0801	0.50	17.50
S14 *	Pooled	236	96	4.37	3.50	2.6670	1.50	16.00
S14	H4	134	48	4.30	3.50	2.0878	1.50	9.50
S14	H6	102	48	4.43	3.25	3.1639	1.50	16.00

* S13 and S14 phenotyping used the performance of the row

Table 6.4 Summary statistics for winter survival in the SBSH-BC_1F_2 population

	Athens 13	Athens 14	Salina 13	Salina 14
No. of surviving plants in Field1	10	117	4	4
No. of surviving plants in Field2	15	50	1	3
No. of surviving plots in both Fields	2	30	1	0

Table 6.5 Relationship of rhizomatousness and survival in Athens 2013 in the SBSH-BC_1F_2 population

	Survival		
RG	None	Survival in one field	Survival in two fields
0	3 (100.0%*)	0 (0.0%)	0 (0.0%)
1	18 (100.0%)	0 (0.0%)	0 (0.0%)
2	60 (92.3%)	5 (7.7%)	0 (0.0%)
3	113 (90.4%)	11 (8.8%)	1 (0.8%)
4	29 (82.9%)	5 (14.3%)	1 (2.9%)

* Percentages are calculated with respect to each row

Table 6.6 Relationship of rhizomatousness and survival in Athens 2014 in the SBSH-BC_1F_2 population

	Survival		
RG	None	Survival in one field	Survival in two fields
0	27 (50.0%)	13 (24.1%)	14 (25.9%)
1	22 (47.8%)	13 (28.3%)	11 (23.9%)
2	39 (41.5%)	17 (18.1%)	38 (40.4%)
3-4	11 (26.2%)	11 (26.2%)	20 (47.6%)

* Percentages are calculated with respect to each row

Table 6.7 Parameters of interval QTL mapping results for rhizome group (**RG)** in the SBSH-BC_1F_2 population

QTL	Env	Peak SNP	Chr	Pos (cM)	LOD	Variance explained (%)	Effect	Left (Mb)	Right (Mb)
qRG.6B.H4.1	A14	S6_45264939~ S6_46574440	6B	55.2	2.6	8.54	-0.71	40.8	47.4
qRG.1C.H6.1	A14	S1_66947494	1C	127.0	5.7	20.57	-1.10	66.9	69.1
qRG.2C.H6.1	A14	S2_66432232	2C	148.4	2.8	11.02	-0.91	66.4	72.7
qRG.4B.H6.1	A14	S4_8893237	4B	53.0	4.4	17.41	1.08	8.9	53.5
qRG.4B.H6.2	A13	S4_57391506	4B	123.0	3.5	13.97	0.91	55.9	58.3
qRG.6A.H6.1	A14	S6_57182238	6A	121.4	2.7	11.41	0.84	56.2	58.4
qRG.6A.H6.1	S14	S6_58440379	6A	143.0	2.7	11.61	1.54	56.6	58.2
qRG.6B.H6.1	S14	S6_4292628	6B	1.0	3.4	13.03	-1.43	3.2	4.3
qRG.6B.H6.1	A13	S6_3185305	6B	19.0	2.6	10.39	-0.65	3.4	4.3
qRG.6B.H6.1	A14	S6_3185305	6B	19	4.6	16.32	-0.96	3.3	3.4
qRG.6B.H6.2	A14	S6_40800373	6B	89.0	5.1	17.61	-0.96	40.8	47.0
qRG.6B.H6.3	S13	S6_47011071	6B	104.9	3.7	15.01	-1.68	45.3	47.2
qRG.8B.H6.1	A14	S8_5265160	8B	54.0	3.1	12.73	-0.81	5.3	6.1
qRG.8B.H6.2	S13	S8_2986253	8B	136.0	3.4	10.22	-1.61	1.3	17.1
qRG.9B.H6.1	S13	S9_51332178	9B	51.4	3.1	12.28	1.53	48.0	53.6
qRG.10B.H6.1	S14	S10_52116222	10B	171.8	2.7	10.94	-1.26	5.7	56.7
qRG.10C.H6.1	S13	S10_8524545~ S10_8605702	10C	82.0	3.4	13.38	-1.71	1.3	12.8

Table 6.8 Parameters of Rhizome Group (**RG**, calculated based on the proportion of presence of rhizomes for each genotype and collapsed into five groups) QTLs from single marker analysis

QTL	Peak SNP	Chr	Pos	P vlaue	Effect	Left	Right
qRG.1.H4.1	S1_71002531	1	71.0	0.0000507	0.73	S1_2042095	S1_71437035
qRG.6.H4.1	S6_60754800	6	60.8	0.0000162	-0.83	S6_36266865	S6_60754800
qRG.1.H6.1	S1_65856526	1	65.9	0.00000123	-1.26	S1_48405347	S1_73686219
qRG.2.H6.1	S2_75611690	2	75.6	0.0000961	1.78	S2_56262076	S2_75611690
qRG.3.H6.1	S3_58273458	3	58.3	0.000101	1.18	S3_3357473	S3_69916601
qRG.4.H6.1	S4_8893237	4	8.9	0.0000963	0.95	S4_8296204	S4_58586683
qRG.6.H6.1	S6_3185305	6	3.2	0.00000399	-1.06	S6_1970777	S6_50273666
qRG.6.H6.2	S6_58864585	6	58.9	0.000167	1.08	S6_52203657	S6_61374763
qRG.9.H6.1	S9_48945128	9	48.9	0.00000902	1.34	S9_44764515	S9_58024049

Table 6.9 Parameters of QTL mapping results for number of rhizomes (**RN**) in the SBSH-BC_1F_2 population

QTL	ENV	Peak SNP	Chr	Pos (cM)	Lod	Variance explained (%)	Effect	Left (Mb)	Right (Mb)
qRN.6B.H4.1	A14	S6_45264939~ S6_46574440	6B	55.2	3.4	11.12	-0.30	40.8	47.4
qRN.1B.H6.1	A13	S1_58336516	1B	24.8	2.7	11.04	-0.24	58.8	60.6
qRN.1C.H6.1	A14	S1_66947494	1C	126.0	3.8	14.76	-0.35	67.5	69.1
qRN.4B.H6.1	A14	S4_8893237	4B	53.0	3.6	14.04	0.34	8.9	53.5
qRN.4B.H6.2	A13	S4_56292024~ S4_56292426	4B	119.0	4.2	16.32	0.36	52.6	58.3
qRN.6A.H6.1	A14	S6_57182238	6A	121.4	3.3	13.61	0.34	56.2	58.2
qRN.6B.H6.1	A14	S6_3185305	6B	18.0	5.6	19.04	-0.37	3.2	47.0
qRN.6B.H6.1	A13	S6_3185305	6B	20.0	3.2	12.85	-0.27	3.4	4.3
qRN.6B.H6.2	A14	S6_42382998~ S6_43160977~ S6_44310800	6B	82.1	5.0	21.20	-0.46	41.9	47.0
qRN.6E.H6.1	A14	S6_3163968	6E	8.0	2.7	11.65	-0.36	3.2	48.3
qRN.8B.H6.1	A14	S8_1293001~ S8_1751335~ S8_2579259~ S8_2580157	8B	123.0	3.4	11.90	-0.36	1.1	17.1
qRN.9B.H6.1	A14	S9_47992439	9B	35.1	2.7	11.10	0.29	48.0	55.8
qRN.9C.H6.1	A14	S9_55440637	9C	89.0	2.9	8.83	-0.34	11.7	55.4

Table 6.10 Parameters of QTL mapping results for Rhizome Distance (**RD**) in the SBSH-BC_1F_2 population.

QTL	Env	Year	Chr	Pos	Peak SNP	LOD total	LOD binary	LOD normal	Var binary	Effect binary	Var normal	Effect normal	Left (Mb)	Right (Mb)
qRD.2D.H4.1	Salina	2014	2D	12.0	S2_7141720~ S2_7705193	3.7	3.3	0.5	9.38	**-1.46**[1]	3.82	1.53	5.7	56.2
qRD.1C.H6.1	Athens	2014	1C	126.0	S1_66947494	3.5	2.4	1.3	9.03	**-1.89**	6.31	-1.00	66.9	69.1
qRD.2C.H6.1	Athens	2013	2C	107.0	S2_8953693	3.2	0.0	3.2	0.17	**4.87**	10.10	3.84	8.4	72.0
qRD.4A.H6.1	Salina	2014	4A	144.0	S4_65720396	5.0	0.0	4.9	0.64	-0.07	10.80	**-3.83**	45.6	65.7
qRD.4B.H6.1	Athens	2014	4B	53.0	S4_8893237	3.4	2.0	1.2	8.76	**1.39**	7.27	1.23	8.9	53.5
qRD.4B.H6.2	Athens	2013	4B	121.8	S4_57391596	3.9	0.7	3.2	3.08	10.08	13.02	**2.57**	55.9	58.6
qRD.4D.H6.1	Athens	2013	4D	39.5	S4_8741878	3.0	0.0	3.0	0.30	4.58	11.71	**3.08**	3.2	65.5
qRD.4D.H6.1	Salina	2014	4D	57.0	S4_62367065	3.3	0.8	2.3	3.36	-0.45	2.25	**0.96**	16.9	65.8
qRD.6A.H6.1	Salina	2014	6A	104.0	S6_56308299	3.5	3.5	0.0	13.93	**2.06**	0.83	1.06	55.4	57.2
qRD.6A.H6.1	Athens	2013	6A	115.0	S6_56234867	3.1	0.1	2.9	0.67	-7.07	12.01	**2.12**	56.2	58.1
qRD.6B.H6.1	Salina	2014	6B	0.0	S6_4292628	5.7	5.1	0.6	19.21	**-1.86**	5.87	-1.98	3.3	4.3
qRD.6B.H6.2	Athens	2013	6B	80.0	S6_42382998~ S6_43160977~ S6_44310800	3.3	0.3	3.0	1.29	-8.37	11.78	**-1.94**	40.8	47.0
qRD.6B.H6.2	Athens	2014	6B	82.1	S6_42382998~ S6_43160977~ S6_44310800	3.9	1.8	2.1	8.14	-1.47	11.02	**-1.23**	40.8	47.0
qRD.6B.H6.3	Salina	2013	6B	104.9	S6_47011071	3.7	3.7	0.0	15.01	**-1.68**	0.18	0.12	45.3	47.2
qRD.8A.H6.1	Salina	2014	8A	81.5	S8_1084774	3.9	3.7	0.2	13.26	**-2.26**	2.08	1.02	1.1	53.7
qRD.8A.H6.2	Salina	2013	8A	210.0	S8_1813552	4.9	0.1	4.9	0.69	-0.07	4.20	**-2.56**	1.7	41.4
qRD.8B.H6.1	Salina	2013	8B	134.0	S8_2986253	9.4	3.4	6.1	11.04	-1.22	15.03	**7.25**	1.3	17.1
qRD.9A.H6.1	Salina	2013	9A	85.0	S9_15158378	5.6	1.2	4.9	3.07	0.61	2.84	**-2.77**	15.2	51.8
qRD.9B.H6.1	Salina	2013	9B	51.4	S9_51332178	3.1	3.1	0.1	12.28	**1.53**	0.58	0.45	48.0	53.6
qRD.10C.H6.1	Salina	2014	10C	86.0	S10_6021229	3.6	3.5	0.1	14.48	**-1.86**	1.30	-0.97	1.2	12.8

[1] Numbers in bold are the allele substitutional effect correspond to the model (binary or normal) with larger LOD scores.

Table 6.11 Comparison of QTLs for rhizome presence in the SBSH-BC_1F_2 populations with PQF_2 and PQRIL populations, and with flowering QTLs in the same population and vegetative branching (VB) QTLs in both the SBSH-BC_1F_2 and the PQRIL population

FL QTL	RG QTL	Rhiz-related QTLs PQ-F2	RZ PQ-RIL	VB QTL BC_1F_2	VB PQ-QTL
	qRG.1.H4.1 (+)	LAR[1], LSR (+)	qRZ1.2 (+)		
	qRG.6B.H4.1 (-[1])	LSR[2], RG[4] (-)		qBRCH.6B.H6.2 (-)	
qFL.1C.H6.1 (+)	qRG.1C.H6.1 (-)	LAR, LSR (+)	qRZ1.2 (+)		
	qRG.2C.H6.1 (-)	LSR (-)			
	qRG.3.H6.1(+)	RG (+)	qRZ3.2 (+)	qTL.3E.H6.1, qBRCH.3E.H6.1(+)	qSR3.1 (+)
	qRG.4B.H6.1 (+)	LSR, RG (+)			
	qRG.4B.H6.2 (+)	LSR, RG (+)		qTL4A.H4.1 (+), qTL4D.H4.1 (+)	qIM4.1 (+)
qFL.6E.H6.1 (+)	qRG.6A.H6.1 (+)			qTL6A.H6.1 (+)	qM1_6.1 (+)
qFL.6B.H6.1 (+)	qRG.6B.H6.1 (-)			qBRCH.6B.H6.1 (-)	
qFL.6B.H6.1 (+)	qRG.6B.H6.2 (-)	LSR, RG (-)		qBRCH.6B.H6.2 (-)	
qFL.6B.H6.2 (+)	qRG.6B.H6.3 (-)	LSR (-)		qTL.6B.H6.1 (-)	
	qRG.8B.H6.1 (-)				qTL8.1, qVG8.1 (+)
	qRG.8B.H6.2 (-)		qRZ8.1 (+), qRN8.1 (+)		qMA8.1, qM1_8.1, qAX8.1, qIM8.1 (+)
	qRG.9B.H6.1 (+)	LSR (+)	qRZ9.1 (+)	qTL.9B.H6.1 (+)	qTR9.1 (+)
	qRG.10B.H6.1 (-)	LSR (+)			
qFL.10A.H6.1 (-)	qRG.10C.H6.1 (-)			qTL.10C.H6.1 (-), qBRCH.10C.H6.2 (-)	

[1] Effects from the non-*S. bicolor* parent

[2] LAR: log(n+1) of the number of rhizomes producing above-ground shoots(LAR)

[3] LSR: log(n+1) of subterranean rhizomatousness (rhizome score; LSR) measured after overwintering

[4] RG: regrowth

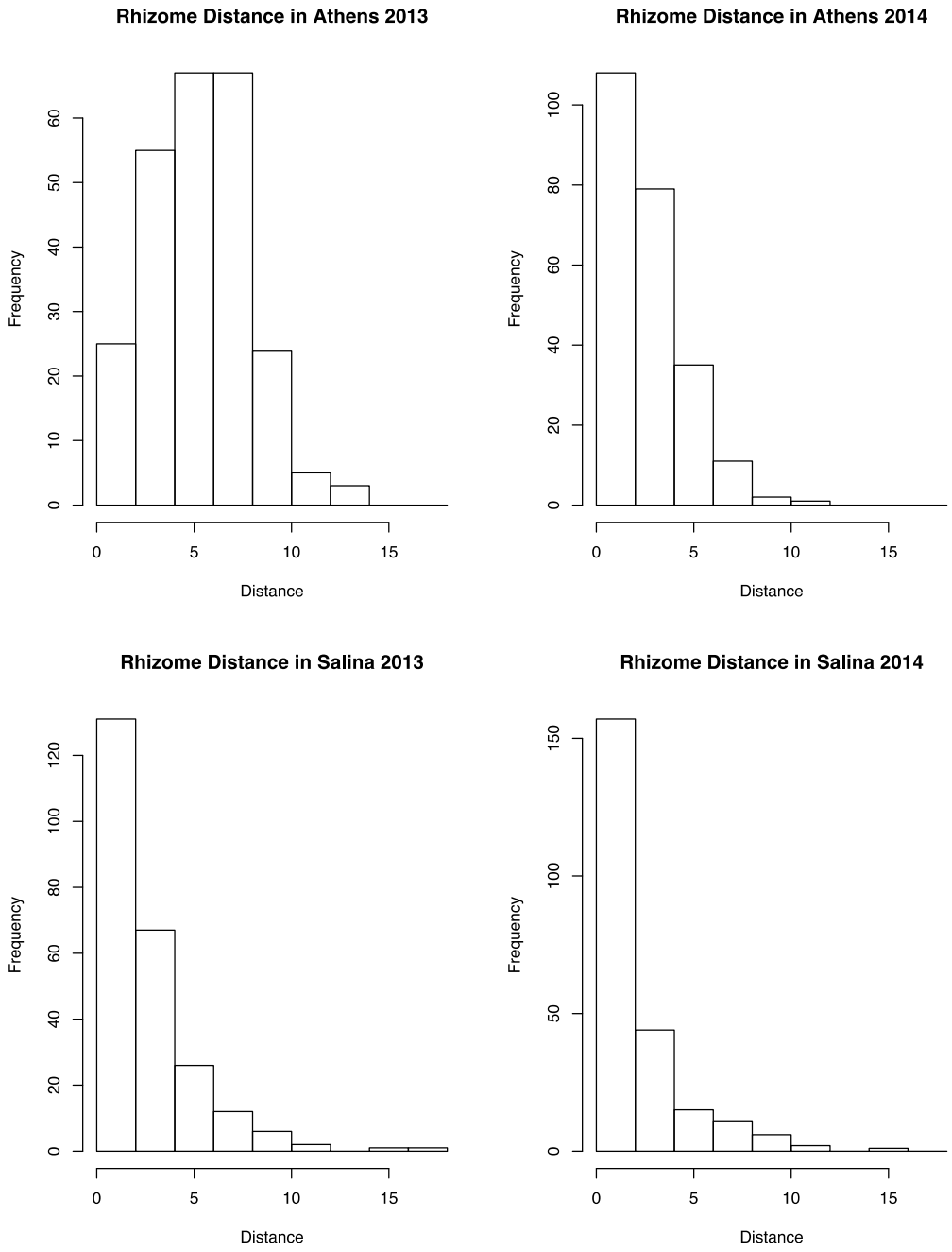


Figure 6.1 Histograms of rhizome distances (**RD**) in four environments.

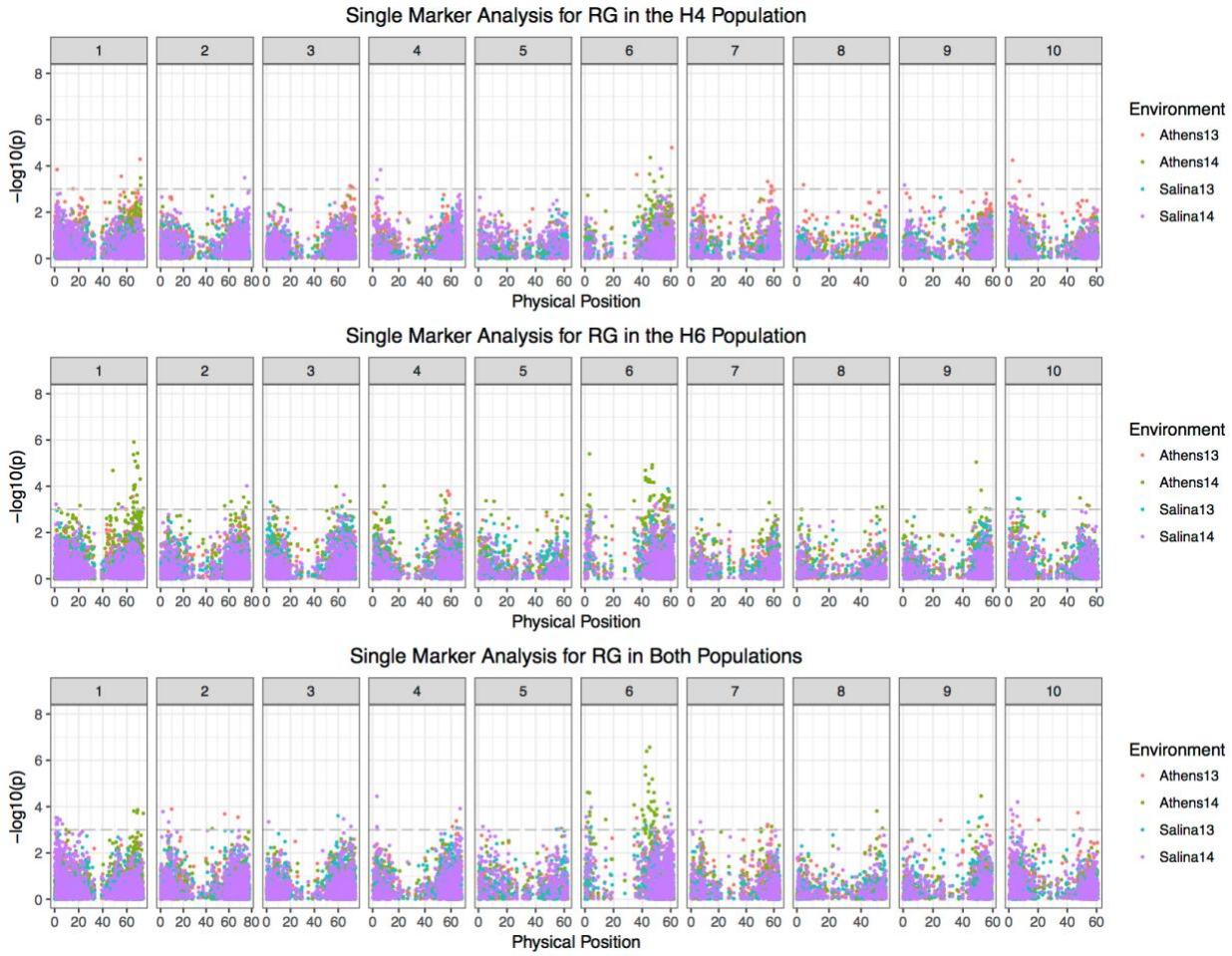


Figure 6.2: Single marker analysis result of **RG** in four environments.

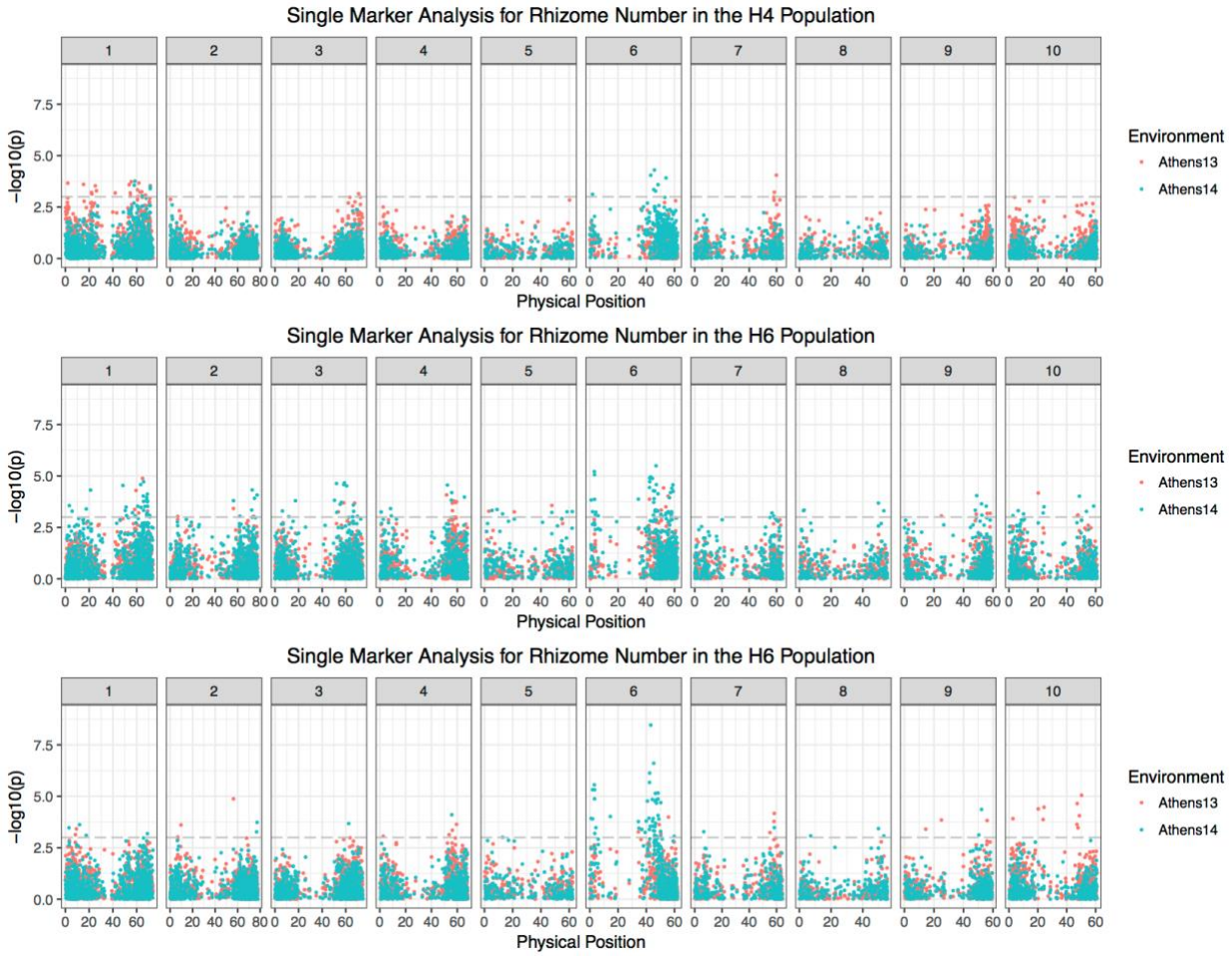


Figure 6.3 Single marker analysis for number of rhizomes (**RN**) in Athens 2013 and Athens 2014.

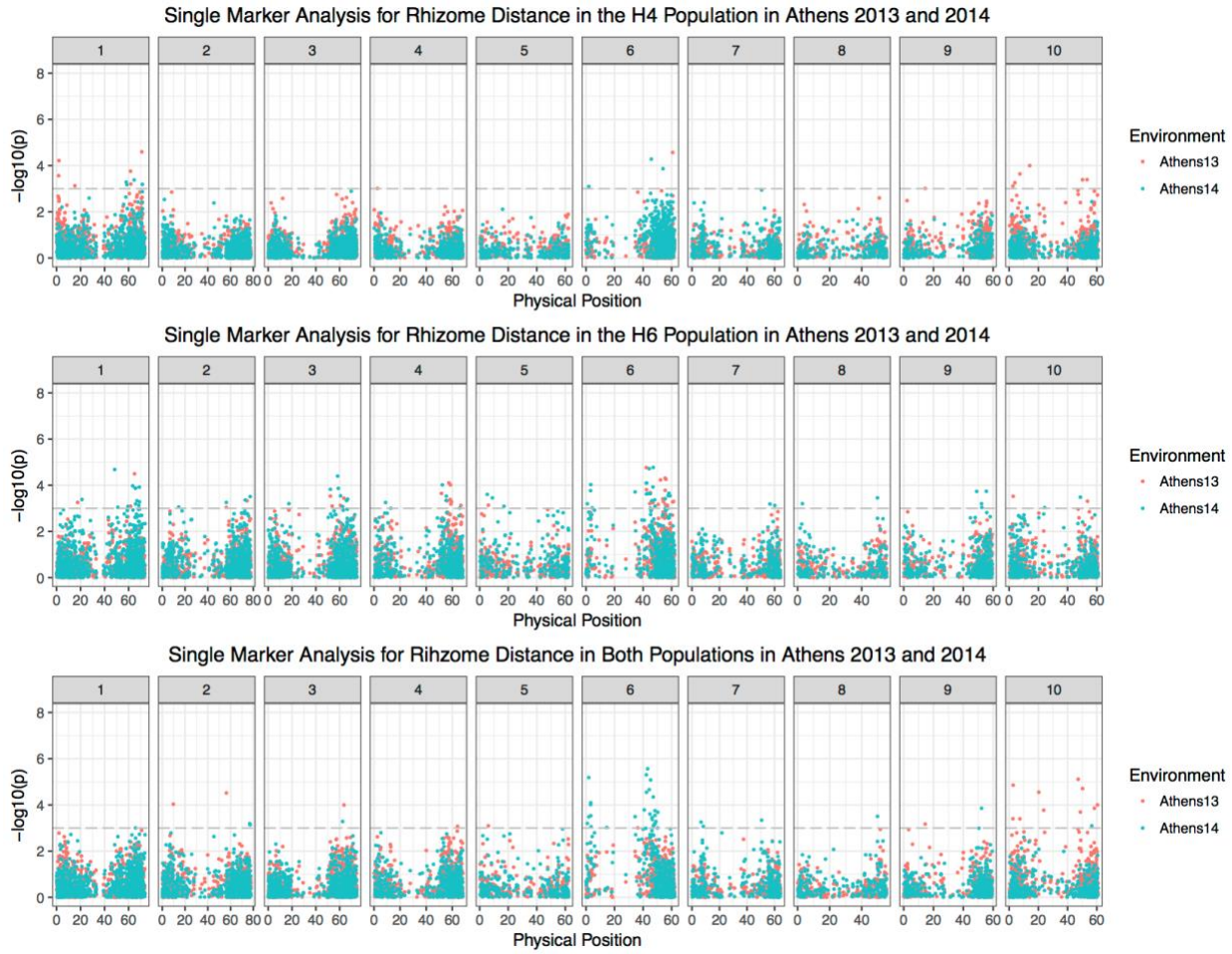


Figure 6.4 Single marker analysis result for Rhizome Distance (**RD**).

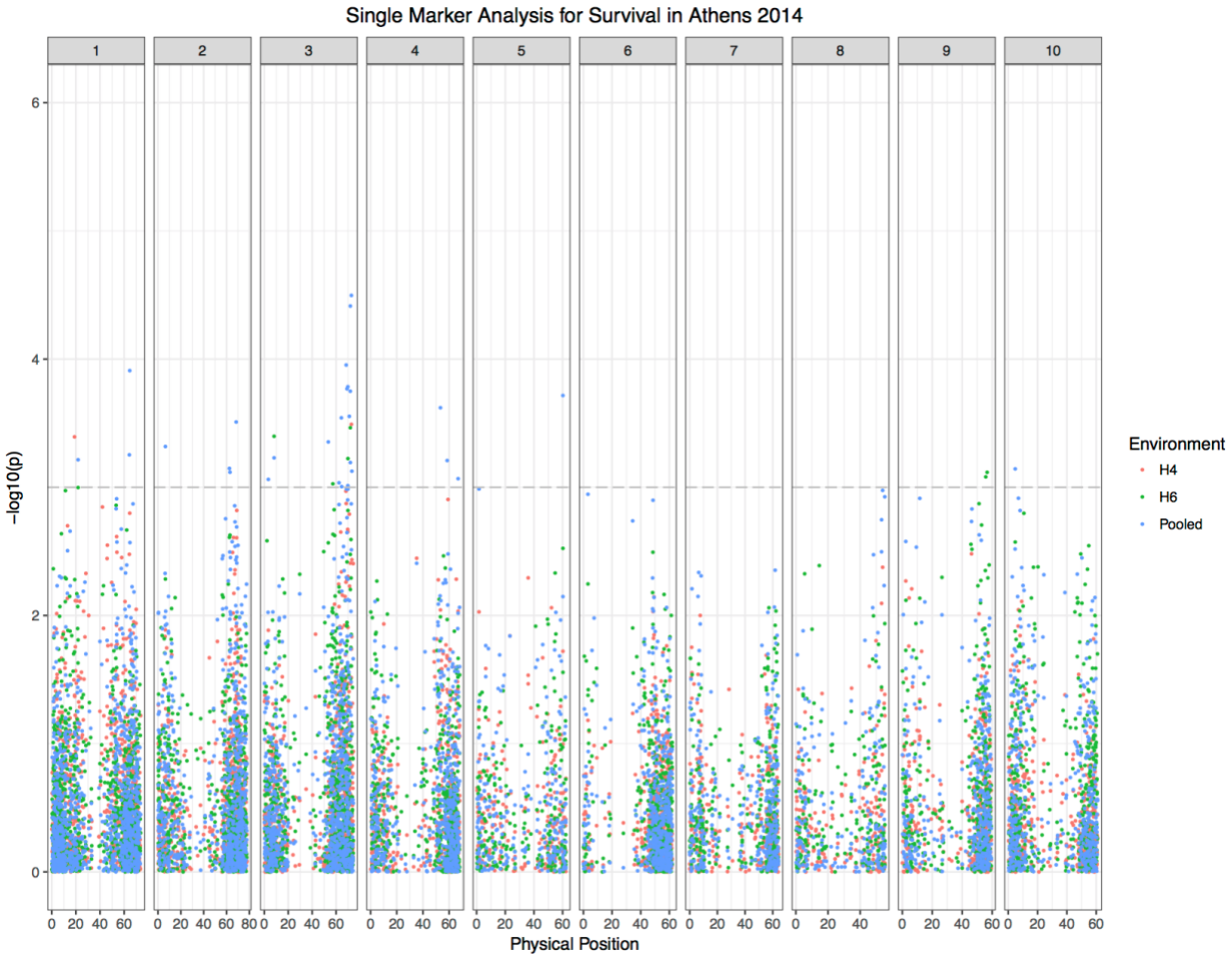


Figure 6.5 Single marker result for Survival (**ISur**)

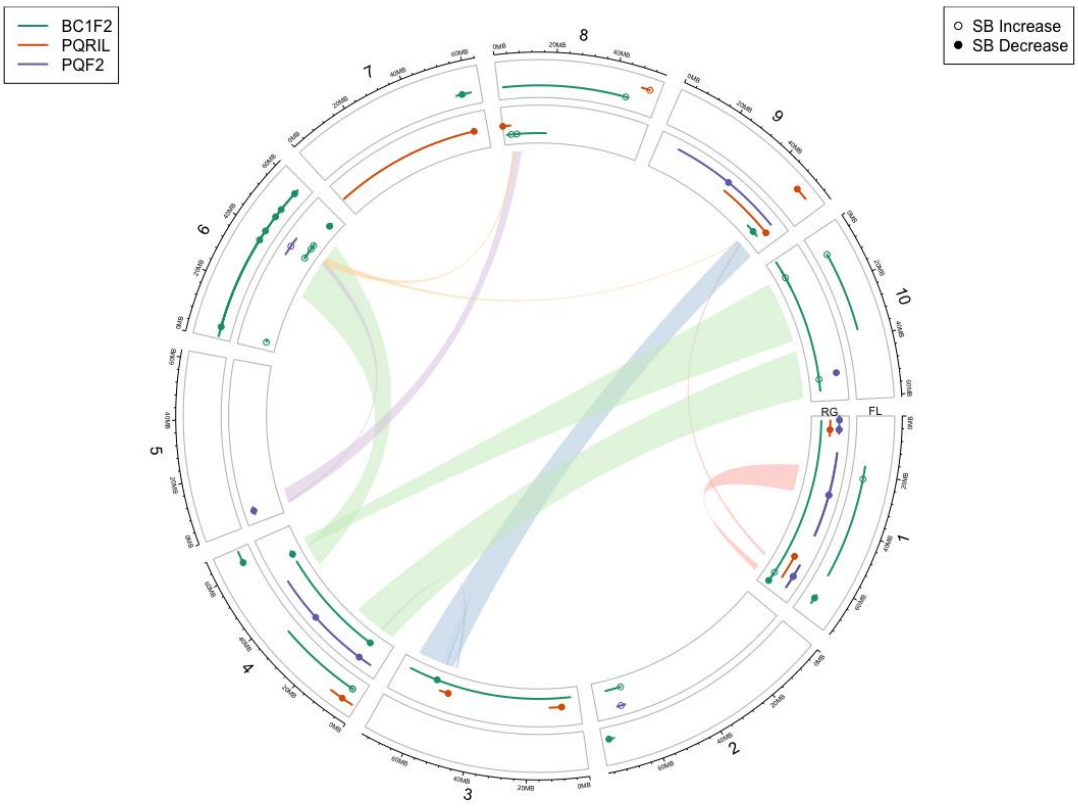


Figure 6.6 Genomic regions related to flowering (**FL**) and rhizomatous (**RG**) QTLs comparison in SBSH-BC_1F_2, PQ-RIL and PQ-F_2 populations. Links in the middle of represent the duplication events in sorghum overlapping with rhizome QTLs (Lee et al. 2013)